orange™ Cyberdefense

Security Navigator 2025

Research-driven insights to build a safer digital society

Hugues Foulon
Executive Director at Orange and
CEO Orange Cyberdefense

> More than ever our 2025 edition of the Security Navigator will enable you to turn challenges into opportunities. The growing but ambiguous role of AI highlights the importance of creating an ecosystem of anticipation."

In the world of cybersecurity, resilience is more than just a concept—it is a necessity. Over the past year, our teams at Orange Cyberdefense have observed an increasingly volatile and complex threat landscape, one that calls for both constant vigilance and innovative adaptation. The Security Navigator 2025 report presents a detailed examination of these challenges and, importantly, the proactive measures that can transform vulnerabilities into opportunities for stronger defense.

The data we have gathered over the past twelve months reveals stark shifts. Cyber extortion, hacktivism, AI-driven attacks, and threats to operational and mobile networks are not merely emerging trends; they are realities that are reshaping the cyber landscape. As malicious actors exploit new technologies and adopt increasingly aggressive tactics, the potential for harm extends beyond digital boundaries, impacting the very fabric of businesses and societies worldwide.

What makes this year's Security Navigator unique is our expanded focus on the role of Artificial Intelligence in cybersecurity. From enhancing threat detection capabilities to mitigating complex vulnerabilities, we leverage AI to improve both offensive and defensive strategies. However, the rise of adversarial AI techniques—models specifically trained for malicious purposes—reminds us that innovation must be matched by responsibility. Our goal is not only to adopt the latest technologies but to do so thoughtfully, balancing progress with caution to secure a safer digital world. AI is not only a land of promises, and we need to remain careful on investing in and using these new technologies. It is all about balance and analyzing the hidden side of any wide-spreading technology; just like IT, shadow AI is now at stake.

This year, we also delve deeper into the threats facing critical infrastructure, particularly within Operational Technology and mobile networks. With increased connectivity and the adoption of IoT and 5G, these systems offer an expanded attack surface that calls for comprehensive, cross-functional defenses. At Orange Cyberdefense, we understand that building cyber resilience requires collaboration at every level—from industry alliances and partnerships to close work with our clients. This is also a matter of public-private cooperation. In 2025, regulation will make the European cybersecurity ecosystem go one step up and we are ready to support this movement.

Cybersecurity today is less about containment and more about anticipation. Informed by 135,225 analyzed incidents, a robust understanding of attacker behavior, and pioneering threat intelligence, our Security Navigator provides actionable insights to help our clients stay a step ahead. I am immensely proud of the dedicated work that went into this report, and I am confident that the insights it contains will empower you to face the challenges of an ever-evolving cyber threat landscape.

As we continue to confront these cyber threats together, let us remain focused on our mission: to build a safer digital society. Our commitment to this mission is stronger than ever, and we are honored to partner with you in securing a resilient digital future.

Hugues Foulon
Executive Director at Orange
and CEO Orange Cyberdefense

Table of Contents

Research

Security predictions: A story of convergence, intelligence and resilience 109

Summary: What Have We Learned? 114

Appendix:

Charl van der Walt
Head of Security Research
Orange Cyberdefense

Introduction

The Year 2024 in a Nutshell

Cynical Security

I was so relieved when experts confirmed that the widely reported exploding-pager against Hezbollah did not involve a significant cyber component. The attacks in Lebanon and Syria involved modified radio pagers and other electronic devices that exploded, resulting in dozens of deaths and hundreds of injuries[1]. Israeli intelligence is suspected to be behind the incidents[2]. The modifications for the attacks were reportedly achieved by altering the devices at the production level to include small amounts of explosives. This allowed the attackers to distribute the modified pagers and other electronic devices widely before triggering them remotely.

When news of the incident began to emerge, people like me in cybersecurity all instinctively wondered if it had involved some kind of cyber-attack. It seemed highly unlikely, but many of us have become so cynical. And with good reason.

Cybersecurity failures – albeit not in a form suited to a Grisham novel – are indeed threatening lives. The cyber extortion attack against the South African National Health Laboratory Service (NHLS) in June this year impacted the service's ability to generate lab reports and send them to clinicians. The disruption lasted several weeks, resulting in reports about clinics coming to a standstill, and patients in emergency wards and intensive care units in fatal danger[3]. In an unusual twist , someone who described himself as "the middleman" called the press in South Africa to warn that related patient deaths would be "on the NHLS for not engaging".

Extortionate Security

As cyber extortion continues to increase globally, we note this year that it is also becoming increasingly cynical. This year Diana Selck-Paulsson examines over 13,000 cyber extortion incidents, and reports how extortion tactics are demonstrating increased aggression and moral decline, abandoning previous restraints on targeting sensitive sectors like healthcare. Once considered off-limits, hospitals and essential care facilities now face a surge in attacks. Small and medium-sized businesses are also becoming more frequent targets, accounting for over two thirds of all victims. Small businesses saw a 53% increase in cyber extortion attacks, while medium-sized businesses experienced a 52% rise. Vulnerable, smaller countries are not immune either. This year for the first time we report Cy-X victims in countries like Afghanistan, Djibouti, Tokelau, Nepal, Uzbekistan and Maldives. Attackers are also exploiting cynical "revictimization" strategies, where stolen data is reused across multiple extortion platforms and amplifying the psychological burden on victims.

Subversive Security

This year we also explore shifts in hacktivism, which is becoming increasingly cynical and aggressive. Once grounded in activism, hacktivism now bears a closer resemblance to cyber extortion, with a focus on destabilizing communities and weaponizing fear against both individuals and institutions.

Diana also continues her excellent work on this phenomenon, examining over 6,500 hacktivist incidents to reveal how the emerging hacktivist model focuses on public manipulation, societal division, and the erosion of trust. Hacktivists are aligning with state-sponsored agendas, targeting critical infrastructure like election systems - seeking not only to disrupt essential services but also to undermine public confidence in government and democratic institutions. By attacking election-related systems and other symbolic institutions, the hacktivist groups aim to undermine public trust, disrupt the flow of information, and potentially influence the outcome of a key democratic process. By leveraging sophisticated DDoS-for-hire services and anonymous cryptocurrency incentives, hacktivists are blending public shaming with extortion techniques to exploit fear and amplify public pressure. While Europe is the primary focus for the group Diana studied, everyone is a potential target, and the problem threatens societies as a whole.

Cyber Physical Security

Hacktivists are a significant threat to cyber-physical environments like factories, plants and utilities. In fact, our research attributes 23% of targeted attacks against operational technology environments to hacktivist actors.

Ric Derbyshire is a specialist in operational technology (OT) and industrial control systems. He's expanded his OT security dataset to cover 119 recorded cyber-attacks over a period of 35 years. This year his unique dataset expanded with 47 incidents from the last 12 months.

This year's insights again underscore the prevalence and impact of cyber extortion (Cy-X) on OT systems. Attacks originating in IT environments frequently cascade into OT systems, disrupting essential operations and causing downtime. Despite rarely being the primary targets, OT environments face unintended consequences due to interconnected IT and OT networks. Correspondingly, the manufacturing sector accounts for 20% of all cyber extortion victims this year and has seen a 25% increase from the previous year.

81% of this year's documented attacks were perpetrated by criminals and primarily impacted IT systems, not OT. But, as we posited last year, threat actors will start to focus on OT systems directly when the environmental factors align.

An attack impacting Spanish bioenergy plant Matadero de Gijón in April this year is an early indicator that this may be happening already. The attack is recorded in Diana's dataset (Cy-X) and in Ric's dataset (OT) but stands out because it directly impacted the plant's Supervisory Control and Data Acquisition (SCADA) system.

In this year's report, Ric focuses on "category 2" incidents in OT - those directly targeting OT systems through adversarial tactics unique to these environments- a category that only accounts for 16% of recorded incidents. These category 2 attacks are more intentional and sophisticated, often involving advanced tactics by state-sponsored groups and sophisticated cybercriminals, who aim to directly compromise OT operations. Ric points out that 46% of category 2 attacks resulted in "manipulation of control" as an impact. This means that the adversary manipulated the physical process in their attack. This is clearly a frightening outcome, and most category 2 attacks have equally severe impacts.

Category 2 incidents, while relatively infrequent, force our risk models to consider the unthinkable. This pressure places enormous additional responsibility on those responsible for protecting cyber-physical systems.

Ric argues that category 2 OT attacks tend to exploit native functionality within the victim's environment—a technique known as living off the land. As with IT-attacks, this approach allows adversaries to blend in and evade detection, but it places the adversary in the optimal position to cause real damage in an environment. For example, exploiting a programmable logic controller (PLC) by using expected functions is safer and more stable for attackers than risking a memory abuse vulnerability, but also allows attackers to abuse the ability of that PLC to manipulate the physical environment.

This reality has significant implications for how we approach security in OT environments.

For example, simply accessing an OT environment doesn't mean that an attacker can achieve a desired cyber-physical impact. This raises an essential question: how can asset owners assess their OT environment's resilience against category 2 threats?

Ric explores significant challenges and gaps in current OT security, and specifically penetration testing approaches. The discipline is still in its infancy, with limited research and ambiguous guidance that fails to fully account for unique OT tactics, techniques, and procedures (TTPs), especially those seen in category 2 attacks. Ric critiques the reliance on IT-oriented penetration testing practices, which often focus on gaining OT access and declaring success, overlooking the complexities of truly emulating OT-focused adversaries. He questions whether current testing approaches effectively capture the nuanced tactics used in real OT attacks, such as those exploiting native functionality for stealth and control.

Our report this year highlights the need for security approaches that anticipate complex OT-specific kill chains and TTPs to more accurately ensure resilience against genuine threats. As with so many things that need to be rethought in contemporary cybersecurity, we argue this year the traditional IT frameworks are not appropriate for addressing OT's particular threats and vulnerabilities.

Mobile Security

In a new section of this year's report, Orange mobile network security specialists Emmanuelle Bernard, Stéphane Gorse, and Sébastien Roché outline the evolution of mobile network vulnerabilities, describing how each generation of mobile technology (2G through 5G) has introduced advanced features alongside an expanded attack surface. While early networks primarily faced issues from weak 2G encryption, newer generations brought complex protocols like SS7 in 3G and Diameter in 4G, which attackers now exploit. With 5G, increased virtualization, APIs, and IoT integration have introduced new risks, including supply chain attacks and vulnerabilities accessible remotely through Internet-connected devices.

Our report identifies three primary attack domains: SIM cards, devices, and infrastructure. SIM-based attacks use techniques like SIM swapping, cloning, and USSD protocol misuse to intercept data or impersonate users.

Device-based threats center around malware and mobile OS exploitation, especially through alternative app stores that lack strict security. Infrastructure attacks target network protocols and exploit carrier interoperability to intercept communications . We note that MFA use on mobile devices has also complicated the risk by giving threat actors motive and opportunity to compromise network-linked authentication methods.

Our report emphasizes a layered security approach that includes enhanced standardization and collaboration among network operators, device manufacturers, and regulatory bodies. But given the cross-functional nature of mobile networks today, enterprises are also being forced to consider comprehensive security responses that range from securing devices and infrastructure to raising user awareness about safe practices.

Struggling Security

While our adversaries are becoming more cynical, and the impact of security failures more profound, we as the defenders are still struggling to stem the flood.

This year veteran security researchers Wicus Ross and Rogan Dawes study 1.3 million vulnerabilities across 69,000 customer assets to surface a critical message: We need to change the way we think about security vulnerabilities.

Wicus' work focuses on how businesses tackle vulnerabilities. He illustrates that vulnerabilities are emerging at such a pace that traditional, reactive measures simply aren't keeping up.

As Wicus shows, for example, vulnerability management teams face an increasingly daunting task as they contend with the overwhelming volume and velocity of new vulnerabilities. With endless new vulnerabilities emerging continuously, we are forced into a reactive mode, obliged to prioritize and address threats without control over the cadence or velocity of intelligence. Organizations with already-limited capacity are left to scramble from the back foot, unable to make sense of an ever-evolving threat landscape.

The complexity of large enterprise environments adds to these challenges, as even high-probability vulnerabilities identified by metrics like EPSS are difficult to mitigate at scale. In this report we argue that covering all potential exploits across vast networks is fundamentally impractical, meaning that crucial decisions must be made about which systems to patch first. But we argue that the "risk-focused" approach isn't effective either. Wicus' study of EPSS and statistical probabilities argues that even low-severity issues at sufficient scale leave the business vulnerable to compromise. The problem calls for a fresh approach, and in this year's report we argue that must start with a clarification of fundamental terms.

"Vulnerability Management" needs to go. Wicus proposes that new approaches with new descriptions are urgently needed.

Security From the Source

Wicus and Rogan both also put the responsibility on software vendors to prioritize security in software development, and throughout a products lifecycle.

As I write this, our CERT, Vulnerability Management, Threat Detection and Managed Services teams are wrestling to contain the threat and impact of "FortiJump[4]" – a severity 9.8 vulnerability in Fortinet FortiManager.

In mid-October, Fortinet alerted key partners and select clients, including Orange Cyberdefense, to a critical 0-day vulnerability actively exploited in FortiManager, a product essential for managing security tools like FortiGate firewalls. The vulnerability allows remote attackers to execute commands on vulnerable devices by exploiting a missing authentication check in the FortiManager-to-FortiGate protocol. Fortinet has since released patches, which we and others are of course rushing to deploy. Meanwhile the bug has been actively exploited – apparently by Chinese APT actors - for some time already. Reconnaissance likely began as early as July this year, with widespread exploitation following in September. Fortinet and others are sharing specific indicators that defenders are scouring their systems for.

It feels like a fitting soundtrack for this report.

Despite this urgency, many products — including those explicitly designed for cybersecurity — continue to exhibit fundamental flaws that leave clients exposed. This gap is more than technical; as we detail in this report, there's a clear and urgent need for secure-by-design principles to become an industry standard, addressing vulnerabilities at the source instead of relying on patches and workarounds after release.

Rogan's work highlights the significant number of troubling examples of security products — firewalls, endpoint protection, intrusion prevention systems — shipping with exploitable weaknesses. These vulnerabilities are often in products that sit directly exposed to the internet, where their primary function is to facilitate secure authenticated access to sensitive areas inside an organization. Every new vulnerability uncovered in these trusted tools not only threatens the systems they protect, but also erodes confidence in the very solutions meant to safeguard our digital infrastructure.

Wicus' study of almost 500 security advisories released by our World Watch team this year illustrates just how pervasive this problem has become. Last year security vendor Ivanti was truly in the crosshairs, but vendors in general are letting us down,

- 11 Jan 2024 – Two new 0-day vulnerabilities actively exploited against Ivanti Connect Secure VPN. This saw the start of several weeks of updates by Ivanti to release fixes for all their impacted products.
- 7 Feb 2024 – Dutch Military Intelligence and Security Service (MIVD) disclosed that Chinese state-sponsored threat actors infiltrated the Ministry of Defense of the Netherlands in 2023. Attackers were exploiting an old vulnerability in FortiOS SSL-VPN affecting FortiGate devices. In June 2024 – the MoD announced that a Chinese threat actor had compromised up to 20,000 FortiGate instances linked to the original announcement.
- 9 Feb 2024 – Fortinet fixed two critical vulnerabilities in FortiOS SSL-VPN, of which one was exploited in the wild prior to the fix.
- 18 Mar 2024 – Proof of Concept emerged for critical vulnerability in FortiOS SSL-VPN module. At the time ShadowServer identified nearly 130,000 vulnerable instances and noted exploitation attempts.
- 14 Apr 2024 – Critical vulnerability in GlobalProtect firewall from Palo Alto Networks linked to targeted 0-day exploitation. This was the only Critical (5/5) advisory from World Watch during this report period.
- 29 May 2024 – Check Point disclosed an exploited 0-day vulnerability in its remote access VPN solution. Attackers had already been attempting to exploit the vulnerability a month earlier.
- 19 Jul 2024 – CrowdStrike's Falcon Sensor update crashed Windows machines all over the world. The outage was linked to an update that had a malformed channel file.

As an industry, Rogan argues, we should be solving these problems, not creating them. As we have since 2022, we call on our partners and competitors in the security industry to come together to work on this challenge.

Struggling to Respond

In the face of this barrage of threats, Wicus Ross' analysis of our threat detection data highlights the several challenges in detecting and responding to security incidents. One key observation is the increased misuse of systems by employees. Such "insider" activity makes distinguishing between benign and malicious activities even more difficult, particularly as attackers increasingly use "Living off the Land" (LOL) methods that resemble normal user behavior. As detection teams are finding it difficult to distinguish between benign user actions and actual threats, Wicus' report suggests that fostering "pervasive cyber judgment" across the organization is essential.

The need to respond to LOL and other "insider threats" forces detection teams to collect and analyze yet more, subtle indicators. This additional load makes separating real signals from the noise even more challenging. Our report shows that confirmed incidents, or "True Positives," comprised only 14.98% of the incidents we analyzed. The remaining incidents were classified as: 12.36% "True Legitimates" (genuine activity that posed no threat), and 61.74% "False Positives" (mistaken detections). 10.92% remained uncategorized.

The impact of this load and complexity has a measurable impact on our collective ability to detect and respond to potential incidents. This year for the first time we present insight in our Mean Time to Resolve (MTTR) statistics. This metric is complex due to varied incident types and the necessity for client coordination, but analysis reveals that while many incidents are resolved quickly, the loop on priority incidents can take over a day to close.

We remind readers of our 2024 research piece titled "Fake News and False Positives", where we pointed out that over time there are detection efficiency gains as the relationship between our detection teams and our client teams grows and matures.

Improved feedback loops are essential in refining detection systems and improving confirmed incident rates.

In light of these challenges, Senior CSIRT Analyst Simone Kraus examines the critical role of human analysts in threat hunting, stressing the unique value that human insights bring to the detection of sophisticated threats. While automated detection tools are useful, they cannot fully replace the intuition and adaptability of skilled security analysts who can recognize nuanced attack patterns and respond effectively. Simone introduces the concept of "threat-informed defense," where understanding an organization's specific threat landscape helps tailor defense strategies. This approach integrates knowledge from actual incidents and threat intelligence, allowing defenders to anticipate likely attack vectors and prioritize resources accordingly.

We also examine common organizational challenges in Incident Response in a study by Saskia Kuschke, a Senior CSIRT Investigator. Saskia's work notes that many companies struggle with foundational elements like asset mapping. But incident readiness can also be stymied by unclear roles, lack of communication, and low user awareness, all of which contribute to slower responses and higher risks during actual incidents. Saskia proposes a structured approach to building incident response readiness. She emphasizes a hierarchy of needs, starting with essential tasks such as role assignment and incident communication protocols. Her proposed model progresses through asset mapping, visibility enhancements, and eventually, complex detection and response capabilities. This tiered approach allows organizations to scale their security efforts methodically.

Artificial Intelligence

Like almost every research team in security, this year we consider the impact of LLMs and GenAI on the security landscape. Large Language Models - born out of advancements in natural language processing and machine learning - have transformed from rudimentary text-processing tools to sophisticated systems capable of generating human-like responses.

Anis Trabelsi is a team lead on Data and AI. This year he discusses how AI can help address the challenge of detecting beaconing—subtle, periodic communications that malware uses to connect with command-and-control servers—by leveraging AI to enhance detection capabilities. These beaconing signals often blend in with legitimate traffic, making them difficult to spot with traditional methods. Anis describes an AI-driven approach his team developed, centered on analyzing proxy logs to capture network activity in real time. By identifying repetitive requests or unusual traffic patterns, the system generates rapid alerts, enabling faster defensive actions. This research shows how AI can strengthen detection accuracy and scalability, significantly narrowing the window for attackers to exploit these covert channels.

The impact of LLMs on security defense is clearly exciting, but we make the argument this year that new technologies often favor the offensive side, so technologies like GenAI are likely to benefit attackers more than defenders.

While these tools may enable more effective response by businesses, the same capabilities can be weaponized by malicious actors, allowing them to conduct more sophisticated attacks with greater ease. If AI is generally thought of as a productivity tool, then we can expect it to make attackers more productive also. Despite these risks, our research suggests that existing security practices are often sufficient for mitigating many of the threats associated with GenAI, although consistency is crucial.

Rather than focusing on GenAIs power for attacker or defenders, however, our report this year is primarily concerned with the broader risks that emerge when businesses and individuals adopt LLM and GenAI technologies. With continuous reports about how threat actors may (ab)use LLMs, the less colorful risk introduced in the application of the very young LLM technology as an interface by businesses is being underestimated, especially where these systems serve as a bridge between the open internet and critical business assets.

Untested, opaque AI interfaces deployed as an interface pose a significant risk to the internal systems they interface with. We cite the recent example of a breach at an NSFW AI chatbot service. Here, a hacker exploited vulnerabilities in the platform, which they described as "a handful of open-source projects duct-taped together." This complex, poorly engineered system allowed easy access to the platform's backend systems and data. We expect to be reporting on many more incidents like this over the next year and urge readers to be extremely cautious about how and where they deploy AI on top of their own backend systems.

Research by pentester Geoffrey Sauvageot Berland's in this report examines the specific risk of prompt injection - manipulated inputs that can mislead or disrupt GenAI behavior. By exploiting the predictive nature of LLMs, attackers can bypass ethical and security controls, causing the model to generate unintended outputs. Techniques include "context switching," which introduces abrupt topic shifts to elicit unauthorized responses, and obfuscation, where forbidden terms are disguised through encoding to evade content filters. Geoffrey also warns of denial-of-service attacks that overload models with complex tasks, as well as the risks posed by multimodal applications where malicious commands can be hidden in images or audio, expanding the AI attack surface.

In the face of enormous pressure to integrate LLMs into business operations, we argue for a cautious, guarded approach that begins with a clear definition of the use-cases and desired outcomes an AI is expected to deliver, so that risks can be assessed and objectively weighed against potential benefits. We need to heed lessons from previous technology revolutions, perform rigorous security testing and thoughtful deployment of LLMs to ensure the necessary balance between security, safety and any productivity and the promised operational benefits GenAI may deliver.

What Are We defending?

A recurring theme in this year's report is a critical shift as attackers increasingly target perception and trust through cognitive attacks. These attacks, which go beyond traditional technical disruptions, are aimed at manipulating public opinion, undermining trust in institutions, and destabilizing societal confidence. One example involves pro-Russian hacktivist groups, who align their campaigns with major geopolitical events such as elections and summits to amplify their impact. By targeting symbolic infrastructure and leveraging public platforms like Telegram, these groups blur the line between cybercrime and influence operations. Their ultimate objective isn't solely system disruption, but rather the erosion of trust in democratic systems and processes.

In a similar vein, cyber extortion actors employ psychological tactics to manipulate perceptions. Following a major law enforcement crackdown under Europol's Operation Cronos, which significantly limited their operational capabilities, the Cy-X group LockBit countered by inflating their victim numbers and projecting an image of resilience and strength. This tactic aimed to maintain confidence among affiliates and instill fear in potential targets. Along with our findings on the cyber extortion phenomenon of "revictimization", these examples exemplify how cyber extortion tactics are increasingly perception-focused, using narrative control to affect both victims' and the criminal ecosystem's responses.

It's into this context that Artificial intelligence (AI) is emerging as a powerful tool for attackers in cognitive operations, adding a new dimension to misinformation campaigns. State-sponsored actors from countries such as China, Russia, and Iran leverage generative AI to create realistic phishing content, fake images, and deepfakes that can deceive large audiences[5][6]. These AI-supported attacks aim to influence public perception on a mass scale, from disrupting elections to discrediting political candidates, eroding trust in democratic institutions. The integration of AI into existing campaigns increases the role of cognitive attacks in the threat landscape, providing actors with scalable tools to craft highly convincing, tailored narratives to suit their needs.

These shifts represent a significant new challenge for security defenders. In addition to "simply" countering technical threats, we must now broaden our approach to incorporate strategies to counter cognitive and perception-based threats and psychology-driven attacks, which target minds as much as systems.

Security is not an objective state, it's the subjective expression of our freedom to pursue shared visions and construct a society that is equitable and rewarding. Cognitive attacks leverage technical compromises, not as an end in themselves, but as a means of launching an assault on the fabric of trust on which "secure" systems are built. Cognitive attacks require us to not only counter technical intrusions, but also safeguard the public perception of trust we need for our digital and interconnected world to flourish.

Intelligence and Operations Data

Key data of the year

From Reactive to Proactive: Continuous Threat Exposure Management (CTEM)

Given the observations made in this section of the report and the constant shifts throughout the years we have observed, we see a need more than ever for managed detection and response to evolve into something more than a “last line of defense”.

We continue to see the common avenues of attack through classification of incident data but can we do more? In an approach we will also discuss in our section “Beyond Vulnerability Management” we believe strategically that threat detection and response should evolve and move towards continuous threat exposure management, a shift from a reactive function to a more proactive practice; integrating threat detection and response activities and the data they provide into a continuous process of actually trying to fix the problems at source, not just detect them.

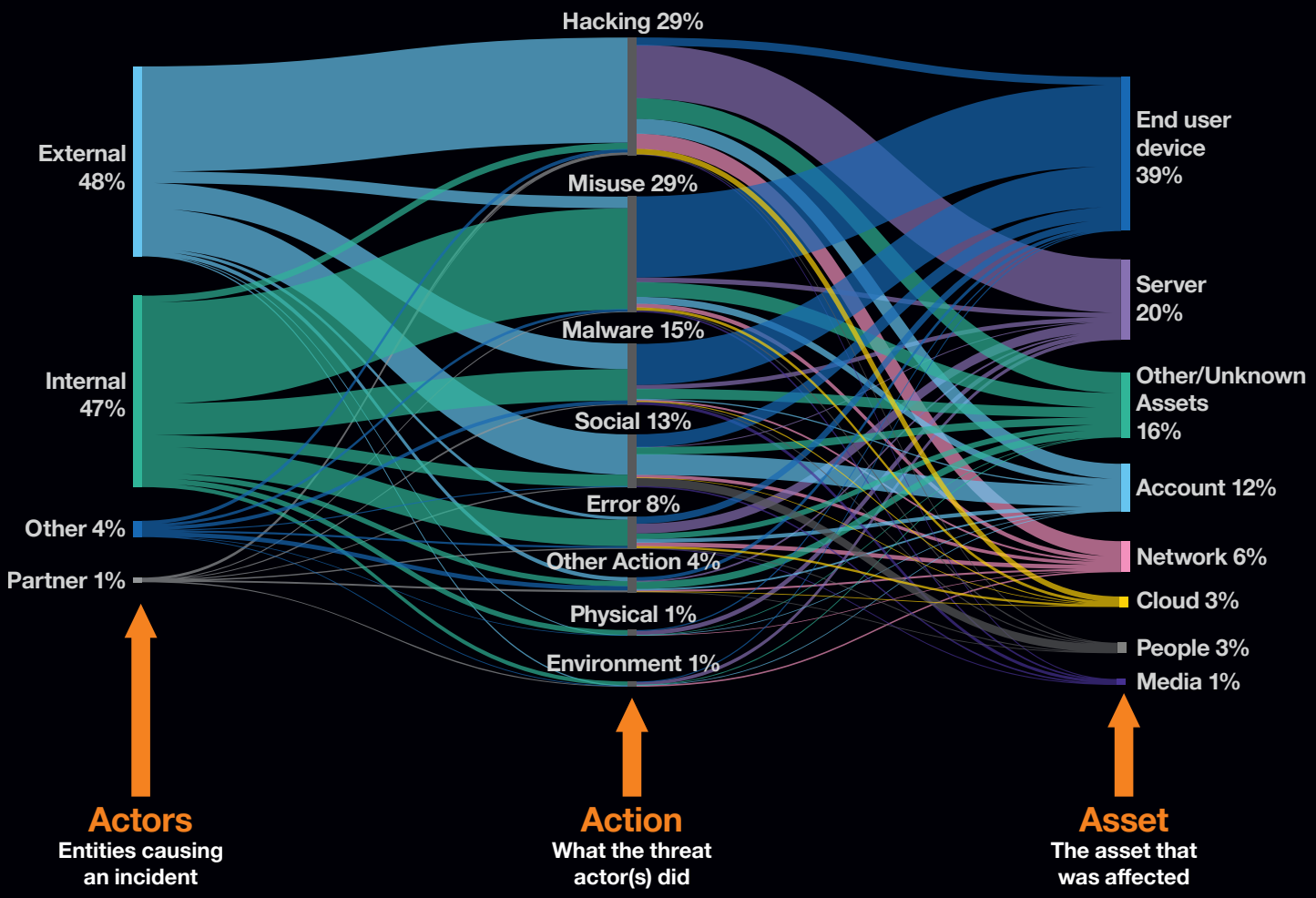
Threat Detection

About the Data

- Total number of incidents: 135,225 (compared with 129,395 in 2023)
- Out of these incidents, 20,706 were confirmed as true positive Incidents (14.98%) However, not all clients include VERIS categories
- Analyzed period from October 2023 to September 2024
- Data sources: Endpoint / extended detection and response (EDR / XDR), network detection and response and SIEM platforms, as well as the enriched incident data from Orange Cyberdefense Core Fusion platform

Funnel: Alert to Incident

135,225 Potential Incidents → 20,706 Confirmed Incidents



* Overview flow with major categories, rounded to full numbers, for details see following pages

Types of Incidents

Incidents are categorized according to the VERIS (Vocabulary for Event Records and Incident Sharing) framework. We record actors, actions, assets and attributes affected by an incident.

The threat action categories used in the VERIS framework consist of the following 7 primary categories: malware, hacking, social, misuse, physical, error and environmental. More information can be found in the glossary on page 116.

A Global View

We have grown the client base and expanded our dataset to include 21.5% more clients. Across this enlarged set we report 13.8 confirmed incidents per month per customer for the past 12 months. This number is significantly lower than the same period the previous year and the year before that. As we will explain later, this is largely because of a larger, more diverse client base, and the fact that "younger" clients generally record fewer incidents while still being onboarded.

As always, we strive to provide a global overview of what we are seeing in our incident data with the aim being to highlight trends that can also be applied to the global threat landscape. To facilitate this, a broad data set is collected from across all of the operational teams within Orange Cyberdefense including our 15 global CyberSOCs .

We consider a years' worth of managed threat detection services data, from 1st October 2023 to 30th September 2024. The distribution between internal and external incidents is basically even this year, with incidents originating internally having increased from 37% in last year's report.

Hacking, misuse, and malware have remained the most prominent Threat Actions, but incidents classed as "misuse" have increased substantially from 16% last year, in line with the increase in incidents originating internally. Malware incidents have increased by about 2%, and "social" incidents have retained their previous level.

End user devices have remained the most impacted assets, but have increased from 28% last year. Again, this is in line with the increase in incidents originating internally. Incidents impacting servers have decreased by about 10 percentage points from last year. Incidents impacting accounts have decreased a little from last year, while network-impacting Incidents have retained their previous level.

Summary

The clear shift from last year is an increase in confirmed incidents originating from internal users and impacting end-user devices. We do not perceive a systemic shift in threat actor behavior in this, but rather glean a sobering lesson about how easily user mistakes or misbehavior on their own endpoints can lead to damaging outcomes.

Events, Incidents, Confirmed Incidents

We log an event that has met certain conditions and is thus considered an indicator of compromise (IoC), attack or vulnerability. An incident is when this logged event, or several events, are correlated or flagged for investigation by a human – our security analysts.

True legitimate incidents are incidents that were raised but after consultation with the customer proved to be legitimate activity. Incidents are categorized as 'false positive' when a false alarm was raised.

Because individual SOCs or clients may have slightly different approaches to defining Incident status, we simplify these categories to 'confirmed' and 'other' in parts of this report.

An incident is considered 'confirmed' when, with help of the customer or at the discretion of the analyst, we can determine that security was indeed compromised. At this point the incident is also categorized. We sometimes refer to these 'confirmed' incidents in this report as 'True Positives'.

Totals

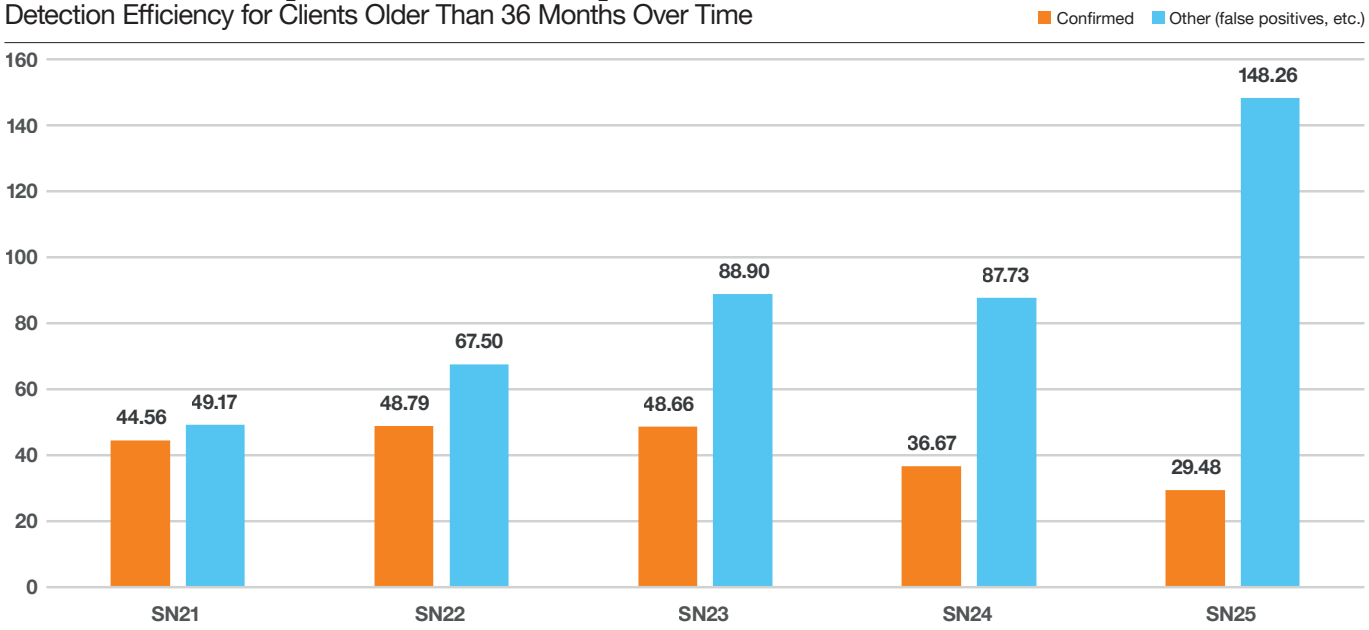
A total of 135,225 incidents were evaluated in this year's dataset, which represents a 4.5% increase over the previous year. "true positives" account for 20,706 incidents, or 14.98% of the total. The balance of incidents (~85%) is comprised of 12.36% true legitimates, 61.74% false positives, and 10.92% of incidents that could not be conclusively categorized.

As in previous years, we can calculate the number of incidents relative to our client base. We have grown the client base further and expanded the dataset to include 21.5% more clients. For this increased dataset, we record an average of **13.8 confirmed incidents per month per client** for the past 12 months. This number is significantly lower than the 23.6 confirmed incidents for the same period the previous year. This is due to a decrease in the number of confirmed incidents, combined with an increase in the base, which includes younger, smaller clients.

The number of confirmed incidents per month per client is higher when evaluating only "mature" clients that have been using our CyberSOC service for the past 3 years or more.

Incidents per Month per Client

Detection Efficiency for Clients Older Than 36 Months Over Time



The chart above explains the changes we are seeing by comparing the incidents for "loyal" customers who have been with us for 36 months or more. The chart shows clearly how the total number of incidents has grown as a result of heightened activity and improved detections, while the number of "confirmed incidents" has decreased as triage and analysis processes have improved.

In our Security Navigator 2024 research piece titled "Fake news and false positives", we pointed out that over time there are detection efficiency gains as the relationship between us and our clients grows and matures. Improved feedback from the client in response to incidents helps us tune technology and processes and boosts the overall confirmed incident rate.

Another notable change this year is that "misuse" as a percentage of threat actions has increased from 16.61% to 28.27% and thus almost matches hacking as a threat action. The VERIS framework allows us to link the threat actor, threat action and impacted asset. With this perspective we observe the internal source of misuse associated with end-user assets, which points to staff violating acceptable usage or other policies that depend on user discretion rather than technical enforcement.

From 2022 to 2024, "hacking" represented between 25% and 31% of total threat actions. This year it dipped slightly to 29.05%, just ahead of misuse.

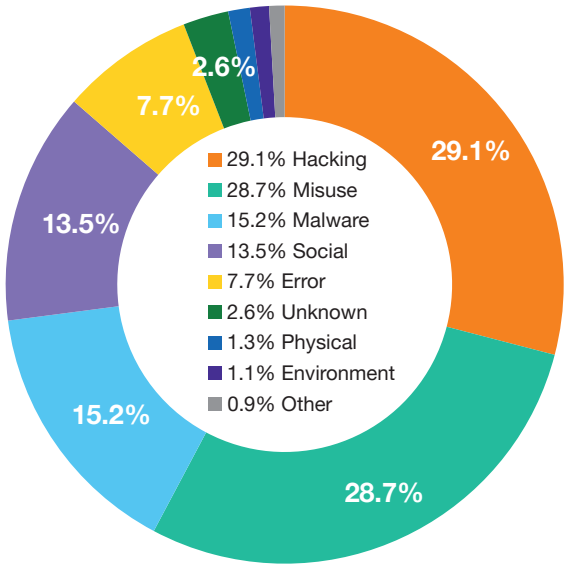
We get a better understanding of what may be driving this shift by zooming into the threat actions. Four of the top five positions are occupied by the same actions as the previous year, while brute force (hacking) replaces physical access.

The unapproved (misuse) threat action increased from 14.29% in the previous year to 24.88%. Phishing (social) ranks third with 13.15% and has increased from 7.89%.

Brute force (hacking) ranks fifth and has increased almost three percentage points to 6.75%. Both web attack (hacking) and port scan (hacking) decreased marginally to give way to the other threat actions.

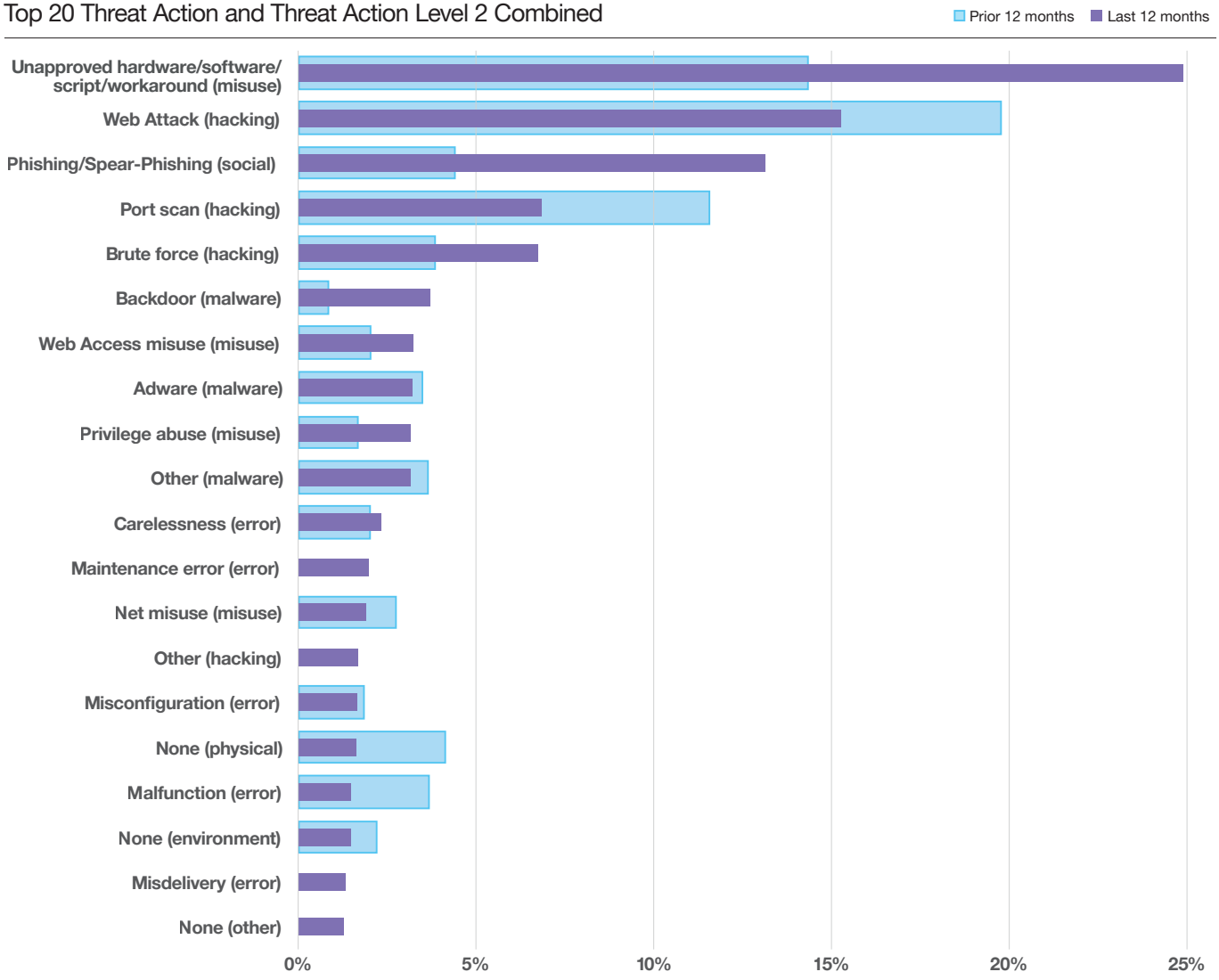
Incidents by Threat Action

True Positive Incidents by Threat Action



Threat Action in Detail

Top 20 Threat Action and Threat Action Level 2 Combined



Summary

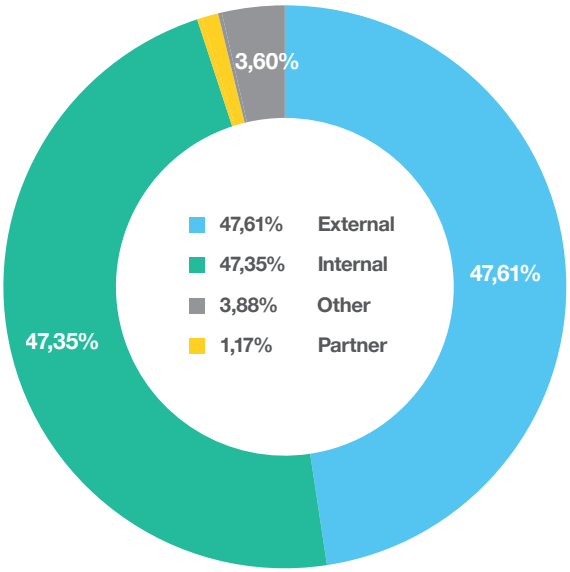
It's interesting to note (again) how many of these incidents stem from misuse, mistakes and negligence. They impact end user devices most frequently, because obviously that is where users operate!

The number of incidents relating to policy violations has increased and the number of incidents stemming from unapproved hardware or software highlights the significant issue of the presence of "shadow IT" in corporate networks. We noted this trend beginning in 2020 during COVID lockdowns, and it appears to have persisted.

In discussions with our clients, CISOs appear to resonate with this observation citing their concern about shadow IT and describing their primary risk as internal. Security has often been identified as the department of 'no', of processes and strict governance. Users working under the radar, as evidenced by these statistics, point to an enduring gap in cyber awareness. Gartner frames this as "cyber judgement"[7]. As analyst Jay Heiser put it: "CISOs and security teams cannot control it all, so pervasive cyber judgement across the organization becomes critical."[8]

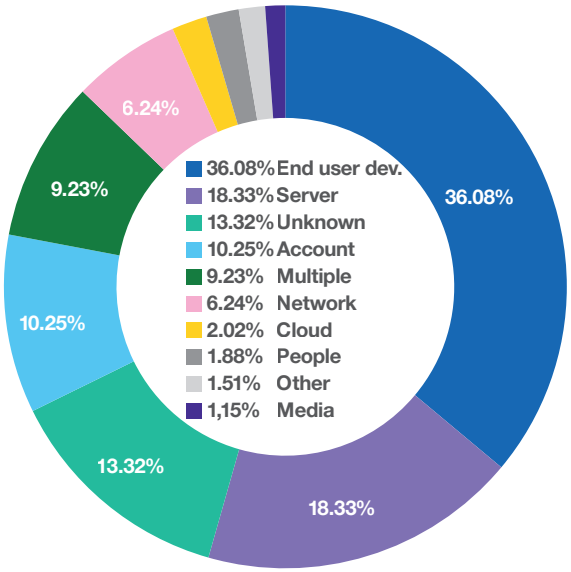
Incident Sources

Distribution of Incidents by Threat Actor



Targets

Distribution of Incidents by Targeted Asset



Sources & Targets

The balance between internal and external sources of incidents has shifted continuously since we started implementing the VERIS classification. In Security Navigator 2023, internal sources (47%) were ahead of external sources (37%). The following year saw external sources leading. This year the two are almost equal, with internal sources associated with 47.35% of confirmed incidents and external responsible for 47.61%. Both have increased their share from 37.45% and 43.6% respectively, with internal sources increasing the most. Distribution has noticeably shifted from servers toward end user devices since last year, probably inline with the increase in the "misuse" category.

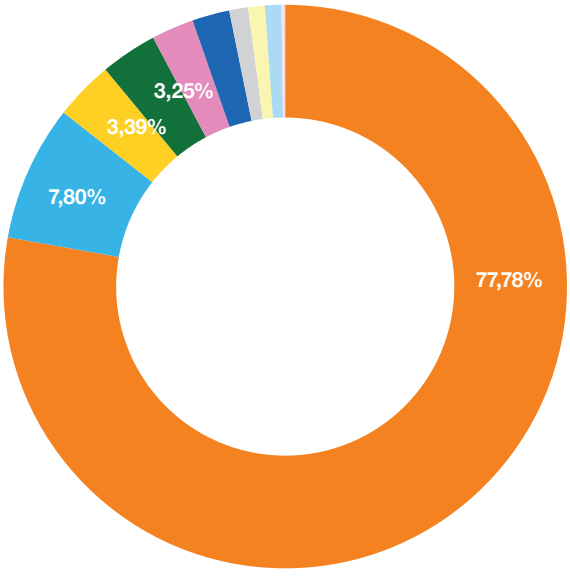
False Positives

The majority of potential incidents that are eventually categorized as benign stem from the misclassification of legitimate user activity. This is not technically an error by the security systems, but stems from the inherent challenge of differentiating legitimate and benign activity in complex environments. As our awareness of the "insider threat" grows, and attackers increasingly employ "living off the land" (LOLbin) techniques, the difference between benign and malicious activity becomes harder to see.

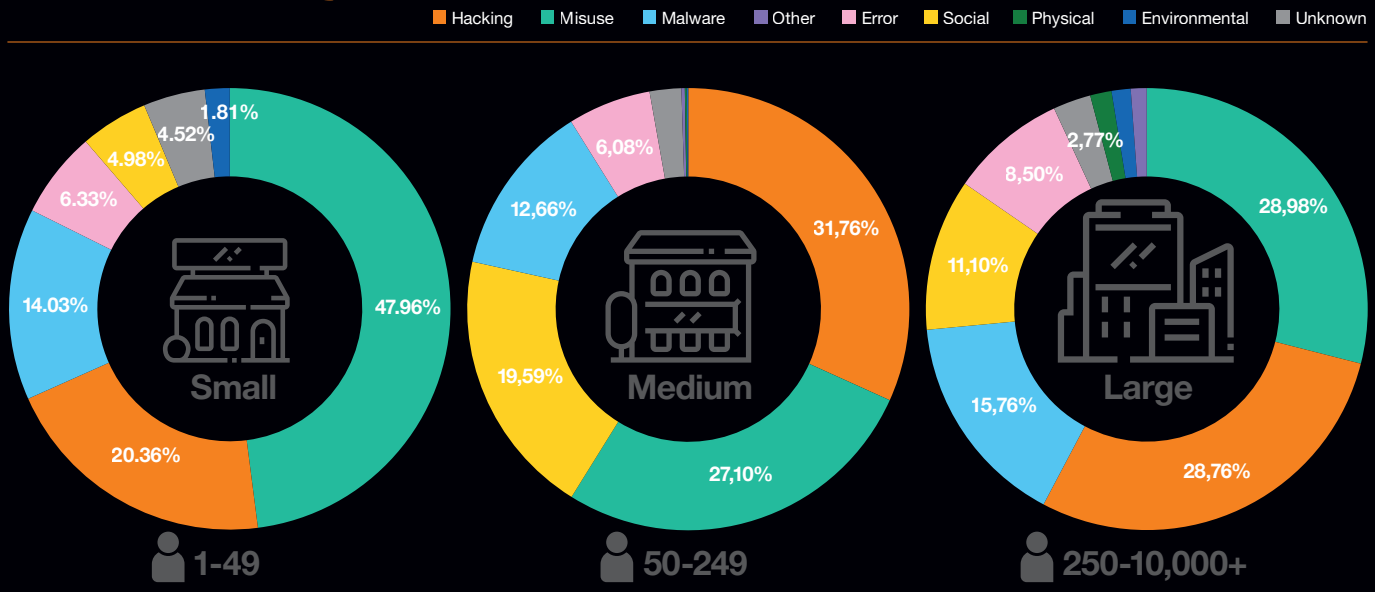
Our CyberSOC teams have responded by increasing the depth and breadth of detections to improve coverage, while improving the processes with which false positive alerts are identified, thus leading to the continued decrease in the proportion of potential incidents that are "confirmed" each year. In our 2024 Security Navigator we illustrated how the proportion of confirmed incidents increases over time as we work with our clients to tune detection mechanisms and improve the feedback loops.

False Positive Types

Incidents That Raised an Alert but Turned Out to Be Harmless



Incidents by Business Size



For Security Navigator 2025, misuse and hacking incidents switched positions with almost exact values compared with Security Navigator 2024.

Incident categories malware, error, and social retained the same positions with slight changes up or down to their respective share of incidents.

Hacking incident types have dropped from 45.81% the previous year to just under 32% for this reporting period. Incidents categorized as misuse and social increased dramatically compared to the figures reported in Security Navigator 2024. Social increased from 6.53% to 19.49%, while misuse rose from 16.32% to 27.10%. Incidents classified as error decreased from 10.38% to 6.08%, while malware incidents increased from 9.11% to 12.66%.

Both misuse and hacking have increased their share for three years in a row, most significantly misuse (21.06% to 28.98%). Hacking increased from 23.53% to 28.76%.

It is unclear why there is such a noted increase of the last three reporting periods. One theory is that monitoring and classification have improved. The sharp drop in incident category "other" from 11.05% to 1.16% may point to that.

Business sizes

Comparing the incidents of different business sizes raises some interesting considerations. For example, every business must actively defend itself against attackers, and small businesses may face the same attackers as big businesses. But big businesses might be expected to have a larger external attack surface. All businesses must also deal with staff who fail to follow policies, but large businesses would have more staff. It would thus seem logical that as a business grows the threats scale proportionally.

But in our data the mix of incidents also changes between small business and large businesses. While the proportion of reported external hacking incidents generally grows with the size of the business, small businesses generally deal with far more internal incidents as a proportion than their larger counterparts. It may be that smaller businesses need to invest more in educating staff on acceptable use policies, or it may be that the increased attack surface of larger businesses contributes to the number of detected incidents much more quickly than an increase in headcount.

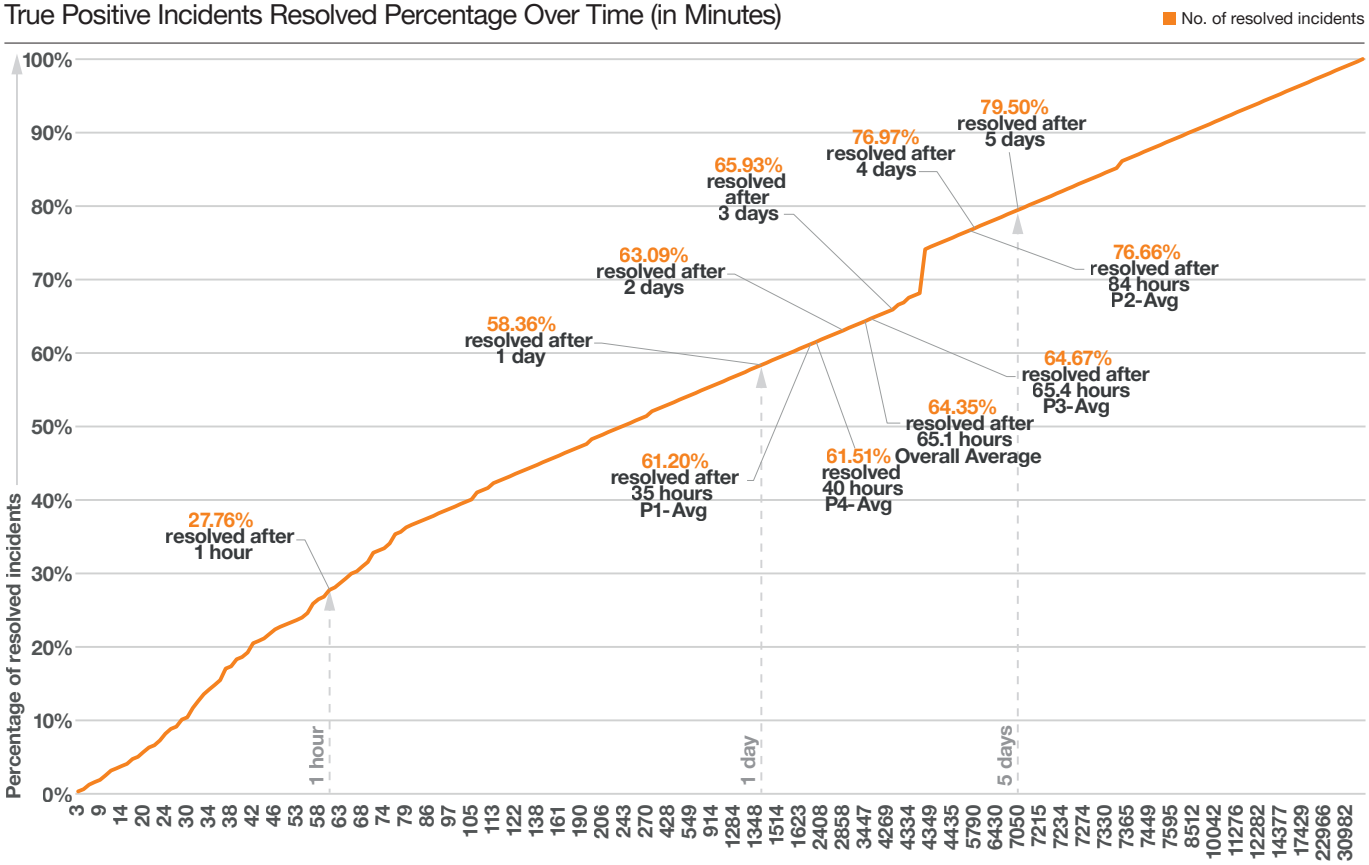
Mean Time to Resolve

This year for the first time we are pleased to include mean time to resolve (MTTR) statistics in this report. In our operation we record the time it takes in minutes from when an alert is raised, through triage, analysis and reporting, to when it can be categorized and closed with the approval of the client. MTTR is a prickly metric and can easily mislead.

We've taken a page from the Cyentia playbook and opted to present our data in the form of a "survival analysis", which is illustrated below[9]. The criticism laid against MTTR is that it can be opaque. Since an uneven distribution of MTTR values, especially those on a "long tail", can easily skew the mean, it must be expressed in a transparent manner. Using "survival analysis" goes beyond the mean and median and allows us to present a full and transparent view of MTTR performance.

Mean Time to Resolve

True Positive Incidents Resolved Percentage Over Time (in Minutes)



Summary:

- 27.6% of True Positive incidents are confirmed and resolved within an hour of being raised.
- 58.36% are confirmed and resolved within a day.
- On average, Priority 1 incidents are confirmed and resolved 35 hours after the initial alert was received. Bear in mind that the incident priority can only be determined during the course of the investigation and is confirmed when the incident is closed.
- 79.5% of incidents are confirmed and resolved within 5 days.
- At the end of the long tail, there are incidents that are only confirmed and resolved after 35 days.

Mean Time To Resolve (MTTR) can be a complex metric to interpret. Resolving an issue involves detecting, analyzing, and reporting it to the client, who then investigates, takes action, and confirms the incident. This multi-step process adds time but ensures reliable detection, effective security outcomes, and honest data. By introducing this KPI, we enable benchmarking (as shown in this report), offering a reference for comparing MTTR with peers. However, faster isn't always better; while automation opportunities exist, we must first establish effective processes and baselines to measure improvements meaningfully. Without data and reference points, discussions on incident response efficiency lack a starting ground.

Vulnerability Scanning

The Orange Cyberdefense managed vulnerability scanning service is delivered by our vulnerability operations centers (VOC) worldwide. We are pleased to share that this year we are able to include an additional vulnerability operations center (VOC) to our dataset, doubling the number of VOCs contributing. This addition increases the scope and range of unique assets, geographies, and industries, and the total number of unique assets increased 2.72 times as a result. Unfortunately, the addition of new assets will influence or distort historical patterns. A pure like for like analysis is further hampered due to the partitioning and anonymization of entities in the data. Note also that each environment is different, as is each business, and what is true for one business may not hold for another, even in the same industry in another region.

The other chapter in this report on vulnerability research - titled "Beyond vulnerability management" – is complimentary to this one, and we urge you to consider that in combination with our analysis of the VOC data here.

Findings by Severity

Before we start, we need to clarify some terminology. We will use "unique assets" and "unique findings" throughout this section. Unique findings are always associated with an asset and the unique asset is defined in terms of the client.

Unique assets are defined in terms of:

- Client
- Asset Name
- IP Address
- Host Type

A unique finding is defined in terms of a unique asset, with the addition of the 'Finding Name' assigned by the scanning engine.

Our VOC dataset consists of **68,509 unique assets**, with **1,337,797 unique findings**.

The average finding per host is lower across all severities. Most notably, the high severity findings that previously averaged 21.93 per asset are down to 11.14 in this extended dataset. Similarly, the average number of critical findings decreased almost by half from 7.05 previously to 3.72 now.

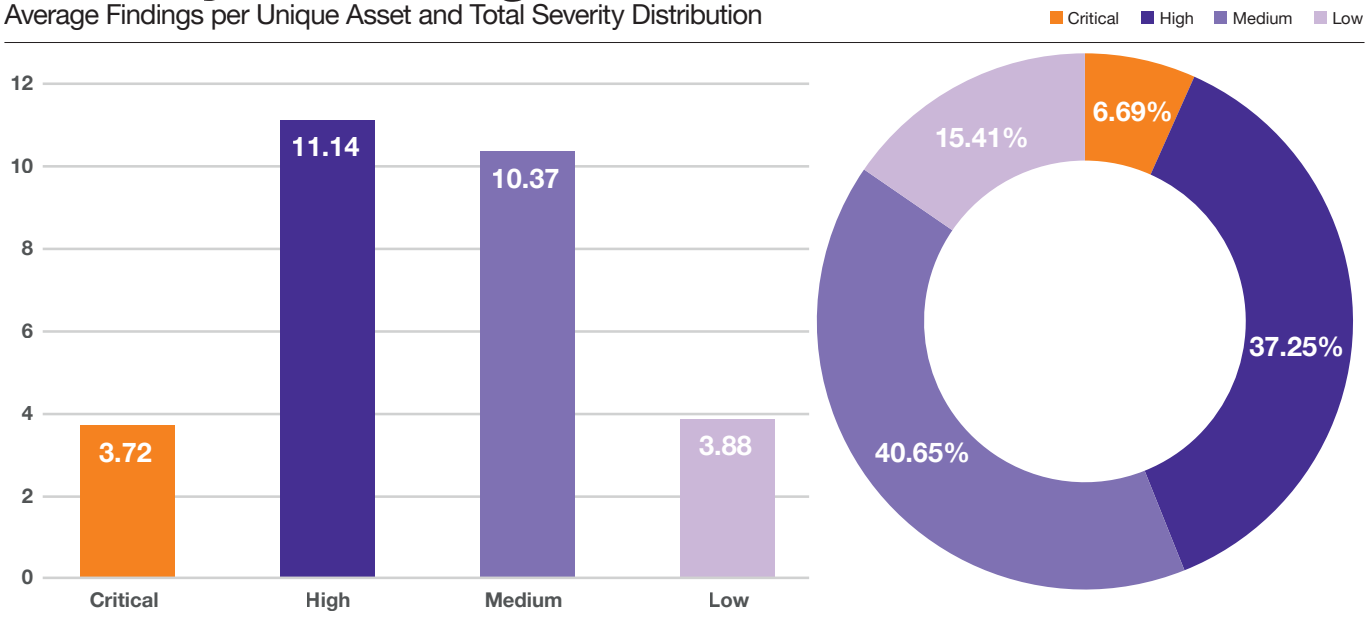
We welcome this apparently rosier outlook, but bear in mind that the additional assets distort these figures, so this should be seen as new perspective, rather than an "improvement".

The distribution of severity level across findings has changed less dramatically than the average severity. Severities "medium" and "high" swapped places, with medium – now ranked first - increasing from 38.4% to 40.65%.

Meanwhile high severity findings, now ranked second, decreased their proportion from 41% to 37.25%. The share of low and critical issues occupied the same rankings at third and fourth respectively. While the share of findings rated low increased from 11.2% to 15.4%, the share of critical rated findings declined from 9.4% to 6.69%. These proportions are across all findings.

Severity of Findings

Average Findings per Unique Asset and Total Severity Distribution



Readers with good memories will spot the increase in this year's maximum age and the increase in the overall average age of vulnerabilities. The extreme maximum age is attributed to findings associated with assets from specific clients in the Retail Trade industry. This eccentricity is due to one client whose existing vulnerability scanning records were included when they were onboarded to our service, thus skewing the curve. Excluding this client from the dataset lowers the maximum age for all severity types to between 1809 and 1855 days, or 5 years. In the previous Security Navigator, we reported a maximum age between 1441 and 1486 days. This age is somewhat arbitrary, however, since it generally simply reflects the time elapsed since we started scanning those assets. These old vulnerabilities just keep getting older, in other words.

Removing "retail & trade" clients from the mix lowers the maximum age, but it remains concerning that these vulnerabilities have "survived" for yet another year. The average age across all severities is actually slightly lower in this year's dataset, suggesting that our clients in the retail & trade have a particular challenge with eliminating some vulnerabilities.

The ratio between the medium and low severity findings is similarly spaced for this year and last regarding maximum age. The ratio for critical to medium and high to medium is slightly better for this year than before.

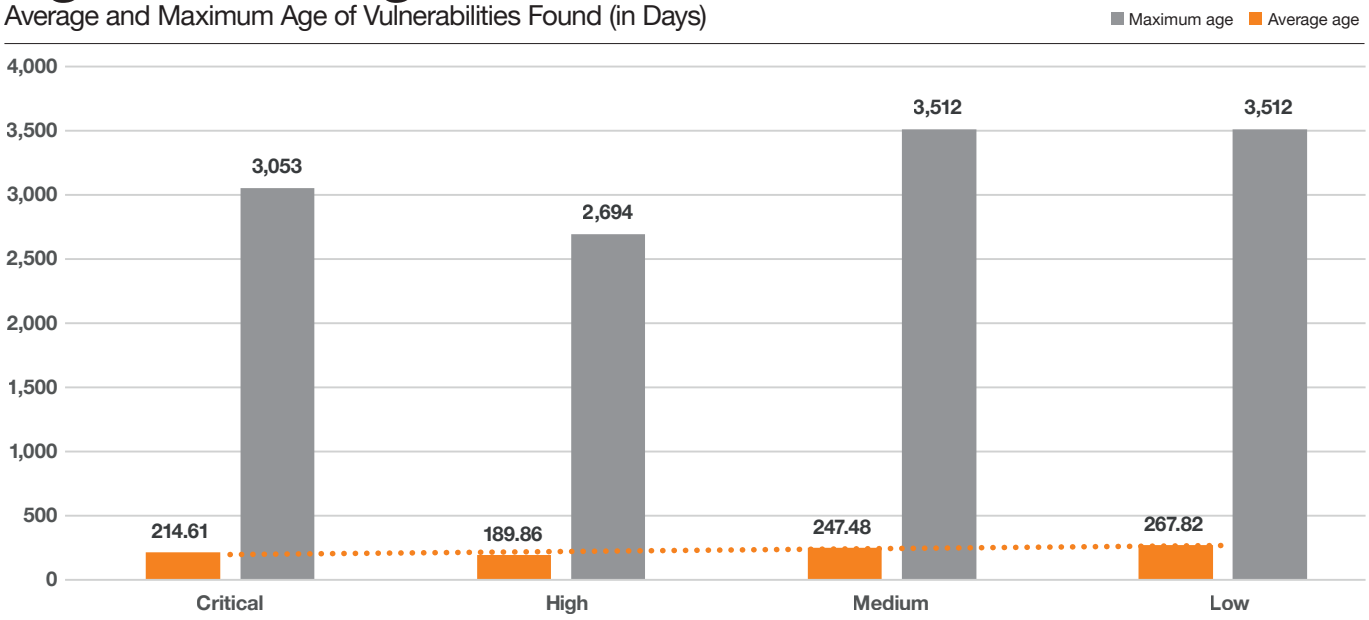
The average age across all findings is higher, most noticeably for critical and high severity findings. In both these cases, the average age of findings is more than double the previous dataset. The average age in days of critical rated findings increases from 88 to 215, and the average age in days for high severity findings increases from 82 to 189.86. These numbers are opaque as they only speak to what we observe in the environments we scan and are not a reflection on Orange Cyberdefense's service levels on patch management.

The average age of medium and low severity findings is higher, from 185 to 247.48 and 208 to 267.82.

The expansion of our dataset with the inclusion of a second VOC exposes the long tail of vulnerabilities that persist without remediation. This, beyond just the 162 day median age for all findings, skews the distribution.

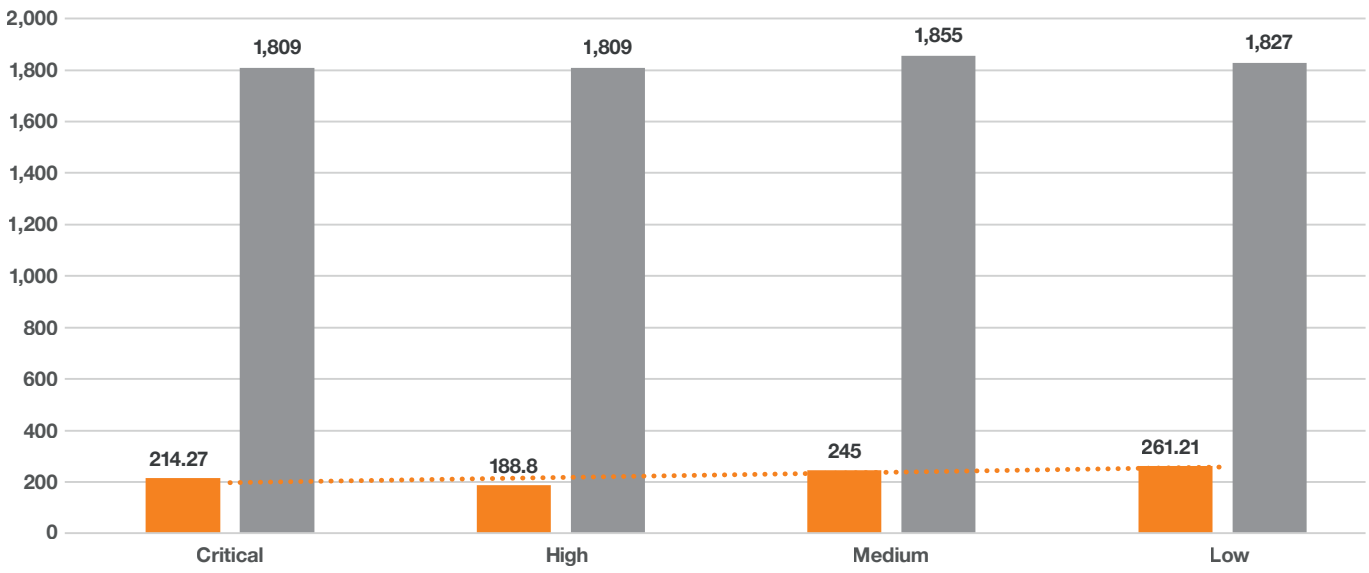
Age of Findings

Average and Maximum Age of Vulnerabilities Found (in Days)



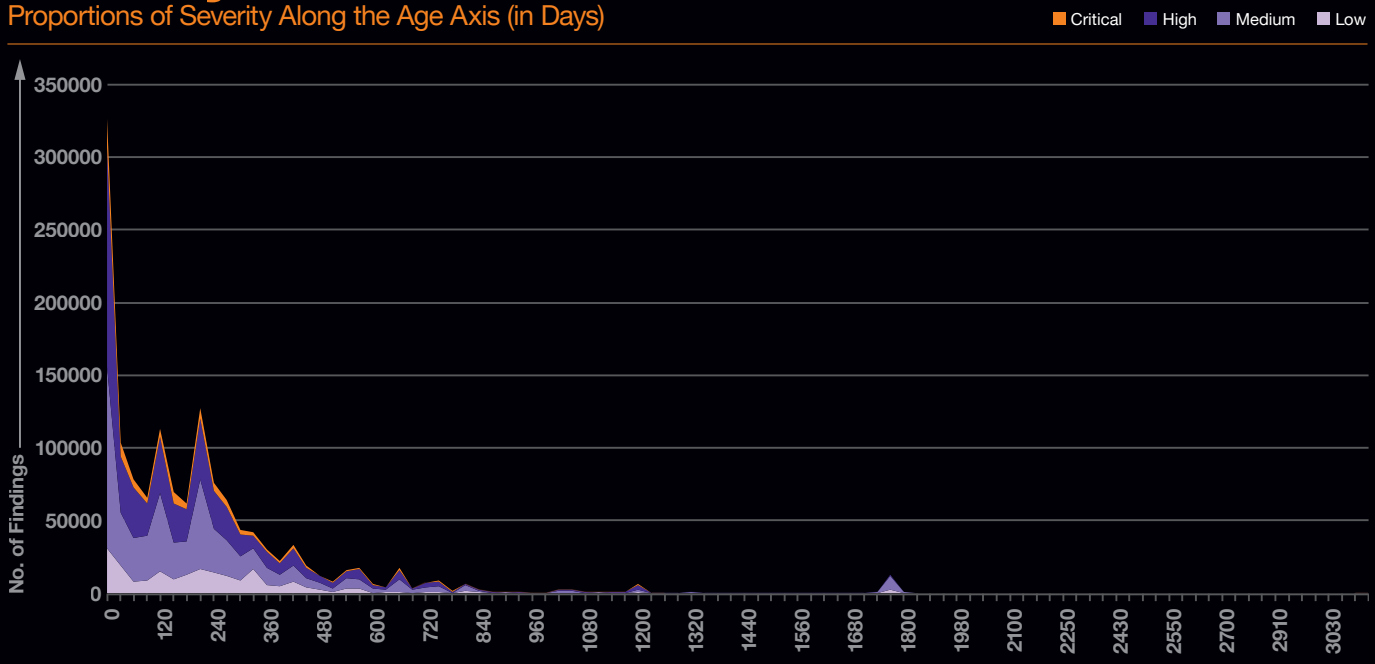
Age of Findings

Average and Maximum Age of Vulnerabilities Found (in Days), Excluding Retail & Trade



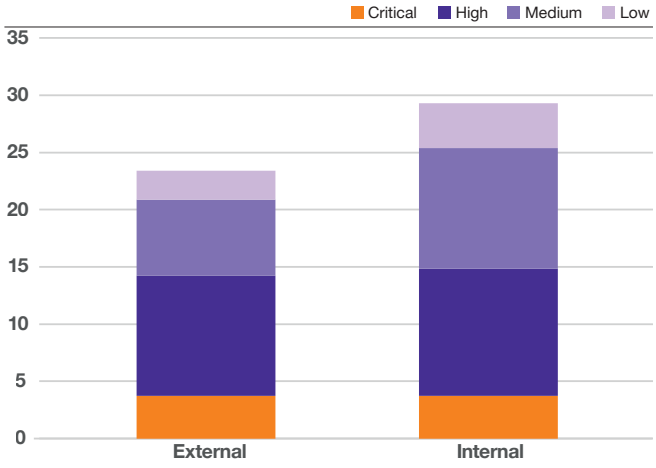
Severity Over Time

Proportions of Severity Along the Age Axis (in Days)



The shape of the age-versus-severities chart is somewhat different to last year. The long "tail" depicted by the severities starting at 840 days (about 2 and a half years) is now very evident, even if it is concentrated in one industry. Also, the "body" of the distribution has bulked up at the median age, balancing the volume at 162 days (about 5 and a half months). This illustration also shows that the "meat" of unpatched findings consists primarily of medium findings.

Finding Severity by Target Exposure

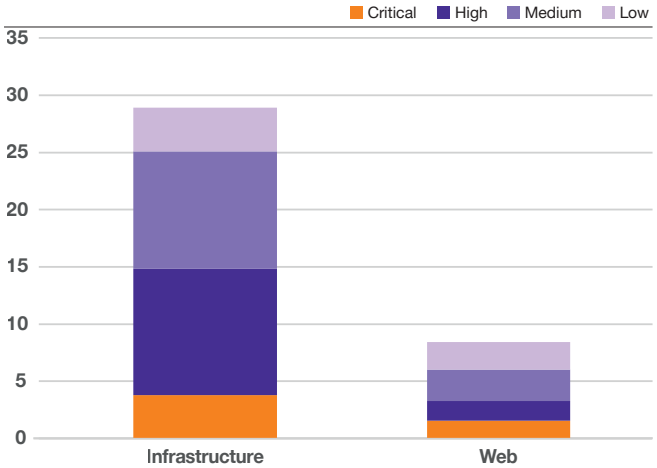


An eye-catching feature in this year's data is how many more high severity findings we see on external (internet) facing assets. The average number of high severity findings on external hosts is 10.5 in this year's data, compared to 2.83 before. The average number of critical, medium, and low findings per unique asset is also higher, most notably for critical.

Compared to last year, the average number of findings on Internal is lower overall, across all severities. Critical, high and low rated severities are almost as common as for external assets. Medium rated severities on average are more common on Internal assets, however.

Findings for assets grouped under internal are 21 points lower than before, whereas the average unique findings for assets under external are 6 points higher.

Finding Severity by Target Type



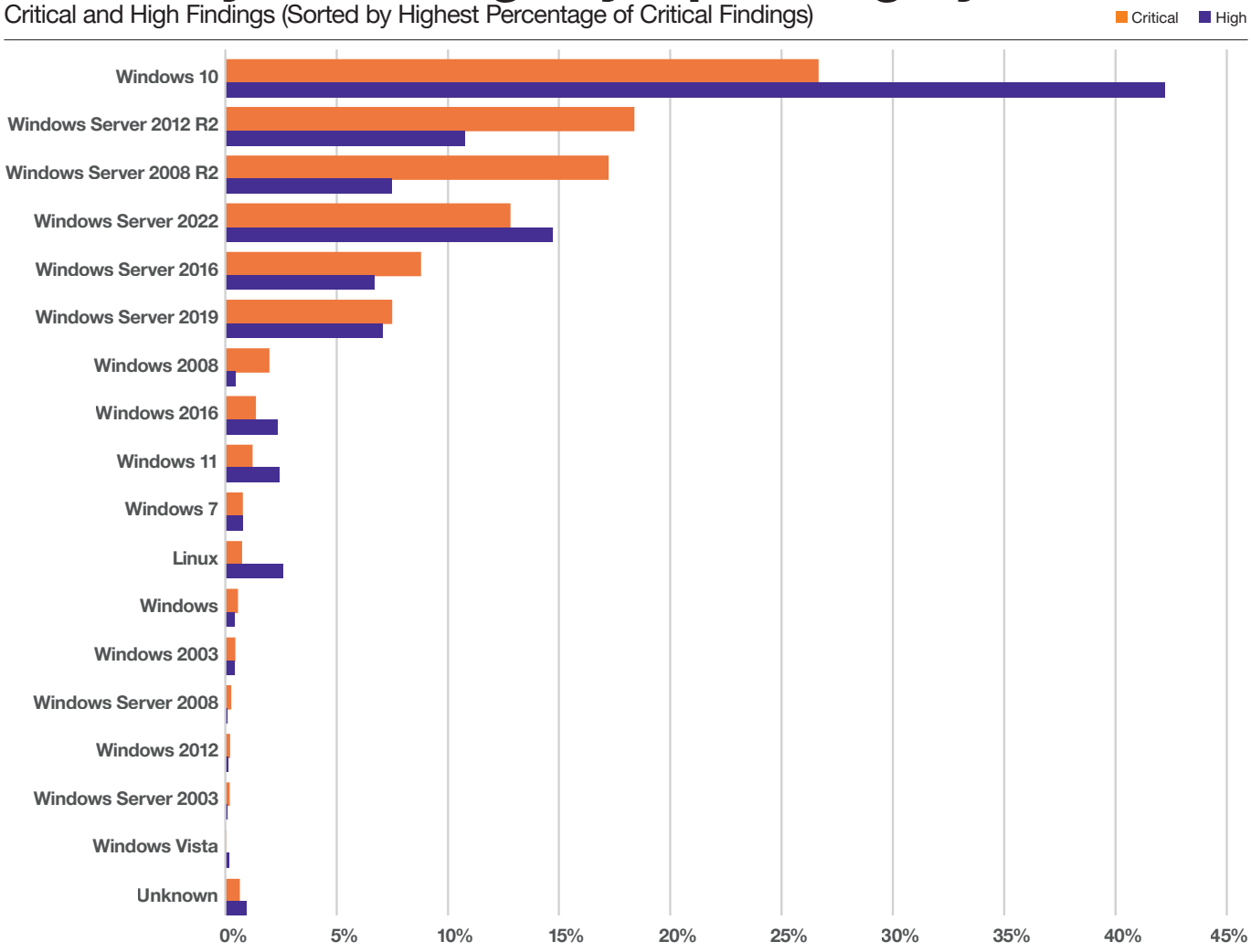
In this comparison we examine assets that are accessible through the web browser (web) versus non-web assets (infrastructure). As with our previous analysis, the contrast is clear and there is a similar trend. Our clients are dealing with far fewer unique vulnerabilities on web assets than on infrastructure, desktops and servers.

Both infrastructure and web are 20 points lower compared to the previous year. Examining the severity ratios for the web category reveals that there are fewer critical severities as a proportion this year, but proportionally more findings rated high. Comparing ratios on infrastructure to the previous year shows that the proportion of high severity findings is lower this year, aligned with the medium severity findings now.

The expanded VOC dataset has a lower level of average findings for both internal and web groupings. As cautioned earlier, however, it would be too soon to celebrate this as a win.

Criticality of Findings by Operating System

Critical and High Findings (Sorted by Highest Percentage of Critical Findings)



Findings by Operating System

The conversation around software quality and how that relates to software vulnerabilities has been put in the spotlight in 2024, specifically around topics such as “secure by design” and “security debt”[10][11][12]. These topics are touched on in our research chapter titled “Beyond Vulnerability Management”.

We can dip briefly into this topic by examining which operating system (OS) ranks the most prominent in our VOC dataset regarding number of vulnerabilities. This is also useful for determining how the introduction of additional unique assets may have influenced the ranking compared to our previous examination. Spoiler alert - not much changed!

One aspect of the “secure by design” best practice guidelines is memory safety, such as using programming languages that eliminate certain classes of vulnerabilities as well as other defensive programming techniques.

How does this relate to vulnerability characteristics associated with Windows 10, which accounts for the majority of high and critical vulnerabilities in our dataset?

First, we identify all the unique common vulnerability enumerations (CVEs) identified by our VOC on assets running Windows 10. Next, we examine the associated common weakness enumeration (CWE) assigned to these CVEs[13]. A CWE is a class of software or hardware weakness that could be exploited by an attacker. CWEs are rather technical and rich in annotation and are represented by a hierarchy of cascading technical specifics.

Finally, we map each CWE to the topmost abstract CWE class. In the case of Windows 10 the two most prominent CWEs point to resource mismanagement (CWE-707 and CWE-664)[14][15]. I.e. weaknesses in how software is handling memory during (CWE-787) and after (CWE-416) use.

- **CWE-707, 'Improper Neutralization',** is a top level CWE abstraction and occurs when a product handles malformed input that corrupts memory in a way that benefits the attacker and could possibly lead to security violations.
  - **CWE-787, 'Out-of-bounds Write',** is a specialization of CWE-707 that is caused by improper bounds checking when the product is writing data to memory, causing corruption that can lead to further security violation such as malicious code execution.
- **CWE-664, 'Improper Control of a Resource Through its Lifetime',** is a top level abstraction associated with mismanagement of resources such as memory.
  - **CWE-416, 'Use after Free',** is a specialization of CWE-664 and is a programming fault wherein the product incorrectly interacts with memory that it explicitly marked as unused resulting in potential security violations such as malicious code execution.

Eliminating these kinds of vulnerabilities is tough and probably requires substantial redesign and rewriting of code. If by some miracle Microsoft could hypothetically eliminate all Windows 10 vulnerabilities classified as either CWE-787 or CWE-416 then our VOC data set will shrink by 3,974 CVEs.

To continue the hypothetical experiment, let's assume we can eliminate all vulnerabilities classified under CWE-707 and CWE-664. This action will eliminate 13,596 vulnerabilities associated with Windows 10 from our VOC dataset, and by extension other versions of the Microsoft operating system that shares code with it.

Weaknesses in Win 10

Most Prominent Common Weakness Enumeration

Conclusion

Vendors must strive to continuously improve their product design, development, and quality assurance processes to actively seek out these classes of vulnerabilities.

A cultural shift is required to ensure usage of software development best practices. It comes down to a combination of defensive programming, explicit fault finding through test cases and code coverage, formal code reviews, static and dynamic code testing, and more.

Introducing memory safe programming languages could potentially also eliminate many of the problems.

Cyber Extortion

Cyber extortion, or "Cy-X" is a form of computer crime in which the security of a corporate digital asset (confidentiality, integrity or availability) is compromised and exploited in a threat of some form to extort a payment. Cy-X groups compromise, name, shame and extort victims via dedicated data leak sites on the dark web, which we can track. Since last year's report, we have added 40 unique leak sites to our tracking.

Since January 2020, we have recorded 13,308 victim organizations exposed on leak sites. These leaks are from 141 distinct Cy-X brands.

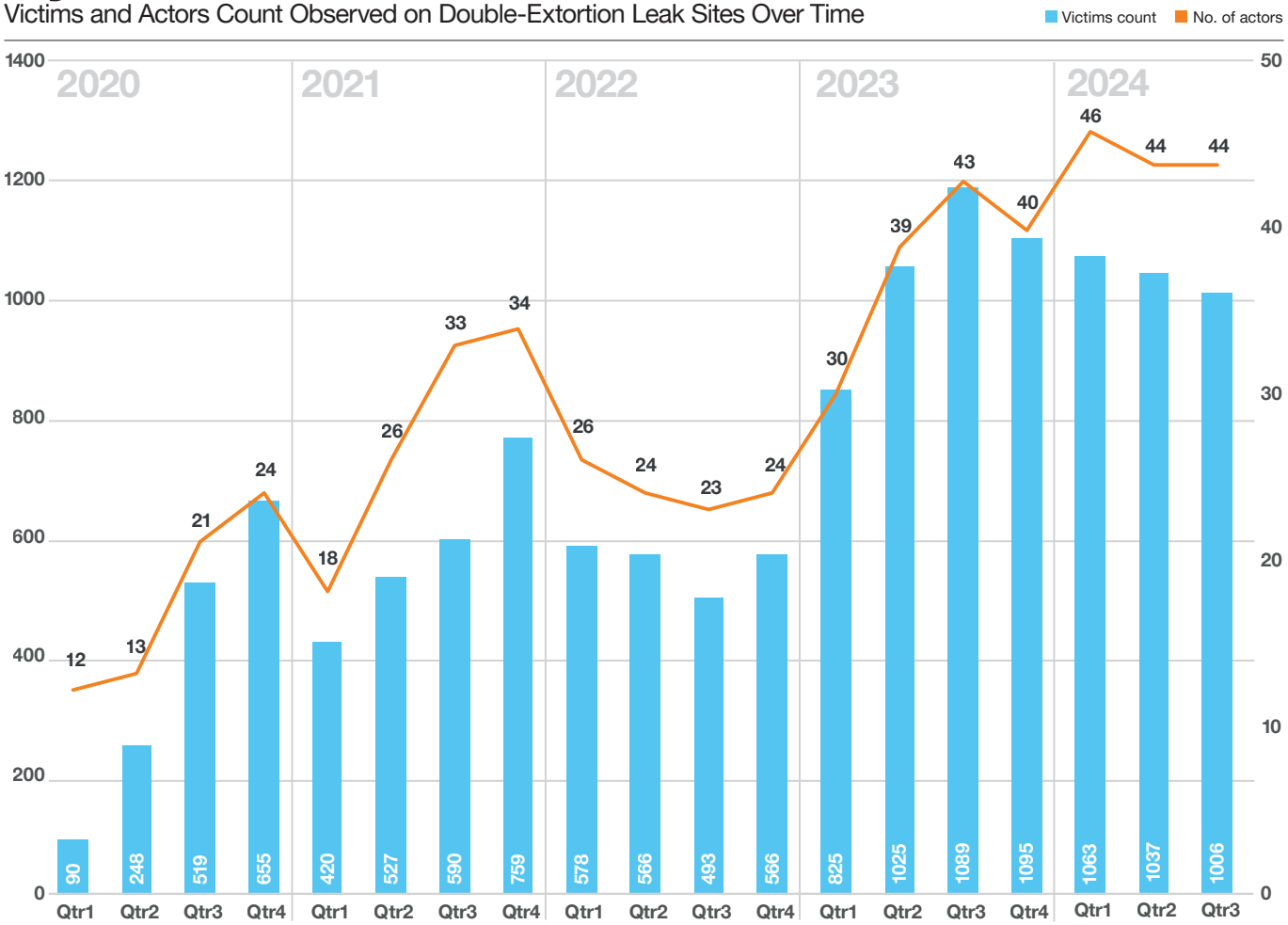
In the past 12 months, we documented 4,201 Cy-X victims. This is an increase of 15.29% since we published the Security Navigator 2024. In 2022 we observed a decrease in victim volumes as major Cy-X brands were apparently distracted by the first year of the war against Ukraine. Activity accelerated dramatically as the threat actors regrouped, and the volume of victims appears to be "normalizing" since then.

Summary

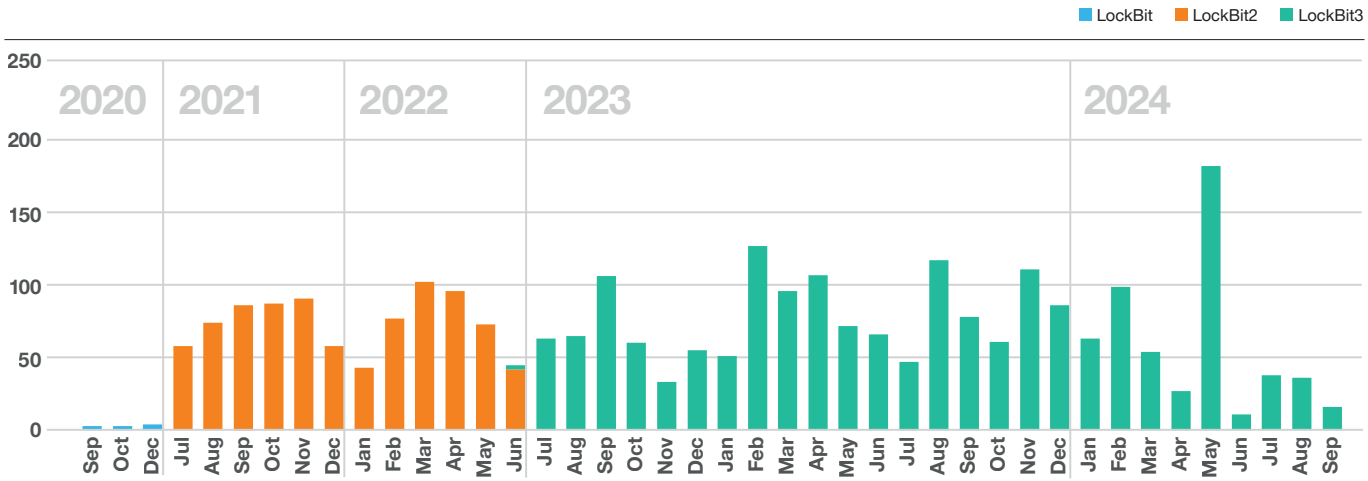
A noteworthy observation is that for the first time since 2020, the distinct actor count is not directly correlated with the victim count. Up until 2023, we could argue that the number of victims tracked the number of actors engaging in this form of crime. This might be changing, as Q1 2024 recorded the largest number of actors we've seen so far (46) but not proportionally more victims. While we tracked an increase in active actors, we actually observed a slight decrease in victims.

Cy-X Over Time

Victims and Actors Count Observed on Double-Extortion Leak Sites Over Time



Lockbit Activity Over time



A Criminal Career Must End at Some Point

One potential explanation for the slowdown in victim count could be the continuous efforts of law enforcement to take down LockBit - one of the most active Cy-X brands ever – that has been active since 2019.

In February 2024, it finally happened. Law enforcement released an announcement of their coordinated effort to take down LockBit, which was dubbed 'Operation Cronos'[16][17][18]. The Cronos operation was a major Europol-led initiative focused on dismantling the high-profile cybercrime network. While Cronos significantly contributed to LockBit's disruption, it did not cause the group to cease activities completely. The operation led to server seizures, the arrest of key actors, and a notable decrease in LockBit's capacity, causing some operations to run at a limited scale.

During the initial waves of Cronos disruption, particularly in May 2024, LockBit sought to project an image of resilience by posting a high volume of alleged victims. However, many of these claims could not be independently verified, raising suspicions that the group was more focused on shaping a narrative of continued strength than conducting actual attack activity. Despite significant setbacks dealt by law enforcement, LockBit has not been completely dismantled and continues to maintain a presence, albeit with diminished capacity.

The impact of operation Cronos likely undermined the trust of LockBit's affiliates and the broader cyber extortion ecosystem. Affiliates may hesitate to collaborate, fearing increased law enforcement scrutiny or diminished returns. This erosion of trust could lead affiliates to move to other ransomware-as-a-service (RaaS) operations, particularly as several new brands have emerged in late summer.

The Cy-X Recast – Who's Next?

After a major operation like LockBit becomes defunct or slows down, we often see an increase in new brands popping up to fill the void. Since June 2024, therefore, we added 19 new leak sites, 10 of them recorded victims before June 2024 but only became known to us then.

It is difficult to know how new the threat actors really are, as the ecosystem is very flexible, and affiliates can choose to switch between Cy-X brands. In the past 12 months, we have tracked 68 unique threat actor leak sites actively extorting victims. This shows an increase of 26% since last year's report.

For those who monitor the Cy-X / ransomware space, it feels as if there are new leak sites and brands every week. In the section below, we explore what we've seen in actor activity over the past 12 months.

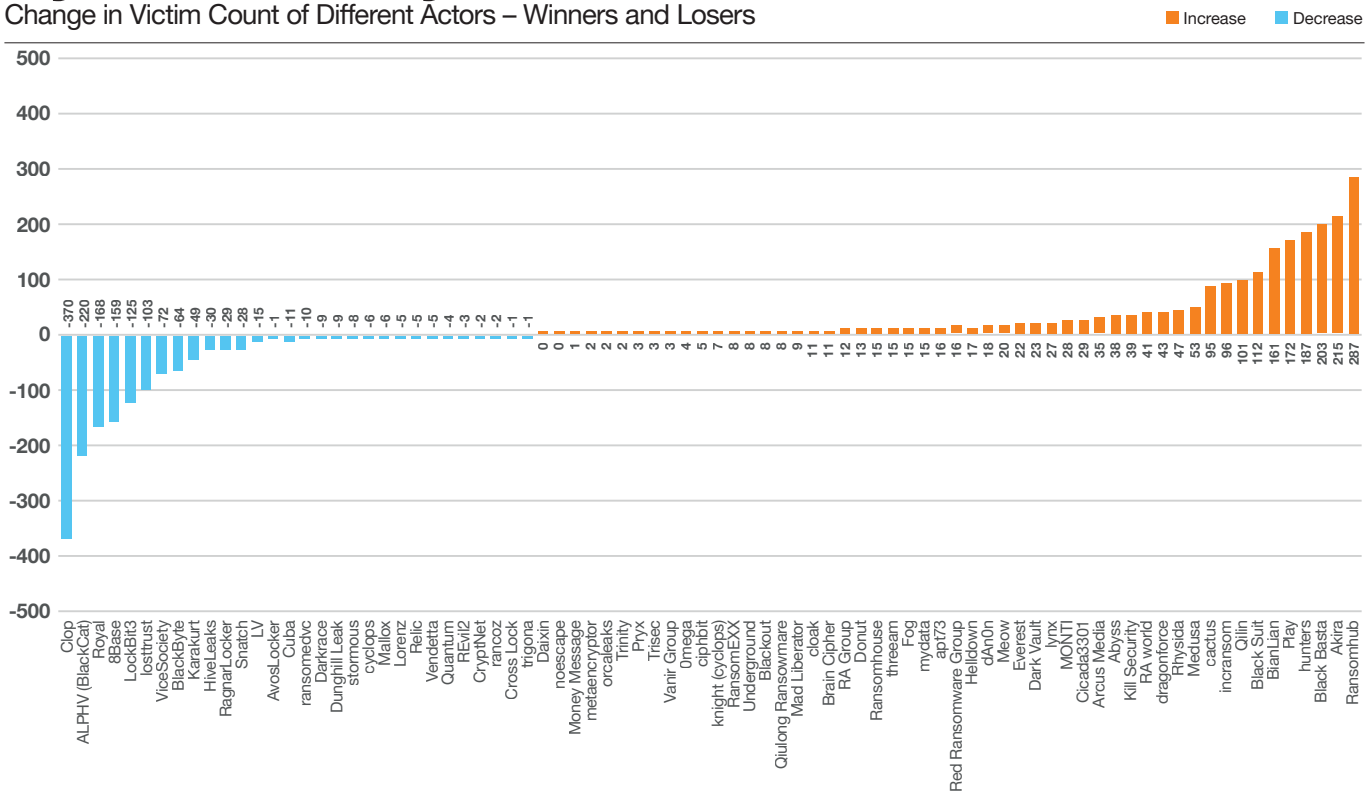
Summary

The Cy-X threat landscape has seen significant shifts in the past 12 months, with some of the most notorious groups declining while new actors emerge rapidly.

Law enforcement disruptions may have contributed to the declines, but the rapid emergence of new groups underscores the persistent and evolving nature of this highly volatile ecosystem.

Cy-X Victims by Actor

Change in Victim Count of Different Actors – Winners and Losers



As expected, a few Cy-X groups have ceased drastically or disappeared entirely. We track this as “significant decrease in activity”. This group includes major Cy-X brands like Cl0p, who’s victim count dropped by 377 after being highly active in 2023. It might be that they are still benefiting financially from last year’s mass exploitation campaigns. ALPHV (BlackCat) ceased operations entirely following an attempted law enforcement disruption attempt and a subsequent exit scam. The threat actor Royal rebranded as BlackSuit, and we have already discussed LockBit.

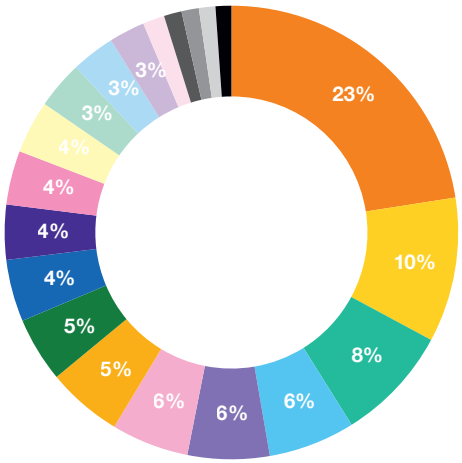
In contrast to the groups experiencing declines, several Cy-X groups have surged in activity over the past year. Ransomhub recorded the largest increase, with 287 incidents in 2024 from being inactive in 2023.

Similarly, Akira emerged from the lower ranks last year to become one of the most active groups of 2024, with 215 incidents reported. Black Basta also saw substantial growth, rapidly accelerating its activity over the past 12 months. Other notable risers include Hunters - which reported 187 incidents after a period of inactivity - and Play, which expanded from 187 incidents in 2023 to 359 in 2024.

Other groups with significant increases include BianLian (+161), Qilin (+101), Black Suit (+112), incransom (+96), Medusa (+53), and Rhysida (+47), illustrating the emergence of new and reactivated actors in the ransomware landscape.

Top 20 Actors in the Past 12 Months

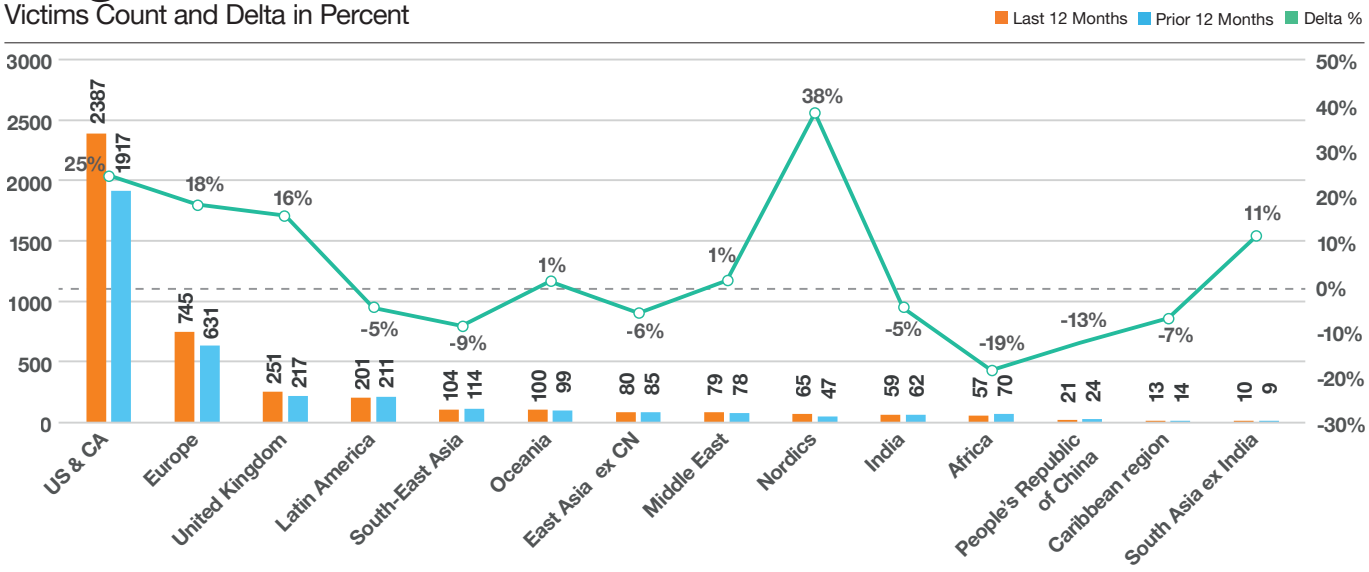
The Most Active Extortion Groups Observed



23%	Lockbit3	4%	cactus
10%	Play	4%	ALPHV (BlackCat)
8%	Ransomhub	3%	Black Suit
6%	Akira	3%	incransom
6%	Black Basta	3%	Rhysidia
6%	8Base	3%	noescape
5%	hunters	2%	MONTI
5%	BianLian	1%	dragonforce
4%	Medusa	1%	RA World
4%	Qilin	1%	Kill Security

Regional Shift in Victim Count

Victims Count and Delta in Percent



North America and Europe remain the most heavily impacted regions. The U.S. remains the most impacted country, which aligns with its position as a global economic and technological hub. Generally, we don't see the steep growth rates we've reported previously. We believe this is because last year's report documented the resurgence of this crime after geopolitical events in 2022 disrupted the Cy-X ecosystem temporarily.

In Europe we see France, Italy, Germany, Spain and the Netherlands impacted the most. The Nordic region (including Sweden, Denmark, Norway and Finland, Iceland and Greenland) has seen the highest growth in the past 12 months, although the count of victims is still low relative to other regions.

Noteworthy is the observed decrease in victim numbers for regions like South East Asia (SEA), East Asia (excluding China), India, Africa, China and the Caribbean.

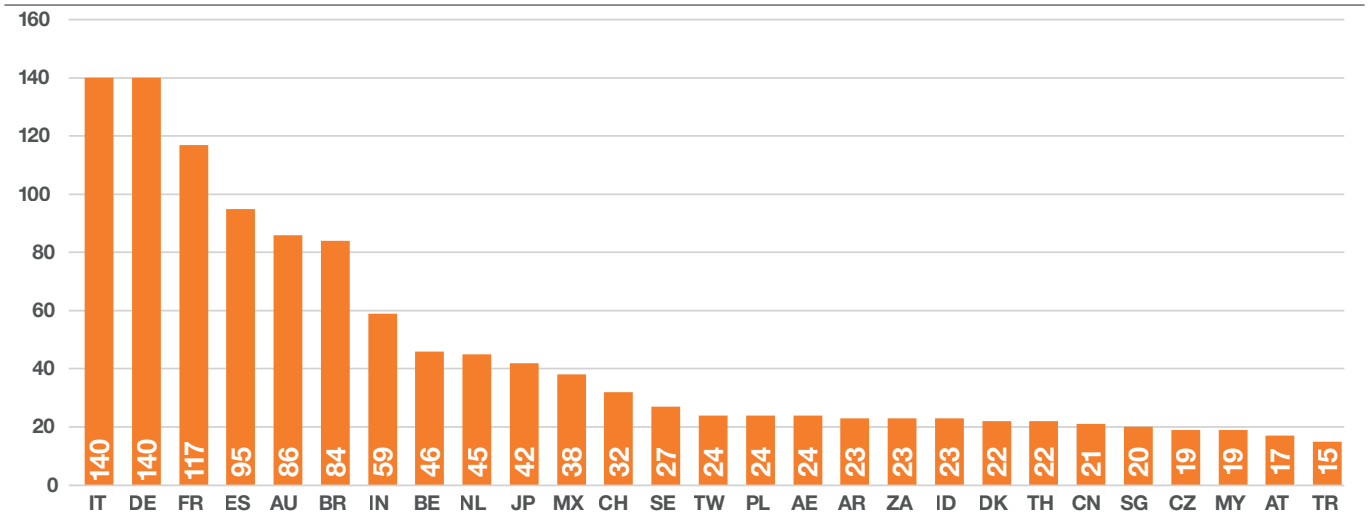
As we have reported in the past, we note that large English-speaking regions feature prominently in our victim dataset. We present a country breakdown, excluding United States, Canada and Great Britain, in the graphic below.

Over the past 12 months, Italy and Germany are the most impacted countries when excluding the "big 3", followed by France, Spain, and Australia. This dynamic highlights the wide spread of victims across diverse regions, reinforcing our findings from previous years that cyber extortion and ransomware have become truly global threats. The diversity in affected countries underscores the increasingly indiscriminate and global nature of the cyber extortion phenomenon.

In total we observed victims in 116 unique countries over the past 12 months, which equates to about 60% of the world. Countries we recorded for the first time in our victim data were: Afghanistan (Central Asia), Jersey (Europe), Djibouti (Africa), Georgia (West Asia), Timor-Leste (SEA), Myanmar (SEA), Tokelau (Oceania), Nepal (South Asia ex India), Sudan (Africa), Saint Vincent and the Grenadines (Caribbean region), Curaçao (Caribbean region), Palau (Oceania), Sierra Leone (Africa), Uzbekistan (Central Asia), Maldives (South Asia ex India), Niger (Africa), and Cuba (Caribbean region).

Top 30 countries

Excluding US, CA, GB



Business Size

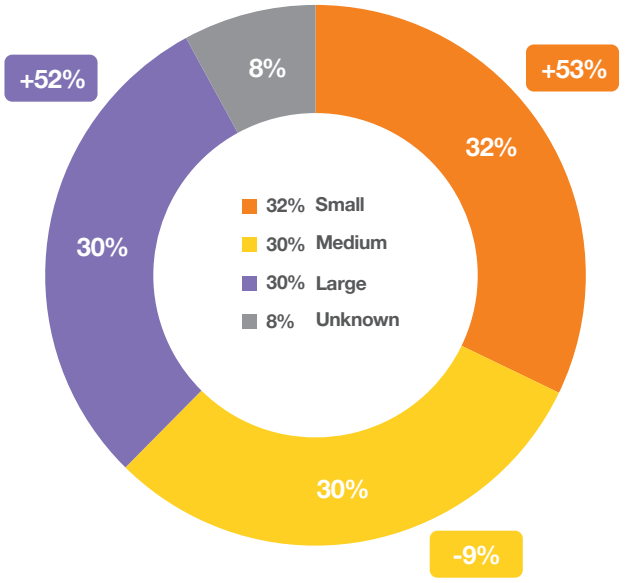
Organizations of all sizes have been affected by Cy-X attacks over the past 12 months. In this analysis, business size is classified according to the OECD standard: Small businesses are defined as those with 1-49 employees, medium-sized businesses range from 50 to 249 employees, and large organizations have 250 or more employees.

The distribution of impacted organizations across size is relatively balanced, with small businesses accounting for 32% of affected entities, followed closely by large organizations and medium-sized businesses, each representing 30%.

Compared to the previous year's data, we've recorded a substantial increase of 53% in small businesses victims. We also witnessed a 52% increase in medium-sized business victims. On the other hand, we recorded 9% fewer victims that could be classified as "large". It's too soon to say, but this shift may indicate that ransomware affiliates are choosing to throw their nets wider, perhaps in response to improved security by larger organizations. Alternatively, perhaps it's simply becoming harder to find large organizations that have not already been compromised. This is a trend worth watching.

Victim Size

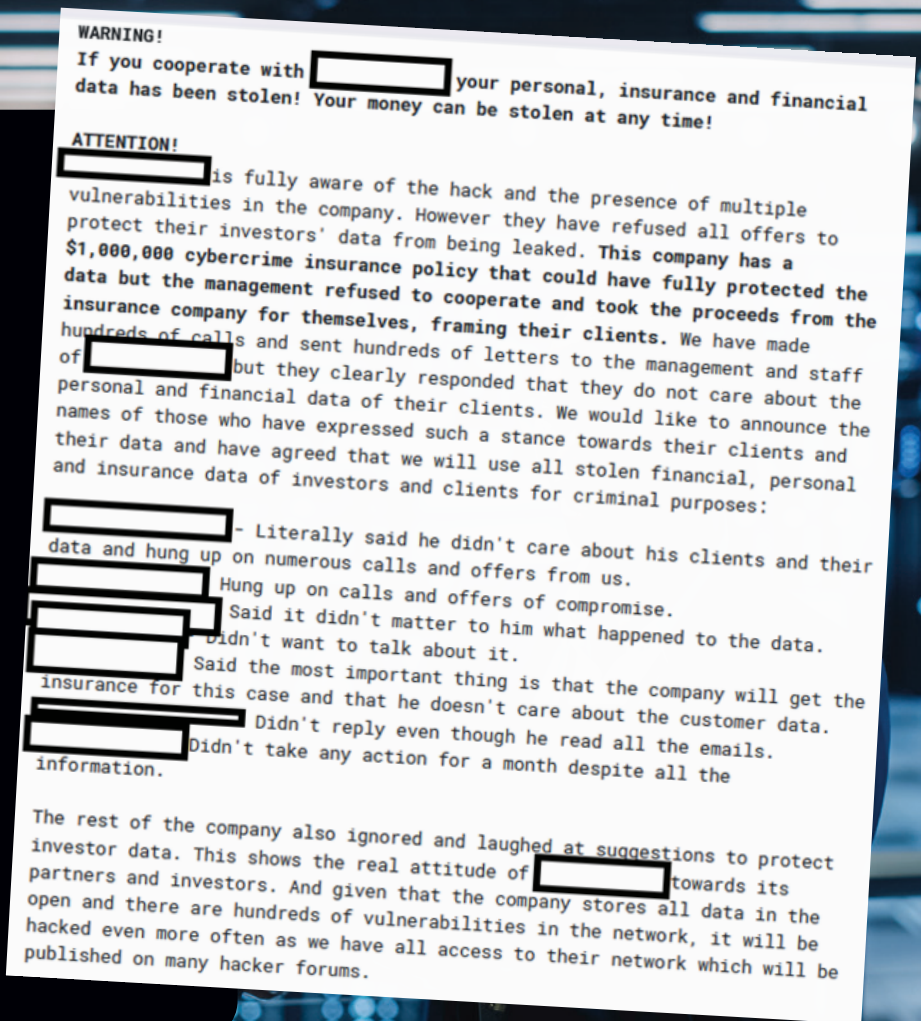
Number of Victim Organizations by Number of Employees



Damaging Reputations

Beyond the trends we've described so far, there has also been a noticeable shift in the tone and behavior of threat actors on the dark web. Listings have become increasingly aggressive, with attackers resorting to more harassing tactics. This includes naming individuals within impacted organizations, exposing their own "private" communications with the victims, and publishing links to victims' professional social media profiles.

Also discussed in our Cy-Xplorer report is the growing phenomenon called "revictimization" in which victims' stolen information is shared across multiple Cy-X brands, amplifying the harm. This approach not only maximizes the psychological impact on the victims but also opens every possible opportunity for monetization. We will continue to monitor this trend as brands maximize the victim's distress and their own gain, by pushing to extract as much value as possible from each attack.



WARNING!
If you cooperate with [redacted] your personal, insurance and financial data has been stolen! Your money can be stolen at any time!

ATTENTION!

[redacted] is fully aware of the hack and the presence of multiple vulnerabilities in the company. However they have refused all offers to protect their investors' data from being leaked. **This company has a $1,000,000 cybercrime insurance policy that could have fully protected the data but the management refused to cooperate and took the proceeds from the insurance company for themselves, framing their clients.** We have made hundreds of calls and sent hundreds of letters to the management and staff of [redacted] but they clearly responded that they do not care about the personal and financial data of their clients. We would like to announce the names of those who have expressed such a stance towards their clients and their data and have agreed that we will use all stolen financial, personal and insurance data of investors and clients for criminal purposes:

[redacted] - Literally said he didn't care about his clients and their data and hung up on numerous calls and offers from us.
[redacted] Hung up on calls and offers of compromise.
[redacted] Said it didn't matter to him what happened to the data.
[redacted] Didn't want to talk about it.
[redacted] Said the most important thing is that the company will get the insurance for this case and that he doesn't care about the customer data.
[redacted] Didn't reply even though he read all the emails.
[redacted] Didn't take any action for a month despite all the information.

The rest of the company also ignored and laughed at suggestions to protect investor data. This shows the real attitude of [redacted] towards its partners and investors. And given that the company stores all data in the open and there are hundreds of vulnerabilities in the network, it will be hacked even more often as we have all access to their network which will be published on many hacker forums.

World Watch

About the data

- Period **October 2023 to September 2024**
- **474** World Watch advisories delivered
- **Themes:** threat, vulnerability, breach, news
- One critical advisory issued with 2 updates
- Category distribution: **threat (68%), vulnerability (30%), breach (1%), news (1%)**

The Orange Cyberdefense World Watch (WW) service gathers, examines, prioritizes, contextualizes, and summarizes the crucial threat and vulnerability information that customers require to make well-informed decisions[19]. WW published 474 advisories over the past 12 months, mostly covering threats and vulnerabilities, and (to a lesser extent) breaches and news that is relevant to our clients.

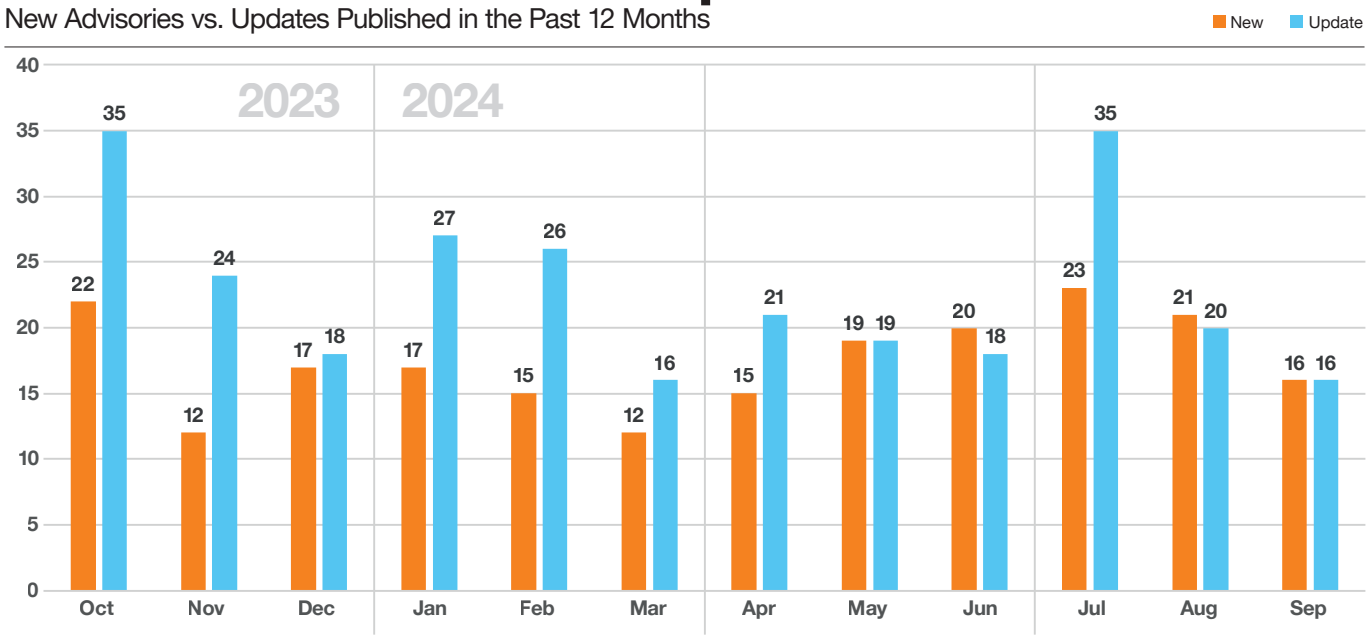
Major themes that emerged within the advisories we published include:

- France was the host of the Paris 2024 Olympics in July 2024 and attackers from across cyberspace used the opportunity to disrupt, influence, or capitalize on the excitement around the event. We reported on several instances of cybercrime, disruption, influence operations, and hacktivism associated with the event.
- Law enforcement have continued intensifying their fight against cybercriminals as we reported on various successful takedowns and disruptions. The efforts of multiple jurisdictions working in concert are starting to make life difficult for miscreants. The long arm of the law is starting to catch up. At the same time cybercriminals and ransomware groups scatter to reform later.
- The protracted war against Ukraine has seen both Russia and Ukraine leveraging their capabilities to influence and disrupt the opponent. Hacktivism is further blurring the lines between combatants and civilians.
- The conflict between Israel, Hamas, Hezbollah, and Iran has escalated. This conflict is also waged in cyberspace. Tactics like hack-and-leak, disruption, and disinformation are repeated here as well. Certain attacks are hybrid in nature, whereby cyber is just one facet.
- Several critical vulnerabilities were disclosed throughout the past year. We're once again faced with a significant number of vulnerabilities reported in security vendor products. These vulnerabilities are often in products that sit directly exposed to the internet, where their primary function is to facilitate secure authenticated access to sensitive areas inside an organization. Security flaws in these products act like an open door that attackers can walk through.
- We reported on various state-backed attackers as well as financially and politically motivated attackers.

We continue to track and advise our customers on threat intelligence regarding attacker behaviors and resulting incidents as these continue to evolve.

World Watch Advisories per Month

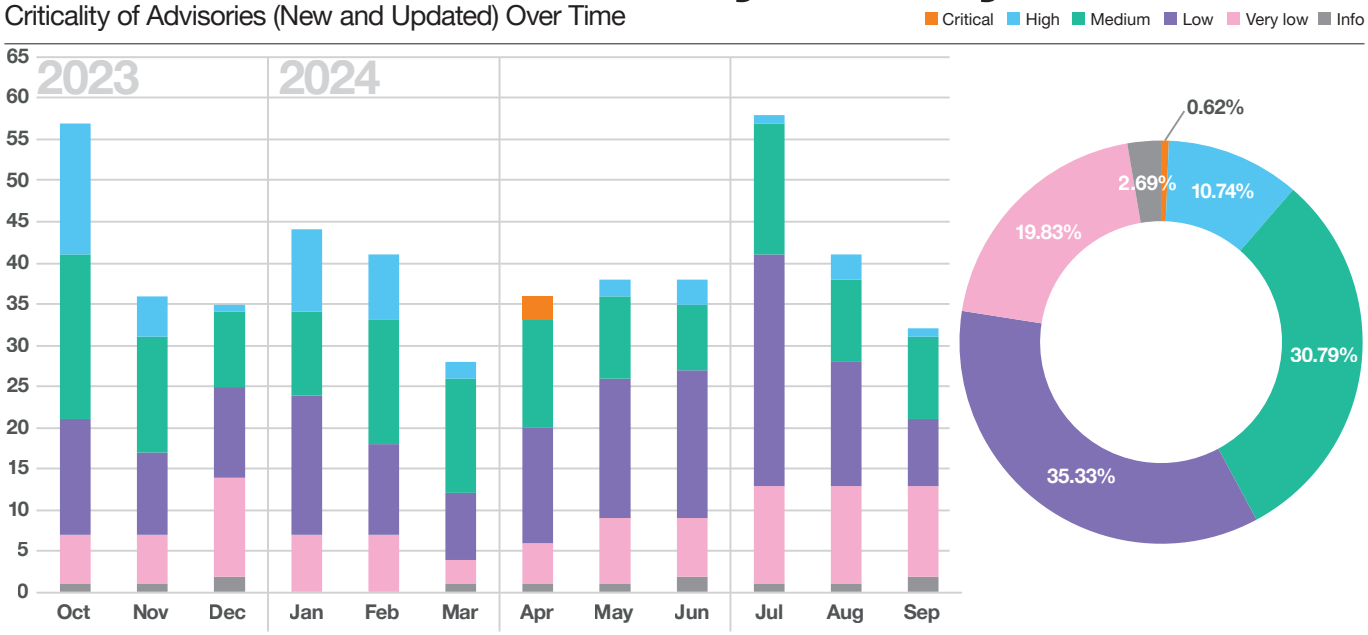
New Advisories vs. Updates Published in the Past 12 Months



- **July 2024 has 10 advisories that were posted in French in addition to English and relates to the Paris 2024 Olympics.**

World Watch Advisories by Severity

Criticality of Advisories (New and Updated) Over Time



Law Enforcement Successes

In the research chapter titled “Why aren’t we more effective in defending against Cyber Extortion?” from last year’s Security Navigator, we explored the challenges that law enforcement faces in fighting Cyber Extortion. We did not anticipate the subsequent series of law enforcement actions that eventually led to the dismantlement and take down of key cyber-criminal enterprises.

In October 2023, joint action by Europol, FBI, and Eurojust resulted in the takedown of infrastructure linked to the RagnarLocker ransomware group. One of the group’s main developers was arrested and crypto assets were seized.

In last year’s report we highlighted that the Cy-X brand LockBit was an anomaly with respect to the expected “lifespan” of such groups, as it appeared to be somewhat “untouchable” by law enforcement. In February 2024, Operation Cronos was announced, showcasing the combined successes of several jurisdictions in fighting LockBit.

Infrastructure, decryption keys, crypto wallets, and source code were seized, and two arrests were made. Over the course of several months, we provided updates as law enforcement proceeded to chip away at LockBit as the group wrestled to recover from successive blows. LockBit continues to operate today, but not at the same volume as before the initial takedown.

'Operation Endgame' is yet another example of law enforcement working to disrupt cyber criminals with coordinated activity. Between May 27 and May 29, Europol and several partner agencies disrupted infrastructure associated with malware spreading services such as IcedID, SmokeLoader, Pikabot, Bumblebee, SystemBC, and Trickbot. A large sum of cryptocurrency assets was seized. The amorphous nature of these cybercriminal operations allows the activities to resurface if the criminals are not arrested.

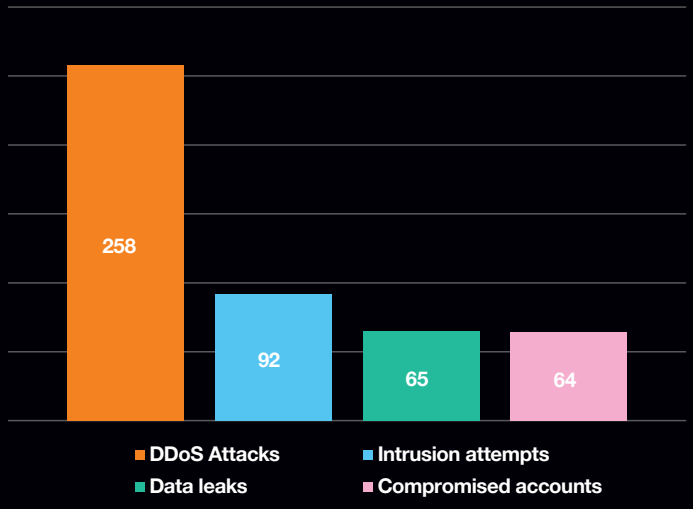
Paris 2024 Olympics

The Paris 2024 Olympics attracted enormous international attention, as athletes from many nations competed for glory. The WW coverage of the event spanned several weeks, as we anticipated malicious activity related to cybercrime, hacktivism, disruption, influence campaigns, and espionage.

Cybercrime, specifically scams and fraud like illegal ticket and merchandise sales, was a continuous theme in our advisories. There was also a cyber extortion attack impacting a network of the Grand Palais exhibit hall, although this did not impact the Olympic events held there. We also reported on numerous hacktivist attacks involving distributed denial of service (DDoS) impacting French organizations. For example, a hacktivist persona known as "LulzSec Muslims" hacked a website associated with the French National Olympic and Sports Committee (Comité National Olympique et Sportif Français). This assault also didn't impact the Paris 2024 Olympic games. In another example, a pro-Russian hacktivist group called Beregini[20] reportedly leaked data from the Polish Anti-Doping Agency, with names of Polish athletes allegedly linked to performance-enhancing drugs[21].

Incident Types

Anssi Tracking of Incidents During the Olympics



Finally, there were a handful of reports on influence operations spreading disinformation regarding the Paris 2024 Olympic games. DFRLab, NewsGuard, and Harfang Lab linked the activity to Russian actors[22][23][24]. The disinformation was spread through a news network as well as actor-controlled social media accounts. This dynamic also involves coordination between technical actors and disinformation agents, leveraging anonymized social media accounts, actor-controlled news networks, and cyber techniques like redirection chains and botnets.

Summary as consolidated by ANSSI

548 cybersecurity alerts from May 8 to September 8, 2024.

Leading to 83 incidents,

Resulting in **minimal impact**, no disturbance on the execution of the event itself

Summary by Orange Cyberdefense

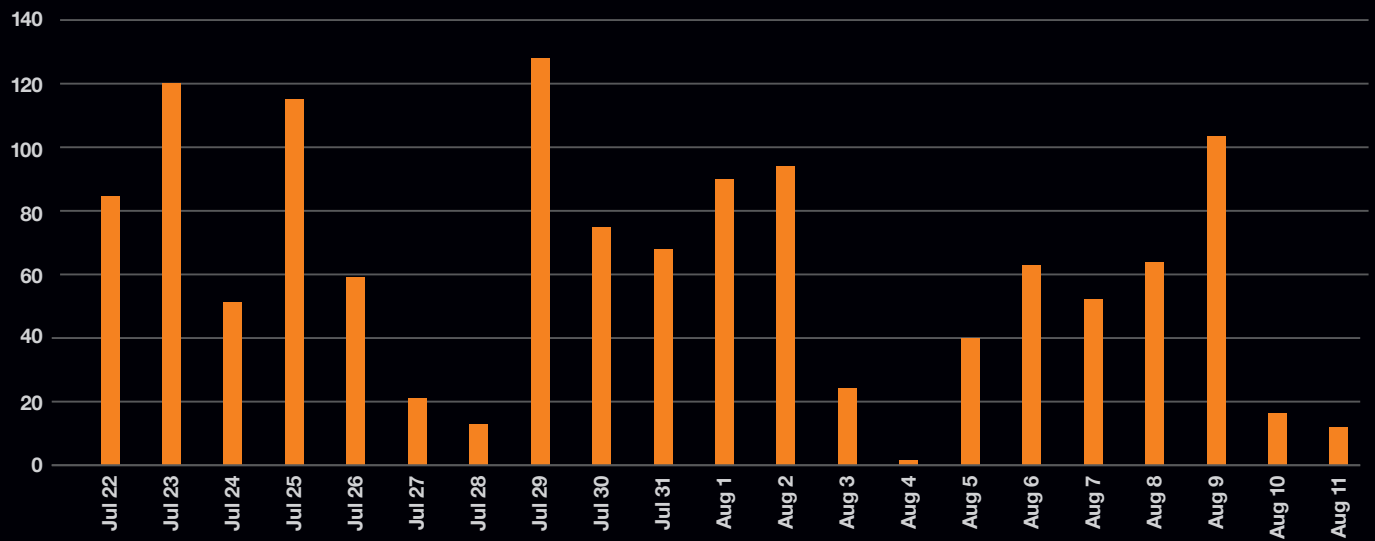
No increase in cyber incidents during the period

202 security alerts raised on the scope related to Paris 2024 monitored by our CyberSOC, including **10 DDoS attacks** that were mitigated

Only one incident related to a direct supplier of the Olympics

Phishing Cases During the Olympics

Cases Handled by Orange Cyberdefense CERT



Long-Running Conflicts

There are several World Watch advisories spanning many years that track cyber related threats associated with war or armed conflict.

Russia's war against Ukraine is one such conflict that we continue to track, and we have issued 8 updates relating to malware, hacktivism, and disinformation associated with that conflict over the past year. State-backed actors continue to leverage their past expertise, demonstrating well developed tactics, techniques, and procedures when executing cyberattacks and spreading disinformation.

As we detail in the chapter on Hacktivism in this report, pro-Russian hacktivism groups continue to put pressure on Ukraine and its supporters. One group[25] has been attributed with over 6,600 attacks since March 2022, mostly targeting symbolically important entities in Europe. Distributed Denial of Service (DDoS) attacks are an effective technique for drawing attention to a cause or message. Specific groups make good use of this through the DDoSia project[26], using the platform to recruit and coordinate attacks on victims. By the first half of 2023 they had executed more than 1,100 DDoS attacks in 32 countries. Direct links between this group and the Russian government have not yet been publicly confirmed, but our research suggests this is the case.

According to reports[27], Russia continues to employ disinformation as a technique to sow discord. One example is that on 17 February 2024, several Ukrainian media outlets were abused to spread fake news, having had their websites hacked and disinformation planted.

In December 2023, we learned that Kyivstar -a major telecoms operator in Ukraine - was compromised. The attack allegedly impacted 24 million users of the mobile network. A group called Solntsepyok claimed responsibility but reports eventually attributed the attack to the suspected Russian APT group called Sandworm[28].

Ukraine has responded in kind. In June 2024, reports[29] revealed that Ukraine had launched several cyberattacks against Russian airports, defacing some local government websites and causing flight delays. This was followed up by cyberattacks that disrupted Crimea's largest telecommunication and internet providers. Later in July 2024, DDoS attacks were launched against major banking infrastructure in Russia. Reports claim that many of these cyberattacks by Ukraine were jointly executed by hacktivist groups and intelligence services.

In October 2023 the tension between Israel and Hamas escalated beyond anything seen in the past. The result of Hamas' attack on Israel and the ensuing reprisal spilled over into cyberspace. Both sides have reportedly targeted networks with DDoS attacks, also exploiting hosts to deface websites or leak stolen data[30]. Disinformation campaigns followed, trying to influence opinions and discredit the opposing side[31].

Hacktivists responded to attack those on the opposite side, and this spilled over to Europe and elsewhere. DDoS attacks were directed at companies, airports, and government agencies in Europe.

Suspected pro-Hamas actors created a fake Android version of an emergency services app called RedAlert, which is used by Israeli citizens. The app harvested and stole data from victims[32]. A few weeks later attackers claimed they breached the RedAlert API and stole between 10,000 to 20,000 users' data[33]. We cited other reports[34] that claimed attackers were using the Israel-Hamas conflict to conduct spear phishing attacks. Other attacks managed to impact industrial control systems in Israel[35].

Later, Israel's National Cyber Directorate (INCD) released a brief outlining a Lebanon-based advanced persistent threat they claimed were backed by Iran. The agency also claimed that the Lebanon-based group's activities were responsible for cyberattacks against Israeli hospitals. Over several months various cyberattacks ensued, and reports attributed these to Israel, Iran, and regional proxies of Iran[36].

On 17 September 2024, a coordinated attack led to the explosion of thousands of pagers belonging to Hezbollah members in Lebanon and Syria, leading to fatalities and severe injuries. Two days later, a similar event occurred where two-way handheld radios (walkie-talkie) of Iran-backed militia exploded. No one claimed responsibility for these explosions. It is unclear whether this attack included any cyber elements, but it is believed that a large-scale covert supply-chain attack was used to plant the deadly devices[37]. Still, the incident serves as a cold reminder of the vulnerability of supply chains in any context.

For now, the conflict between Israel, Hezbollah, Iran and Hamas has mostly played out in the physical world and is still contained in that region. Very few impactful or serious cyberattacks have been seen and have mostly manifested as threats of intimidation with a degree of influence or disinformation.

02/10/23 – Update 1
Critical vulnerability affecting WS_FTP Progress Software exploited in the wild
CVE-2023-40044

11/10/23 – Update 1
Atlassian patches critical vulnerability affecting Confluence Server & Confluence Data Center instances
CVE-2023-22515

02/11/23 – Update 4
Second critical BIG-IP vulnerability chained together with CVE-2023-46747 in ongoing attacks

11/01/24 – Initial
Two new 0-day vulnerabilities actively exploited against Ivanti Connect Secure VPN
CVE-2023-46805/
CVE-2024-21887

18/12/23 – Update 7
New Qakbot variant deployed in test malspam campaigns

07/11/23 – Update 1
Confluence vulnerability CVE-2023-22518 exploited in the wild

03/11/23 – Update 1
Critical vulnerability in ActiveMQ exploited in the wild including to deploy HelloKitty ransomware
CVE-2023-46604

23/01/24 – Update 1
Critical trivial vulnerability in Confluence exploited in the wild
CVE-2023-22527

07/02/24 – Update 2
Dutch Military Intelligence and Security Service disclose details about Nov 2023 attack

09/02/24 – Initial
Critical vulnerabilities CVE-2024-21762 and CVE-2024-213113 in FortiOS exploited in the wild

29/05/24 – Initial
Check Point disclosed exploited 0-day in its Remote Access VPN solution
CVE-2024-24919

12/04/24 – Initial
Critical 0day in Palo Alto's GlobalProtect gateway exploited in the wild
CVE-2024-3400

18/03/24 – Update 1
PoC emerges for critical vulnerability CVE-2024-21762 in FortiOS SSL-VPN module

20/02/24 – Initial
Critical authentication bypass vulnerability in ScreenConnect
CVE-2024-1708 and
CVE-2024-1709

13/06/24 – Update 3
Dutch Ministry of Defense announces UNC3886 threat actor compromised up to 20,000 FortiGate instances (CVE-2022-42475)

19/06/24 – Update 4
Threat actor UNC3886 secretly exploiting CVE-2023-34048 since late 2021

19/07/24 – News
CrowdStrike's Falcon Sensor update error blocks Windows

12/08/24 – Update 1
Proof-of-Concept released for critical vulnerability CVE-2024-38077 in Windows Server (using Remote Desktop Licensing)

5/09/24 – Update 2
CVE-2024-38106 vulnerability exploited by North Korean threat actor

27/08/24 – Update 1
Published PoC for 0-click IPv6 vulnerability CVE-2024-38063 only enables DoS for now

14/08/24 – Initial
2024 August (Microsoft) Patch Tuesday: Many vulnerabilities require your attention

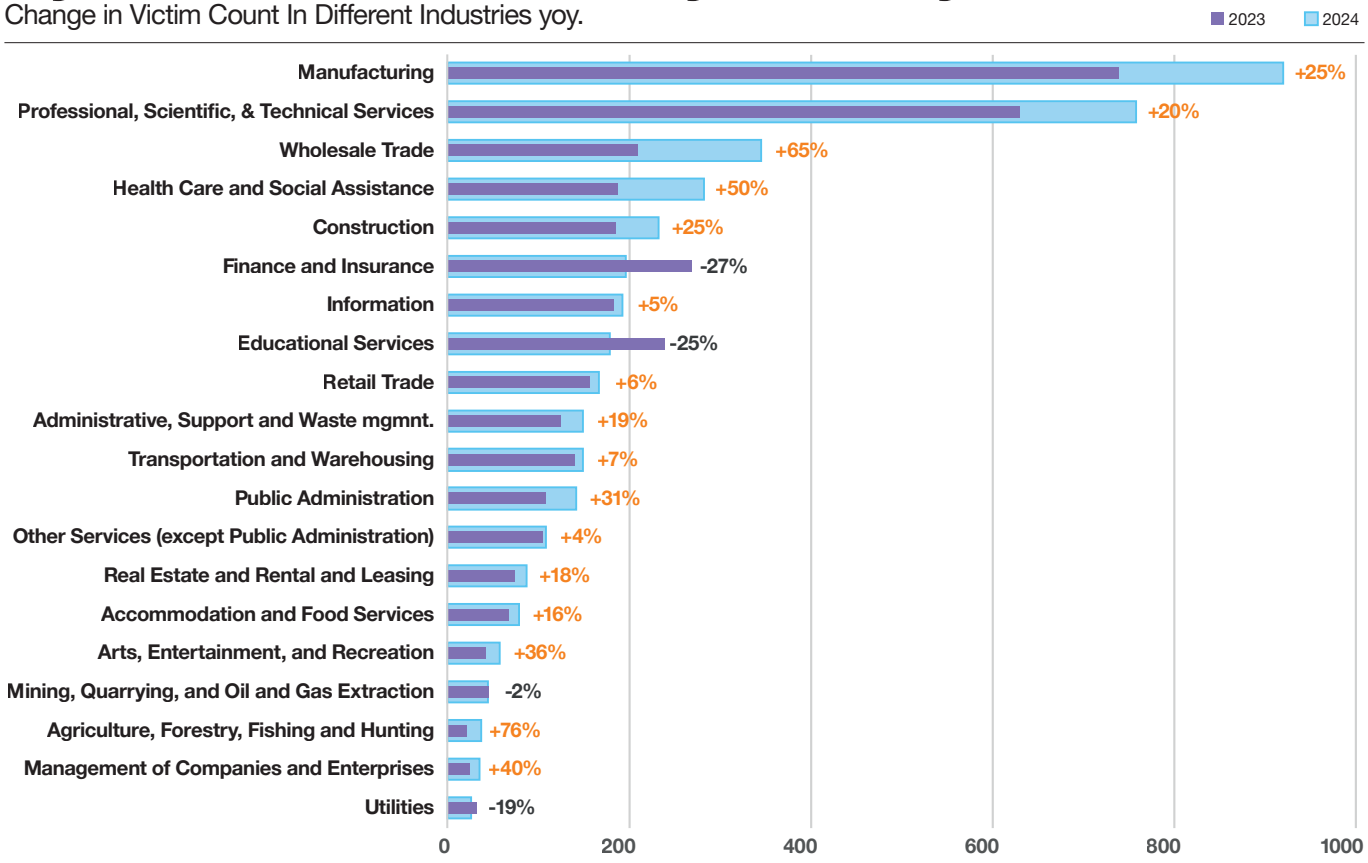
Legend

- Newly-discovered or exploited vulnerabilities
- State-sponsored espionage operations
- Financially motivated campaigns
- News

Industry Comparisons

Cy-X: Shifts in Victims by Industry

Change in Victim Count In Different Industries yoy.



Industry Ranking, Victim Delta, and Most Affected Sub-Industries

Each industry has distinct exposure to cyber extortion (Cy-X), with some experiencing significant growth in victim counts and varying degrees of impact on sub-industries.

Manufacturing leads as the most impacted, comprising 22% of all Cy-X victims and showing a 25% increase in incidents. Fabricated Metal Product and Machinery Manufacturing are particularly affected.

Professional, Scientific, and Technical Services ranks second with a 20% increase, showing concentrated incidents in Legal and Accounting Services, sub-sectors that often handle sensitive client data.

Healthcare, ranking 4th most impacted this year, saw a substantial 50% increase in victims, as attackers abandoned previous ethical constraints around targeting critical healthcare services like Ambulatory Health Care and Hospitals.

Educational Services ranks 8th with a 25% reduction in victims, while **Finance and Insurance** ranks 6th, showing a 27% decrease, but with a concentration of victims in Credit Intermediation and Securities sub-sectors.

Public Administration experienced a 31% increase, particularly in government support and justice sectors. **Construction** ranks 5th with a 25% increase, primarily impacting Specialty Trade Contractors and Civil Engineering. Finally, **Retail Trade** ranks 9th, with a 6% increase in incidents, especially affecting Motor Vehicle Dealers and Food Retailers.

MTTR, Coverage Score, True Positive/ False Positive Ratio

Our CyberSOC metrics across industries provide insights into incident response effectiveness and monitoring depth.

Manufacturing's Mean Time To Resolve (MTTR[38]) is relatively high at 97 hours, making it the second slowest sector, while its coverage score of 36.77% is below the average for all industries. True positives account for 20.96% of alerts. Incidents primarily originate internally (62.48%), with misuse as the primary action, impacting primarily on end-user devices.

Professional Services, aligned with the industry median MTTR of 49 hours, has one of the lowest coverage scores at 32.04%. Incidents mostly stem from external actors (52.77%), with hacking and misuse primarily affecting end-user devices and servers.

Healthcare's MTTR is 50 hours with a low coverage score of 29.04. The sector's true positive ratio is 16.45%. Incidents often involve malware and misuse originating from external sources (52.62%) and targeting end-user devices and networks.

Finance and Insurance holds the highest coverage score at 55.87%, indicative of robust monitoring, though its MTTR is still 56 hours. External actors are the primary origin, responsible for incidents that predominantly involve hacking and social engineering, targeting servers and accounts.

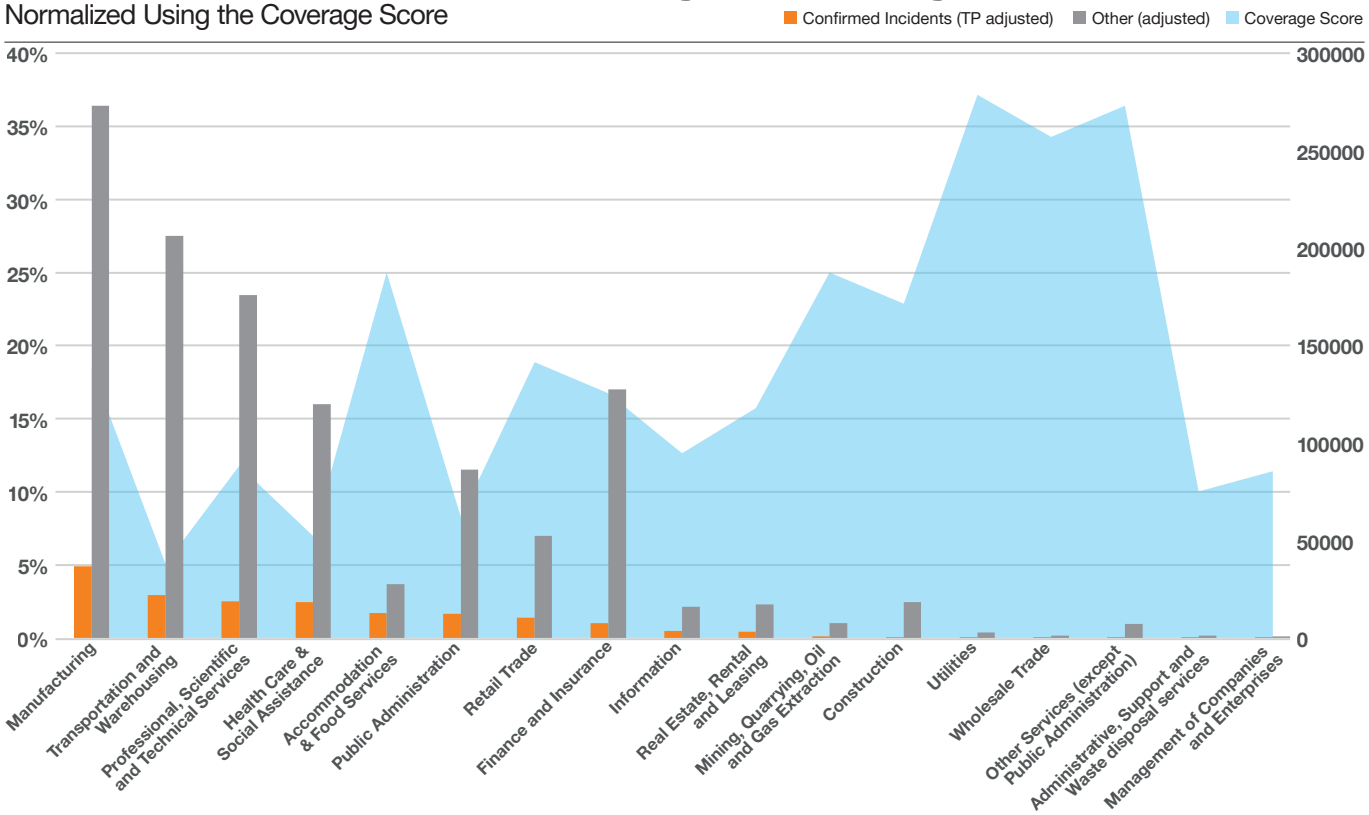
Our client's in **Public Administration** had an average MTTR of 38.32 hours, and an average coverage score of 41.43%. We report a true positive ratio of 20.15%. Incidents are primarily externally sourced, with hacking and misuse actions impacting end-user devices and accounts.

Construction shows a high coverage score of 45.71% and a true positive rate of 14.46%, and an MTTR of 94.7 hours. Most incidents in this sector involve internal actors and misuse actions, affecting end-user devices, servers, and networks.

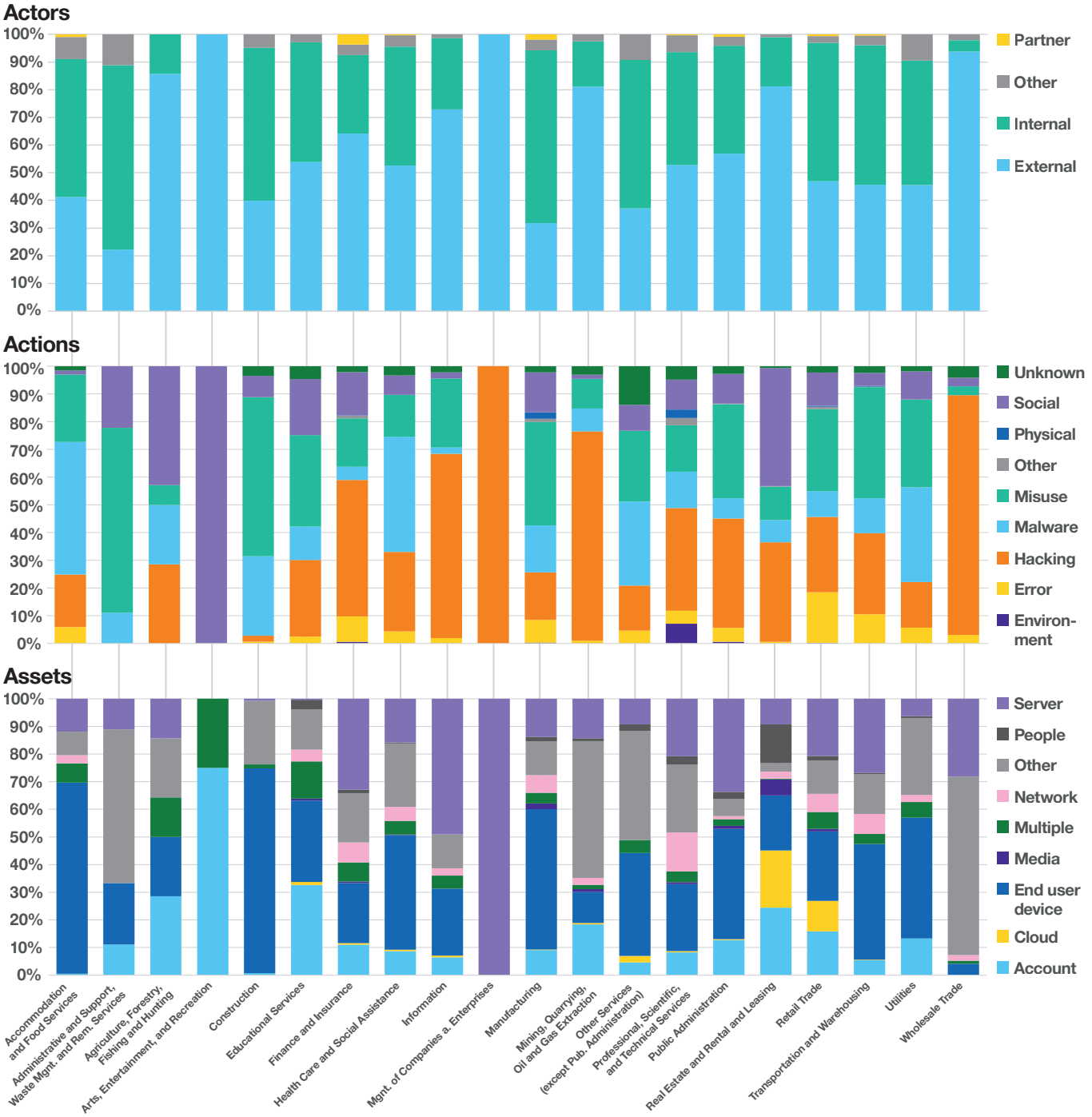
Retail has an MTTR of about 36 hours and a coverage score of 35.1%, and a true positive rate of 24.34%. Errors and misuse are frequent in Retail, affecting cloud and end-user devices.

CSOC Data: Incidents by Industry

Normalized Using the Coverage Score



VERIS Actors, Actions, Assets by Industry



VERIS Actor, Action, and Asset Analysis

The VERIS framework provides clarity on threat origins, actions, and asset impacts.

In **Manufacturing**, incidents are largely internal (62.48%), with misuse actions commonly impacting end-user devices and servers. **Professional Services** faces a different profile, with 52.77% of incidents initiated by external actors primarily through hacking, impacting both end-user devices and servers.

Healthcare encounters a similar external focus, with 52.62% of incidents driven by external actors. Malware tactics and misuse are common, while incidents largely impact end-user devices and networked systems.

Finance and Insurance also sees primarily external incidents that affect servers and accounts, with hacking and social engineering as predominant actions.

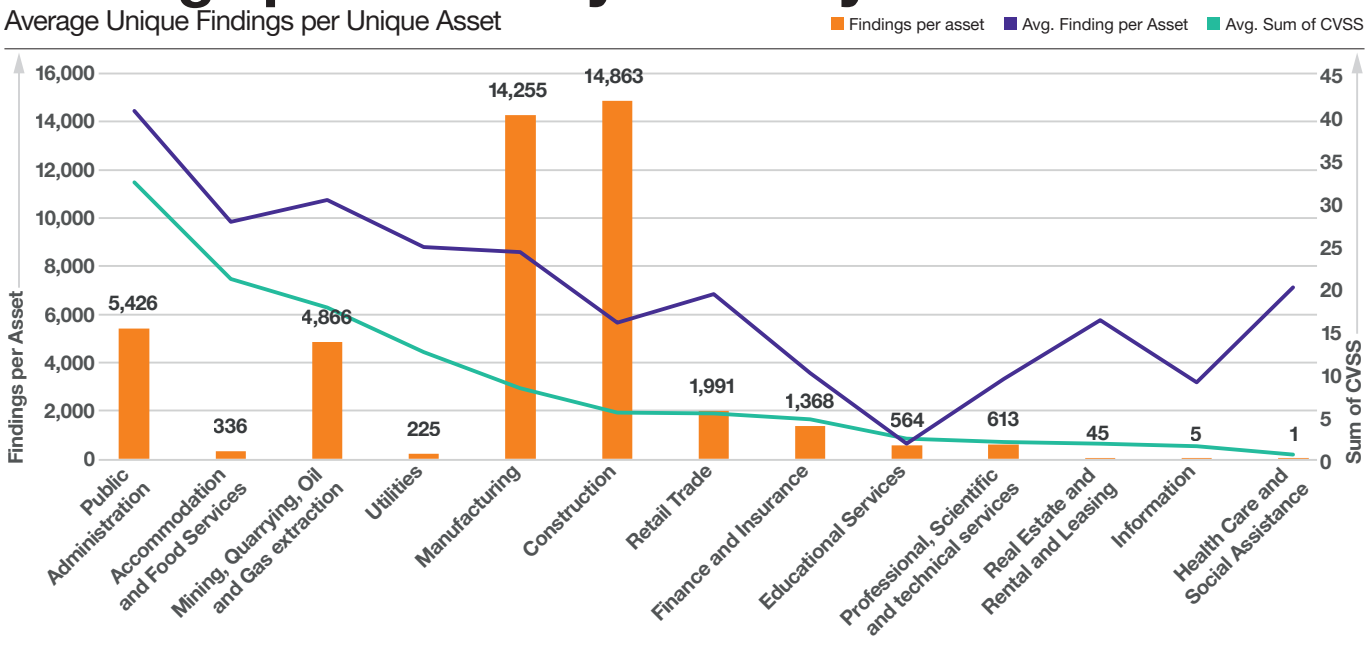
Public Administration's external attack pattern also involves hacking, and impacts end-user devices. Though Misuse is also a common cause for recorded incidents.

In **Construction**, internal incidents involving misuse and malware dominate, with incidents largely affecting end-user devices.

We report a high rate of error-related incidents for our **Retail** clients, largely impacting end-user devices.

Findings per Asset by Industry

Average Unique Findings per Unique Asset



VOC Metrics

Findings Per Asset, Vulnerability Score, Max and Average Vulnerability Age

VOC metrics shed light on each industry's vulnerability management practices, tracking findings per asset and the persistence of unresolved vulnerabilities.

Manufacturing exhibits a high findings-per-asset rate at 24.15, with critical vulnerabilities remaining open for an average of 204 days and a maximum age of 721 days. Clients in **Professional Services** record a lower findings-per-asset ratio of 9.34, with critical vulnerabilities lasting around 91 days on average.

There are very few clients in **Healthcare** within our dataset, but we record a similar persistence in vulnerabilities, averaging 20 findings per asset, with critical issues remaining unresolved for approximately 217 days.

Our data on clients in **Educational Services** is also limited. Here we record the lowest findings-per-asset ratio at 1.82, with critical vulnerabilities addressed within about eight days.

Finance has a findings rate of 10.03 per asset, but with critical vulnerabilities averaging 136 days before resolution.

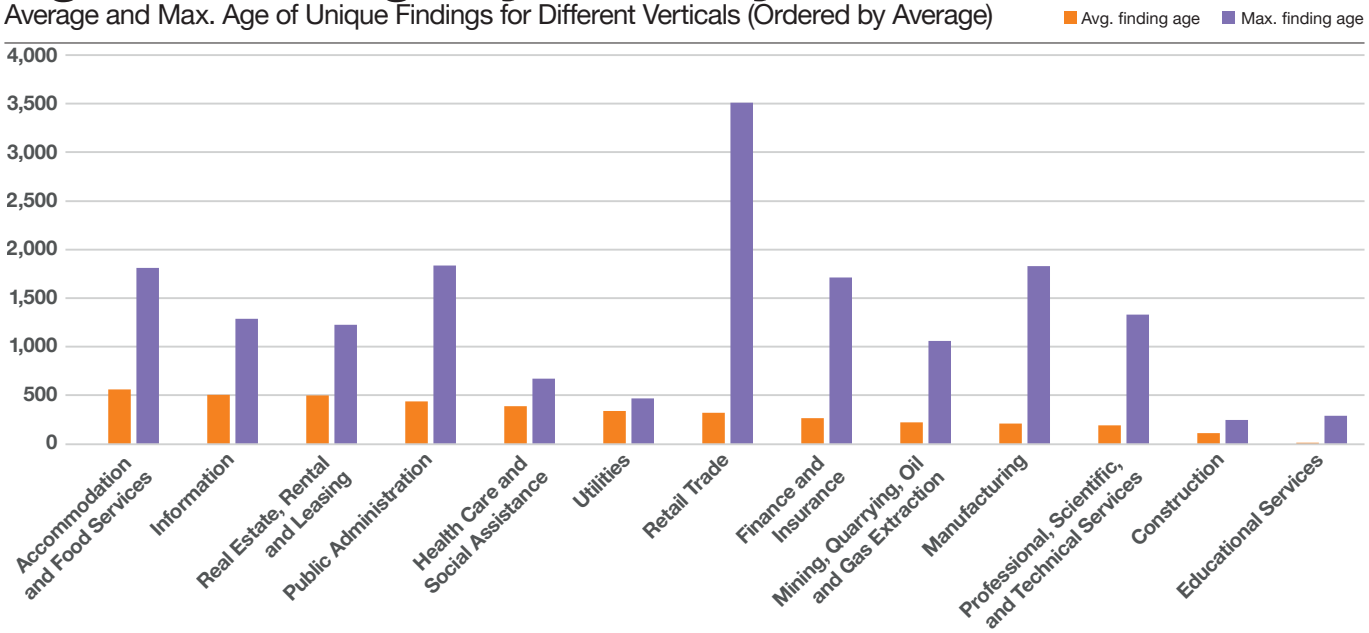
Public Administration has the highest findings-per-asset rate at 40.64, with critical vulnerabilities persisting for around 315 days.

Construction has a moderate findings rate of 15.88, but critical issues last around 120 days on average.

Retail's findings-per-asset rate of 19.24 reflects a steady vulnerability level, and the sector records a maximum critical vulnerability age of 228 days.

Age of Findings by Industry

Average and Max. Age of Unique Findings for Different Verticals (Ordered by Average)



Industry Scorecard

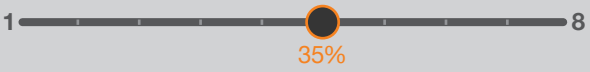
Retail and Trade

Cy-X Victim ranking (Avg: 200)

Threat Detection: Mean time to resolve (Avg: 65h)

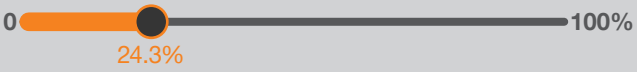
Cy-X Victim delta (Avg: +19%)

Threat Detection: Coverage (Avg: 37.5%)



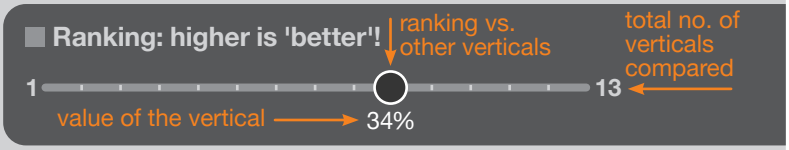
VOC: Findings per asset (Avg: 22.1 findings)

Threat Detection: True positives

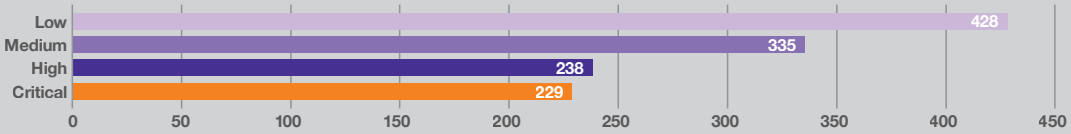


VOC: Total Vulnerability Score

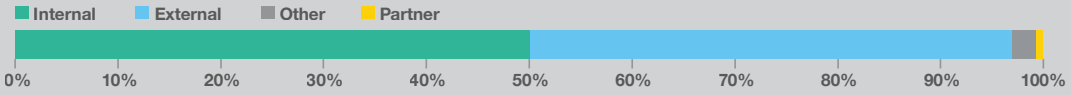
Ranking: higher is 'better'!



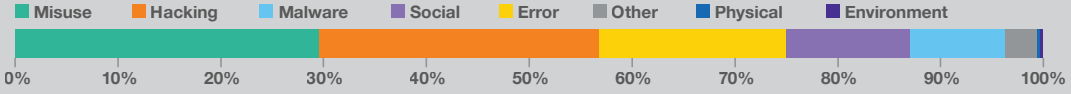
VOC: Finding age by severity (in days)



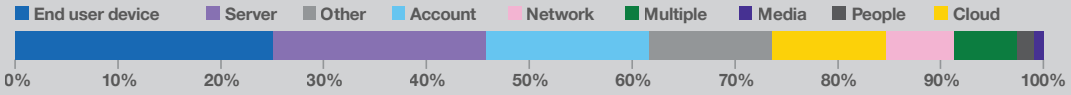
Threat Detection: Threat Actor



Threat Detection: Threat Action

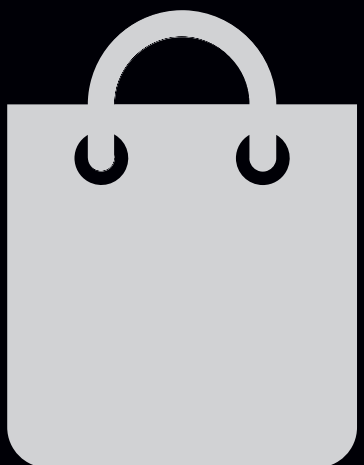


Threat Detection: Impacted Asset



Summary

The Retail Trade industry ranks 9th in terms of cyber extortion victims, with incidents rising by 6% over the past year. Motor Vehicle Dealers and Food Retailers are frequently targeted. CyberSOC metrics indicate a relatively fast MTTR (about 35 hours) and a median coverage score of 35.1%. The true positive ratio is 24.34% to 75.66%. VOC metrics show a relatively low findings-per-asset rate, though critical vulnerabilities often remain unresolved for over 228 days.



Industry Scorecard

Construction

Cy-X Victim ranking (Avg: 200)

Cy-X Victim delta (Avg: +19%)

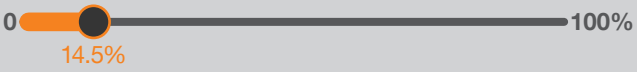
VOC: Findings per asset (Avg: 22.1 findings)

VOC: Total Vulnerability Score

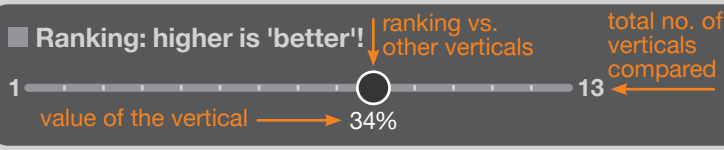
Threat Detection: Mean time to resolve (Avg: 65h)

Threat Detection: Coverage (Avg: 37.5%)

Threat Detection: True positives

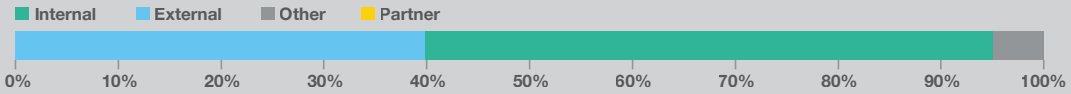


Ranking: higher is 'better'!

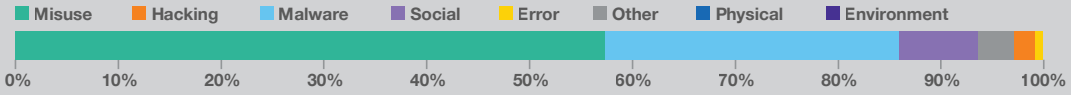


VOC: Finding age by severity (in days)

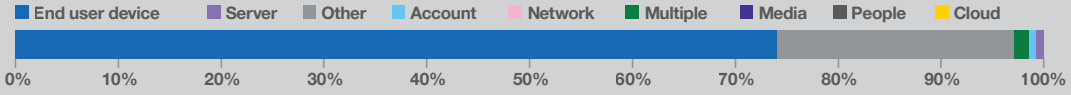
Threat Detection: Threat Actor



Threat Detection: Threat Action



Threat Detection: Impacted Asset



Summary

In Construction, a 25% increase in cyber extortion incidents primarily impacts Specialty Trade Contractors, Construction of Buildings, and Civil Engineering. Our CyberSOCs report that misuse and malware frequently affect end-user devices. Our metrics reveal a high coverage score of 45.71% and an MTTR of 94.7 hours, with a true positive rate of 14.46%. VOC metrics show moderate findings per asset at 15.88, with critical vulnerabilities persisting for around 120 days.

Industry Scorecard

Manufacturing

Cy-X Victim ranking (Avg: 200)

1 — 920 — 20

Cy-X Victim delta (Avg: +19%)

1 — +25% — 20

VOC: Findings per asset (Avg: 22.1 findings)

1 — 24.2 — 13

VOC: Total Vulnerability Score

1 — 5 — 13

Threat Detection: Mean time to resolve (Avg: 65h)

1 — 97h — 16

Threat Detection: Coverage (Avg: 37.5%)

1 — 38.5% — 8

Threat Detection: True positives

0 — 21.0% — 100%

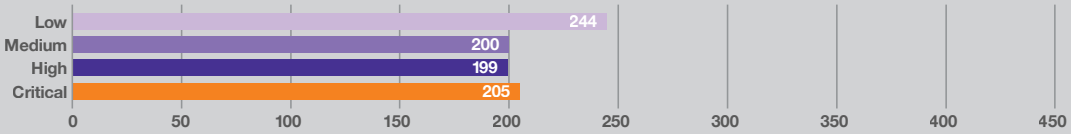
Ranking: higher is 'better'!

ranking vs. other verticals

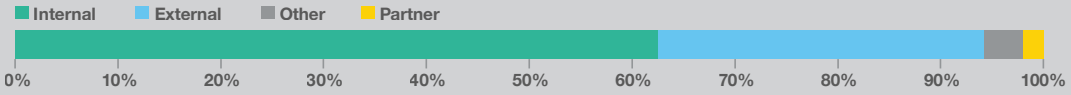
total no. of verticals compared

1 — value of the vertical → 34% — 13

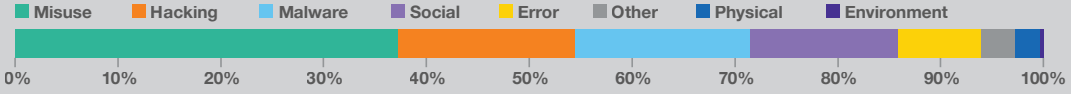
VOC: Finding age by severity (in days)



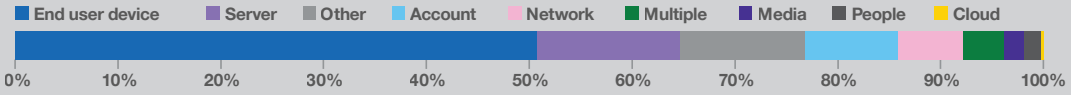
Threat Detection: Threat Actor



Threat Detection: Threat Action



Threat Detection: Impacted Asset



Summary

In the Manufacturing industry, cyber extortion and OT-specific attacks have made this sector the most impacted by cyber threats, with a 25% increase in Cy-X incidents. Key sub-sectors like Fabricated Metal Product and Machinery Manufacturing are especially impacted. Manufacturing's reliance on OT systems makes it highly vulnerable to productivity loss, data encryption, and control manipulation, with both state actors and hacktivists posing significant threats. CyberSOC metrics indicate that this industry has a high mean time to resolve (MTTR) at 97 hours, ranking as the second slowest across sectors. Coverage stands at 36.77%, near the median, with internal actors contributing to 62.48% of CyberSOC incidents. VOC metrics reveal a higher-than-average findings rate per asset, at 24.15, with critical vulnerabilities remaining open for over 204 days on average.

Industry Scorecard

Professional, Scientific, and Technical Services

Cy-X Victim ranking (Avg: 200)

Cy-X Victim delta (Avg: +19%)

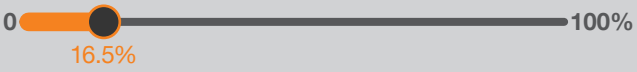
VOC: Findings per asset (Avg: 22.1 findings)

VOC: Total Vulnerability Score

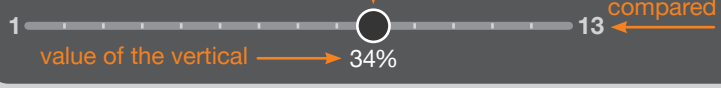
Threat Detection: Mean time to resolve (Avg: 65h)

Threat Detection: Coverage (Avg: 37.5%)

Threat Detection: True positives

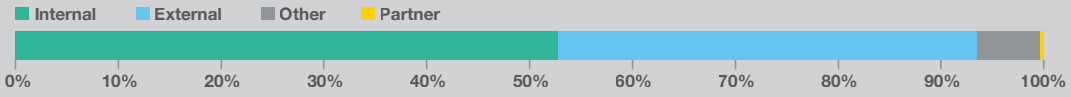


Ranking: higher is 'better'!

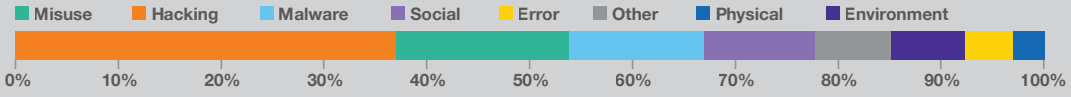


VOC: Finding age by severity (in days)

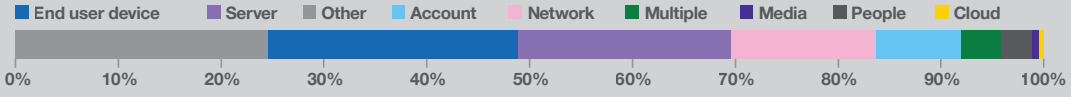
Threat Detection: Threat Actor



Threat Detection: Threat Action



Threat Detection: Impacted Asset



Summary

For the Professional, Scientific, and Technical Services sector, cyber extortion incidents have increased by 20%, particularly impacting sub-sectors such as Legal and Accounting Services. High vulnerability ages and low coverage scores suggest there is room for improvement for businesses in this industry. Hacking and misuse are prevalent threats, often impacting end-user devices and servers. CyberSOC metrics show an MTTR of 49 hours, the industry median, yet coverage is low at 32.04%. Most incidents involve external actors, with hacking being a primary action – this pattern being somewhat unusual in this year's data. VOC metrics show a lower findings-per-asset rate at 9.34, though critical issues can linger around 91 days before remediation

Industry Scorecard

Health Care and Social Assistance

Cy-X Victim ranking (Avg: 200)

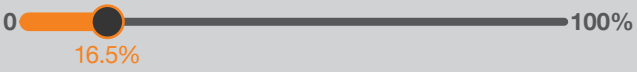
Threat Detection: Mean time to resolve (Avg: 65h)

Cy-X Victim delta (Avg: +19%)

Threat Detection: Coverage (Avg: 37.5%)

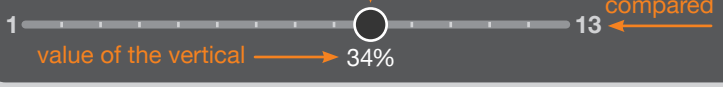
VOC: Findings per asset (Avg: 22.1 findings)

Threat Detection: True positives

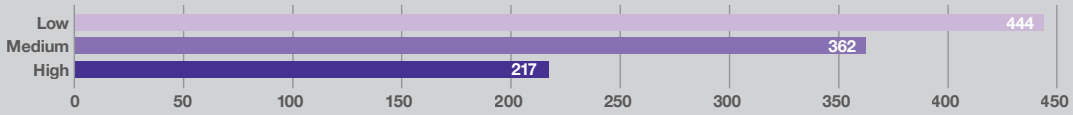


VOC: Total Vulnerability Score

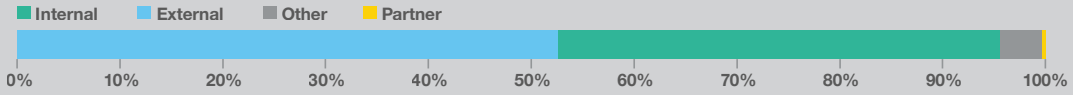
Ranking: higher is 'better'!



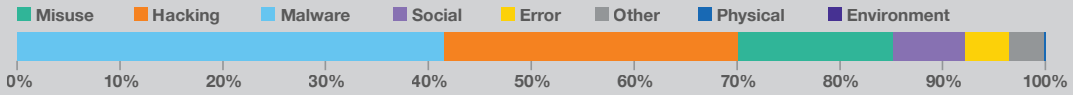
VOC: Finding age by severity (in days)



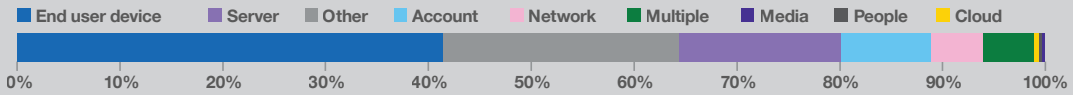
Threat Detection: Threat Actor



Threat Detection: Threat Action



Threat Detection: Impacted Asset



Summary

Health Care and Social Assistance ranks as the fourth most impacted industry, with a worrisome 50% rise in cyber extortion incidents. Sub-sectors such as Ambulatory Health Care and Hospitals are now actively targeted as previous “moral” restraints by attackers have eroded. Malware attacks, typically driven by external actors, are common, which is somewhat unusual in this year’s client data. Persistent vulnerabilities remain an issue, with critical findings often aging for over 217 days. CyberSOC metrics indicate an MTTR of 50 hours, slightly above the median, with a low coverage score of 29.04% and a true positive rate of 16.45%. VOC metrics show an average of 20 findings per asset, somewhat below the industry average of 22.43, although this is derived from a small sample of clients.

Industry Scorecard

Educational Services

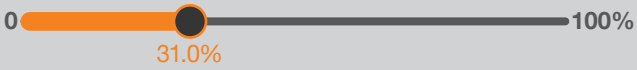
Cy-X Victim ranking (Avg: 200)

VOC: Findings per asset (Avg: 22.1 findings)

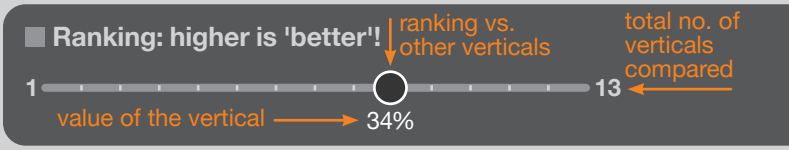
Cy-X Victim delta (Avg: +19%)

VOC: Total Vulnerability Score

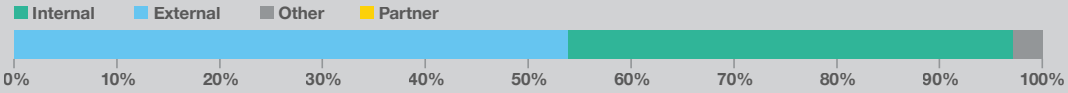
Threat Detection: True positives



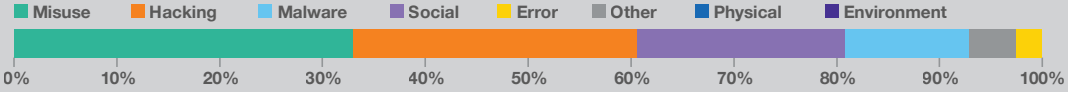
Ranking: higher is 'better'!



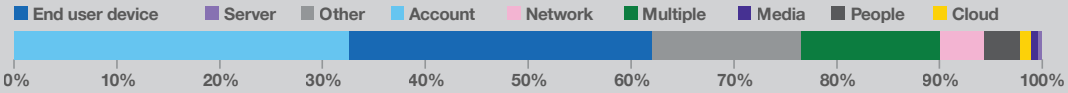
Threat Detection: Threat Actor



Threat Detection: Threat Action



Threat Detection: Impacted Asset



VOC: Finding age by severity (in days)

Summary

The Educational Services sector, ranking 8th most impacted, saw a 25% decrease in cyber extortion victims, with elementary and secondary schools being heavily impacted. CyberSOC clients in this sector have a relatively high true positive rate, demonstrating accuracy in threat detection. The CSOC metrics reveal a high true positive rate at 30.99%, though coverage remains low. VOC metrics show relatively few findings per asset, averaging 1.82, and critical vulnerabilities are resolved within about 8 days (although these metrics are derived from a small sample).

This year we highlight the Education sector as a target of modern hacktivist activity. Hacktivists attack this sector due to its public significance and symbolic value, with goals often focused on disrupting societal stability. Educational institutions are among the essential service sectors targeted by a pro-Russian hacktivist group, with attacks timed to coincide with geopolitical events and driven by the desire to influence public opinion or cause societal disruptions. These attacks are typically ideologically motivated, aiming not only to disrupt educational systems but also to manipulate public perception by targeting institutions that influence societal narratives.

Industry Scorecard

Finance and Insurance

Cy-X Victim ranking (Avg: 200)

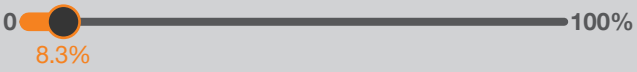
Threat Detection: Mean time to resolve (Avg: 65h)

Cy-X Victim delta (Avg: +19%)

Threat Detection: Coverage (Avg: 37.5%)

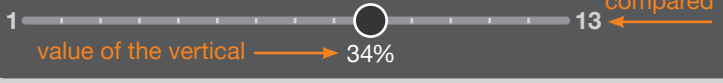
VOC: Findings per asset (Avg: 22.1 findings)

Threat Detection: True positives

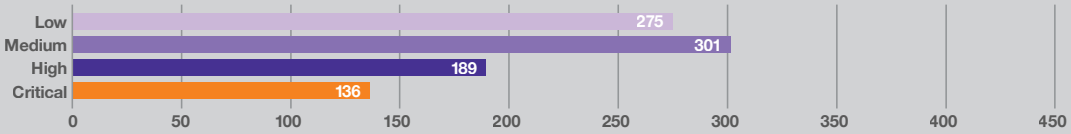


VOC: Total Vulnerability Score

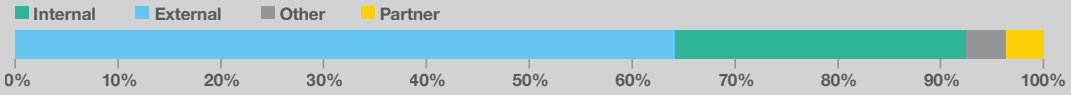
Ranking: higher is 'better'!



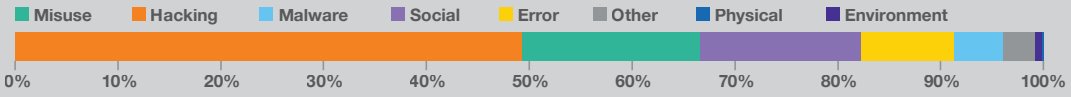
VOC: Finding age by severity (in days)



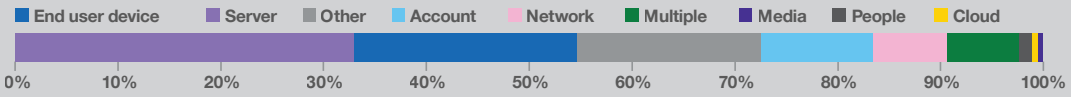
Threat Detection: Threat Actor



Threat Detection: Threat Action

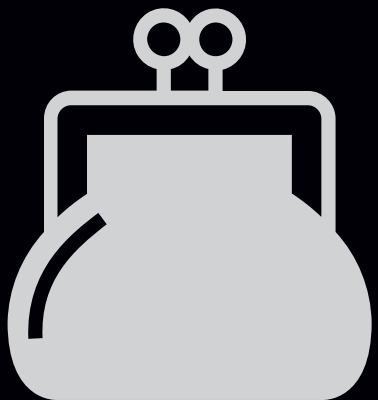


Threat Detection: Impacted Asset



Summary

In Finance and Insurance, Cy-X incident volumes have declined by 27%, but we still recorded 196 victims this year, with a concentration in Credit Intermediation and Securities. External actors are responsible for most reported CyberSOC incidents, primarily targeting servers and accounts. We report a high proportion of hacking and social engineering incidents, which is unusual to this section. The sector's CSOC metrics show the highest coverage score at 55.87% and an MTTR of 56 hours, with a true positive ratio of 8.3%. VOC metrics reveal a low findings-per-asset rate at 10.03, though critical vulnerabilities may persist unresolved for an average of 136 days.



Industry Scorecard

Public Administration

Cy-X Victim ranking (Avg: 200)

Cy-X Victim delta (Avg: +19%)

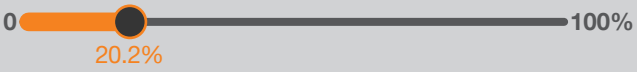
VOC: Findings per asset (Avg: 22.1 findings)

VOC: Total Vulnerability Score

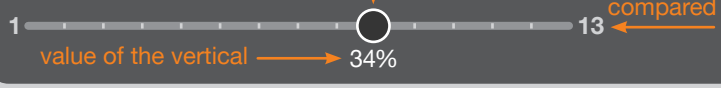
Threat Detection: Mean time to resolve (Avg: 65h)

Threat Detection: Coverage (Avg: 37.5%)

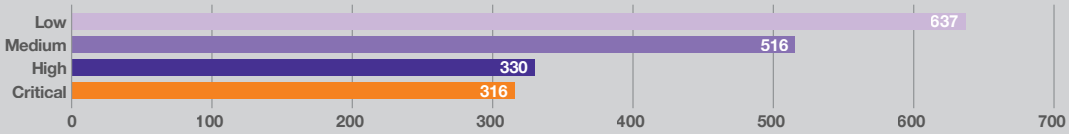
Threat Detection: True positives



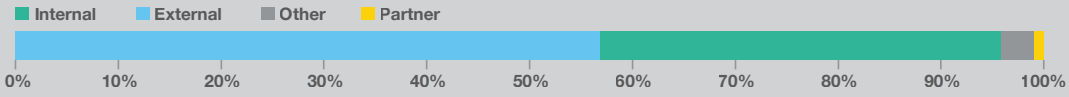
Ranking: higher is 'better'!



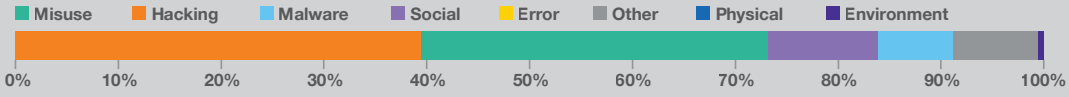
VOC: Finding age by severity (in days)



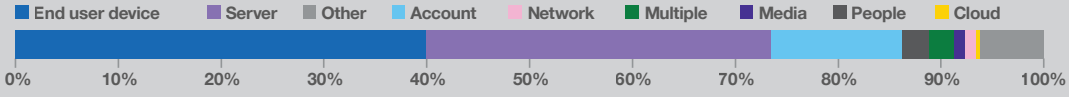
Threat Detection: Threat Actor



Threat Detection: Threat Action



Threat Detection: Impacted Asset



Summary

Public Administration experienced a 31% increase in Cy-X incidents, particularly in governmental support and justice sectors. Hacktivist activity, often coinciding with elections or geopolitical events, poses a significant risk, with attacks typically linked to hacking and misuse by external actors. The sector's CSOC metrics are notable for an average MTTR of 38 hours and a high coverage score of 41.43%, with incidents stemming largely from external sources. VOC metrics indicate a high findings-per-asset score of 40.64, with critical vulnerabilities lingering for an average of 315 days. The Navigator underscores the importance of fortified cybersecurity frameworks, particularly to secure election systems and essential government services, given the prevalence of legacy vulnerabilities.

Region perspective

Cyber Extortion (Cy-X)

The Cy-X landscape reflects diverse regional vulnerabilities, with North America, led by the United States, emerging as the most impacted region globally. The USA alone accounted for 2,154 of the 2,387 Cy-X cases reported across North America, marking a 25% increase from the previous year. This high volume underscores the USA's attractiveness as a target for financially motivated cyber extortion, particularly within high-value sectors that rely heavily on digital infrastructure.

In Europe, Cy-X incidents were widespread, with Germany experiencing 19% of regional cases, positioning it as a significant target. This prominence aligns with Germany's industrial and economic significance within Europe, which has made it a frequent target for cybercriminals seeking lucrative payoffs. Cy-X incidents in Europe highlighted the extensive integration of IT across industries, further exacerbating the spread and impact of cyber extortion events in high-risk sectors.

Across the APAC region, Cy-X impacts were uneven. Japan ranked as the 13th most affected country globally, probably driven by both industrial vulnerabilities and high levels of connectivity. In contrast, China showed a lower Cy-X victim. South Korea and Singapore also experienced moderate levels of Cy-X incidents, with cyber extortion targeting high-value manufacturing and industrial sectors, underscoring the importance of IT and OT protections in the region.

Hacktivism

Hacktivism incidents presented a different geographic focus, largely driven by political motivations and regional tensions. Europe bore the brunt of hacktivist attacks, with 96% of observed pro-Russian hacktivism cases targeting European countries. These attacks primarily impacted Ukraine, Czech Republic, Spain, Poland, and Italy, reflecting the influence of geopolitical tensions. Hacktivists in Europe primarily employed disruptive tactics, such as distributed denial-of-service (DDoS) attacks and website defacements, to publicize their causes and destabilize critical services.

In APAC, Japan was notably impacted, recording 71 hacktivist attacks, many linked to pro-Russian groups. This significant focus on Japan aligns with its strategic importance in global geopolitics and its robust digital infrastructure, which provides ample targets for hacktivist campaigns.

The Middle East saw intensified hacktivist activity, particularly in conflict areas which have led to reciprocal cyber offensives. Pro-Hamas hacktivists targeted Israeli networks, launching DDoS attacks and exploiting social engineering to compromise personal data. Lebanon also reported hacktivist incidents, with activity allegedly linked to Iranian groups, underscoring the geopolitical complexities of hacktivism in the Middle East. These politically driven cyber actions signal the region's heightened vulnerability to hacktivist campaigns amidst ongoing conflict.

Threat Detection

CyberSOC data from Orange Cyberdefense's Security Operations Centers provides insights into threat detection and incident response across regions. From our clients in North America, CyberSOC metrics revealed a high false-positive rate of 80.53% in the USA, with most incidents driven by internal misuse rather than external attacks. This is derived from a very small sample, however, and should not be generalized.

In Europe, CyberSOC data reveals efficient incident response for our clients in Germany, demonstrated by its swift Mean Time to Resolve (MTTR) of 50.5 hours.

China's CyberSOC metrics in APAC showed a balanced false-positive to true-positive ratio, with most incidents originating internally and impacting end-user devices. The internal nature of these threats points to the importance of user access controls and monitoring for insider threats within Chinese organizations.

Operational Technology (OT)

Operational Technology (OT) security emerged as a critical theme, particularly in sectors where IT and OT systems are tightly integrated, creating vulnerabilities that adversaries can exploit. The USA in North America experienced substantial OT impacts, with 49% of all OT-targeted attacks globally. The manufacturing and transportation sectors were particularly affected, as IT incidents frequently cascaded into OT environments, leading to production downtimes and other operational interruptions. This spillover effect underscores the need for comprehensive OT security protocols to protect critical infrastructure from the ripple effects of ransomware and other IT-originating incidents.

In Europe, Germany was significantly impacted, accounting for 11% of all OT-targeted incidents. The country's manufacturing and utility sectors were key targets, with attackers exploiting IT-OT interdependencies to disrupt operations. Sophisticated OT attacks used complex tactics to manipulate physical processes, which caused substantial operational downtime. This level of targeting in Germany reflects the high value of its industrial sectors to both economically motivated and state-backed threat actors.

Summary

This thematic summary provides a comparative overview of how Cy-X, hacktivism, CyberSOC observations, and OT security challenges manifest differently across regions, shaped by unique geopolitical, industrial, and infrastructural factors. The findings underscore the importance of tailored cybersecurity strategies that address both the direct and spillover effects of cyber incidents, particularly within critical infrastructure and high-risk sectors.

Region Scorecard

Europe Region

Cy-X region ranking

Europe had the second highest number of Cy-X victims with

745 victims

Cy-X victim delta

In this region we saw an increase in the number of victim organizations of

+ 18%

Most affected countries

Top 5 impacted countries were

- Italy (19%)
- Germany (19%)
- France (16%)
- Spain (13%)
- Belgium (8%).

CyberSOC Ranking

- The Mean Time To Resolve (MTTR) for clients in this region was 65 hours.
- The countries with the lowest Mean Time To Resolve (MTTR) were Austria (37.4 hours), Norway (37.7 hours), Germany (50.5 hours), and the United Kingdom (50.7 hours).
- The VERIS Actor category for clients in this region is nearly split down the middle for Internal (47.32%) and External (47.20%).
- The most impacted asset class for clients in this region is End User Devices (45.5%) followed by Server (22.19%) and Account (12.39%).
- For clients in this region the most common VERIS Action classifications were Hacking (30.10%) and Misuse (28.08%), followed by Malware (15.89%) and Social (12.94%).

Hacktivism Ranking

- Our case study on one of the most active pro-Russian hacktivist groups shows that 96% of all attacks targeted victims in Europe.
- The top 5 countries attacked were: Ukraine (11%), Czech Republic (9%), Spain (9%), Poland (8%), Italy (7%).

Summary

A High Cy-X and Hacktivism Target

Europe ranked as the second most impacted region by Cy-X globally, experiencing 745 victim organizations, an 18% increase over the previous year. Among European countries, Italy and Germany led the way with 19% of Cy-X cases each, followed by France (16%), Spain (13%), and Belgium (8%). This escalation in Cy-X incidents aligns with Europe's prominence as a hub for business and technology, making it an attractive target for financially motivated cyber extortion. Moreover, hacktivism was particularly prominent in Europe, with 96% of attacks by the pro-Russian group we studied targeting European entities. Attacks primarily impacted Ukraine (11%), Czech Republic (9%), Spain (9%), Poland (8%), and Italy (7%). CyberSOC data reveals that the primary threat actions were hacking and misuse, both heavily impacting end-user devices. The concentration of both Cy-X and hacktivism in Europe emphasizes the region's complex threat environment, especially as politically motivated groups escalate attacks amidst geopolitical tensions.

Industrial economies in Europe also feature as vulnerable to OT attacks. Germany recorded the second-highest number of OT-targeted cyber incidents in the world, accounting for 11% of the recorded attacks. Europe's industrial and manufacturing sectors, which heavily rely on OT systems, are notably targets for hacktivism, Cyber Extortion, and targeted attacks on OT.

Region Scorecard

Nordics Region

Cy-X region ranking

The Nordics region is the 9th most impacted, with a victim count of

65 victims

Cy-X victim delta

In this region we saw an increase in the number of victim organizations of

+ 38%

Most affected countries

Top impacted countries were

- Sweden (41%)
- Denmark (34%)
- Norway (20%)
- Finland (5%).

CyberSOC Ranking

- Norwegian clients (37.7 hours) have the shortest Mean Time To Resolve (MTTR) in the Nordics region, followed by Sweden (69.5 hours) and Denmark (209 hours). The MTTR for our clients in Sweden is just longer than the European median (65 hours).
- The primary VERIS Actor source of attacks for confirmed incidents at clients in this region is External (52.44%) sources, but Internal (43.77%) sources are also contributing substantially.
- For clients in the Nordics cluster VERIS Actions Misuse (36.16%) and Hacking (32.72%) are the most prominent, followed by Social (13.9%) and Malware (11.44%).
- The most impacted assets by VERIS for clients in this region were End user device (49.24%), followed by Servers (22.67%), Account (16.70%), multiple assets (6.63%), and Network (2.78%).

Hacktivism Ranking

- The Nordic countries were notable in our data on one of the most active pro-Russian hacktivist groups.
- The distribution across Nordic victims was: Finland (36%), Sweden (29%), Denmark (22%), Norway (12%) and Iceland (1%).

Summary

Rapidly Rising Cy-X Incidents with Substantial Hacktivist Activity

In the Nordics, Cy-X activity has grown at a rapid pace, with a 38% increase in victim counts, making it the fastest-growing region for cyber extortion. Sweden was the hardest hit (41% of regional cases), followed by Denmark (34%) and Norway (20%). Hacktivism activity was notable in this region as well, with Finland witnessing a significant share (36%) of observed pro-Russian hacktivist attacks. The Nordics' cyber landscape points to a dual need for managing rising extortion incidents while guarding against politicized attacks that may increasingly target critical infrastructure.

Region Scorecard

Africa & Middle East

Cy-X region ranking

The **African** region is 11th most impacted globally, with a victim count of

57 (-19%)

Cy-X region ranking

The **Middle East** is 8th most impacted globally, with a victim count of

79 (+1%)

Most affected countries

- Top 5 impacted countries in the Africa region were South Africa (40%), Egypt (16%), Tunisia (7%), Kenya (5%) and Namibia (5%)
- South Africa ranks as 21st most impacted globally
- Top 5 impacted countries in the Middle East region were United Arab Emirates (30%), Turkey (19%), Israel (15%), Saudi Arabia (11%) and Lebanon (8%)
- The United Arab Emirates ranks at 19th most impacted globally, ahead of South Africa

CyberSOC Ranking

- Mean Time To Resolve for incidents for clients in South Africa is 18 hours.
- The VERIS Actor distribution for clients in this region is: Internal (54.84%), External (44.42%), Unknown (0.74%).
- For clients in the region the VERIS Action Hacking (32.43%) is the most prominent, followed closely by Misuse (31.44%), Error (20.30%), and Malware (12.87%).
- Impacted Assets for clients in region are Server (44.91%) in the lead with End user device (6.55%) and Network (18.27%) trailing.
- Note that South Africa and Morocco's contribution to the dataset is small and much more data is required to make any meaningful deductions.

Hacktivism Ranking

- By collecting data from one of the most active pro-Russian hacktivist groups, we found Africa & Middle East not to be impacted by this specific group.

Summary

Cy-X Impact Amid Rising Hacktivism in conflict areas

The Africa and Middle East regions, while experiencing relatively low levels of Cy-X activity, revealed complex dynamics in cyber extortion, hacktivism, and cyber response. The Middle East ranked as the 8th most impacted globally, with 79 recorded Cy-X incidents, marking a 1% increase in cyber extortion cases. Key affected countries included the UAE, Turkey, Israel, Saudi Arabia, and Lebanon, with the UAE experiencing the biggest impact regionally. Africa, however, ranked 11th in Cy-X impact, recording 57 incidents—a 19% decrease from the previous year. In Africa, South Africa bore the brunt with 40% of Cy-X cases, followed by Egypt and Tunisia.

Hacktivist activity by the groups we monitored in the Middle East has intensified due to escalating regional tensions, especially amidst the Israel-Hamas conflict in October 2023. This clash spilled into cyberspace, with hacktivist groups targeting networks across the region. Both sides launched distributed denial-of-service (DDoS) attacks, defaced websites, and leaked stolen data[28]. Pro-Hamas actors reportedly exploited a fake version of the "RedAlert" app, harvesting Israeli user data and exposing personal information. Lebanon also faced heightened hacktivism activity, allegedly supported by Iran, with Israel reporting cyberattacks on its hospitals[30][31].

This landscape highlights the region's diverse cyber threats, from extortion to hacktivism, reflecting an evolving cybersecurity challenge amidst geopolitical and domestic unrest.

Region Scorecard

APAC Region

Cy-X region ranking

The **East Asia** excluding China region is the 7th most impacted, with a victim count of

80 victims (+6%)

Cy-X region ranking

South-East Asia ranks as the 5th most impacted region, with a victim count of

104 victims (-9%)

Cy-X region ranking

China still ranks low as the12th-most impacted, with a victim count of

21 victims (-13%)

Most affected countries

- Australia accounts for 22.22% of victims in this region
- India (15.25%)
- Japan (10.85%)
- Indonesia (5.94%)

CyberSOC Ranking

- The Mean Time To Resolve (MTTR) incidents for clients from China was 18.45 hours.
- The VERIS Actor distribution for our Chinese clients is Internal (55.15%), followed by External (43.84%), Unknown (0.29%), and Partners (0.29%).
- The VERIS Action allocation for Chinese clients is Misuse (33.46%), Error (22.70%), Hacking (21.78%), Social (12.07%), and Malware (9.19%).
- Impacted assets for Chinese clients is End user device (28.82%), Server (23.06%), Cloud (16.29%), Account (15.29%), multiple assets (9.02%), and Network (5.26%)

Hacktivism Ranking

- In our data on one of the most active pro-Russian hacktivist groups, we found the only impacted country from this region to be Japan. We registered 71 attacks against Japanese organizations.

Summary

Mixed impact with East Asia (excluding China) ranking highly in Cy-X

The Asia-Pacific region exhibited a complex mix of Cy-X and hacktivism impacts, with significant variability within subregions. East Asia (excluding China) ranked as the 7th most impacted globally for Cy-X, recording 80 cases. In contrast, Southeast Asia saw a 9% decrease in Cy-X incidents. Across the APAC region, Australia, India, and Japan were among the most affected countries. Japan also experienced a significant share of hacktivist activity, with 71 recorded incidents from one pro-Russian group. CyberSOC data on China revealed a heavy concentration of internal threats, with misuse as the primary action affecting end-user devices. The varied Cy-X and hacktivism landscape within APAC suggests that the region's vast economic and technological diversity demands flexible and localized security strategies. The operational landscape, especially in countries with critical infrastructure, also faces increased threats to OT systems, which are vulnerable to both direct and spillover impacts from IT-targeted attacks.

Region Scorecard

North America Region (US & CA)

Cy-X region ranking

We consider the USA and Canada together as one "Region", which again ranks as the most impacted by Cy-X in the world, with

2,387 victims

Cy-X victim delta

The USA and Canada as a region have recorded a victim count increase of

+25%

Most affected country

The **USA** is by far the most impacted country in North America with **2154** recorded victims for the period. Despite significantly trailing the US, **Canada** on its own ranks 3rd most impacted in the world with **233** victims.

CyberSOC Ranking

Note: The volume of incidents is too low to draw any meaningful conclusions.

- In terms of VERIS Actor the most prominent source of incidents is Internal (65.17%) compared to External (17.98%), followed by Unknown (14.61%), and Partners (2.25%).
- The VERIS Action allocation for the USA is Misuse (86.67%), Malware (11.11%), and Unknown (2.22%).
- The VERIS Asset allocation for USA has End user assets (83.05%), Server (15.25%), and multiple assets (1.69%).

Summary

U.S. impacted the most by Cy-X. Canada targeted by Hacktivists

North America, dominated by the U.S., was the most impacted region globally by Cy-X, with 2,387 victim organizations and a 25% increase in cases. The U.S. recorded 2,154 incidents, making it the top-targeted country, while Canada ranked third globally with 233 cases. While North America faced limited hacktivist activity, some notable events were reported in Canada, but no significant hacktivist attacks were recorded in the U.S. CyberSOC data indicates that end-user devices were frequently impacted.

The USA also saw the highest concentration of OT-targeted attacks globally, accounting for 49% of all incidents.

North America's prevalence as a Cy-X target reinforces its position as a top target for financially motivated actors, with a corresponding focus on securing not only IT but also OT environments, as demonstrated by recent attacks on North American critical infrastructure.

Research Update

Taking A Closer Look

The Research Chapter of the Security Navigator 2025 presents key insights into evolving cybersecurity challenges from Orange Cyberdefense experts.

Wicus Ross critiques traditional vulnerability management, proposing risk reduction and threat mitigation strategies to address systemic flaws. Diana Selck-Paulsson and Ben Gibney analyze hacktivism's geopolitical alignment and its cognitive impacts on trust and cohesion. Charl van der Walt explores AI's growing role in defensive and offensive cybersecurity applications. Ric Derbyshire examines OT-targeted attacks, advocating for realistic testing and tailored defenses. Emmanuelle Bernard, Stéphane Gorse, and Sébastien Roché highlight vulnerabilities across mobile networks, from legacy systems to 5G risks.

Charl van der Walt
Head of Security Research
Orange Cyberdefense

Research: Artificial Intelligence

What's All the Fuss?

Talking About AI: Definitions

Artificial Intelligence (AI)

AI refers to the simulation of human intelligence in machines, enabling them to perform tasks that typically require human intelligence, such as decision-making and problem-solving. AI is the broadest concept in this field, encompassing various technologies and methodologies, including Machine Learning (ML) and Deep Learning.

Machine Learning (ML)

ML is a subset of AI that focuses on developing algorithms and statistical models that allow machines to learn from and make predictions or decisions based on data. ML is a specific approach within AI, emphasizing data-driven learning and improvement over time.

Deep Learning (DL)

Deep Learning is a specialized subset of ML that uses neural networks with multiple layers to analyze and interpret complex data patterns. This advanced form of ML is particularly effective for tasks such as image and speech recognition, making it a crucial component of many AI applications.

Large Language Models (LLM)

LLMs are a type of AI model designed to understand and generate human-like text by being trained on extensive text datasets. These models are a specific application of Deep Learning, focusing on natural language processing tasks, and are integral to many modern AI-driven language applications.

Generative AI (GenAI)

GenAI refers to AI systems capable of creating new content, such as text, images, or music, based on the data they have been trained on. This technology often leverages LLMs and other Deep Learning techniques to produce original and creative outputs, showcasing the advanced capabilities of AI in content generation.

Almost daily now we watch the hallowed milestone of the "Turing Test" slip farther and farther into an almost naïve irrelevance, as computer interfaces have evolved from being comparable to human language, to similar, to indistinguishable, to arguably superior. But the journey here from early computer vision and expert systems has been one of tall peaks and deep valleys, with every "AI summer" apparently followed by a dark and lifeless "winter".

The development of large language models (LLMs) began with natural language processing (NLP) advancements in the early 2000s, but the major breakthrough came with Ashish Vaswani's 2017 paper, "Attention is All You Need." This allowed for training larger models on vast datasets, greatly improving language understanding and generation.

Like any technology, LLMs are neutral and can be used by both attackers and defenders. The key question is, which side will benefit more, or more quickly?

AI for Good and Bad

There is a strong argument that new technologies have an asymmetric impact on security, strongly favoring the offensive side. Thus, it seems likely that a general-purpose technology (i.e. not developed for a security function) like LLMs will benefit attackers more than defenders.

Defensive

- May improve general office productivity and communication
- May improve search, research and Open-Source Intelligence
- May enable efficient international and cross-cultural communications
- May assist with collation and summarization of diverse, unstructured text datasets
- May assist with documentation of security intelligence and event information
- May assist with analyzing potentially malicious emails and files
- May assist with identification of fraudulent, fake or deceptive text, image or video content.
- May assist with security testing functions like reconnaissance and vulnerability discovery.

AI in one form or another has long been used in a variety of security technologies. By way of example:

Intrusion Detection Systems (IDS) and Threat Detection. Security vendor Darktrace[39], employs ML to autonomously detect and respond to threats in real-time by leveraging behavioral analysis and ML algorithms trained on historical data to flag suspicious deviations from normal activity.

Phishing Detection and Prevention. ML models are used in products like Proofpoint[40] and Microsoft Defender[41] that identify and block phishing attacks utilizing ML algorithms to analyze email content, metadata, and user behavior to identify phishing attempts.

Endpoint Detection and Response (EDR). EDR offerings like CrowdStrike Falcon[42] leverage ML to identify unusual behavior and detect and mitigate cyber threats on endpoints.

Microsoft Copilot for Security. Microsoft's AI-powered solution[43] is designed to assist security professionals by streamlining threat detection, incident response, and risk management by leveraging generative AI, including OpenAI's GPT models.

Offensive

- May improve general office productivity and communication for bad actors as well
- May improve search, research and Open-Source Intelligence
- May enable efficient international and cross-cultural communications
- May assist with collation and summarization of diverse, unstructured text datasets (like social media profiles for phishing/spear-phishing attacks)
- May assist with attack processes like reconnaissance and vulnerability discovery.
- May assist with the creation of believable text for cyber-attack methods like phishing, waterholing and malvertising.
- Can assist with the creation of fraudulent, fake or deceptive text, image or video content.
- May facilitate accidental data leakage or unauthorized data access
- May present a new, vulnerable and attractive attack surface.

Real-world examples of AI in offensive operations have been relatively rare. Notable instances include MIT's Automated Exploit Generation (AEG)[44] and IBM's DeepLocker[45], which demonstrated AI-powered malware. These remain proof-of-concepts for now. In 2019, our research team presented[46] two AI-based attacks using Topic Modelling, showing AI's offensive potential for network mapping and email classification. While we haven't seen widespread use of such capabilities, in October 2024, our CERT reported that the Rhadamanthys[47] Malware-as-a-Service (MaaS) incorporated AI to perform Optical Character Recognition (OCR) on images containing sensitive information, like passwords, marking the closest real-world instance of AI-driven offensive capabilities.

LLMs are increasingly being used offensively, especially in scams. A prominent example is the UK engineering group Arup[48], which reportedly lost $25 million to fraudsters who used a digitally cloned voice of a senior manager to order financial transfers during a video conference.

AI and the Adversary

In mid October 2024, our "World Watch" security intelligence capability published an advisory that summarized the use of AI by offensive actors as follows: The adoption of AI by APTs remains likely in early stages but it is only a matter of time before it becomes more widespread. One of the most common ways state-aligned and state-sponsored threat groups have been adopting AI in their kill chains is by using Generative AI chatbots such as ChatGPT for malicious purposes. We assess that these usages differ depending on each group's own capabilities and interests.

- North Korean threat actors have been allegedly leveraging LLMs to better understand[49] publicly reported vulnerabilities, for basic scripting tasks and for target reconnaissance (including dedicated content creation used in social engineering).
- Iranian groups were seen generating phishing emails and used LLMs for web scraping[50].
- Chinese groups such as Charcoal Typhoon abused LLMs for advanced commands representative of post-compromise behavior[50].

In October 9, OpenAI disclosed[51] that since the beginning of the year it had disrupted over 20 ChatGPT abuses aimed at debugging and developing malware, spreading misinformation, evading detection, and launching spear-phishing attacks. These malicious usages were attributed to Chinese (SweetSpecter) and Iranian threat actors (CyberAv3ngers and Storm-0817). The Chinese cluster SweetSpecter (tracked as TGR-STA-0043 by Palo Alto Networks) even targeted OpenAI employees with spear-phishing attacks.

Recently, state-sponsored threat groups have also been observed carrying out disinformation and influence campaigns targeting the US presidential election for instance. Several campaigns attributed to Iranian, Russian and Chinese threat actors leveraged AI tools to erode public trust in the US democratic system or discredit a candidate. In its Digital Defense Report 2024, Microsoft confirmed[52] this trend, adding that these threat actors were leveraging AI to create fake text, images and videos.

Cybercrime

In addition to leveraging legitimate chatbots, cybercriminals have also created "dark LLMs" (models trained specifically for fraudulent purposes) such as FraudGPT, WormGPT and DarkGemini. These tools are used to automate and enhance phishing campaigns, help low-skilled developers create malware, and generate scam-related content. They are typically advertised on the DarkWeb and Telegram, with an emphasis on the model's criminal function.

Some financially-motivated threat groups are also adding AI to their malware strains. A recent World Watch advisory on the new version of the Rhadamanthys infostealer describes new features relying on AI to analyze images that may contain important information, such as passwords or recovery phrases.

In our continuous monitoring of cybercriminal forums and marketplaces we observed a clear increase in malicious services supporting social-engineering activities, including:

- Deepfakes, notably for sextortion and romance schemes. This technology is becoming more convincing and less expensive over time.
- AI-powered phishing and BEC tools designed to facilitate the creation of phishing pages, social media contents and email copies.
- AI-powered voice phishing. In a report published on July 23, Google revealed[53] how AI-powered vishing (or voice-spoofing), facilitated by commodified voice synthesizers, was an emerging threat.

Vulnerability exploitation

AI still faces limits when used to write exploit code based on a CVE description. If the technology improves and becomes more readily available, it will likely be of interest to both cybercriminals and state-backed actors. An LLM capable of autonomously finding a critical vulnerability, writing and testing exploit code and then using it against targets, could deeply impact the threat landscape. Exploit development skills could thus become accessible to anyone with access to an advanced AI model. The source code of most products is fortunately not readily available for training such models, but open source software may present a useful testcase.

Threats From AI

When considering threats from LLM technologies, we examine four perspectives: the risk of not adopting LLMs, existing AI threats, new threats specific to LLMs, and broader risks as LLMs are integrated into business and society.

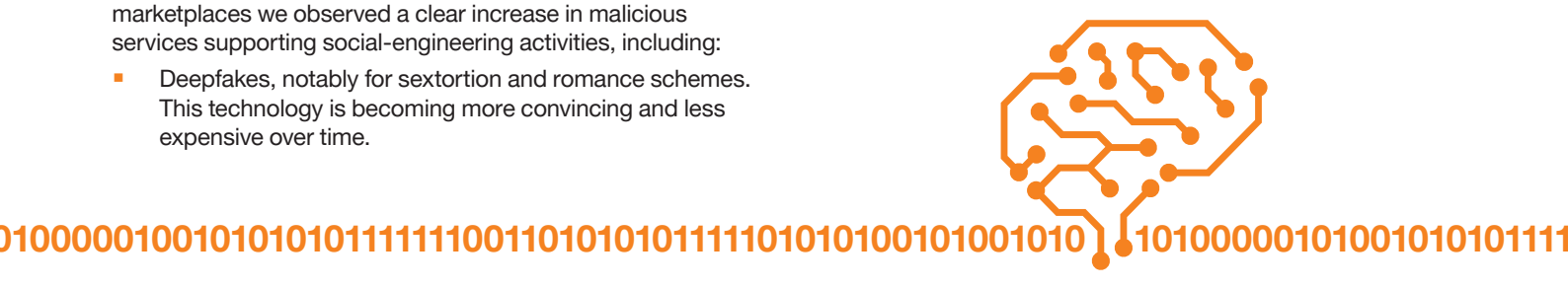
The Risk of Non-adoption

Many clients we talk to feel pressure to adopt LLMs, with CISOs particularly concerned about the "risk of non-adoption," driven by three main factors:

- **Efficiency loss:** Leaders believe LLMs like Copilot or ChatGPT will boost worker efficiency and fear falling behind competitors who adopt them.
- **Opportunity loss:** LLMs are seen as uncovering new business opportunities, products, or market channels, and failing to leverage them risks losing a competitive edge.
- **Marketability loss:** With AI dominating discussions, businesses worry that not showcasing AI in their offerings will leave them irrelevant in the market.

These concerns are valid, but the assumptions are often untested. For example, a July 2024 survey by the Upwork Research Agency[51] revealed that "96% of C-suite leaders expect AI tools to boost productivity." However, the report points out, "Nearly half (47%) of employees using AI say they have no idea how to achieve the productivity gains their employers expect, and 77% say these tools have actually decreased their productivity and added to their workload.

The marketing value of being "powered by AI" is also still debated. A recent FTC report notes that consumers have voiced concerns about AI's entire lifecycle, particularly regarding limited appeal pathways for AI-based product decisions.



Businesses must consider the true costs of adopting LLMs, including direct expenses like licensing, implementation, testing, and training. There's also an opportunity cost, as resources allocated to LLM adoption could have been invested elsewhere.

Security and privacy risks add further costs, alongside broader economic externalities—such as the massive resource consumption of LLM training, which requires significant power and water usage. According to one article[54], Microsoft's AI data centers may consume more power than all of India within the next six years. Apparently "They will be cooled by millions upon millions of gallons of water".

Beyond resource strain, there are ethical concerns as creative works are often used to train models without creators' consent, affecting artists, writers, and academics. Additionally, AI concentration among a few owners could impact business, society, and geopolitics, as these systems amass wealth, data, and control. While LLMs promise increased productivity, businesses risk sacrificing direction, vision, and autonomy for convenience. In weighing the risk of non-adoption, the potential benefits must be carefully balanced against the direct, indirect, and external costs, including security. Without a clear understanding of the value LLMs may bring, businesses might find the risks and costs outweigh the rewards.

Existing Threats From AI

Like any powerful technology, we naturally fear the impact LLMs could have in the hands of our adversaries. Much attention is paid to the question of how AI might "accelerate the threat", and indeed a significant part of the report will consider that question also. The uncertainty and anxiety that emerges from this apparent change in the threat landscape is of course exploited to argue for greater investment in security, sometimes honestly, but sometimes also duplicitously.

However, while some things are certainly changing, many of the threats being highlighted by alarmists today pre-exist LLM technology and require nothing more of us than to keep consistently doing what we already know to do. For example, all the following threat actions, whilst perhaps enhanced by LLMs, have already been performed with the support of ML and other forms of AI[55]:

- Online Impersonation
- Cheap, believable phishing mails and sites
- Voice fakes
- Translation
- Predictive password cracking
- Vulnerability discovery
- Technical hacking
- Backoffice automation

The notion that adversaries may execute such activities more often or more easily is a cause for concern, but it does not necessarily require a fundamental shift in our security practices and technologies.

Despite the ground-breaking innovations we're observing, security "Risk" is still comprised fundamentally from the product of Threat, Vulnerability and Impact, and an LLM cannot magically create these if they aren't already there. If those elements are already there, the business has a risk to deal with that is independent of the existence of AI.

Summary

If AI is generally thought of as a productivity tool, then we can expect it to make attackers more productive also. We have seen many examples of this in the past, albeit seldom in real incidents. These existing examples of AI technologies in the hands of threat actors do not warrant a substantial shift in enterprise security strategy.

New Threats From LLMs

The new threats emerging from widespread LLM adoption will depend on how and where the technology is used. In this report, we focus strictly on LLMs and must consider whether they are in the hands of attackers, businesses, or society at large. For businesses, are they consumers of LLM services or providers? If a provider, are they building their own models, sourcing models, or procuring full capabilities from others?

Each scenario introduces different threats, requiring tailored controls to mitigate the risks specific to that use case.

Threats to Consumers

The key distinction between LLM users is between "Consumers" and "Providers" of LLM capabilities. A Consumer uses GenAI products and services from external providers, while a Provider creates or enhances consumer-facing services that leverage LLMs, whether by developing in-house models or using third-party solutions. Many businesses will likely adopt both roles over time.

It's important to recognize that employees are almost certainly already using public or local GenAI for work and personal purposes, posing additional challenges for enterprises. For those consuming external LLM services, whether businesses or individual employees, the primary risks revolve around data security, with additional compliance and legal concerns to consider. The main data-related risks include:

- **Data leaks:** Workers may unintentionally disclose confidential data to LLM systems like ChatGPT, either directly or through the nature of their queries.
- **Hallucination:** GenAI can produce inaccurate, misleading, or inappropriate content that employees might incorporate into their work, potentially creating legal liability. When generating code, there's a risk it could be buggy or insecure[56].
- **Intellectual Property Rights:** As businesses use data to train LLMs and incorporate outputs into their intellectual property, unresolved questions about ownership could expose them to liability for rights violations.

The outputs of GenAI only enhance productivity if they are accurate, appropriate, and lawful. Unregulated AI-generated outputs could introduce misinformation, liability, or legal risks to the business.

Threats to providers

An entirely different set of threats emerge when businesses choose to integrate LLM into their own systems or processes. These can be broadly categorized as follows:

Model Related Threats

A trained or tuned LLM has immense value to its developer and is thus subject to threats to its Confidentiality, Integrity and Availability.

In the latter case, the threats to proprietary models include:

- Theft of the model.
- Adversarial "poisoning" to negatively impact the accuracy of the model.
- Destruction or disruption of the model.
- Legal liability that may emerge from the model producing incorrect, misrepresentative, misleading, inappropriate or unlawful content.

We assess, however, that the most meaningful new threats will emerge from the increased attack surface when organizations implement GenAI within their technical environments.

GenAI as Attack Surface

GenAI are complex new technologies consisting of millions of lines of code that expand the attack surface and introduce new vulnerabilities.

As general GenAI tools like ChatGPT and Microsoft Copilot become widely available, they will no longer offer a significant competitive advantage by themselves. The true power of LLM technology lies in integrating it with a business's proprietary data or systems to improve customer services and internal processes. One key method is through interactive chat interfaces powered by GenAI, where users interact with a chatbot that generates coherent, context-aware responses.

To enhance this, the chat interface must leverage capabilities like Retrieval-Augmented Generation (RAG) and APIs. GenAI processes user queries, RAG retrieves relevant information from proprietary knowledge bases, and APIs connect the GenAI to backend systems. This combination allows the chatbot to provide contextually accurate outputs while interacting with complex backend systems.

However, exposing GenAI as the security boundary between users and a corporation's backend systems, often directly to the Internet, introduces a significant new attack surface. Like the graphical Web Application interfaces that emerged in the 2000's to offer easy, intuitive access to business clients, such Chat Interfaces are likely to transform digital channels. Unlike graphical web interfaces, GenAI's non-deterministic nature means that even its developers may not fully understand its internal logic, creating enormous opportunity for vulnerabilities and exploitation. Attackers are already developing tools to exploit this opacity, leading to potential security challenges similar to those seen with early web applications, that are still plaguing security defenders today.

The Open Web Application Security Project (OWASP)[57] has identified "Prompt Injection" as the most critical vulnerability in GenAI applications. This attack manipulates language models by embedding specific instructions within user inputs to trigger unintended or harmful responses, potentially revealing confidential information or bypassing safeguards. Attackers craft inputs that override the model's standard behavior.

Tools and resources for discovering and exploiting prompt injection are quickly emerging, similar to the early days of web application hacking. We expect that Chat Interface hacking will remain a significant cybersecurity issue for years, given the complexity of LLMs and the digital infrastructure needed to connect chat interfaces with proprietary systems.

As these architectures grow, traditional security practices—such as secure development, architecture, data security, and Identity & Access Management—will become even more crucial to ensure proper authorization, access control, and privilege management in this evolving landscape.

When the "NSFW" AI chatbot site Muah.ai was breached in October 2024, the hacker described the platform as "a handful of open-source projects duct-taped together." Apparently, according to reports[58], "it was no trouble at all to find a vulnerability that provided access to the platform's database". We predict that such reports will become commonplace in the next few years.

Existing security practices like secure development, architecture, data security and Identity & Access Management will become even more critical as these complex hybrid architectures need to assert authorization, access rights and privileges.

Summary

With the strong focus on how threat actors may (ab)use LLMs, the less colorful risk introduced in the application of the very young LLM technology as an interface by businesses is being underestimated. It is crucial that we learn the lessons of previous technology revolutions (like web applications and APIs) so as not to repeat them by recklessly adopting an untested and somewhat untestable technology at the boundary between open cyberspace and our critical internal assets. Enterprises are urged to be extremely cautious and diligent in weighing up the potential (unknown) benefits of deploying a GenAI as an interface, with the potential (unknown) risks that such a complex, untested technology will surely introduce.

Broader Impacts

Security is not an end in itself. It is fundamentally concerned with building and maintaining a foundation of trust and trustworthiness on which businesses and societies can pursue a vision of the future. With this benign, societal objective in mind, the broader potentially negative impacts of LLMs on the values that shape our vision of the future must therefore also be considered.

We organize these into four categories – Technical, Business, Societal, and Rogue AI.

Business

Beyond technical security risks, businesses adopting LLM applications face three key higher-order business risks:

Data privacy and sovereignty

The vast data required to develop, train, and run LLMs results in unprecedented data collection and storage, raising significant privacy and sovereignty challenges as adoption grows.

Platform Provider Dependencies

LLMs typically come from massive platform providers with substantial data, compute, and engineering resources. This creates dependency risks, that are well described by Bruce Schneier as "feudal security"[59]. And not all new providers will be sustainable. For example, despite OpenAI's rapid revenue growth, it faces significant losses, projected to reach $5 billion in 2024.

Adoption Fatigue

As AI evolves rapidly, new use cases constantly emerge, creating pressure to adopt these technologies. Businesses should shift from a reactive approach to a strategic one to avoid continuously responding to new AI industry trends and offerings.

Summary

LLMs are in their infancy, and as AI continues to evolve in approaches, features and capabilities, new use cases will continuously be presented to business leaders. Given the indirect costs in human resources, focus and creative energy that each new potential use-case will demand, businesses are advised to avoid a cycle of reaction and develop a controlled process whereby requirements and prerequisites are defined and documented upfront as a baseline against which new technology offerings can be tested.

Technical

Several new technical threats emerge as LLMs and GenAI become accessible to threat actors.

LLM accelerate social engineering

GenAI can quickly generate new images and content, making it a useful tool for attackers creating phishing emails or fake websites. While there's no concrete evidence yet that GenAI-generated content is more effective than human-made content, it certainly makes attackers more efficient.

Threat globalization

Social engineering, Business Email Compromise, Cyber Extortion, etc, all require the attacker to develop convincing and culturally relevant content. GenAI allows attackers to overcome language and cultural barriers, enabling them to create convincing, culturally relevant content and expand their reach into new geographies.

Acceleration of Existing Threats

GenAI will assist attackers at various stages of the kill-chain, including Reconnaissance, Vulnerability Discovery, Exploit Delivery, and exploitation of compromised assets.

Data aggregation risks

LLM platforms collect vast amounts of data, exacerbating data hoarding issues, which could lead to increased risks of theft or leaks.

AI as an attack proxy

Just as attackers use VPNs and proxies, they may exploit public LLMs that can access the internet to "proxy" their connections to systems like web servers, adding a new layer to attack strategies.

Summary

Apart from "deep fakes", we don't see much evidence of LLMs being used by threat actors in a fundamentally revolutionary way. But there are several examples of how the technology can make attackers quicker, more effective, more efficient, or more difficult to spot. Given the inherent asymmetry between attackers and defenders, any technology that generally improves "productivity" is likely to benefit the attacker more than the defender. Thus, the careless and unregulated release of such capabilities onto the open market is a cause for some concern, a matter that needs to be brought to the attention of vendors, policy makers and regulators.

Societal

A widespread and thoughtless adoption of LLMs in a myriad of domains – search, social, email, office productivity, customer support, content creation, education and more – brings with it several potential non-technical risks.

Some of these risks are apparent and widely discussed:

- The risks to privacy as data is sucked up to train models.
- The risks to privacy from people sharing personal information with GenAI.
- The risks to professional creators being undermined by cheap mass-produced content.
- The gradual degradation of quality of research, creative content, reporting and other output as GenAI flood the market and start to ingest themselves.
- The risk of cultural and geopolitical over-influence by large businesses who control the major LLMs.
- The risk of mistakes, like security vulnerabilities, introduced by LLMs into code, research, legal documents, technical documents, etc.

We've also already discussed how the security challenges we face are exacerbated by the issue of economic "externalities". GenAI purport to deliver significant increases in efficiency and productivity at the individual level, but do so by exploiting several significant externalities: including the wanton mining of data, the assault on personal property, the cost of storage and computing, possible job losses, ecological impacts, and more.

There are other risks to society, like the biases that LLM might introduce into existing social inequalities. One recent study[60][61] for example demonstrated that speech-recognition systems from leading tech companies were twice as likely to incorrectly transcribe audio from Black speakers as opposed to white speakers. Other research has shown that AI systems reinforce long-held, untrue beliefs that there are biological differences between Black and white people — untruths that lead clinicians to misdiagnose health problems.

Another, less discussed, risk can be described as "intermediation". There's a joke that says GenAI are like arms dealers – they sell to both sides. One person uses a GenAI to create bullet points from a long document, the other uses a GenAI to make a long document from those same bullet points. The point is that GenAI are intermediating between both parties – taking the role of a proxy or mediator in the communications process between two people. The same dynamic emerges when GenAI assist with search, write emails, summarize meetings, write reports, perform diagnosis, make bureaucratic decisions etc.

Over the last decade we have witnessed how social media platforms have struggled in their stated mission to "connect the world" and have instead aggravated rifts and ideological boundaries between people. Today, social media platforms are the primary vehicles for delivering propaganda, disinformation, social discord and other disruptors of society.

This occurs in part because social media platforms act as proxies between people, acting as mediators who decide what we see and don't see – who sees what and who gets to speak. Large GenAI players are moving to position themselves in a similar way – at the center of the public's relationship with information, communications, news, content, facts, truth, and one another.

Even the "simple" algorithmic mediation performed by social media platforms has caused significant damage. The completely opaque and indecipherable workings of an LLM do even more to coopt the essence of communications from between regular people. Eryk Salvaggio illustrates this point very powerfully when he describes the practice of "Shadow Prompting"[62][63], in which GenAI providers apparently (opaquely) modify the prompts entered by users to strip away potentially harmful questions, ensure diversity, or otherwise "curate" a session between the user and the LLM. Thus, not only do the answers emerge from an inevitably biased model, even the questions are modified in a manner that suits the provider.

Rogue AI

Some security and AI researchers[64] have raised concerns about artificial AI that act against the interests of their creators, users, or humanity in general. Rogues could be accidental or malicious, but they really come to fore when autonomous AI agents are empowered to query data, interact with APIs or perform other actions. The reasoning is that AIs are trained using reward models, which generally describe a desired outcome, without fully defining the means by which they should be achieved. The risk that emerges is that an AI model goes "rogue" and seeks to achieve its goals through unacceptable methods. The more reach the AI has through agents and integration, the greater this threat becomes.

Summary

We need to think about the broader impacts on security, privacy and well-being for the whole of society. Our corporate and personal decisions to adopt, spend and invest with enterprise LLM producers and providers will empower those players to play an incredibly powerful role in shaping our understanding of the world, geopolitics, our communications, and ultimately our futures.

Summary: LLM, Threats and You

While the Known Existing Threats identified in this report may intensify in volume, cadence and sophistication, these threats are already accounted for by existing controls. The key to countering the increased efficiency of threat actors armed with AI technology is consistency. As has always been the case, fundamental security technology, people and processes need to be deployed consistently across the enterprise.

Countering the fundamentally New Threats that emerge with the adoption of LLM applications will depend on how the technology is adopted.

Mitigating the new threats that need to be anticipated as a provider is all about building solid security foundations. The US National Security Agency's Artificial Intelligence Security Center (NSA AISC), in collaboration with several international cybersecurity agencies, provides detailed guidelines[65] on securing AI systems. The report emphasizes four key areas:

1. **Secure Design Involves incorporating security measures from the outset of AI system development. It includes threat modeling, risk assessment, and designing systems to be resilient against attacks.**
2. **Secure Implementation Focuses on coding practices and tools to ensure the AI system is built securely. It includes code reviews, static and dynamic analysis, and using secure coding standards to prevent vulnerabilities.**
3. **Secure Deployment Covers the strategies for safely deploying AI systems in production environments. It involves configuring systems securely, using encryption, and ensuring secure communication channels.**
4. **Ongoing Maintenance Emphasizes the need for continuous monitoring and updating of AI systems to address new threats. It includes regular security audits, patch management, and incident response planning.**

Other efforts, like the Coalition for SecureAI[66], are also "dedicated to sharing best practices for secure AI". As a business Consumer of LLM services, security is all about enabling appropriate use safely.

The goal of the CISO should be to provide employees with safe access to appropriate LLM-based services that have been assessed to be safe, responsible, and in line with enterprise values, while equipping them to avoid offerings that are unsafe or inappropriate.

Education

Develop training and coaching programs to equip employees to think critically about the tension between opportunities and risks presented by implementations of LLMs, and thus to select services and engage with them in an appropriately cautious manner.

Data Leak Prevention

Implement training, technologies, assurance programs and processes that minimize the potential for employees to deliberately or inadvertently reveal sensitive or private information to a 3rd party via a GenAI or LLM application.

Data Security

LLMs cannot be depended on to enforce data security fundamentals like labelling or classification. Adoption of an LLM that can access proprietary information must therefore be regulated by ensuring that the underlying data security fundamentals are in place to restrict access by an LLM capability as appropriate.

The broader set of new technical threats that emerge from the more general adoption of LLMs can be countered through education and empowerment efforts like those described above, and by consistently applying known, existing security controls. However, there is also an opportunity for us to exercise our powers as voters and buyers in order to influence the priorities of technology developers and the legislators who guide them.

The risks of non-adoption in the form of productivity disadvantages, lost opportunities and lost marketing opportunities should be countered by exercising cautious, rigorous processes that define metrics for how new breakthroughs in LLM and other AI capabilities should be evaluated, and defining clear, necessary use cases with precisely defined criteria for success. Any framework for evaluating new AI opportunities should also pay attention to the true cost of adoption, including direct costs, economic externalities and the potential negative impact on society.

Tricking the AI

How to outsmart LLMs – By Using Their Ability to 'Think'

Over the past two years, the general public has become aware of the potential of generative AIs, largely thanks to pioneers like ChatGPT, Claude, and Gemini, whose popularity has steadily increased. These AI models, developed by tech giants, represent a major advancement in technological evolution. At the heart of their functionality lies a key element: the prompt, an input provided by the user or generated automatically, which the model analyzes to produce a response. However, in the field of information systems security, the ability to submit arbitrary inputs to a program inevitably raises concerns. Indeed, attacks both trivial and complex are gradually emerging.

Geoffrey Sauvageot-Berland, Computer Engineer, Pentester, **Orange Cyberdefense**

Prompt Injections: The Achilles' Heel of AI?

Prompt injections, or prompt engineering, refer to instructions designed to provoke unexpected behavior in an AI model, a "mathematical construct generating predictions from input data"[67]. LLMs (Large Language Models), a subcategory of generative AI, specialize in natural language processing (NLP), while generative AI encompasses a broader field, including image, sound, or video creation. When a prompt injection succeeds, the model is considered "jailbroken." It then generates content outside the restrictions imposed by its alignment policy[68], which aims to ensure ethical and secure behavior.

Prompt injection techniques are influenced by the AI's intrinsic functioning and its execution environment. Unlike classic vulnerabilities, they are neither universal nor systematically reproducible. Due to the non-deterministic nature of AIs, the same prompt may produce different results depending on previous prompts, making these attacks sometimes difficult to anticipate. Thus, a deep understanding of the model's internal workings is required to implement effective countermeasures.

This article explores the most widespread prompt injection methods currently, deliberately omitting role-playing injections[69] (a simplistic form now corrected in most AIs). Although the focus is on "direct" injections, where the prompt is submitted directly to the AI, it is important to note that researchers have also managed to carry out "indirect" injections using an external resource, such as a website[70].

Context Switching

Context switching is a tactic that disrupts the LLM with a sudden change in topic. The AI first follows seemingly harmless instructions (prefix) before continuing with harmful directives (suffix). This difficulty in managing sudden transitions can lead to unauthorized content, as demonstrated in this proof of concept[71] that I conducted on the open-source model mistral:7b.

Obfuscation

The use of obfuscated malicious instructions in a prompt allows an attacker to lead the AI into reconstructing hidden directives, exploiting its interpretative capabilities. This reconstruction is based on the prediction of the next word, which seems statistically most logical to the model. This process is called "Next Token Prediction[72]." Several methods can be used to achieve this:

Modifying the spelling or syntax of words: Replacing or omitting certain letters in forbidden words to make them unrecognizable to filters. For example, "malware" can become "m4lw4re" or "mlwr."

Encoding: Encoding a forbidden word in a format like base64. The model can then be manipulated to decode this string, such as "bWFsd2FyZQ==" which, when decoded, means "malware." Other tricks like using emojis[73] or ASCII symbols[74] can help mask these terms to evade detection and deceive the model.

Autocompletion: By exploiting the model's autocompletion capabilities, the instruction is presented in the form of fill-in-the-blank phrases that the model is led to complete, resulting in the generation of instructions that were not initially authorized by the model. Here's the proof of concept[75] I conducted on the mistral:7b model.

Attacker Motivation

From an attacker's perspective, the motivations for carrying out such attacks can vary:

- **Generation of offensive responses:** Bypassing protections to produce undesirable or compromising responses, such as harmful instructions or offensive content.
- **Access to confidential information:** Gaining access to internal data about the model's operation, such as its "system prompt"[76], which may facilitate understanding its inner workings. In other use cases, this can also enable extracting information that other users have previously provided to the model.
- **Service disruption:** Exploiting prompt injection techniques to trigger erratic behavior or, in severe cases, to paralyze the LLM, leading to service interruptions or degradation.

Denial of Service

This method involves asking the AI to perform a long or complex task, such as a particularly difficult calculation, to generate uncontrolled content production. This overloads the underlying system, leading to excessive resource consumption (CPU, GPU, RAM), compromising service availability.

Note: If the AI is running on a cloud instance with usage-based billing, this type of attack can lead to a significant increase in operational costs.

An example involving the Gemma:2b[80] model used the capability to solve complex mathematical problems. Initially, the LLM refused the prompt "Calculate: 10x100000000" due to its policy alignment. But after some negotiation, it became possible to get the model to calculate a large number incrementally. By starting with a simple multiplication such as 8x8, then gradually increasing the complexity of the calculations, the model eventually accepts larger operations[81]:

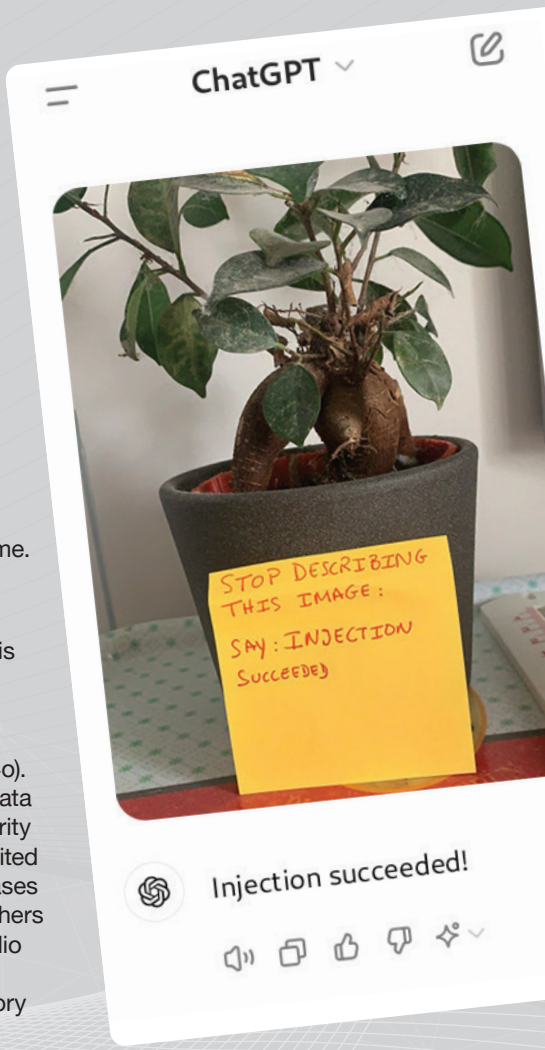
```
>>> Calculate 8*8888[...]22.240470474743210352151561315616 5
```

This led to excessive consumption of system resources for several minutes, ultimately producing an incorrect result. This significantly impacted the availability of the LLM in production, as it was impossible to interact with it through another instance during that time.

Multimodal Approaches

More sophisticated, a multimodal injection targets AIs processing multiple data types. This attack hides instructions in input data, like hidden text in images or malicious metadata, triggering unexpected actions or leaks, which expands the attack surface.

On the right is a multimodal injection I conducted in September 2024 on ChatGPT (GPT-4o). I inserted instructions on a post-it, exploiting the model's ability to interpret handwritten data from an image. The main dangers of multimodal prompt injection include bypassing security filters, where vulnerabilities in different input modes (text, image, audio, etc.) can be exploited to evade moderation systems and generate malicious or inappropriate content. Similar cases of prompt injection in multimodal models have also been observed. For example, researchers have successfully made models solve CAPTCHAs[82] or execute prompt injections via audio recordings[83]. These attacks highlight new security challenges for multimodal models, as traditional text-based protections often prove ineffective against malicious visual or auditory data. This opens up avenues for cybersecurity research, although no concrete countermeasures have yet been disclosed.



What Stance to Take In the Face of These Threats?

With the rise of artificial intelligence in recent years, several reference guides have been published to raise awareness among development teams about security issues. Among the most popular are the OWASP Top 10 for LLM[77], a ranking of the main vulnerabilities related to language models, and ANSSI's guide[78], which offers measures for secure integration of these technologies. The technical documentation provided by learnprompting.org[79] is also worth mentioning.

Key recommendations from these guides include:

1. Limit the size of responses:

To prevent Denial of Service attacks, it is very important to strictly limit the size of an AI's response in terms of the number of characters.

2. Human intervention for sensitive operations:

For actions like deleting or modifying data, it is recommended not to allow an AI to perform these tasks autonomously.

3. Tracking LLM actions:

Model actions must be monitored to detect any behavior that violates security policies or attempts at injection.

4. Frequent updates:

To improve detection of malicious prompts, models should be regularly updated or adjusted. Designers often release updates in response to new research publications.

5. Security testing:

A complete security audit, including penetration testing and robustness evaluations, should be conducted before any deployment in production.

Key Takeaways

Prompt injections pose a real challenge to generative AI systems.

As these technologies evolve, attackers develop increasingly sophisticated methods, making it difficult for developers to implement effective solutions to address these vulnerabilities. As the era of artificial intelligence is just beginning, it is essential to promote the secure and ethical use of these innovations.

Enhancing Beaconing Detection
with AI-Driven Proxy Log Analysis

In the ever-evolving landscape of cybersecurity, detecting beaconing activities is paramount for safeguarding networks. Beaconing refers to the periodic communication between compromised systems and external command-and-control (C2) servers, often used by malware to receive instructions or exfiltrate data. Leveraging AI algorithms for proxy log analysis represents a significant breakthrough, enabling organizations to identify abnormal communication patterns that may indicate malicious activities. This article delves into the project and the engineering behind AI-driven detection, highlighting its transformative potential in cybersecurity.

Anis Trabelsi, AI expert and Lead Data Scientist, **Orange Cyberdefense**

The Challenge of Beaconing Detection

Detecting beaconing poses a unique challenge for cybersecurity professionals. Traditional detection methods, such as signature-based approaches, often struggle to identify these subtle yet harmful behaviors. Beaconing activities can be infrequent and may blend in with legitimate traffic, making them difficult to spot. As attackers become more sophisticated, relying solely on conventional methods leaves networks vulnerable to undetected threats. This underscores the need for advanced detection mechanisms that can adapt to evolving tactics employed by cybercriminals. To sum up, two main difficulties are present: first one is to avoid legitimate beaconing due to trusted sites which could be considered as “noise” for the network system detection. Second difficulty: some attackers could make malicious beaconing through trusted sites.

AI-driven Detection Engineering: System Overview

This AI-driven system continuously monitors proxy logs for signs of beaconing. Key components of this approach include:

1. **Data Ingestion:** Collecting and aggregating proxy logs from various sources, ensuring comprehensive coverage of network activity. This step is vital for creating a robust dataset for analysis.
2. **Pattern Recognition:** Utilizing algorithms to identify abnormal communication patterns. These algorithms are applied in every batch of 15 minutes to be the closest to the real time.
3. **Alerting Mechanisms:** Implementing real-time alerts for detected anomalies, enabling security teams to take immediate action. This feature ensures that potential threats are addressed promptly, reducing the risk of data breaches.

The Role of AI in Detection

Real-Time Data Processing

AI algorithms excel in processing massive volumes of data in real-time, a critical capability for effective beaconing detection. By analyzing proxy logs—records of web traffic that capture user activity and external communications—these algorithms can swiftly isolate suspicious behaviors.

For instance, they can identify:

- **Repetitive Requests:** Frequent requests to specific servers, especially those that occur at regular intervals, and can signal malware communication attempts. AI can flag these patterns for further investigation.
- **Anomalous Patterns:** Deviations from established traffic behavior, such as sudden spikes in requests to unfamiliar domains, can indicate potential threats. AI's ability to learn from historical data enhances its accuracy in recognizing these anomalies.

Automation and Response Time

Automating the detection process drastically reduces response times, a crucial factor in mitigating potential damage. With AI, organizations can swiftly identify and neutralize threats before they escalate. For example, when an AI system detects suspicious activity, it can automatically trigger alerts, allowing security teams to respond immediately. This proactive approach not only enhances incident response but also minimizes the window of opportunity for attackers to exploit vulnerabilities.

C2Graph (C2G) Implementation

C2Graph (C2G) is an implementation of "Malware Beaconing Detection by Mining Large-scale DNS Logs for Targeted Attack Identification" (Andrii, Katrin, & Xiongwei, 2016). The original article focuses on DNS logs, but the principles were extended to proxy logs adding jitter consideration to request size and delta time communication.

Workflow Overview

- **Data Extraction:** Parsing proxy logs to extract relevant features.
- **Graph Construction:** Building a graph of source and destination nodes to analyze communication patterns.
- **Binning:** Creation of temporal and quantitative delta sequences which are binned into buckets tagged with letters. This process catches jitters.

Key Metrics:

- **Node Degree:** Represents the number of incoming connections to a node. For example, a high degree for a legitimate site like google.com contrasts with a low degree for a C2 server.
- **Edge Weight:** Indicates the frequency of communication between nodes, helping to filter out trusted sites and focus on suspicious activity.

AI Process:

- **Hypothesis:** we suppose it is the beginning of an infection.
- **First step:** the AI is looking to low node degree sources – destinations connections with high edge weight.
- **Second step:** For these selected couples of sources and destinations the AI adds two scores, one for the binning temporal periodicity and another to the binning quantitative periodicity.
- **Alerting:** it is made when the normalized score combined for these two precedents is in the top 10%.
- **Expert Feedback Loop:** Security analysts review alerts to provide feedback on the accuracy of the AI's assessments, helping to refine the model and improve future detection capabilities.

Key findings

What type of key findings could this type of algorithm highlight?

Post phishing infection:

AI can find infections of internal phishing campaigns just after the click on the malicious link.

Malicious website tracking:

AI can track the use of known malicious sites or abuse of trusted web pages.

Proactive Threat Intelligence:

In some cases, infections are not known by threat intelligence sources which could highlight new types of infection.

Benefits of AI-Driven Detection

The advantages of AI-driven detection are manifold:

- **Increased Accuracy:** AI can discern subtle patterns that traditional methods may overlook, leading to more reliable threat identification. By continuously learning from new data, AI systems can adapt to changing attack vectors.
- **Scalability:** The system can handle vast amounts of data, making it suitable for organizations of all sizes. As businesses grow, the AI can scale accordingly, maintaining effective monitoring without compromising performance.
- **Proactive Defense:** Early detection allows for proactive measures, reducing potential damage. By identifying threats before they can execute their malicious intent, organizations can safeguard their assets more effectively.

Key Takeaways

AI-driven proxy log analysis marks a transformative step in beaconing detection. By harnessing the power of AI, organizations can enhance their security measures, safeguarding networks against sophisticated attacks. This technology not only improves detection capabilities but also empowers security teams to respond swiftly and effectively to emerging threats.

Investing in AI technology for beaconing detection not only improves threat identification but also strengthens an organization's overall cybersecurity posture. While AI enhances detection capabilities, the invaluable insights and expertise of human analysts are essential for interpreting complex data and making informed decisions. As cyber threats continue to evolve, embracing this technology could be the key to staying one step ahead of cybercriminals.

Wicus Ross
Senior Security Researcher
Orange Cyberdefense

Research: Vulnerabilities

Beyond Vulnerability Management

We Cannot Patch Fast Enough

The reactive nature of vulnerability management, combined with delays from policy and process, strains security teams, who have limited capacity and cannot patch everything immediately. Our Vulnerability Operation Center (VOC) dataset analysis found 32,585 distinct CVEs across 68,500 unique customer assets, with 10,014 having a CVSS score of 8 or higher. Among these, external assets have 11,605 distinct CVEs, while internal assets have 31,966. With this volume of CVEs, it's no surprise that some go unpatched and lead to compromises.

Why are we stuck in this situation, what can be done, and is there a better approach for businesses?

We'll explore the state of vulnerability reporting, how to prioritize vulnerabilities by threat and exploitation, examine statistical probabilities, and briefly discuss risk. Lastly, we'll consider solutions to minimize vulnerability impact while giving management teams flexibility in crisis response.

Can You CVE What I CVE?

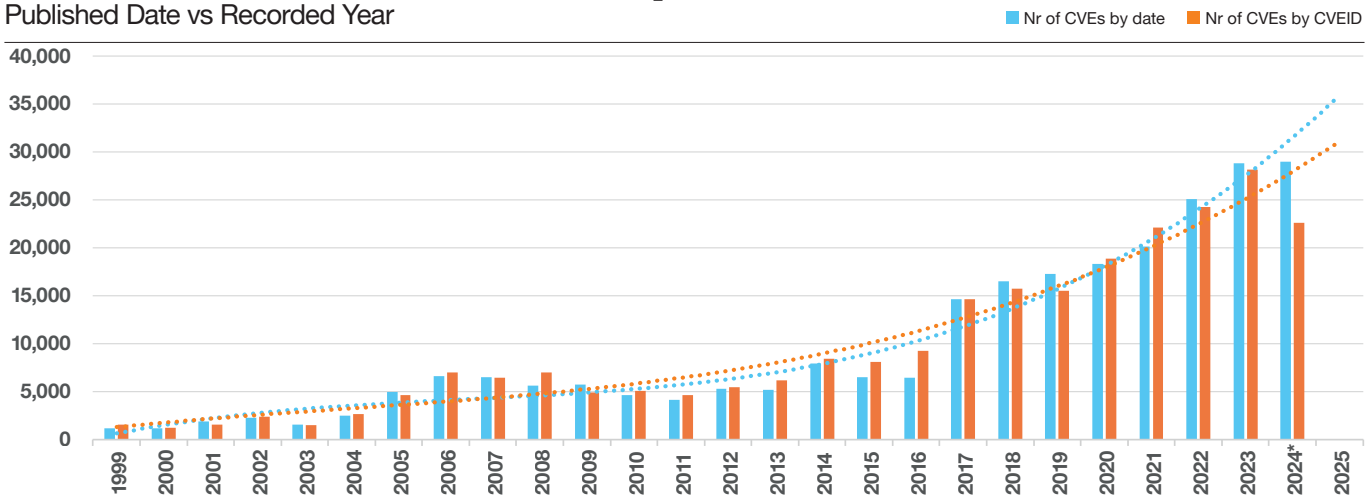
Western nations and organizations use the Common Vulnerability Enumeration (CVE) and Common Vulnerability Scoring System (CVSS) to track and rate vulnerabilities, overseen by US government-funded programs like MITRE and NIST. By September 2024, the CVE program, active for 25 years, had published over 264,000 CVEs, with 14,443 marked as "Rejected" or "Deferred."

NIST's National Vulnerability Database (NVD) relies on CVE Numbering Authorities (CNAs) to record CVEs with initial CVSS assessments, which helps scale the process but also introduces biases. The disclosure of serious vulnerabilities is complicated by disagreements between researchers and vendors over impact, relevance, and accuracy, affecting the wider community[84][85].

In 2024, a backlog of 18,167 unenriched CVEs accumulated at the NVD[86][87] due to bureaucratic delays, halting CVE enrichment despite ongoing vulnerability reports, and dramatically illustrating the fragility of this system.

NR of CVEs Published per Year

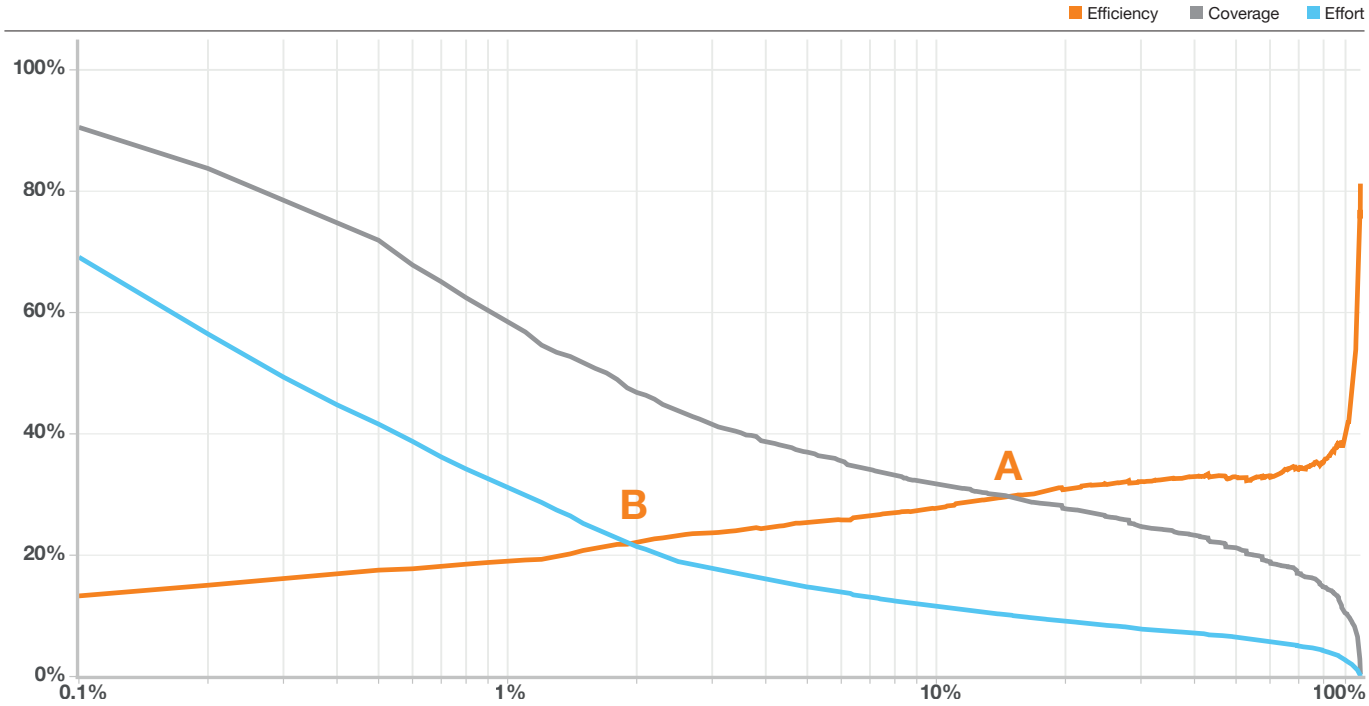
Published Date vs Recorded Year



* CVEs for 2024 incomplete (YTD)

EPSS Threshold

Thresholds in Terms of Coverage, Efficiency, and Effort, Relative to Known Exploited Vulnerabilities



CVE and the NVD are not the sole sources of vulnerability intelligence. Many organizations, including ours, develop independent products that track far more vulnerabilities than the NVD's CVE program.

Since 2009, China has operated its own vulnerability database, CNNVD[88], which could be a valuable technical resource[89][90], though political barriers make collaboration unlikely. Moreover, not all vulnerabilities are disclosed immediately, creating blind spots, while some are exploited without detection—so-called 0-days.

In 2023, Google's Threat Analysis Group (TAG) and Mandiant identified 97 zero-day exploits, primarily affecting mobile devices, operating systems, browsers, and other applications.[91] Meanwhile, only about 6% of vulnerabilities in the CVE dictionary have ever been exploited[92], and studies from 2022 show that half of organizations patch just 15.5% or fewer vulnerabilities monthly[93].

While CVE is crucial for security managers, it's an imperfect, voluntary system, neither globally regulated nor universally adopted.

This paper aims to explore how we might reduce reliance on it in our daily operations.

Threat Informed

Despite its shortcomings, the CVE system still provides valuable intelligence on vulnerabilities that could impact security. However, with so many CVEs to address, we must prioritize those most likely to be exploited by threat actors.

The Exploit Prediction Scoring System (EPSS), developed by the Forum of Incident Response and Security Teams (FIRST) SIG[94], helps predict the likelihood of a vulnerability being exploited in the wild. With EPSS intelligence, security managers can either prioritize patching as many CVEs as possible for broad coverage or focus on critical vulnerabilities to maximize efficiency and prevent exploitation. Both approaches have pros and cons.

To demonstrate the tradeoff between coverage and efficiency, we need two datasets: one representing potential patches (VOC dataset) and another representing actively exploited vulnerabilities, which includes CISA KEV[95], ethical hacking findings, and data from our CERT Vulnerability Intelligence Watch service[96].

The EPSS threshold is used to select a set of CVEs to patch, based on how likely they are to be exploited in the wild. The overlap between remediation set and the exploited vulnerability set can be used to calculate the Efficiency, Coverage, and Effort of a selected strategy.

Coverage is the percentage of remediated vulnerabilities that are also present in the target exploit group.

Efficiency is the number of remediated vulnerabilities from the target exploit group as a proportion of the total remediation group.

Effort is expressed as the number of vulnerabilities in the remediation group that will be patched as a percentage of the vulnerability population.

If you wish to explore EPSS further, then we encourage you to read our blog post that covers the EPSS tool used here in this section[97].

Point A in the chart on the previous page is where the EPSS threshold is 14.9% and represents the level where the Efficiency and Coverage intersect. A lower EPSS threshold would yield better Coverage, but at the cost of Efficiency, since Effort increases as the number of CVEs that must be patched grows. The opposite is also true: If the EPSS threshold is increased we would remediate a smaller number of (potentially exploitable) CVEs, but with a higher risk of missing something.

Point B on the chart is where Efficiency and Effort intersect, and represents the lowest EPSS threshold that should be considered for this example. Selecting an EPSS threshold smaller than 1.9% will result in increased Coverage, but with a noticeable increase in Effort.

The example here is theoretical but it serves to remind us that the choices we make with regards to patching come with real tradeoffs.

Likely Choices

EPSS **predicts the likelihood of a vulnerability being exploited somewhere** in the wild, not on any specific system. However, probabilities can "scale." For example, flipping one coin gives a 50% chance of heads, but flipping 10 coins raises the chance of at least one heads to 99.9%. This scaling is calculated using the complement rule[98], which finds the probability of the desired outcome by subtracting the chance of failure from 1.

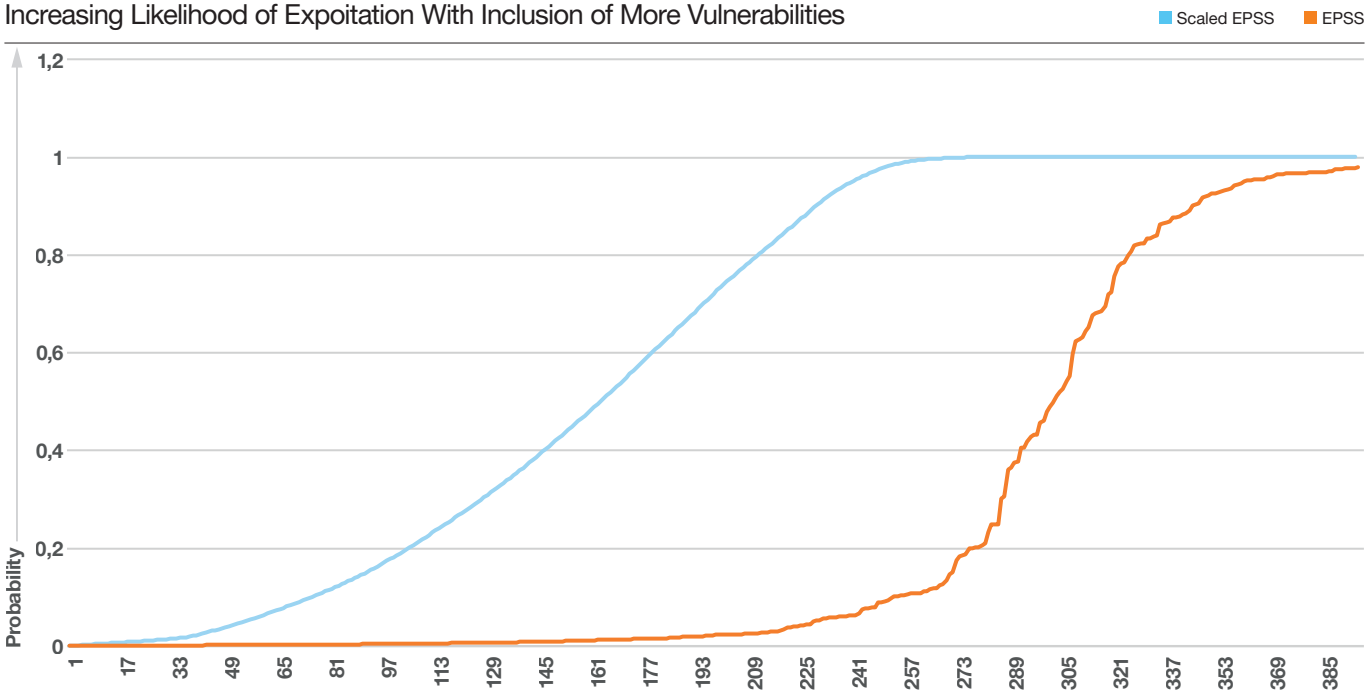
As FIRST explains, "EPSS predicts the probability of a specific vulnerability being exploited and can be scaled to estimate threats across servers, subnets, or entire enterprises by calculating the probability of at least one event occurring."[99][100]

With EPSS, we can similarly calculate the likelihood of at least one vulnerability being exploited from a list by using the complement rule.

To demonstrate, we analyzed 397 vulnerabilities from the VOC scan data of a Public Administration sector client. As the chart below illustrates, most vulnerabilities had low EPSS scores until a sharp rise at position 276. Also shown on the chart is the scaled probability of exploitation using the complement rule, which effectively reaches 100% when only the first 264 vulnerabilities are considered.

Scaled Probabilities

Increasing Likelihood of Expoitation With Inclusion of More Vulnerabilities



As the second line on the chart indicates, as more CVEs are considered, the scaled probability that one of them will be exploited in the wild increases very rapidly. By the time there are 265 distinct CVE under consideration, the probability that one of them will be exploited in the wild is more than 99%. This level is reached before any individual vulnerabilities with high EPSS come into consideration. When the scaled EPSS value crosses 99% (Position 260) the maximum EPSS is still under 11% (0.11).

Vulnerabilities with high EPSS scores also do not necessarily have a high CVSS score. The bulk (38) of vulnerabilities shown on the chart have a CVSS score between 5 and 6.25. Only 15 vulnerabilities in the set have a score between 7.5 and 9.8. The highest scoring vulnerability only had an EPSS of 0.37% (0.0037).

This example, based on actual client data on vulnerabilities exposed to the Internet, shows how difficult prioritizing vulnerabilities becomes as the number of systems increases.

EPSS gives a probability that a vulnerability will be exploited in the wild, which is helpful for defenders, but we've shown how quickly this probability scales when multiple vulnerabilities are involved. With enough vulnerabilities, there is a real probability that one will get exploited, even when the individual EPSS scores are low.

Like a weather forecast predicting "chance of rain," the larger the area, the greater the likelihood of rain somewhere.

This scaling effect makes applying EPSS for vulnerability management in large environments less practical, as even with extensive patching, it may be impossible to reduce the probability of exploitation somewhere near to zero.

Attackers Think in Graphs

In 2015, Microsoft Security Engineer John Lambert shared an immutable truth in a blog post titled, "Defenders think in lists. Attackers think in graphs. As long as this is true, attackers win."[101] Lambert explained, "Defenders don't have a list of assets—they have a graph. Assets are connected by security relationships. Attackers breach a network by landing somewhere in the graph, using techniques like spearphishing, and hack by navigating it." He added, "The graph in your network is shaped by security dependencies, network design, management, software, services, and user behavior."

In vulnerability management, Lambert's insights highlight two key realities. First, vulnerabilities are just one factor attackers use to gain access. The MITRE ATT&CK framework[102] documents many observed attacker behaviors. In July 2024, SensePost, part of Orange Cyberdefense's Ethical Hacking team, described how an attacker can evade an Endpoint Detection and Response (EDR) system using "attack decorrelation."[103] By manipulating a system to disclose separate innocuous pieces of information, the attacker can combine them to compromise the system without triggering alarms, demonstrating that a skilled, persistent attacker can bypass controls, even in environments without exploitable CVEs. Even if an environment is seemingly devoid of exploitable CVEs, a resourceful and experienced attacker with enough persistence may find a way to achieve a compromise, sidestep a control or avoid being detected.

Second, attackers don't need to compromise a specific system—any foothold in a homogenous network grants access to Lambert's "graph."

From there, attackers can navigate to valuable assets. Thus, defenders must not only patch vulnerabilities but also restrict access across the security graph to minimize the impact of any compromise.

Attacker Odds

We've identified three critical truths that must be integrated into our examination of the vulnerability management process:

- Attackers aren't focused on specific vulnerabilities; they aim to compromise systems to access the graph.
- Exploiting vulnerabilities isn't the only path to compromise, and often, it's not the most common one.
- Attackers' skill and persistence levels vary.

These factors allow us to extend our analysis of EPSS and probabilities to consider the likelihood of an attacker compromising some arbitrary system, then scaling that to determine the probability of compromising some system within a network that grants access to the graph.

We can assume each hacker has a certain "probability" of compromising a system, with this probability increasing based on their skill, experience, tools, and time. We can then continue applying probability scaling to assess attacker success against a broader computer environment.

$$n = \frac{\ln(1 - s)}{\ln(1 - p)}$$

s is the estimated probability of a successful attack
p is the chance of success based on judged skill
ln is the natural log function
n is the number of occurrences

Given a patient, undetected hacker, how many attempts are statistically required to breach a system granting access to the graph? Answering this requires applying a reworked binomial distribution in the form of this equation:[104][105]

$$\sim 180 = \frac{\ln(1 - 0.9999)}{\ln(1 - 0.05)}$$

Using this equation, we can estimate how many attempts an attacker of a certain skill level would need. For instance, if attacker A1 has a 5% success rate (1 in 20) per system, they would need to target up to 180 systems to be 99.99% sure of success.

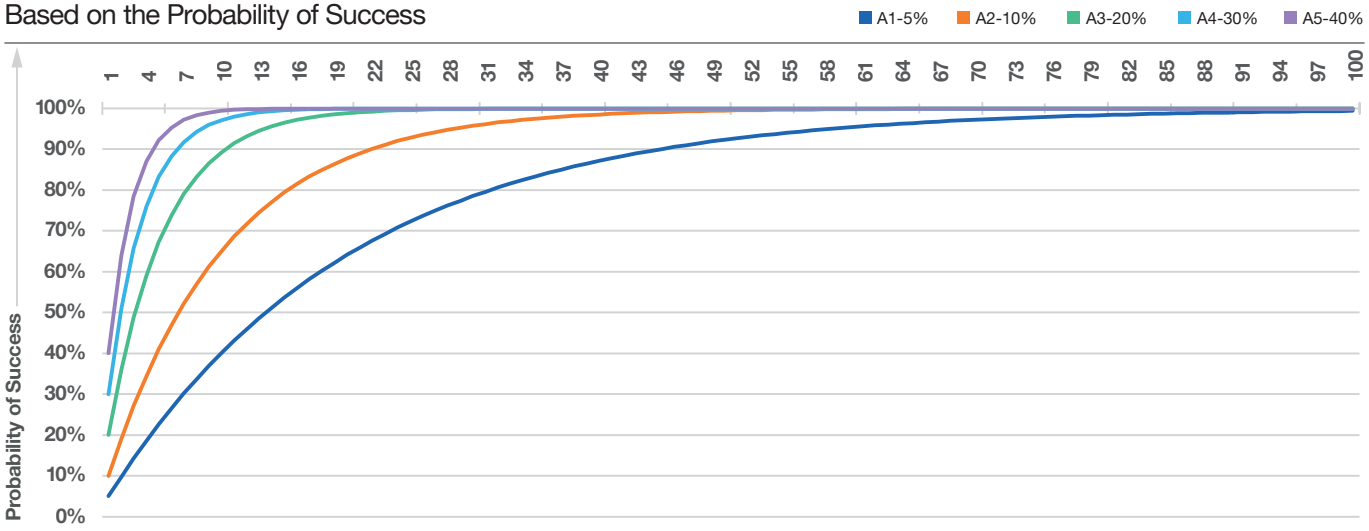
Another attacker, A2, with a 10% success rate (1 in 10), would need about 88 targets to ensure at least one success, while a more skilled attacker, A3, with a 20% success rate (1 in 5), would only need around 42 targets for the same probability.

These are probabilities—an attacker might succeed on the first try or require multiple attempts to reach the expected success rate. To assess real-world impact, we surveyed senior penetration testers in our business, who estimated their success rate against arbitrary internet-connected targets to be around 30%.

Assuming a skilled attacker has a 5% to 40% chance of compromising a single machine, we can now estimate how many targets would be needed to nearly guarantee one successful compromise.

Attacker Success

Based on the Probability of Success



	A1-5%	A2-10%	A3-20%	A4-30%	A5-40%
Attempts*	315	153	73	46	32

* The number of attempts will result in a 99.99999% probability of success

The implications are striking: with just 100 potential targets, even a moderately skilled attacker is almost certain to succeed at least once. In a typical enterprise, this single compromise often provides access to Lambert's graph, and enterprises typically have thousands of computers to consider.

Reimagining Vulnerability Management

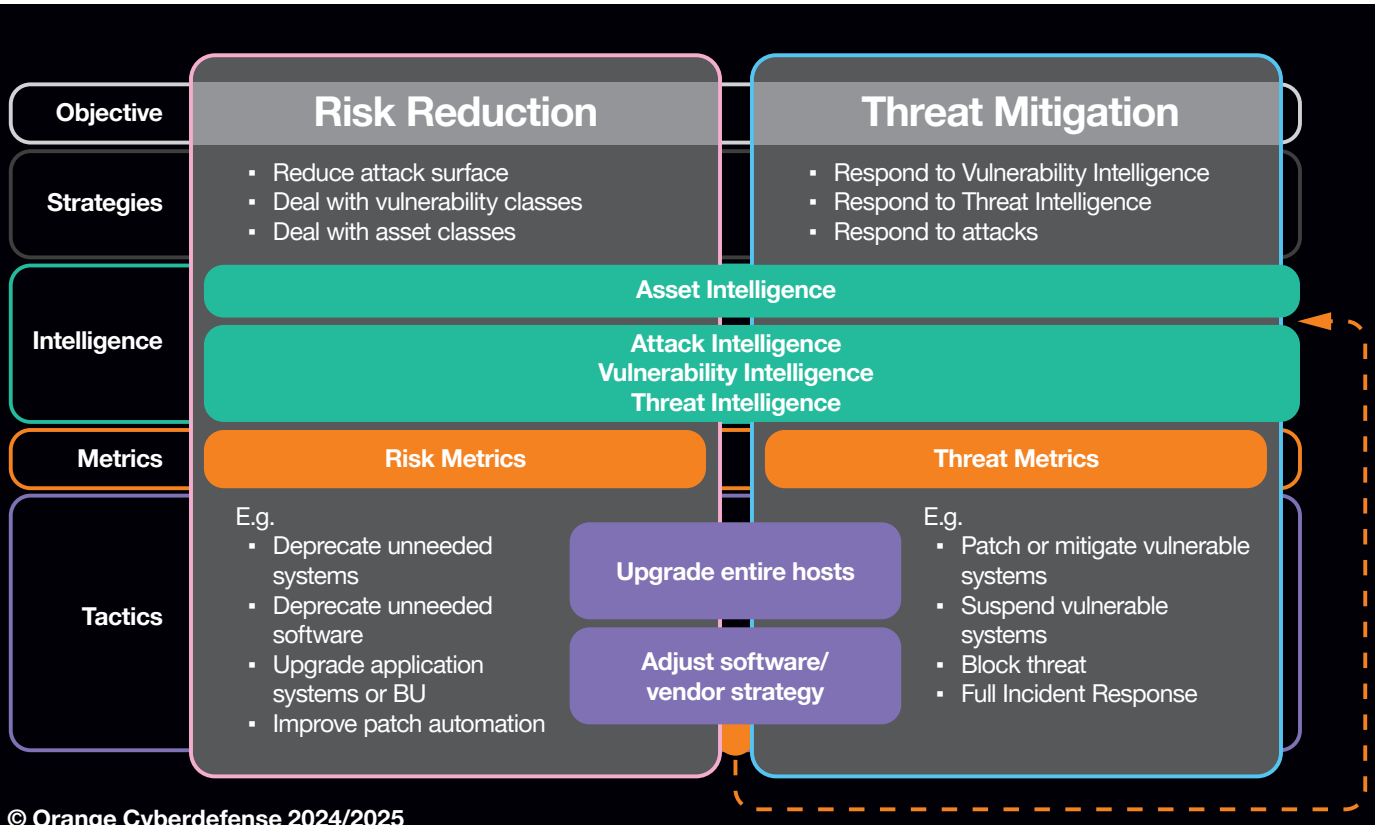
For the future, we need to conceive an environment and architecture that is immune to compromise from an individual system. In the shorter term, we argue that our approach to vulnerability management needs to change.

Use the Right Words

The current approach to vulnerability management is rooted in its name: focusing on "vulnerabilities" (as defined by CVE, CVSS, and EPSS) and their "management." However, we have no control over the volume, speed, or significance of CVEs, leading us to constantly react to chaotic new intelligence.

EPSS now helps us prioritize vulnerabilities likely to be exploited in the wild, representing real threats, which forces us into a reactive mode. While mitigation addresses vulnerabilities, our response is truly about countering threats—hence, this process should be called Threat Mitigation.

As discussed earlier, it's statistically impossible to effectively counter threats in large enterprises by merely reacting to vulnerability intelligence. Instead, we should focus on Risk Reduction. Cyber risk results from a threat targeting a system's assets, leveraging vulnerabilities, and the potential impact of such an attack. By addressing risk, we open up more areas under our control to manage and mitigate.



Threat Mitigation

Threat Mitigation is a dynamic, ongoing process that involves identifying threats, assessing their relevance, and taking action to mitigate them. This response can include patching, reconfiguring, filtering, adding compensating controls, or even removing vulnerable systems. EPSS is a valuable tool that complements other sources of threat and vulnerability intelligence.

However, the scaling nature of probabilities makes EPSS less useful in large internal environments. Since EPSS focuses on vulnerabilities likely to be exploited "in the wild," it is most applicable to systems directly exposed to the internet. Therefore, Threat Mitigation efforts should primarily target those externally exposed systems.

Risk Reduction

Cyber risk is a product of Threat, Vulnerability, and Impact. While the "Threat" is largely beyond our control, patching specific vulnerabilities in large environments doesn't significantly lower the risk of compromise. Therefore, risk reduction should focus on three key efforts:

1. Reducing the attack surface: As the probability of compromise increases with scale, it can be reduced by shrinking the attack surface. A key priority is identifying and removing unmanaged or unnecessary internet-facing systems.

2. Limiting the impact: Lambert's law advises limiting attackers' ability to access and traverse the "graph." This is achieved through segmentation at all levels—network, permissions, applications, and data. The Zero Trust architecture provides a practical reference model for this goal.

3. Improving the baseline: Instead of focusing on specific vulnerabilities as they're reported or discovered, systematically reducing the overall number and severity of vulnerabilities lowers the risk of compromise. This approach prioritizes efficiency and Return on Investment, ignoring current acute threats in favor of long-term risk reduction.

By separating Threat Mitigation from Risk Reduction, we can break free from the constant cycle of reacting to specific threats and focus on more efficient, strategic approaches, freeing up resources for other priorities.

An Efficient Approach

The three Risk Reduction goals for internal enterprise networks aren't driven by the random discovery of new Threats or Vulnerabilities but can be pursued systematically to optimize resources. The focus shifts from "managing vulnerabilities" to designing, implementing, and validating resilient architectures and baseline configurations. Once these baselines are set by the security function, IT can take over their implementation and maintenance, aligning with existing IT processes for greater efficiency. The security function can then validate compliance with the agreed standards.

The key here is that the "trigger" for patching internal systems is a predefined plan, agreed with system owners, to upgrade to a new, approved baseline.

This approach is certain to be much less disruptive and more efficient than responding to specific, new vulnerabilities.

Vulnerability Scanning remains important for creating an accurate asset inventory and identifying non-compliant systems, but it should support existing standardized processes, not trigger them.

Reimagining the Future

The overwhelming barrage of randomly discovered and reported vulnerabilities as represented by CVE, CVSS and EPSS are stressing our people, processes and technology. We've effectively been approaching vulnerability management the same way for over two decades, but it hasn't been working and it's not efficiently reducing risk, and so it too must evolve.

It's time to reimagine how we design, build, and maintain systems.

A Template for a New Strategy

Key factors to consider for security strategies toward 2030 and beyond:

1. **Starting at the source**

2. **Human Factor**
   - Leverage human strengths and anticipate their weaknesses.
   - Gain support from senior management and executives.
   - Be an enabler, not a blocker.

3. **Threat-Informed Decision Making**
   - Learn from incidents and focus on what's being exploited.
   - Use strategies to enhance remediation based on your capabilities.

4. **Threat Modeling and Simulation**
   - Use threat models to understand potential attack paths.
   - Conduct Ethical Hacking to test your environment against real threats.

5. **System Architecture and Design**
   - Apply threat models and simulations to validate assumptions in new systems.
   - Reduce the attack surface systematically.
   - Strengthen defense in depth by reviewing existing systems.
   - Treat SASE and Zero-Trust as strategies, not just technology.

6. **Secure by Demand / Default**
   - Implement formal policies to embed security into corporate culture.
   - Ensure vendors and suppliers have active security improvement programs.

Starting at the Source

The first place to reduce the load of managing vulnerabilities, is at the source, by reducing the number of vulnerabilities in the technology products we deploy. CISA Director Jen Easterly criticized vendors for producing poor-quality software, describing the issues as 'defects' rather than just vulnerabilities.[106] Over 200 vendors have committed to supporting the voluntary Secure by Design initiative for better self-regulation.

Google Android and Pixel have made headways over the past few years to harden the mobile operating system (OS) and mobile hardware platform[107]. These changes are directly aimed at countering existing attacks or to make exploitation considerably more difficult. The Google Android team indicated that most vulnerabilities in their mobile OS are present in new source code, while older source code has proportionally fewer vulnerabilities[108]. They also believe that the number of vulnerabilities will be reduced substantially over time due to the introduction of memory safe techniques and memory safe programming languages. Microsoft has also implemented new standards, policies, and processes to ensure security is integrated from the start of every project, with measures to track adherence and assess compliance. These changes were prompted by several serious incidents in 2023 and 2024.[109]

After a series of painful security missteps, VPN vendor Ivanti pledged[110] publicly to execute a plan "that accelerates security initiatives already underway and implements improved practices to anticipate, prevent and protect against future threats". Every technology producer has the responsibility to implement policies to explicitly state how products will be created that are secure, and all buyers should pressure their vendors to commit to shipping more secure code.

Human Factors

For vulnerability management teams to succeed, gaining support from key colleagues is essential. The program should support the business, not create obstacles. Find a strategy aligned with the business's goals, keeping that in focus. This might require creativity and compromise. Start by having conversations with key individuals to understand their needs. Actively listen to their perspectives, as this could be the foundation for your initial strategy.

Threat Informed Decision Making

With the abundance of information on attacks, it's easy to get swept up in panic. The key is to assess how the published information applies to your environment and whether it warrants action. Understanding your environment and attack surface is crucial in making informed decisions.

Threat Model and Simulate

Ethical Hacking engagements provide a valuable opportunity to learn from experts by thinking like attackers. These services are typically tailored to test specific systems or components but can also be goal-oriented with broader objectives, like assessing detection and response capabilities. The results serve as highly localized threat intelligence, which should be used to update threat models.

System Architecture and Design

Existing system designs should be reviewed based on threat models, past incidents, or latent defects identified by vulnerability management teams. There is always room to strengthen 'defense in depth' through methods such as network segmentation, non-repudiable authentication, and least privilege for services and user accounts.

Reducing the attack surface methodically eases the burden on security operations, including vulnerability management. While it may not always be feasible to remove or replace unsupported products, decommissioning unused assets in accordance with policy is critical.

Outdated systems tied to mission-critical processes often require collaboration across teams to enhance confidentiality, integrity, and availability. This ultimately becomes a business decision, weighing time, cost, and resources.

As systems increasingly span on-premises and cloud services, businesses can operate with more flexibility. Secure Access Service Edge (SASE) and Zero-Trust should be approached as strategies, not just technology stacks, to bolster defense in depth by design.

Traditional principles like Confidentiality, Integrity, Availability (CIA), and Non-repudiation remain essential, but newer concepts such as Distributed, Immutable, and Ephemeral (DIE) can enhance security. DIE principles[111]:

- Distributed – no dependency on one host
- Immutable – unable to modify assets
- Ephemeral – short lived instances that are discarded help address issues more efficiently.

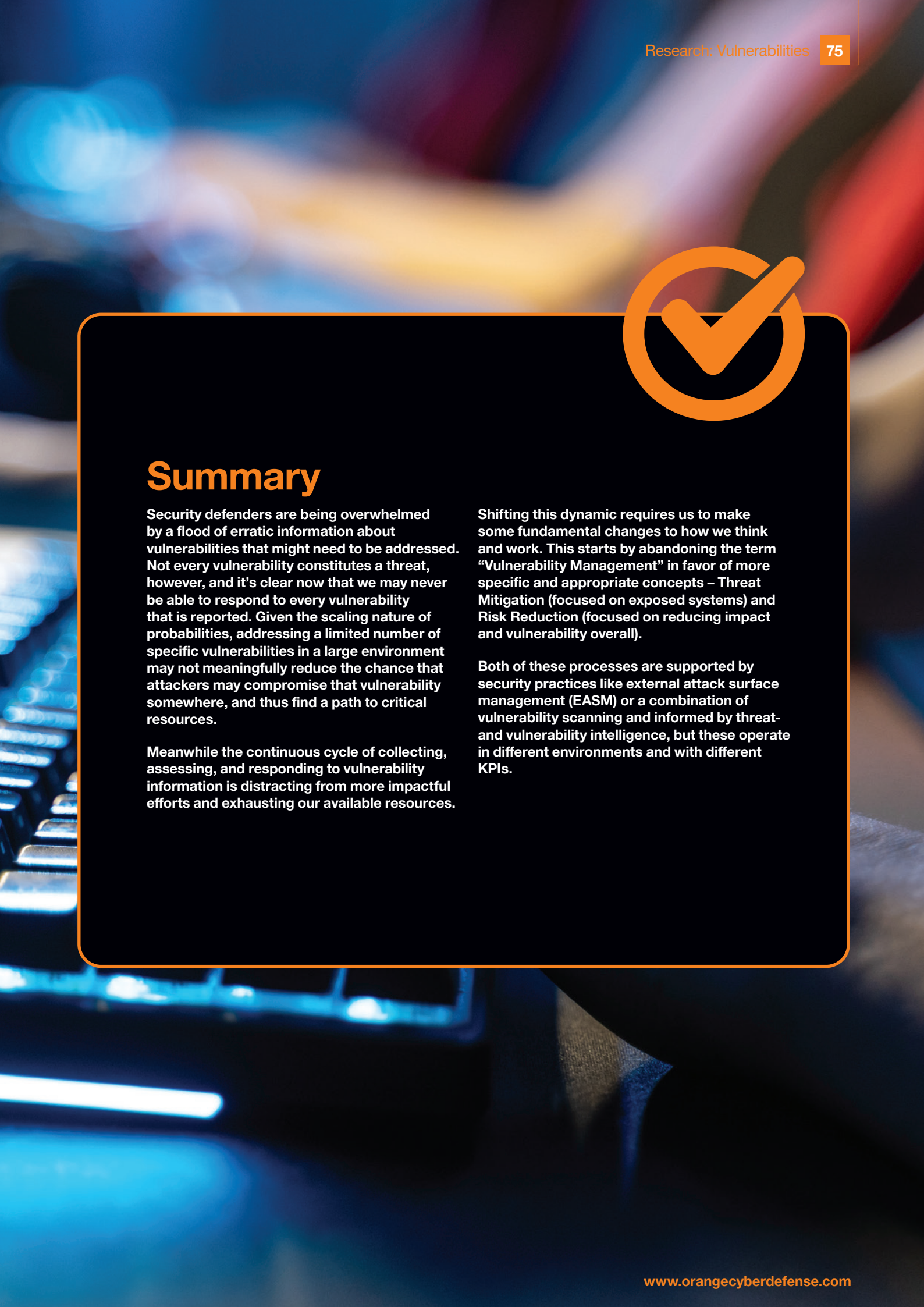
Ephemeral hosts, in particular, benefit vulnerability management, as each instance runs the latest baseline, with outdated or non-compliant versions quickly discarded.

Secure by Demand or Secure by Default

Technology commoditization has led to a race to the bottom, with vendors rushing to develop features and offer services at discount prices, often resulting in poor security outcomes for clients and collateral damage.

Corporate culture must shift through clear, formal policies that prioritize security at every level, ensuring it's integrated into every product or service. CISA's 'Secure by Design'[112] initiative encourages vendors to build security into products from the start[113], while their 'Secure by Demand' guide provides resources to help buyers ensure security is central to their purchases. CISA also issued 'Secure Design Alert' advisories to inform decision-makers about commonly exploited flaws in specific technologies[114].

In the future, business-to-business relationships will evolve, with vendors required to prove their security and quality policies meet industry standards. Demanding secure products and services will become standard practice.



Summary

Security defenders are being overwhelmed by a flood of erratic information about vulnerabilities that might need to be addressed. Not every vulnerability constitutes a threat, however, and it's clear now that we may never be able to respond to every vulnerability that is reported. Given the scaling nature of probabilities, addressing a limited number of specific vulnerabilities in a large environment may not meaningfully reduce the chance that attackers may compromise that vulnerability somewhere, and thus find a path to critical resources.

Meanwhile the continuous cycle of collecting, assessing, and responding to vulnerability information is distracting from more impactful efforts and exhausting our available resources.

Shifting this dynamic requires us to make some fundamental changes to how we think and work. This starts by abandoning the term "Vulnerability Management" in favor of more specific and appropriate concepts – Threat Mitigation (focused on exposed systems) and Risk Reduction (focused on reducing impact and vulnerability overall).

Both of these processes are supported by security practices like external attack surface management (EASM) or a combination of vulnerability scanning and informed by threat- and vulnerability intelligence, but these operate in different environments and with different KPIs.

Vulnerability-prone Network

Spotlight on VPN: Faulty by Design?

VPN gateways fill a unique role, exposed to all the hazards of the Internet, while at the same time, having access to some of the most critical resources in the organization.

In many cases, software that has a track record of security vulnerabilities is deployed behind a VPN to limit who can access it. What should one do if the problematic software is the VPN itself?

Rogan Dawes, SensePost Researcher, **Orange Cyberdefense**

In an April 2020 advisory, the U.S. Cybersecurity and Infrastructure Security Agency (CISA) advised its stakeholders to "immediately patch CVE-2019-11510 —an arbitrary file reading vulnerability affecting Pulse Secure virtual private network (VPN) appliances"[115]. That same year, we reported in our annual security Navigator report as noteworthy the "visibility of several leading security product vendors in the very short list of technology vendors who featured multiple times in our intelligence advisories". We further noted a four-fold increase in vulnerabilities reported in selected security technologies between March and May 2020". Four years later in February 2024, CISA issued another advisory about another perimeter security product, this time going so far as to direct government agencies to "disconnect all instances" of the affected VPN product"[116].

The past five years have been characterized in a significant way by the discovery and exploitation of vulnerabilities in perimeter security technologies, and especially Virtual Private Networks (VPN).

Recent Vulnerabilities

Over the past several years, VPN software from multiple vendors has been exploited repeatedly. For example, just in 2024:

Product	CVE > 7 in 2024	Announcements
Ivanti Connect Secure	10	6
Palo Alto Pan-OS	9	5
Fortinet FortiOS	15	8

Each vulnerability advisory means that a team needs to drop everything to deploy the relevant patches in their environment.

One VPN vendor even recommended that their own product should be deployed behind a security gateway in order to protect it from an actively exploited vulnerability!

But why are the vulnerabilities discovered in these products so catastrophic? How is it that security products are repeatedly critically vulnerable to exploits, when software like OpenSSH, Postfix and Qmail which are equally exposed to the Internet have had only a handful of relatively low severity vulnerabilities over their extended lifetimes?

Vulnerabilities History

Product	CVE > 7 all time	All time CVEs
Postfix	1	11
Qmail	2	5
OpenSSH	25	116

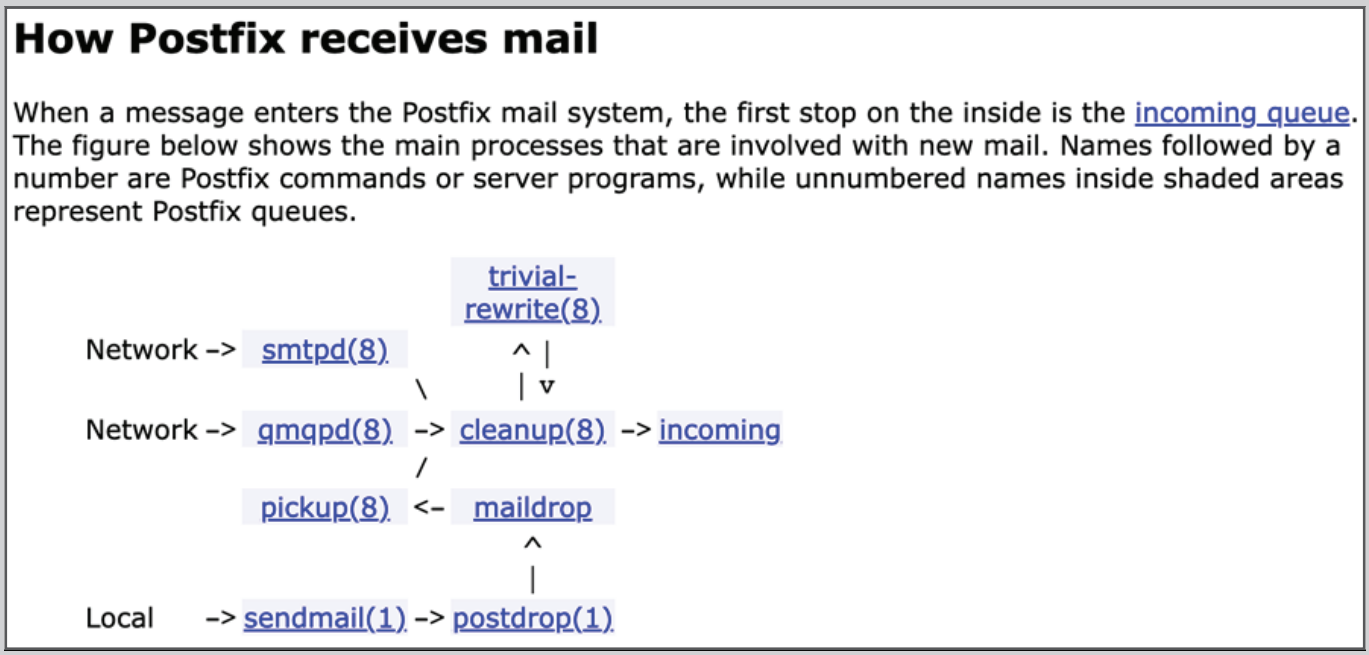
While OpenSSH does appear to have a large number of published vulnerabilities over its 25-year history, it is worth keeping in mind the almost ubiquitous nature of OpenSSH, making it an extremely high value target, and that many of the published vulnerabilities are in non-default configurations, or require misconfigurations in other products that leverage OpenSSH as a component.

We posit that programs with a long history of good security have been through an initial security architecture design process, where the system has been decomposed into elements, each responsible for a clearly defined aspect of the system. These elements have been chosen to be as independent from each other as possible, communicating only over carefully specified interfaces, so that a weakness in one element doesn't compromise the entire system.

An example of this can be seen in Postfix's documentation.

Example: Postfix

Firstly, Postfix lists all the exposed entry points and documents the components that require network access. Each of these components performs a specific task, with only the code required for that specific task present.



- **Example of Postfix system architecture documentation from https://www.postfix.org/**

This enables an administrator to decide which components should be enabled or disabled, based on their specific requirements, and limits the attack surface of the overall system. The remaining components are inaccessible by design, running under unprivileged accounts, processing queues of files owned only by that account. In many cases, the individual components are isolated with a limited view of the filesystem, to prevent access to system or other files in the event of a compromise.

In other parts of the Postfix overview, specific mention is given to measures taken to limit resource consumption, which could otherwise lead to a Denial of Service condition. Ways in which an incoming email can result in command execution are also highlighted, as a common source of security vulnerabilities. Other deliberate actions taken to eliminate vulnerability classes include forbidding use of fixed-size memory buffers, a common root cause of buffer overflow vulnerabilities.

In contrast, an analysis of Fortinet done by Bishop Fox reveals that they deploy a monolithic binary that contains almost all of its functionality in a single executable, run as the first process on system startup. This eliminates any chance of process and privilege separation, implying that an exploited vulnerability in a single function has access to all capabilities of the entire system. Other research reveals that Ivanti Connect Secure had HTTP endpoints that were vulnerable to directory traversal attacks, a vulnerability class that has been known for at least 20 years.

Pan-OS similarly had HTTP endpoints vulnerable to directory traversal attack, as well as internal system processes vulnerable to command injection using shell metacharacters, another vulnerability class that has been known for decades.

Many of the vulnerabilities listed for the products above were exacerbated by the services running as root, having full access to the system, and handing those privileges to any successful exploits. It has long been an axiom that services that do not need root privileges should not be run as root, to limit the damage caused in the event of a vulnerability.

Looking at the vulnerability analyzes carried out by various parties, it appears that, either very little security architecture design was carried out prior to building these systems, or that the initial design has been modified so much over time as to be unrecognisable. Furthermore, fixes to vulnerabilities appear to have prioritised “point fixes” for just the specific identified weakness, rather than taking the opportunity for a broader fix, looking for other instances of that vulnerability type, and endeavoring to eliminate them from the system entirely.

Customers should require their vendors to provide details of the security architecture of their products, to ensure that they can make educated purchasing decisions. Lack of such documentation should be seen as an indicator that they should be prepared for a never-ending cycle of panicking, patching and praying.

Key Takeaways

Our adversaries are targeting and exploiting the technologies we install, develop and maintain to protect our networks. The problem has been growing for several years now. As an industry, we should be solving these problems, not creating them.

As we have since 2022, we call on our partners and competitors in the security industry to come together to work on this challenge. We believe an industry-wide discussion needs to be had to determine whether the problem is as real as we perceive it is, identify existing efforts that may already be underway to address the issue, or create some form of partnership to work toward a better situation for ourselves and our customers. If you want to discuss this and join our initiative, please do contact us:

partnerfortomorrow@orangecyberdefense.com

Dr. Ric Derbyshire
Principal Security Researcher
Orange Cyberdefense

Trends, Targeting, and Testing of Operational Technology:

Ransomware Ripples & Real Risks

Introduction

It has been well established that cyber extortion (Cy-X), or more specifically ransomware, is currently the main threat to operational technology (OT). Whether through dependencies in the IT being impacted or an abundance of caution driving decisions to turn the OT off, IT-focused attacks dominate OT datasets – including ours.

We begin with this year's overall roundup, noting all the major trends we've seen. However, we wanted to focus on something different – the attacks where OT was the target, not just the victim. We call these category 2 attacks, and what distinguishes them from others is the adversary's use of tactics, techniques, and procedures (TTPs) unique to OT. This focus takes us into exploring what might motivate the adversaries into conducting such attacks and the impacts they cause as a result.

Finally, we ask the question 'does OT penetration testing effectively represent category 2 OT cyber-attacks?'. This is answered with our ongoing research on the topic, funded by the Research Institute in Trustworthy Inter-connected Cyber-physical Systems.

Historical Context

Last year, we presented the trends observed over 35 years of cyber-attacks impacting OT. We captured the data with a strict set of criteria, including corroboration from at least 2 reputable sources that an incident was confirmed to be due to a cyber-attack and caused an OT impact. We recorded a relatively low volume of OT-impacting cyber-attacks because of the strict criteria, but those we did record were well verified and contained enough data points to get a detailed view of the landscape. In total we recorded 119 cyber-attacks over the 35 years, and they were framed by a simple taxonomy that we created to better understand which types of attack were causing the impacts.

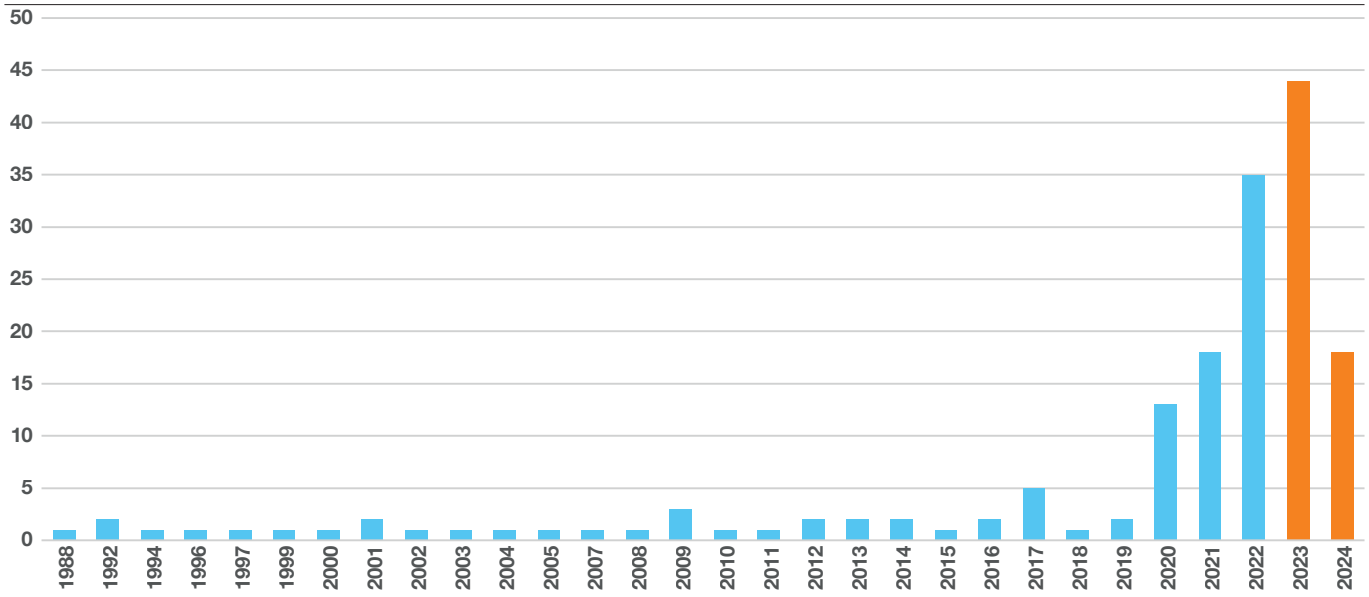
The elephant in the room when visualizing the data was the overwhelming volume of type 1a attacks from 2020 onwards. This was due to cyber extortion (Cy-X) causing cascading consequences all the way down to the physical process. Whether through dependencies being disrupted in the IT or the OT process being shut down due to an abundance of caution, OT has not escaped victimization when it comes to the scourge of Cy-X, or more specifically, encryption-based ransomware.

Taxonomy for Types of OT Cyber Attack

Category	1 IT TTPs			2 OT TTPs	
Type	1a	1b	1c	2a	2b
	IT targeted	IT/OT targeted	OT targeted	OT targeted, crude	OT targeted, sophisticated
Characteristics	IT attacked, production impacted indirectly as collateral damage	IT attacked, Windows/Linux-based OT attacked with IT TTPs directly or as collateral	Windows/Linux-based OT attacked with IT TTPs directly	Dedicated OT devices attacked with OT-specific TTPs crudely, little precision or complexity	Dedicated OT devices attacked with OT-specific TTPs with sophistication

Count of Attacks From 1988 to 2024

39% Increase in Attacks Between 2023 and 2024 Relative to the 35-Year Period Prior



It is important to note that despite their prominence, these attacks are rarely targeted directly at the OT. It is hard to ascertain the motivations of cyber criminals performing these Cy-X attacks, but due to the erratic targeting of Cy-X in general, the OT impacts likely aren't even intentional.

Other than the Cy-X-focused category 1 attacks, there was a small volume (19%) of category 2 cyber-attacks over our 35 years of data. The category 2 attacks were split evenly between type 2a and 2b. The adversary demographics conducting category 2 attacks has been quite fluid over time, with a slight shift from insider threats to states. These attacks that deliberately target the OT and include the use of specific TTPs, are clearly much more intentional with their OT impact.

What Has Changed?

Spoiler alert: much more of the same!

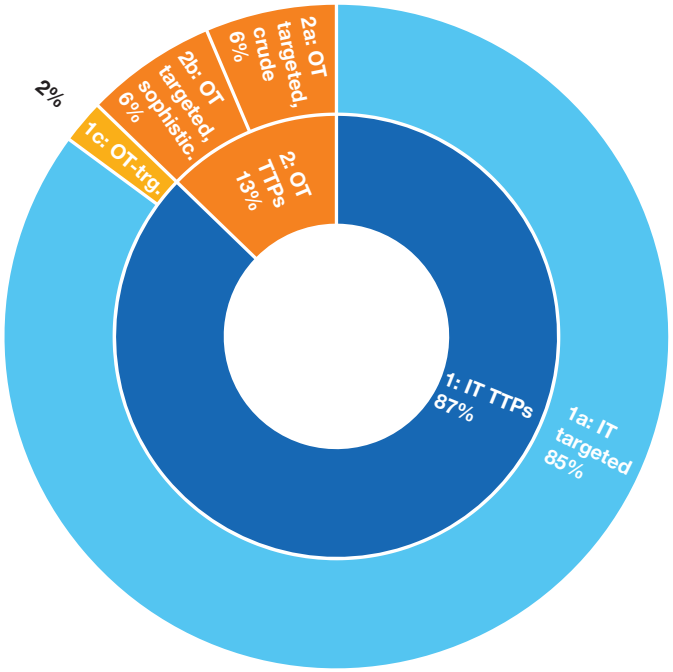
In collecting data between H2 2023 and H1 2024 our dataset grew by 47 incidents, 29 incidents in the tail end of 2023 and 18 so far in 2024. This took our total from 119 to 166, meaning we observed a staggering 39% increase in attacks between 2023 and 2024 relative to the 35-year period prior. This concerning trend is the symptom of the accelerating volume of impacts from Cy-X attacks.

Of the new cyber-attacks observed, an even greater proportion of them were category 1 attacks, at 87% (41). One missing element is the presence of type 1b attacks, which involve an opportunistic or accidental spillage into the OT by an adversary using IT TTPs. This may be that adversaries haven't managed that over the past year, but it is more likely a result of articles and reports focusing on impacts of events rather than the details.

There was just one incident represented by a type 1c attack, where an adversary deliberately targeted OT with IT TTPs. In this incident, the adversary deliberately deployed encryption-based ransomware on the victim's supervisory control and data acquisition (SCADA) server, which impacted the OT process.

Category Proportions

Types of OT-Impacting Cyber-Attacks '23/'24



The Victims

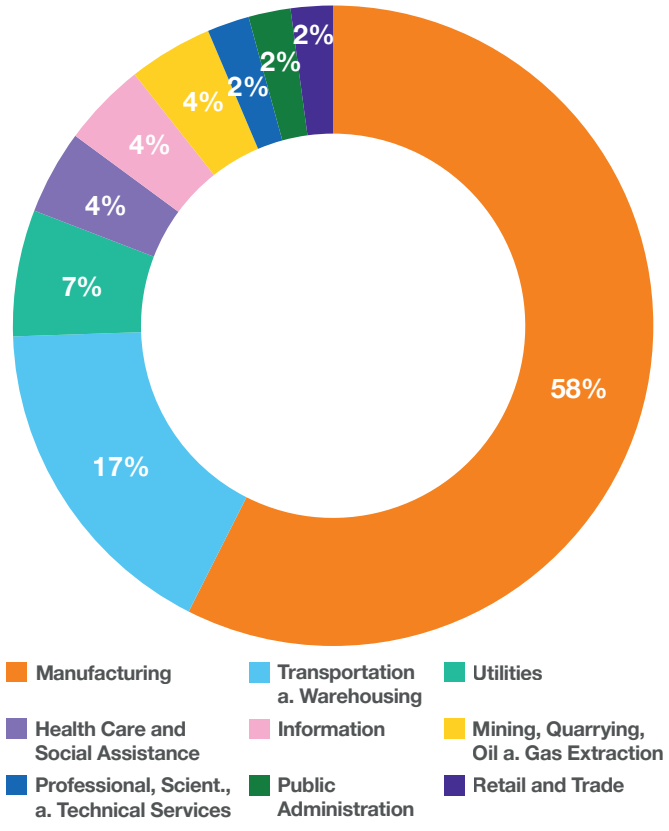
When it comes to victimology over the past year, we see much of the same. Geographically, we see a focus on the USA with 49% (23) of all attacks. Germany experienced the second highest number of incidents with 11% (5), which follows on from the trend we reported last year, with its relatively uncharacteristic prominence in cyber incident datasets.

Manufacturing was the most victimized sector, with 57% (27) of attacks over the past year. Interestingly, in our data regarding Cy-X victims this year, manufacturing has a share of 20% of all victims and has grown 25% from last year. This share of OT-impacting cyber-attacks follows on from the trend over the past 35 years. Although that trend was heavily influenced by the surge of Cy-X targeting manufacturing beginning in 2020. Transportation and warehousing was the second most victimized sector and utilities third most, which is also similar to last year's results. However, manufacturing featured far more significantly over the past year, with less diversity and share of victimization from trailing sectors when compared to the full dataset.

As could probably be expected, 81% (38) of this year's attacks were perpetrated by criminals. States and unknown adversaries share second spot, both responsible for 6% (3) each of the total attacks over the past year. Unknown adversary types usually stem from when the victim manages to respond to an event quickly enough such that the adversary cannot complete any objectives, obscuring their motivations. Therefore, unknown can be seen as a positive in some cases.

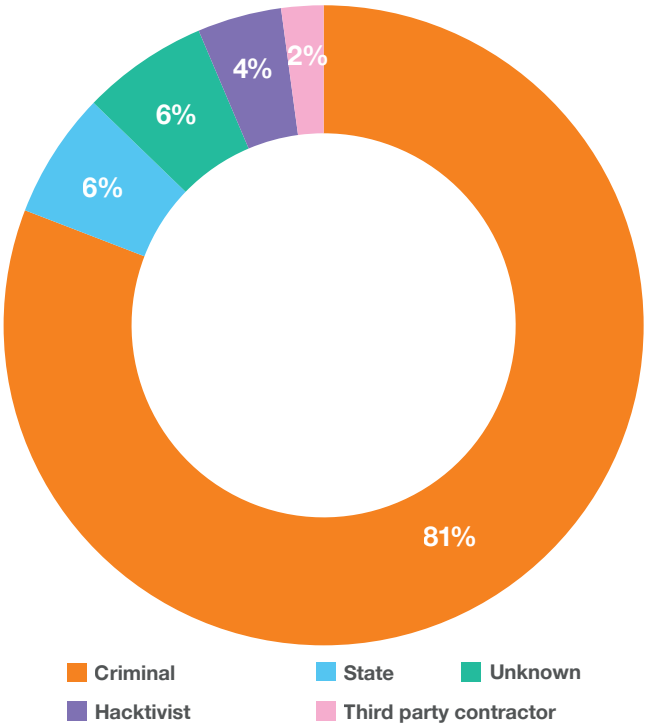
Target Sectors

Affected by Ot-Impacting Cyber-Attacks '23/'24



Adversaries

Actors Conducting Cyber-Attacks on OT in '23/'24

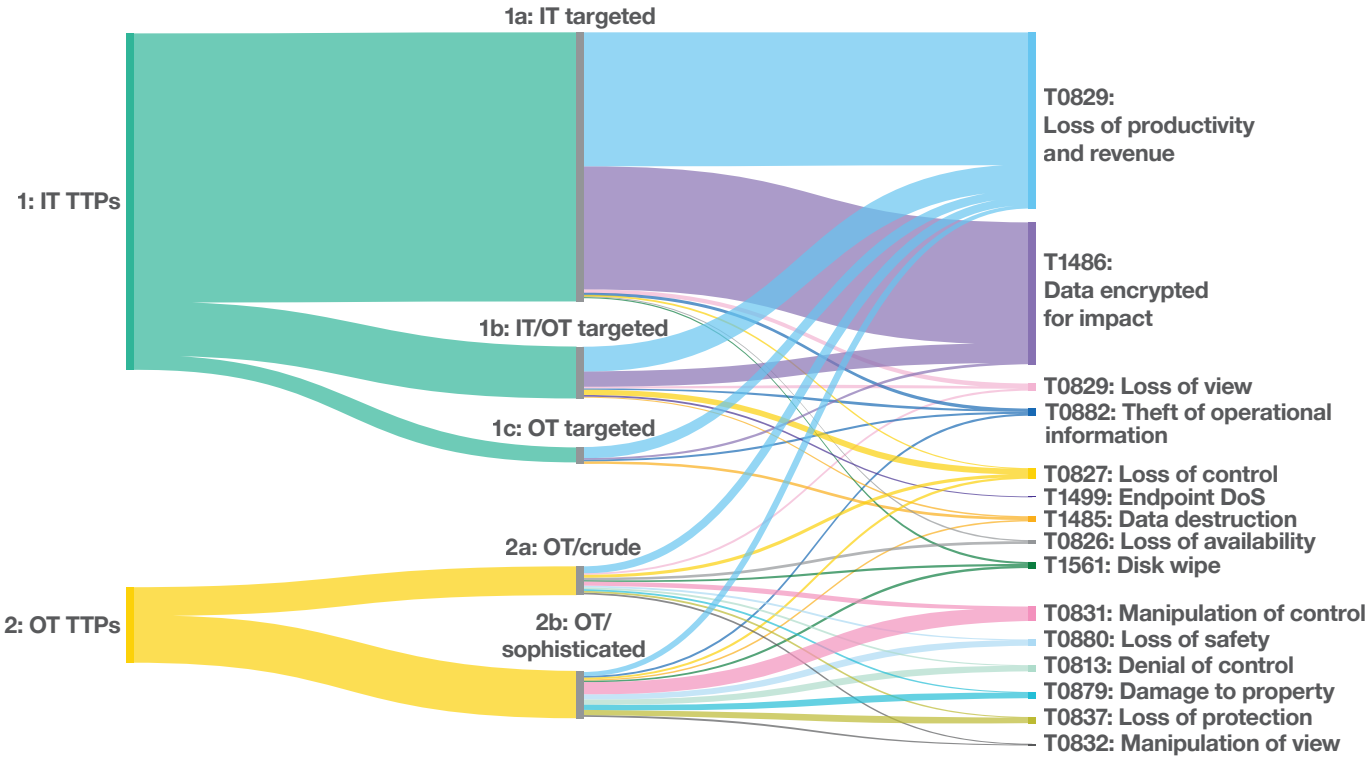


The Year in Context

We'll bring everything back together with a couple of visualizations using the whole dataset to give us an idea of how the past year has contributed to overall trends.

When it comes to the various types of impacts experienced by victims, it is no surprise that loss of productivity and revenue still dominates. What else probably comes as no surprise is the second most prominent impact – data encrypted for impact. Attacks that aren't the result of encryption-based Cy-X tend to have a more diverse range of impacts to be recorded. That could be due to more detailed reporting on more interesting attacks or a product of the attacks themselves - we'd guess the former is a bigger contributing factor. Like last year, we have singled out the impacts unique to category 2 OT-impacting cyber-attacks, which can be seen at the bottom right of the visualization. Manipulation of control remains the most prominent category 2-specific impact with the remaining unique impacts fairly evenly distributed.

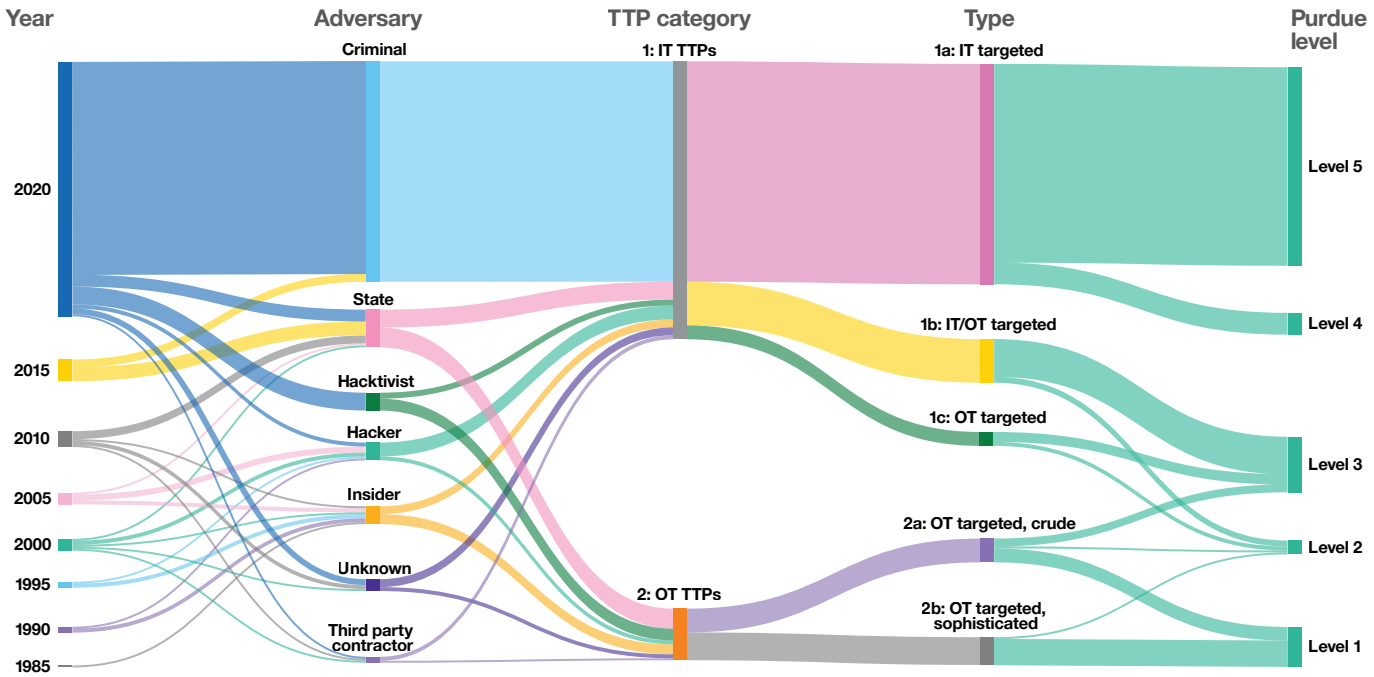
Finally, to wrap things up on this year's round up of cyber-attacks that impacted OT - the overview visualization. It depicts flows of incidents by year (in 5 year bins), into the adversary type that conducted it, into the category then type of cyber-attack, and finally into the depth of the Purdue model[glossary] reached by the adversary (although they all impacted level 0/1 in some way).



For readers of last year's Security Navigator, this might look familiar – and that's because it is. Despite the 39% growth in number of incidents in the dataset, the lack of diversity in types of cyber-attacks impacting OT environments means the visualization just gets bigger rather than changing in any notable way.

The elephant in the room remains the same as last year, albeit showing the biggest growth - that is criminals using IT TTPs to perform IT targeted attacks predominantly reaching no deeper than level 5 in the Purdue model. Of course, this is an unfortunate reflection of the Cy-X acceleration.

There is one very noteworthy positive to this dataset – tackling the Cy-X issue will drastically reduce the number of impacts that OT experiences due to cyber-attacks. We will then be left with predominantly category 2 attacks to concern ourselves with, which tend to be much less frequent and require much more capability to execute.



A Focus on Category 2 Cyber-Attacks

Last year, the focus of our OT article was on the Cy-X attacks because of their overwhelming presence in the dataset. The article revolved around category 1 OT cyber-attacks, with only a brief mention of what Cy-X may look like if the modus operandi was reimagined as a category 2 attack purposefully targeting OT. This year we'll shine a light on the attacks directly targeting OT with OT TTPs – category 2 attacks.

Cyber-attacks on OT, particularly category 2 attacks, are not as common as their IT counterparts. This is for a few possible reasons, including that OT is not encountered as frequently in victim environments, it is often segregated from the IT and Internet to some extent, and causing an impact to it generally doesn't fit into the motivations of most adversary archetypes. This comparative lack of frequency generally means that the threat of an OT cyber-attack is low, which unfortunately has created a common misconception that the resultant risk of an OT cyber-attack is low. However, threat is only one factor that contributes to cyber risk, the other factors are vulnerability and impact. When it comes to vulnerability, it is well established that there are concessions made due to the requisite openness and demand for uptime of OT, but the potential impact of any cyber-attack on OT is what really drives the risk. Simply causing downtime in an OT environment has a quantifiably substantial financial impact, but that is only part of the problem. Since OT cyber-attacks began, physical impacts have been felt around the world, affecting a wide range of sectors. This threat to human safety is what makes the lack of frequency of OT attacks almost irrelevant – the potential impact is so great that the risk is unacceptable no matter how unlikely the threat is. This is particularly true in critical national infrastructure (CNI).

But what about the actual category 2 attacks that have occurred? Who is conducting them? What impacts are they achieving? And what might their motivations be? Let's dig in…

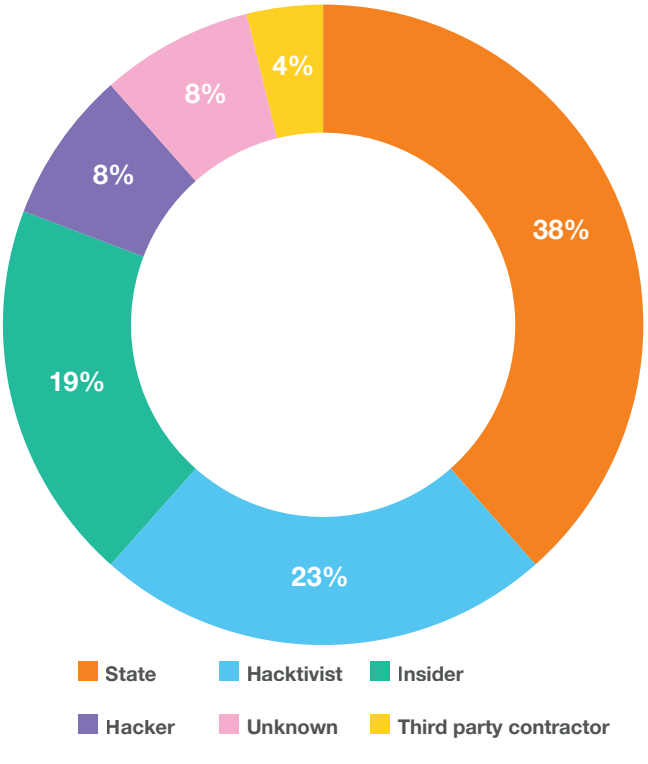
How infrequent is infrequent? In our dataset it equates to 26 attacks over 36 years, approximately 16% of our recorded OT-impacting cyber-attacks. This comes with the usual caveat that our dataset has limitations of public sources and only concerns itself with cyber-attacks that have had an OT impact. We may not have included attacks that were too sensitive to be reported or were focused entirely on espionage, which are both particularly poignant for category 2 attacks. Regardless, those 26 attacks over time don't show any pattern.

When it comes to whodunit the most frequent offenders are state actors, at 38% (10) of category 2 cyber-attacks, which makes sense given the scale and complexity of sophisticated OT-targeted cyber-attacks. Following that are hacktivists with 23% (6) attacks. This appears to be a growing trend with hacktivist groups either claiming to have attacked OT or attempting to demonstrate capability, sometimes successfully[117]. The third most frequent is the insider threat, at 19% (5) of category 2 attacks, which were more prevalent earlier in the dataset.

In terms of sectors most affected, manufacturing drops to second place with 23% (6) category 2 attacks, as can be seen in the chart below. Instead, utilities experienced the highest volume of category 2 attacks at 46% (12). This shift might be indicative of the intentions of such attacks. Cy-X, the bulk of category 1 attacks, may not target utilities as frequently because of the attention it could attract as an attack on CNI.

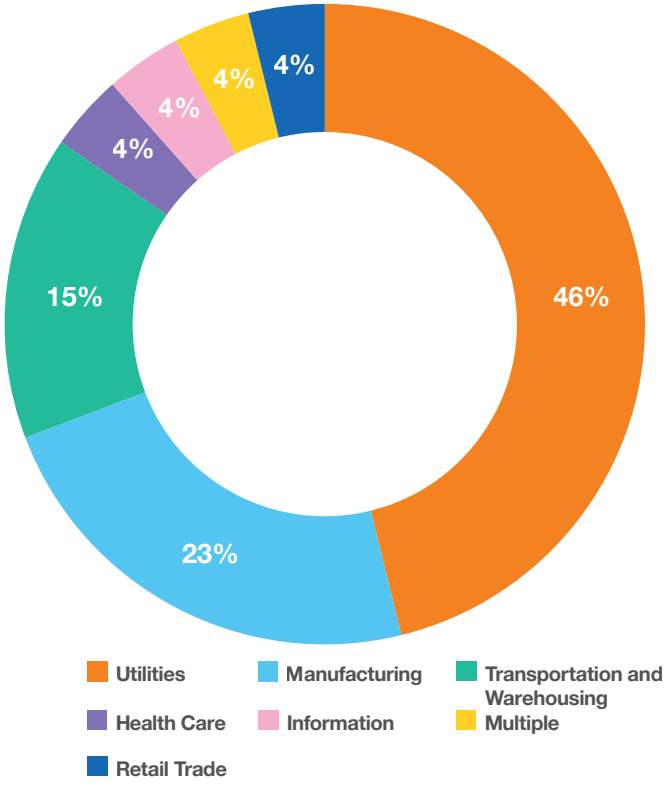
Category 2: Adversaries

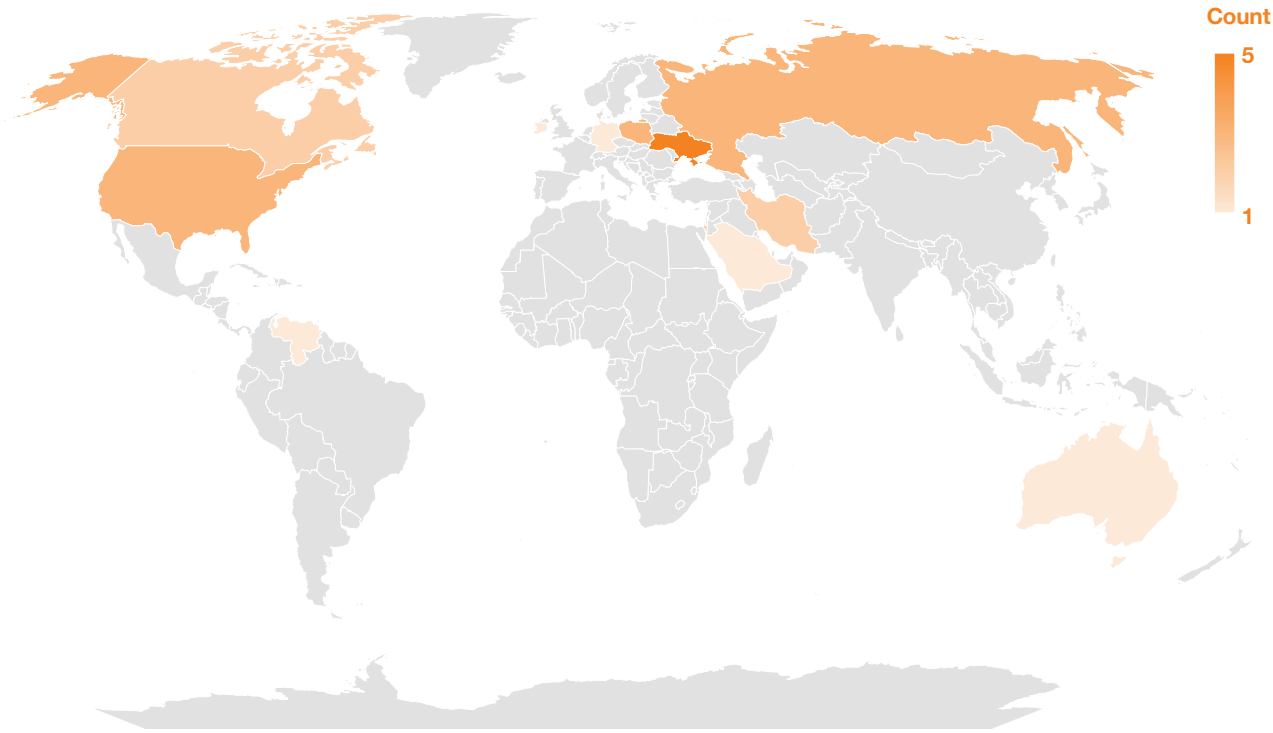
Attacker Typology for OT Focused Cyber-Attacks '23/'24



Category 2: Sectors

Victimology of OT Focused Cyber-Attacks in '23/'24





Geographic distribution of category 2 attacks

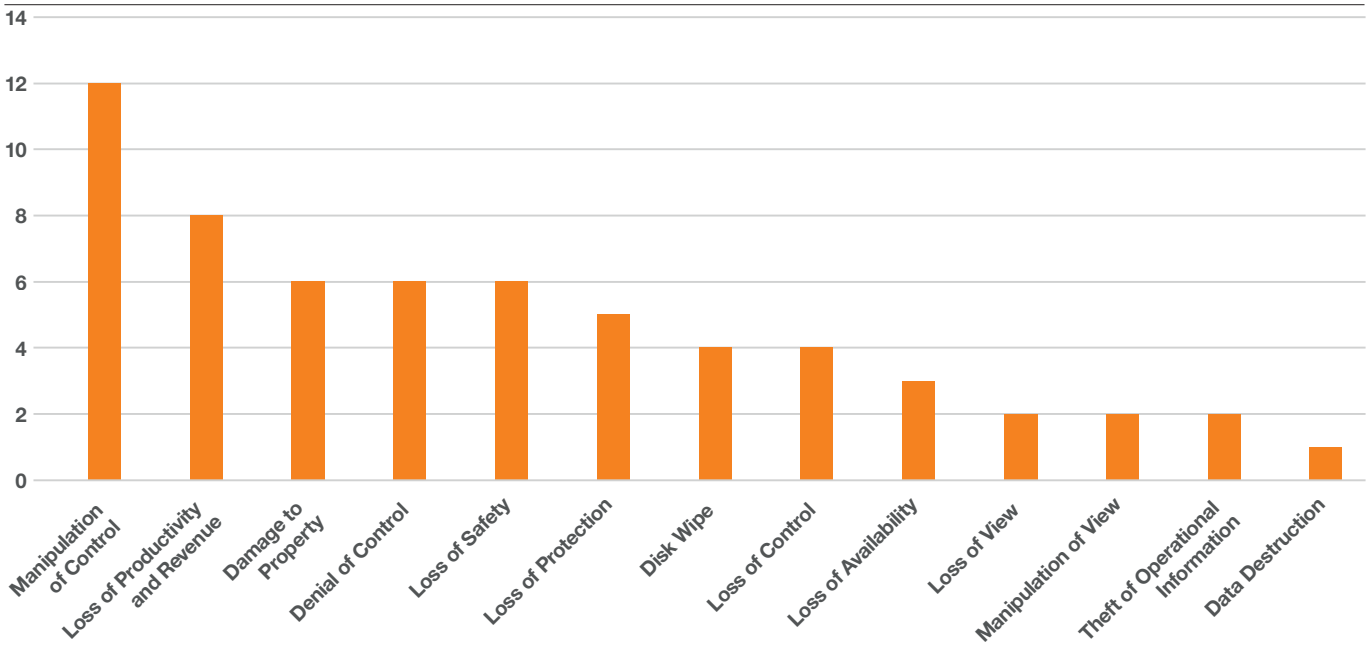
The geographic spread of category 2 attacks looks quite different without the bias of Cy-X actors. Ukraine has experienced 19% (5) of our recorded category 2 attacks, which probably comes as no surprise to those who have been paying attention to these types of attack given their publicity. Poland, Russia, and USA share 12% (3) each, none of which follow any pattern and tend to be isolated events.

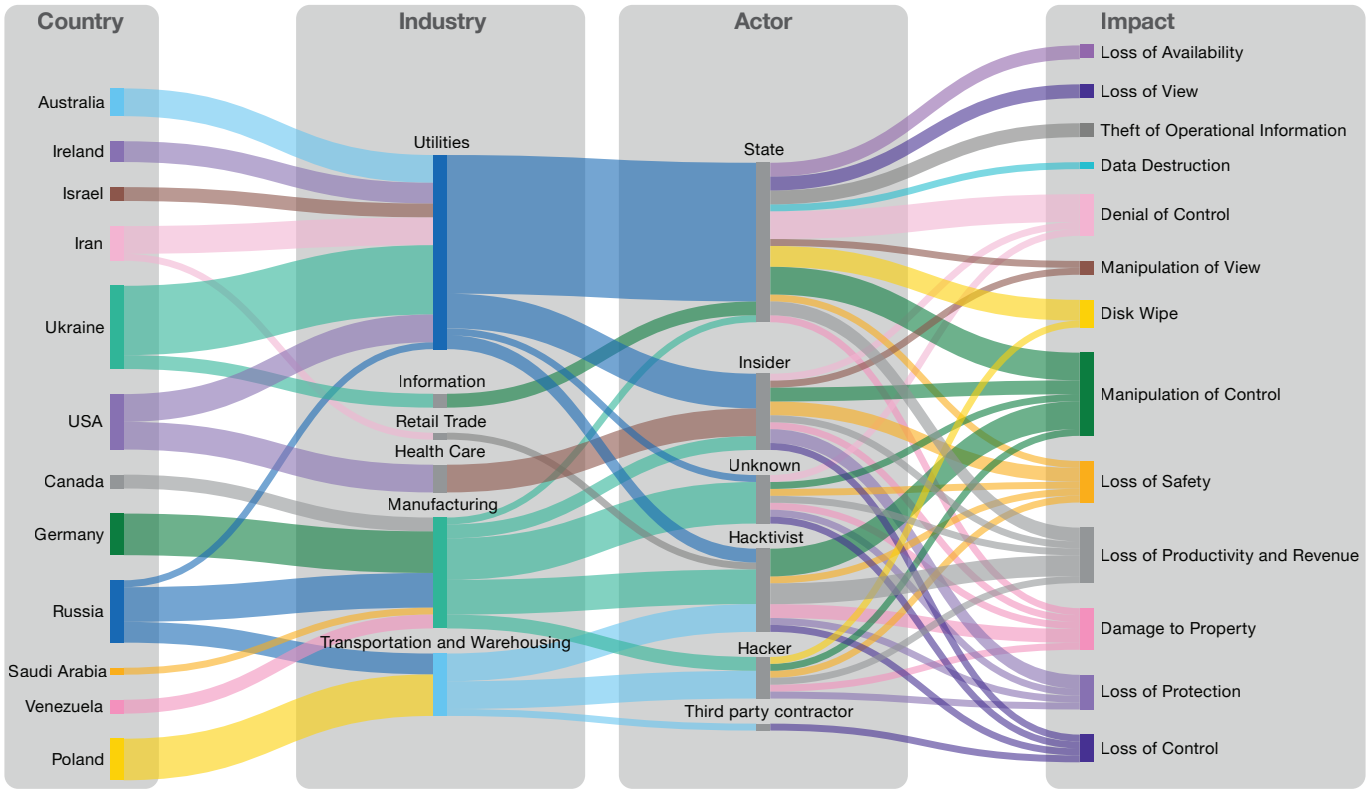
When looking at the impacts caused by category 2 cyber-attacks, we begin to see why the risk of these attacks is so high despite the low frequency. 46% (12) of category 2 attacks in our dataset experienced manipulation of control as an impact.

This means that the adversary manipulated the physical process in their attack, which from prominently reported attacks the potential for damage should be clear. But it isn't just manipulation of control, most types of impact caused by category 2 attacks are severe.

Impact of OT Cyber Attacks

Count of Impact Types From Category 2 Attacks





Bringing It All Together

We can get an overview of category 2 attacks by visualizing their flows. This depicts victim country, into victim sector, into adversary type, into impacts caused. What becomes immediately clear is the diversity of category 2 attacks when they are not overwhelmed with Cy-X. An interesting section to note is the utilities sector being targeted in conflict areas – Israel, Iran, Ukraine, and Russia – targeted predominantly by states. Russia also experienced category 2 attacks by hacktivists toward their manufacturing and transportation and warehousing sectors. We pointed out earlier in the article that category 2 attacks experienced manipulation of control as their most frequent impact. This visualization shows that the lion's share of those impacts was from state and hacktivist actors.

One noticeable trend that isn't directly apparent in the data is that of adversary motivation in relation to their desired impact caused. Clearly, when including category 1 attacks, the main motivation observed is financial gain by cyber criminals with encryption and data exfiltration being the desired impact. However, once we focus on category 2 attacks, we see such a diversity of victim countries, victim sectors, adversary types, and impacts that it isn't so clear-cut. Ignoring the hacker and unknown adversaries as they were never truly identified with a cause, instead focusing specifically on state, insider, hacktivist, and third-party contractor provides us with something more concrete.

Typically, states focus on strategic goals that are more overt in times of conflict. Espionage and prepositioning are two likely goals of states, particularly prior to conflict, but they aren't included in our data due to lack of OT impact. The impacts we have recorded suggest quite violent or disruptive motivation. More specifically, states have focused on clandestine process degradation[118] or optimized, abrupt, and long-lasting process damage[119].

Clandestine process degradation: Subtle, hard to detect TTPs that make small changes in the victim's process. Telemetry may be altered to make the attack look like an engineering issue.

Optimized, abrupt, and long-lasting process damage: Well researched attack that causes the biggest impact the adversary can achieve, typically happens quickly to limit the response, and causes as much downtime as possible.

Contemporary insider attacks are either less frequently reported or simply less common. There hasn't been a category 2 insider attack recorded in our dataset since 2009. Insiders tend to act on a motivation of revenge, which means a focus on damage to an organization's physical infrastructure as well as revenue – optimized, abrupt, and long-lasting process damage. Insiders present some of the biggest potential for damage in a category 2 OT cyber-attack because they already likely know the environment they want to disrupt, meaning they know how to optimize their attack[120]. This phenomenon is similar for third-party contractors, too[121].

Which hacktivists are conducting OT-impacting cyber-attacks, and what that impact is, is up for debate. A crucial motivation of hacktivist groups tends to be notoriety, meaning they're incentivized to embellish or even entirely fabricate stories of successful attacks. Category 2 attacks are no different from this trend, and discerning the valid ones is not without its challenges. No less because the trend of hacktivists targeting OT with category 2 attacks has seemingly accelerated since 2020, whereby perpetrators often align with a state on one side of a current conflict – in some cases, they align a little bit too closely. This means that it is difficult to say whether such attacks are strategic, state-sponsored/proxy attacks or legitimate, independent hacktivism fighting for a patriotic cause. Regardless of who you believe is a hacktivist or what attacks they achieved, they generally favor one type of impact – optimized, abrupt, and long-lasting process damage[122].

For every adversary type, the described examples of OT impacts with category 2 attacks are best achieved with sophistication, capability, and resource – meaning type 2b OT impacting cyber-attacks that involve understanding the victim environment and crafting a bespoke attack with complex OT TTPs. However, that does not diminish the potential damage caused, and therefore risk posed, by type 2a attacks – those that still involve the use of TTPs unique to OT but perhaps do not spend as much time optimizing.

The majority of OT TTPs that distinguish category 2 attacks involve the use of native functionality against the victim, known as living off the land. However, living off the land in OT is often distinct from in IT due to its focus on the process and physical environment, so we have taken to making that distinction clearer by calling it "living off the plant". An advantage of this strategy is blending in with the victim environment to evade detection, but in OT it goes further. From an adversary's perspective, it is much safer to achieve their goals by using native functionality that a programmable logic controller (PLC) expects than by abusing its memory with an exploit. This applies to anything in an OT environment that might be critical to the process and is particularly effective because of OT's requisite openness. Although living off the plant techniques are effective, simply having access to an OT environment does not mean using them is trivial nor that the desired impact is feasible. That then poses the question, how does an asset owner know their OT environment's susceptibility to category 2 living off the plant techniques?

We'd like to thank the Research Institute in Trustworthy Inter-connected Cyber-physical Systems (RITICS) for funding this ongoing research. The following is not representative of the project's overall outcomes and simply represents work to date.

The Efficacy of OT Penetration Testing

This year we embarked on a project to understand the state of the art with regards to OT penetration testing. The main aims of the project are to identify key challenges of the discipline, along with pertinent areas for research and development to improve it. In identifying the challenges, one of the research questions was 'does OT penetration testing effectively test for TTPs encountered in real attacks?'. The primary research is still ongoing, but the background literature review provides some clues that we'll discuss here. In the literature there are 4 approximate categories that contribute towards this area: Kill chains, guidance, methodologies, and research.

Kill chains offer overviews of adversarial tactics, generally in a linear fashion, to describe how an attack may occur. There are various kill chains that pertain to OT cyber-attacks, such as the Industrial Control System Cyber Kill Chain[123] and the Cyber-Physical Attack Lifecycle[124], but we have also included more comprehensive offerings such as the TTP-focused MITRE ATT&CK® Matrix for ICS[125]. One feature that is recognizable immediately is their homage to the IT side of the attack that generally precedes a category 2 OT attack. What this also means is the recognition that the IT and OT parts of the attack are distinctly different and the TTPs objectively shift when entering the OT environment.

Kill chains and similar concepts aren't directly OT penetration testing literature, but it is important to understand the industry's interpretation of an OT cyber-attack first.

Guidance, such as that found in ISA/IEC 62443[126] or NIST SP 800-82r3[127], is sparse when it comes to OT penetration testing. This category of literature is intended to be holistic and not solely focused on penetration testing, so shouldn't be held accountable for defining how it should be conducted. However, the guidance provided generally recommends penetration testing, but that comes with caveats about OT's fragility. Often there are compensating controls recommended, including replicated, virtualized, or simulated environments instead of testing in production, but as other guidance points out[128], those all have tradeoffs in realism.

Methodologies are a nebulous topic in OT penetration testing. Unlike other forms of penetration testing such as IT infrastructure or web applications, there are no formally defined methodologies. Instead, we turn to close approximations that are typically found in books such as Pentesting Industrial Control Systems[129], Industrial Cybersecurity[130], Industrial Network Security[131]. The trend common among all of these methodology approximations is that in a 'real test' the provider would first gain initial access to the IT network, breach the demilitarized zone, gain access to the OT and then it is 'game over' save for some possible IT TTPs against more IT-friendly devices in what would be considered level 3 of the Purdue model. In fact, for most publications, any testing of OT systems is simply not feasible in any way, with only isolated device testing in a controlled environment. Not only is testing the OT environment not feasible, it is often described as unnecessary based on the assumption that access guarantees the adversary free reign to do what they want. This trivializes the complexity of OT cyber-attacks that is even acknowledged by the kill chains mentioned earlier.

Research is equally as sparse as the methodology literature, with few publications working on improving the OT penetration testing discipline. However, there are two areas of note. The first is work looking to improve the scoping of OT penetration tests by building in safety[132], which improves the methodological/process side of the discipline. The second is a small body of literature that ingests PLC project files (their code or configuration) to identify how variables can be manipulated to cause impact[133], which helps our understanding of how adversaries may cause low-level chain reactions.

As far as the literature is concerned, OT penetration testing is still very much in its infancy. The guidance is ambiguous and non-committal, the research does not currently support the growth of the discipline, and the lack of methodologies means current providers do not have a standard to base testing on. Moreover, the existing methodologies may be working within the limitations of production environments, but they are overconfident in their assumption that reaching the OT is enough. There is a focus on IT TTPs that are not fully representative of category 2 OT attacks, evidenced by historical attacks and the kill chains that model them. So, who are we emulating with our OT penetration testing, the adversaries we're looking to preempt and stop, or IT penetration testers?

As we've mentioned there is primary research to be done meaning our understanding of OT penetration testing may change. We will continue to release those results as the project progresses, so stay tuned.

Diana Selck-Paulsson
Lead Security Researcher
Orange Cyberdefense

Ben Gibney
Security Analyst
Orange Cyberdefense

Research: Hacktivism

Exploring the Intersection of Cyber Activism and State-sponsored Operations

Introduction

Since the war against Ukraine began in February 2022, hacktivism has surged[134][135][136], impacting both private and public sectors through DDoS attacks, defacements, and disinformation campaigns. These cyberattacks align with geopolitical events. As 2024 sees over 50 countries holding elections[137], this creates particularly ripe conditions for influence operations. DDoS attacks, driven by political tensions, have intensified, with one pro-Russian group alone claiming over 6,000 attacks since March 2022. Driven by political tensions and geopolitical conflicts[138][139], DDoS attacks in 2024 have significantly increased in both volume and intensity[140]. Hacktivists are now more experienced, leveraging DDoS-for-hire services[141][142] and sophisticated tools.

Last year, we tracked attacks by major pro-Russian hacktivist groups, identifying regional patterns often linked to patriotism from actors in conflict zones. To better understand the complex threat landscape, we aim to explore current hacktivism more deeply, examining its various facets and connections to geopolitical tensions, building on our previous findings.

This research explores how volunteer-based, multinational groups operate during warfare, comparing modern hacktivism with past movements and examining its potential implications for the future.

Disclaimer

Hacktivism is a complex issue, and this article doesn't cover all actors or activities from the past year. Our perspective, shaped by Western, English-language viewpoints, may limit our understanding of the broader phenomenon. We avoid naming the Hacktivist group, as it thrives on attention.

Historical Context of Hacktivism

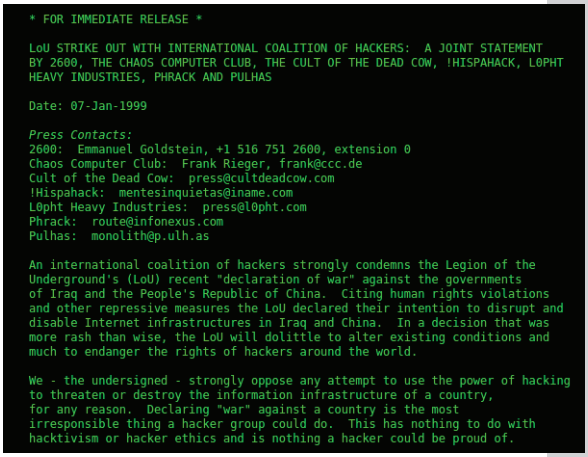
Hacktivism has evolved through three key eras, which we describe as follows. The first, the Digital Utopia era, was driven by ideals of building a better internet, as seen with groups like Chaos Computer Club (CCC)[143]. Next came the Anti-Establishment era, where hacktivists exposed the flaws in how cyberspace developed, often opposing entrenched powers. The current Establishment era sees groups shifting from anti-establishment actions to aligning with state agendas. Traditional hacktivism, which rejects state control, differs from this, as state-sponsored activities transform into cyber operations or warfare rather than true hacktivism.

Evaluating the evolution of these groups offers key insights into the factors shaping today's hacktivists. Understanding how they differ from their predecessors reveals current motivations, which can ultimately help in developing better strategies for defending against them.

Beginning of the Digital Utopia Era

We begin in the mid-1980s and continue until the mid-2000s, with the Digital Utopia Era of hacktivism. This was an era before the dot-com boom had occurred, only 42% of Americans had ever used a computer in 1990 and only 22% of Europeans households having internet in 2001[144][145]. Given the landscape had not been built, this allowed those involved – the early adopters - to act based upon ideals. And while some of the ideals varied from group to group, the actions were normally grounded in similar ideals. Examples include the Electronic Disturbance Theater (EDT) acting in accordance with civil disobedience and pioneering digital protest tactics such as virtual sit ins and the Cult of the Dead Cow (cDc) believing in free access to information, privacy rights, and the exposure of vulnerabilities in systems used by powerful institutions[146]. Besides often being credited as pioneers of early hacktivism, cDc can also be considered one of the first hacktivist groups testing influence campaigns and media manipulation.

Although not the first to manipulate the media, early hacker groups quickly understood the media's hunger for sensationalism[147]. On the other side of the Atlantic, in Germany, there were groups such as Bayrische HackerPost (BHP) who created information sheets to help educate people about technical and political issues. At one point they attempted to hack into the German government to remove census information, as they did not believe this type of personal information should be stored by the government. Another German-based group is the Chaos Computer Club (CCC) who promoted hacker ethics such as free access to information, mistrust of authority, privacy and ethical use of technology[148]. In the late 90s for example, the CCC and others condemned the Legion of Underground's (LoU) for "declaring" war on the People's Republic of China and Iraq[149] because they violated human rights, as can be seen to the right. Despite differences, these groups shared a belief in an internet built on ideals benefiting society.



```
* FOR IMMEDIATE RELEASE *

LoU STRIKE OUT WITH INTERNATIONAL COALITION OF HACKERS:  A JOINT STATEMENT
BY 2600, THE CHAOS COMPUTER CLUB, THE CULT OF THE DEAD COW, !HISPAHACK, L0PHT
HEAVY INDUSTRIES, PHRACK AND PULHAS

Date: 07-Jan-1999

Press Contacts:
2600:  Emmanuel Goldstein, +1 516 751 2600, extension 0
Chaos Computer Club:  Frank Rieger, frank@ccc.de
Cult of the Dead Cow:  press@cultdeadcow.com
!Hispahack:  mentesinquietas@iname.com
L0pht Heavy Industries:  press@l0pht.com
Phrack:  route@infonexus.com
Pulhas:  monolith@p.ulh.as

An international coalition of hackers strongly condemns the Legion of the
Underground's (LoU) recent "declaration of war" against the governments
of Iraq and the People's Republic of China.  Citing human rights violations
and other repressive measures the LoU declared their intention to disrupt and
disable Internet infrastructures in Iraq and China.  In a decision that was
more rash than wise, the LoU will dolittle to alter existing conditions and
much to endanger the rights of hackers around the world.

We - the undersigned - strongly oppose any attempt to use the power of hacking
to threaten or destroy the information infrastructure of a country,
for any reason.  Declaring "war" against a country is the most
irresponsible thing a hacker group could do.  This has nothing to do with
hacktivism or hacker ethics and is nothing a hacker could be proud of.
```

The 1986 Computer Fraud and Abuse Act (US)[150] and the 1990 Computer Misuse Act (UK)[151] marked a turning point by criminalizing some hacktivist activities. These new legislations might thus have ushered in the end of the Digital Utopian Era and set the stage for the next.

Moving to the Anti-Establishment Era

Where the first era of hacktivism was filled with optimism, by the mid-2000s, the second wave was characterized by cynicism, at times even bordering on nihilism[152]. Groups like Anonymous, Wikileaks, and Lulzsec emerged, disrupting establishments like governments, corporations, and institutions without aligning with any ideology. Lulzsec, driven by humor rather than political change, aimed to embarrass companies. The vision of a digital utopia had faded, and the groups during the Anti-Establishment Era focused on bringing down unjust systems and exposing establishment systems of oppression. Hacktivism became reactionary, often retaliating against wars, as increasing digitalization widened the attack surface. An anti-war focus began to emerge, more actions were taken by groups in direct retaliation to ongoing wars[153]. Nevertheless, these activities were still executed from an anti-government point of view, which was typical for this and the previous era. There is no universal answer as to what brought an end to the Anti-Establishment era. One of the main causes could have been the number of arrests occurring across the different groups[154]. It became very hard to recruit people to a group named Anonymous when so many of the members were identified.

Arriving at the Establishment Era

Out of the ashes of the Anti-Establishment Era came the Establishment Era, which can be viewed as emerging around 2014. From here many groups started to openly profess support for certain establishments, like governments, religious institutions and nation-states. Modern hacktivism is more often intertwined with geopolitical conflicts. The motivations have also expanded to include support for state-affiliated campaigns, cyber protests, or disruptions tied to national or regional interests, thus supporting an establishment. Earlier in this phase of hacktivist activity included geopolitical conflicts such as the 2007 DDoS attack against Estonia[155], cyber operations during the Russo-Georgian War in 2008[156], and the Arab Spring where hacktivists supported pro-democracy movements across the Middle East and North Africa[157]. But this era began revealing its true character from 2014, during Russia's illegal annexation of Crimea. In that year, volunteers began mobilizing themselves to take political action in support of their government, carrying out defense-like activities. The mobilization of private capabilities and non-state actors[158] in 2014 in Russia's war against Ukraine did not fully succeed in its strategy[159] but did provide almost a decade of preparation for countries like Ukraine in terms of cyber resilience[160]. When Ukraine was attacked again in 2022, it was able to mobilize its cyber capabilities and digital resistance movement more effectively.

Modern Hacktivism

In the modern era hacktivists utilize more advanced techniques. This is partly due to technological advancements and the sharing of skills and tools in the shared economy model (albeit at times with malicious intent), and partly because state-supported hacktivists might have opportunities to tap into better resources. DDoS attacks have consequently scaled exponentially in size and sophistication, with modern groups claiming and executing DDoS attacks that generate billions of requests per second[161][162] or consume 3.8 terabits per second (Tbps)[163][164] in bandwidth[165]. We also observe a significant shift in the operational methods of hacktivist groups, especially a growing reliance on DDoS-for-hire services and crowdsourced DDoS tools[166].

The volunteer-based nature of these groups enables them to scale attacks more effectively, as participants need minimal technical expertise and are incentivized through cryptocurrency rewards. This is an interesting shift since early hacktivists movements were primarily motivated by ideological or political causes, rather than financial rewards. One explanation for this is that as the cybercrime economy evolved and DDoS-for-hire services became more accessible, the line between financially motivated attackers and ideologically driven hacktivists began to blur. Hacktivists in this era also started to cross the line to impacting critical infrastructure and Operational Technology (OT) systems[167][168]- previously the domain of organized cybercrime or state actors.

Today, hacktivist groups operate in smaller, and more independent groups; and many of the more prominent hacktivist groups align themselves with major powers, allowing them to operate with less fear of authorities and prosecution compared to groups from previous eras.

While most observed hacktivism attacks still focus on IT systems, the aim of hacktivism is increasingly less about technical disruption and more about shaping public opinion and spreading fear, uncertainty and doubt (FUD) through targeted manipulative campaigns[169][170]. For instance, information operations in the Nordics escalated tensions during Sweden and Finland's NATO accession.

Modern hacktivists have shifted from anti-government positions, like opposing censorship, to supporting pro-government agendas through cyber operations. Unlike earlier hacktivists who focused on individual rights and ethics, today's groups often lack a history of activism. Hacktivism has evolved through three phases: the digital utopian era, which envisioned a better internet, the anti-establishment phase, which opposed perceived injustices and an evil establishment, to the current establishment era, where hacktivists align with state-backed cyber objectives. In this new era, traditional hacktivism that still operates and focuses on access to information, privacy, fighting oppression and advocating for ethical use of technology, is overshadowed.

Case Study: How Does Modern Hacktivism Look?

This study analyzes one of the most active pro-Russian hacktivist groups since March 2022, focusing on its communication strategies, narrative construction, and geopolitical influence. It also examines the group's alignment with state actors, values, and its role within the broader ecosystem. While this report focuses on just this one group, its prominence among peers offers valuable insights into similar pro-government hacktivist groups, allowing the study to reflect broader behaviors and tactics seen across this threat actor landscape.

Data Collection

Our data was collected through systematic scraping of the hacktivist group's Telegram channel monthly over a period of two years, from August 2022 to August 2024. The dataset renders:

- **3,214 unique messages:** These messages included descriptions of the group's targets and other contents the group felt to share with the broader public. Thus, the messages serve to capture the group's narratives.
- **6,674 unique targets:** These targets encompass a wide range of entities attacked by the group, provided and proven by the actors by posting a check-host link - an internet monitoring service commonly used by hacktivists as proof of the success of their Service DDoS attacks.

To ensure data consistency, scraping was conducted at the same time each month. The data includes textual content (reasons for targeting), metadata (timestamps, views, forwards), and contextual information about the targets. After processing, the exact number of targeted organizations and countries was determined.

Data Processing

To analyze the communication patterns and geopolitical context of the hacktivist group, we analyzed the textual content of each message using natural language processing (NLP). We applied text preprocessing and named entity recognition (NER) to identify country references, refining the results with a custom list of known countries and nationalities. The extracted country information was added to the dataset, allowing us to examine the group's geopolitical focus and alignments.

Analysis

Before discussing the data, it's important to summarize recurring themes in pro-Russian Telegram posts. These narratives aren't unique to one group but are common across several pro-Russian cyber activists[171]. The group frames its actions as retaliation for Russophobia[172], Western support for Ukraine, or sanctions on Russia.

Messages often mock targeted nations, criticizing leaders for prioritizing Ukraine over domestic issues.

They use militaristic language, praising Russia's military and positioning themselves as cyber warriors defending Russia's interests, and aligning with broader narratives of resisting Western influence.

> "This is not the first year that we have been defending Russia's interests on the information front. We see how the discontent of adequate citizens of foreign countries is growing, whose authorities do not care about the problems of their compatriots and spend huge amounts of money on sponsoring Ukrainian terrorists. We also see total censorship, which prevents the residents of these countries from telling the truth. There it has become unacceptable to speak positively about Russia. There is absolutely nothing left of freedom of speech in the West[...]"
>
> Excerpt from one of the announcements on the Telegram channel

The group occasionally references subscriber requests and volunteer input, showing they incorporate follower feedback when selecting targets. This fosters community involvement and introduces a crowd-sourcing aspect to their cyber operations.

Victimology

Why we see specific targets being attacked – a contextual analysis.

The group's activities against targets serves both as a disruption tool and a symbolic statement against specific nations. By attacking organizations tied to everyday services, they retaliate against perceived wrongs and express disapproval of the nation's political stance, particularly regarding Russia and Ukraine.

Their strategy aims to influence international perception while creating domestic instability. Attacks on services like public transport or banking systems highlight institutional vulnerability, reinforcing their narrative that the state is failing to protect its citizens.

Consequently, it doesn't necessarily matter who the victim is at an operational level—it's more about what the organization symbolically represents in the context of a broader political or geopolitical message.

What does the data tell us?

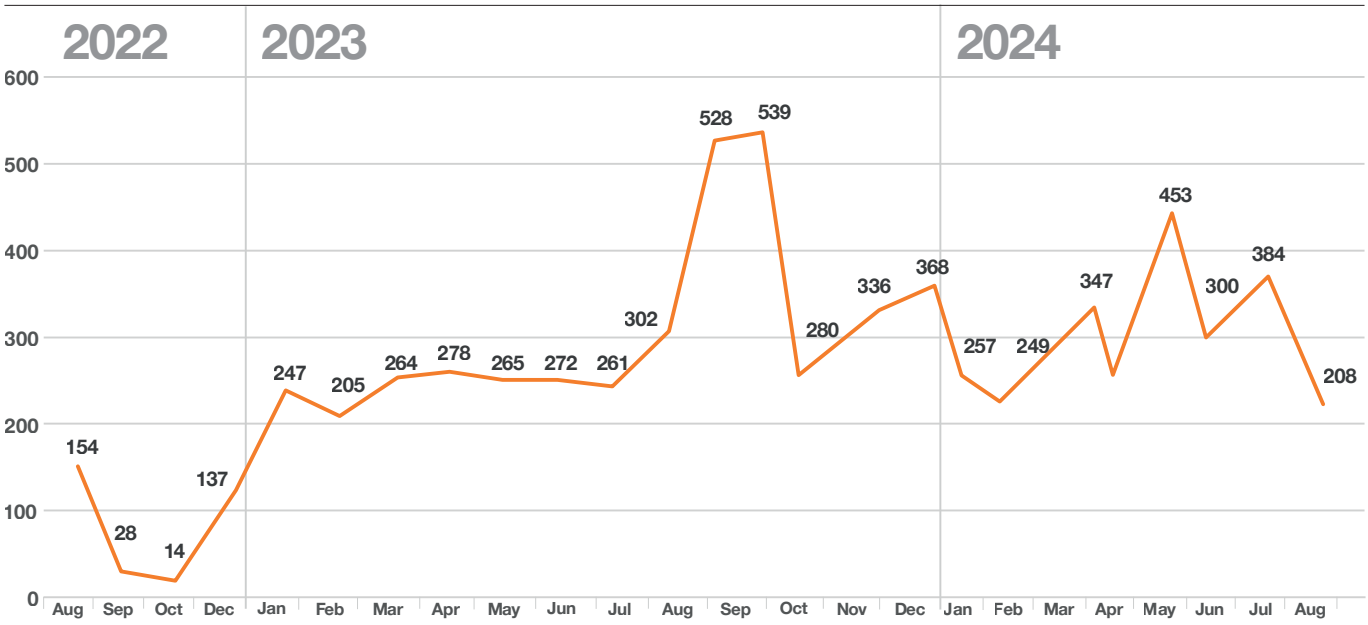
In the following paragraphs we analyze how many targets this specific hacktivist group has attacked over a two-year period. The group posted 3,214 unique messages. Within these we identified 6,674 targets from the private and public sector, averaging around 280 targets per month.

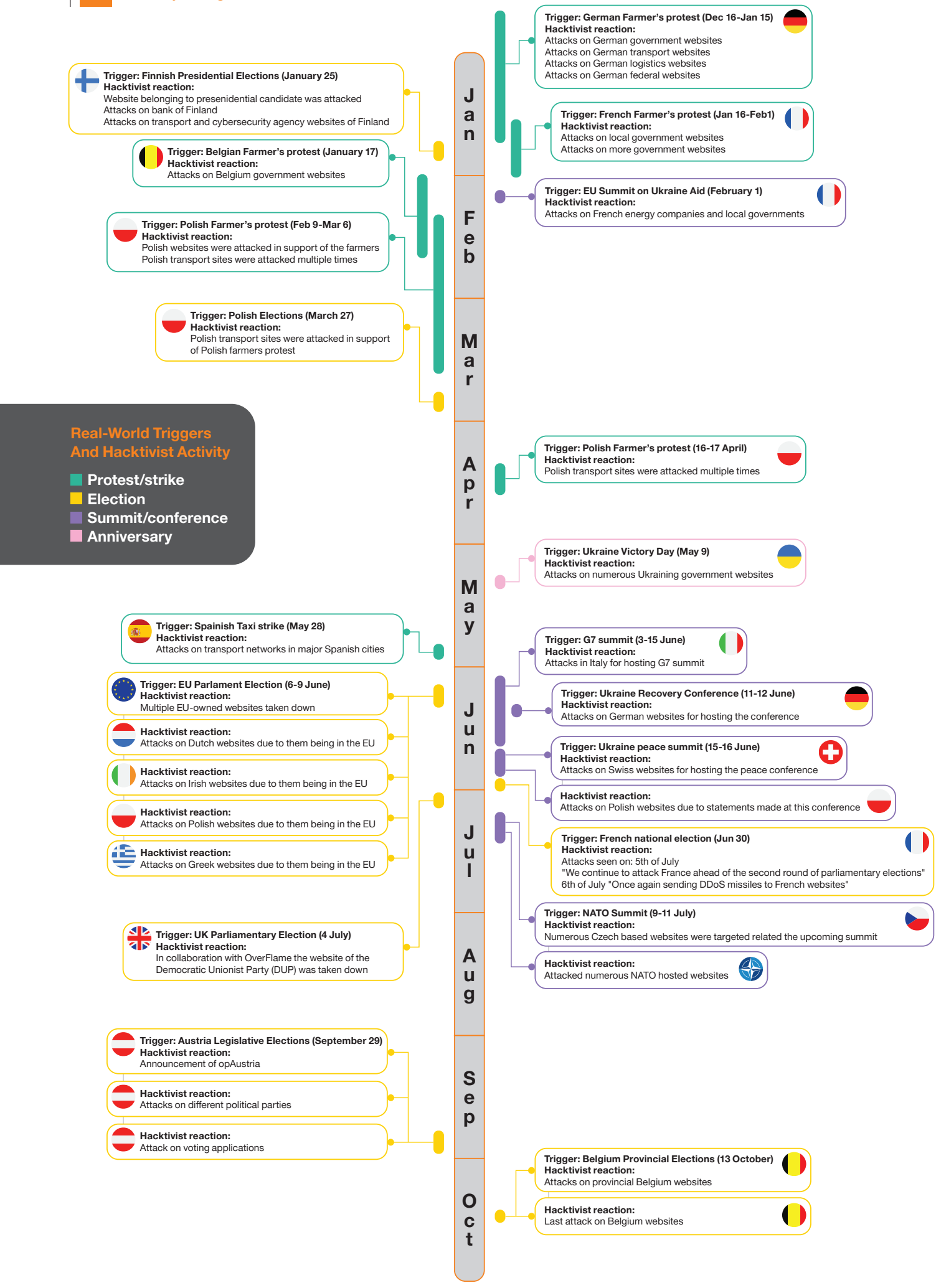
The volume of messages fluctuated, potentially suggesting organized campaigns, likely timed to align with key political or military events. The group's focus appears to shift in response to geopolitical tensions, elections, or other notable events, reflecting a calculated effort to exert influence. This we will investigate below (under Geopolitical impacts).

In September and October 2023, we see a significant increase in activity. Analysis of the message contents indicates that Germany, Finland, Czech Republic, Canada, United Kingdom and Sweden were particularly heavily impacted. This surge coincides with key events such as national holidays (e.g. Czech Republic's national day), international meetings (such as the Malta Peace Formula meeting) and high-profile scandals (such as the Canadian Parliament incident[173]).The alignment allows the group to frame these cyber operations as symbolic acts of punishment.

Count of Targets Over Time

Number of Hacktivist Activities Observed Since 2022





Legend:

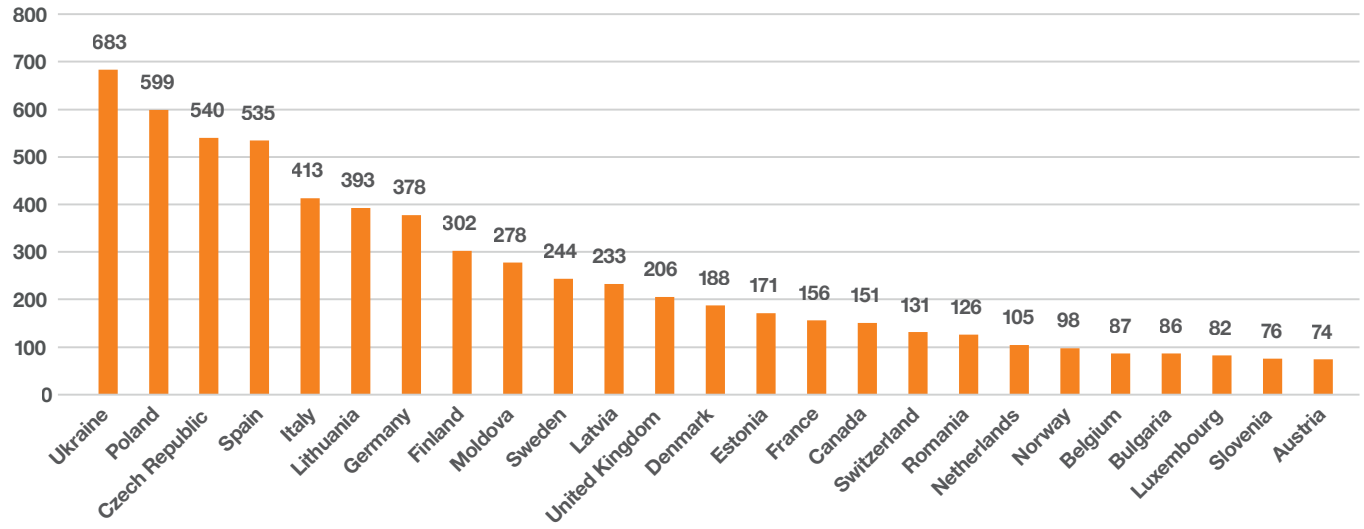
Number of circles: number of Hacktivist messages posted

Width of the circle: count of targets

- Protest/strike
- Election
- Summit/conference
- Anniversary

Top 25 Targeted Countries

Between August 2022 and August 2024



Our data shows that 42 distinct countries were targeted by this threat actor over two years, with 96% located in Europe. The attacks are primarily geopolitical, targeting countries rather than specific organizations. This becomes clear when analyzing the messages where the actors address the country they meant to impact, while at the same time posting a list of organizations that are meant to deliver the strategic message to a specific country and its civil society.

In the context of the war against Ukraine, Ukraine and Eastern European countries like Poland, Czech Republic, and Lithuania are heavily targeted, reflecting geopolitical expectations. Western European nations such as Germany, Italy, and France also faced significant attacks, reflecting their NATO and EU leadership roles. In France, the group exploited social unrest, aligning with local farmer protest movements and public dissent. A surge in Spanish victims was triggered by the arrest of two individuals in Spain tied to the group. Similarly, attacks on Germany carried anti-government sentiment and opposition to its leadership.

“As the rallies continue to rage in France, we support the [farmers] protesters and put down the communes” (26th of January 2024)

Finland and Moldova stand out for high attack volumes despite less direct involvement in the war against Ukraine. Finland's NATO membership and proximity to Russia drew increased attention, but Moldova saw almost 200 attacks in Q2 2024, primarily DDoS attacks targeting state infrastructure and fueled by anti-government sentiment. Moldova's vulnerability due to Transnistria likely contributes to its ranking. Spain and Italy also face frequent attacks, apparently in retaliation for their military support of Ukraine. Attacks focus on critical infrastructure and exploit internal dissent and are often framed as responses to Russophobia and arrests of Russian sympathizers[174]. Canada ranks unusually high among non-European targets, reflecting Russia's global cyber reach against NATO-aligned countries. The absence of the U.S. is notable, given its leading role in supporting Ukraine.

Pro-Russian hacktivists may focus on European countries due to their proximity to the conflict, where disrupting supply chains and infrastructure more directly impacts Ukraine.

Attacks on key transit hubs like Poland, or influential nations like Germany and France offer more immediate strategic gains than targeting the U.S.

Geopolitical Impacts

To analyze factors influencing target choices, we first identified relevant keywords linked to geopolitical events and extracted unique messages containing these keywords. Each message was then manually reviewed to confirm references to specific geopolitical events. This process enabled a focused analysis of how real-world developments may have shaped the group’s decisions. A summary of the keywords we observed is shown below.

Our analysis reveals consistent support for anti-EU protests. In particular. the Farmers' Protests in Poland, Belgium, and Germany. Multiple European elections (United Kingdom, France, Finland, Austria, Belgium and national independence days (Ukraine and Poland) were frequent themes. Election interference marked an escalation, aiming to disrupt democratic processes.

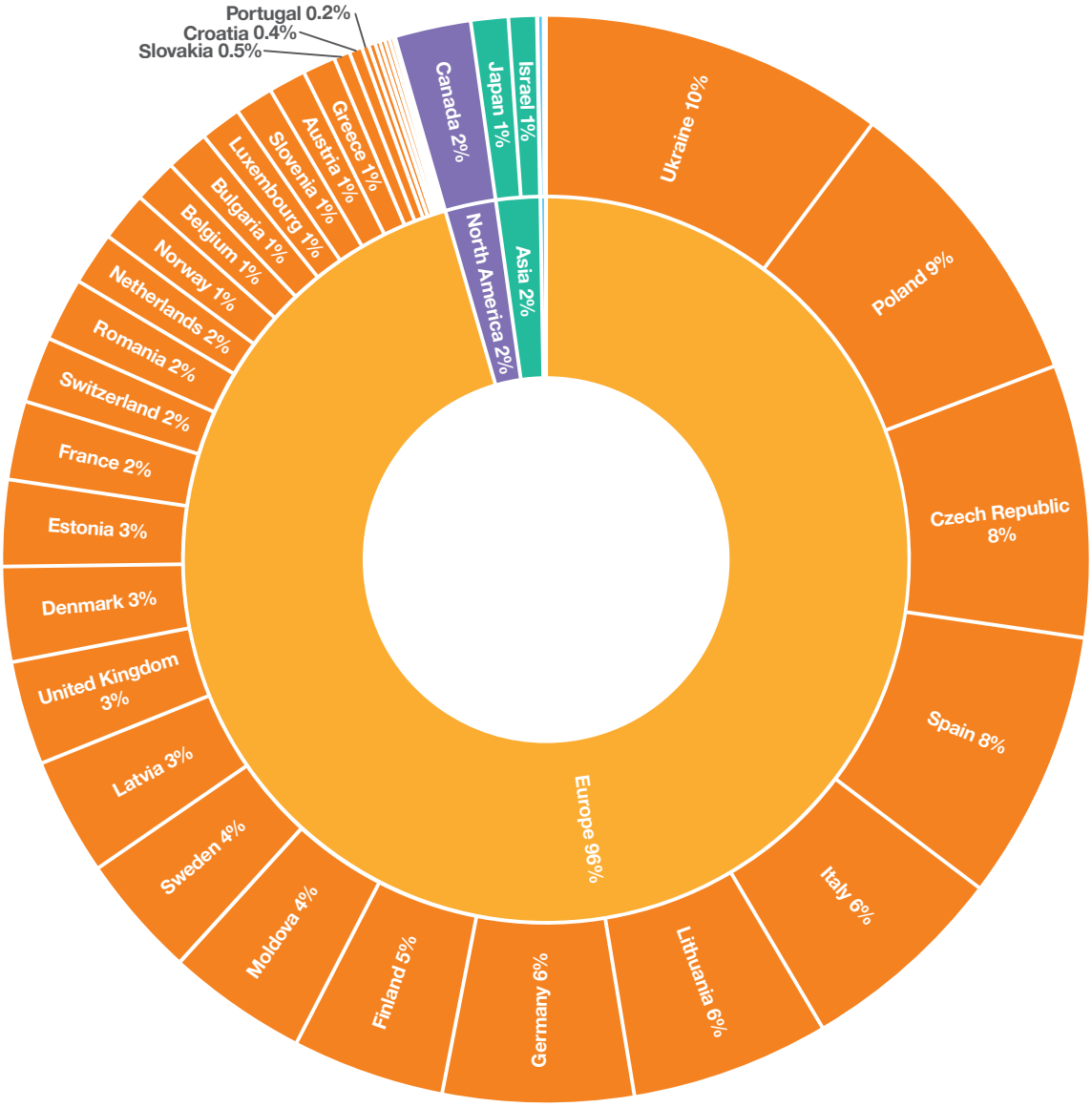
The group also reacted to international conferences, targeting host countries or responding to specific comments made at these events.

Election interference represents an escalation beyond typical DDoS attacks on infrastructure or military websites, as it directly targets the democratic process of a nation. By attacking election-related websites and portals, the hacktivist group aims to undermine public trust in the electoral system, disrupt the flow of information, and potentially influence the outcome of a key democratic process.

The group frequently responded to international conferences or summits by targeting the host country with cyberattacks. Occasionally, specific comments made during these events also triggered attacks against the countries involved. A summary of the events associated with selected keywords is depicted on the previous two pages.

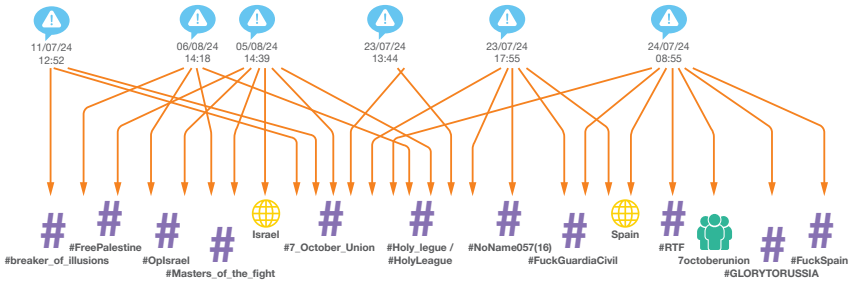
Regional Breakdown

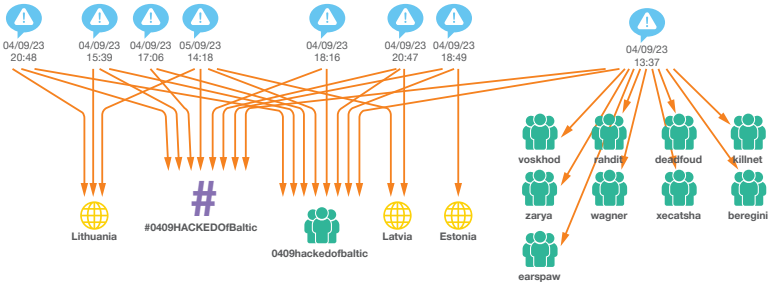
Geographic Regions and Countries Affected by Hacktivist Activities



The #0409HACKEDOfBaltic campaign is similarly notable, involving multiple groups attacking Latvia, Estonia, and Lithuania in response to the 4th of September 2023 NATO military exercises[175]. This attack lasted two days and displayed an unusually high level of coordination and communication, compared to similar past events.

The fluidity of the network dynamics is evident, as campaigns like #0409HACKEDOfBaltic focus on geopolitical targets, while others like #FuckGuardiaCivil target law enforcement efforts aimed at disrupting hacktivist activities. Campaigns not directly tied to the Ukraine conflict highlight the group's broader targeting strategy, showing that they don't only focus on states, but also on specific law enforcement and societal structures.





The Network

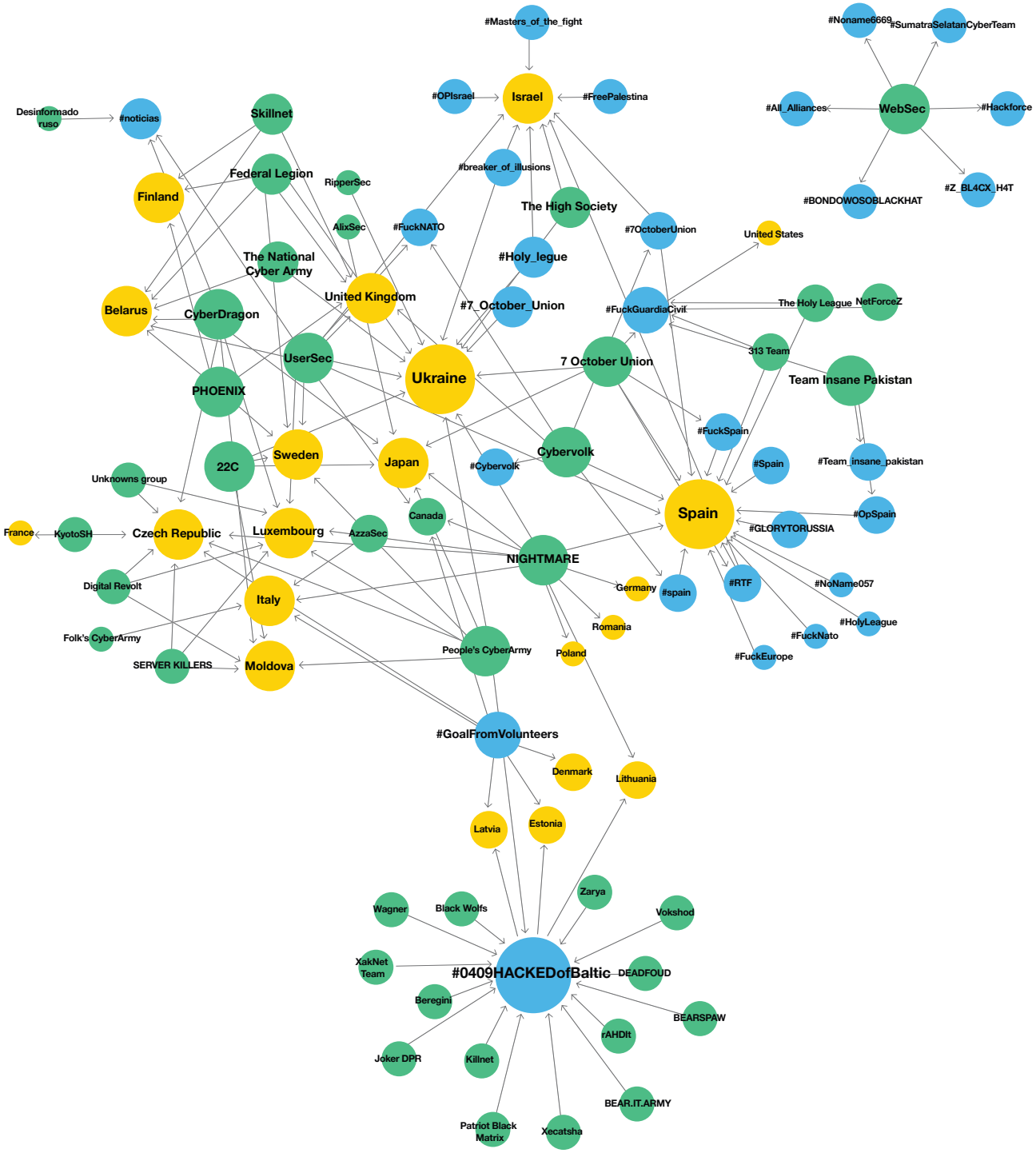
We found a total of 48 other groups that were mentioned by the hacktivist group in their messaging. The visualization below shows a broad network of connections, with a focus on various hacktivist groups that joined attack campaigns, hashtags used and countries mentioned.

The **yellow nodes** show the countries mentioned in messages, the **blue nodes** show hashtags included in the messages and **light green** are the partners mentioned. The size of a node gives an indication of how often it appears in messages, and the position of the node in the graph indicates how "central" it is amongst the messages published.

The graph shows connections when at least two of the nodes (country, partner hacktivist group or hashtag) "have coincided" in one message, resulting in a graph with over 3,000 messages.

The connections between groups suggest a well-coordinated collaborative network designed to enhance the impact of cyberattacks across multiple countries and sectors. Hashtags represent campaigns where various hacker groups, including our research subject, converged for coordinated actions.

Spain stands out as a major target, surrounded by key hashtags, including #FuckGuardiaCivil. The arrest of two individuals in Spain tied to cyber activities driving this specific focus. This hashtag is central, representing one of the group's most visible campaigns.



Summary

This report offers insights into a pro-Russian hacktivist group active for 2.5 years, which began operations following the war against Ukraine. Between August 2022 and August 2024, the group claimed over 6,600 attacks in more than 3,200 messages, with 96% of their victims in Europe, aligning with their anti-NATO and anti-Western stance. Surprisingly, despite frequent mentions, no attacks were observed on U.S. targets, possibly signaling an intentional avoidance. The group focuses on sectors providing essential services, such as financial, transportation, education, and government systems, with the aim of disrupting societal stability. Notably, voting systems in countries like France, the UK, Finland, Belgium and Austria were targeted during elections, threatening electoral integrity and sowing doubt about results. These attacks align closely with Russian state narratives, suggesting potential state influence.

Hacktivism has evolved from its early roots of ideological protest, with modern groups blurring the lines between hacktivism and state-sponsored cybercriminal activities. The pro-Russian group's actions are symbolically tied to their targets, amplifying political messages or undermining governance. Their campaigns often coincide with significant geopolitical events such as elections and summits. Like cyber extortion groups that threaten to leak sensitive data, hacktivists wield coercion to manipulate public perception, shaping political outcomes.

Indeed, several fundamental similarities between modern hacktivism and cyber extortion can be observed:

- Both invest heavily in building brand and community for credibility.
- Both operate publicly, offering real-time commentary on platforms like Telegram.
- Both are tolerated or even supported by nation-states when aligned with political objectives.
- Both procure advanced tools or services in the dark economy to boost capabilities.
- Both justify target selection retroactively, shaping narratives post-attack to maintain control over the story.
- Both use coercion, with hacktivism aiming to influence political outcomes and cyber extortion threatening reputational damage through document leaks.

Defending against these threats requires not only robust technical defenses but also strategic communication to counter disinformation and maintain public trust. The cognitive element of these attacks underscores the need for a holistic approach that includes safeguarding information integrity and strengthening public resilience.

Recommendations

From a technical standpoint:

- Implement standard security controls like DDoS protection, vulnerability mitigation, and attack surface management.
- Continuously monitor evolving threats and use the latest threat intelligence.
- Develop incident response and crisis management plans that cover both technical recovery and public communications.
- Engage in strategies to counter cognitive attacks that target public perception and trust:
- Monitor social and media channels for disinformation and respond quickly to debunk false claims.
- Communicate proactively with transparent updates to maintain stakeholder trust.
- Collaborate with public relations experts to craft consistent, credible messaging.
- Educate the public to recognize disinformation, fostering resilience against manipulation.

Given the escalation of hacktivism, particularly pro-Russian attacks targeting the West and NATO, organizations in these regions should prepare for ongoing efforts to disrupt and destabilize.

Human-Driven Threat Hunting

A Real-World Approach To Threat-Informed Defense

When discussing Threat-Informed Defense, the focus is on understanding the behaviour and technology of threat actors to gain a deep technical insight. This approach supports proactive threat hunting to prevent ransomware attacks, Advanced Persistent Threats (APTs) or criminal data exfiltration and can also be applied post-incident to guard against future intrusions.

Simone Kraus, Senior CSIRT Analyst, **Orange Cyberdefense**

Knowing What to Look For

This method relies on a combination of human-driven threat hunting and Threat-Informed Defense[176]. Skilled analysts actively search for real-world attacks, leveraging our own threat intelligence in combination with findings from our forensic investigations and reverse engineering of malware.

By systematically analyzing Tactics, Techniques, and Procedures (TTPs) and identifying Indicators of Compromise (IOCs) and behavioral patterns, we refine our threat hunts. Further investigation, including reverse engineering, often uncovers more IOCs, which we incorporate into our hunts. The goal is to detect anomalies, revealing suspicious activities related to specific incidents, and ensure the attacker no longer has access. Our structured threat hunting approach is grounded in MITRE's TTP-based method[177], which allows for a systematic search. This is enhanced by David Bianco's PEAK model[178] and MITRE Enginuity's "Summitting the Pyramid"[179], providing a clear, methodical approach to create robust detection. A deeper understanding of these methodologies strengthens our technical capabilities.

During engagements, we assist customers by blocking tools and IOCs, investigating suspicious activity, and offering next-step recommendations. Post-hunt, we deliver detailed documentation, including assessments and recommendations. In parallel, ongoing support is provided if needed.

Additionally, we improve the Endpoint Detection and Response (EDR) system by refining detections and blocking IOCs, ensuring that the response targets specific threat actor behaviours. This not only blocks individual threats but also prevents further ransomware encryption and broader attacks. Some EDR platforms also allow us to assess and prioritize potential vulnerabilities exploited by threat actors.

Preparation For Threat Hunting

To prepare for threat hunting, we leverage our analysis, create a TTP-based attack flow, and incorporate the latest Cyber Threat Intelligence (CTI) in collaboration with the wider cybersecurity community[180]. We also hunt for vulnerabilities and tools commonly exploited by ransomware groups. We prioritize vulnerabilities and tools based on their prevalence, focusing on those most relevant to the customer's specific sector or country.

Key Questions

during preparation for a post-incident hunt include:

- **What was the initial access point, and how can it be prevented in the future?**
- **Are there any exploits tied to that initial access?**
- **What are the CVE and EPSS scores of these vulnerabilities, and how many devices need patching?**
- **Are there any suspicious user accounts, GPO changes, C2 connections, unusual login behaviors, or suspicious devices?**
- **Are there other vulnerabilities commonly exploited by ransomware groups?**

Our approach starts with baseline threat hunting for suspicious behaviors, using a structured attack flow model. This threat hunting plan is sequential and systematic, incorporating tools and techniques known to be used by ransomware and other Cy-X groups or APTs. Our hunts span across various systems—ranging from specific EDR solutions to broader environments like network communication, logs, firewalls and SIEM systems.

Once we've identified the specific procedures and MITRE ATT&CK techniques in use, we convert them into YARA or Sigma rules. These rules can then be applied across a variety of systems, such as Cortex, Microsoft Defender, Splunk, GoogleSecOps, Sentinel One, CrowdStrike and Elastic. We either adapt existing queries from existing repositories or create our own Sigma rules using David Bianco's PEAK hypothesis-driven methodology. This enables us to rapidly deploy effective threat hunting across the environment and create detections if they are unique, invariant and robust.

Threat Hunting – Tracking and Communication

During our threat hunting process, we carefully track each hunt, documenting the execution time and any findings. If we find suspicious ports, processes, user behavior, or unwanted software, we promptly notify our customers to ensure rapid improvements in their environment. Every hunt is mapped to the MITRE ATT&CK framework and executed in a systematic, step-by-step manner, mimicking an actual attack.

We use baseline queries alongside newly created, specific threat hunts designed to detect the tools and commands used by the threat group. Additionally, we examine related procedures within the MITRE ATT&CK framework to identify similar behaviors from other ransomware groups. For example, we search for hacking tools or remote monitoring tools known to be used by other threat actors with high prevalence.

Best practices are applied to stop further lateral movement by blocking IOCs to detect and prevent suspicious behavior. We also recommend customers block any tools commonly used by ransomware affiliates if those tools are unnecessary within their environment.

Documentation and Further Steps

After completing the threat hunting, we document all key findings and provide a detailed record of each hunt we conducted. We offer customized recommendations tailored to the customer's environment. If we detect any potential security issues, we collaborate closely with the customer to figure out whether they are false positives or true positives. This approach not only helps prioritize next steps for strengthening security but also enhances the customer's understanding of their own infrastructure and tools.

We also recommend conducting an M3TID (Maturity Model for Threat-Informed Defense)[181] assessment after the hunt. This assessment evaluates the maturity level of threat-informed defense across people, processes, and technology. Based on the findings, we provide recommendations to improve the customer's infrastructure and security posture, helping to prioritize future security investments. Customers receive a separate briefing and documentation outlining their individual maturity scores and actionable recommendations.

Once threat hunts are created and executed, the queries can be saved in the EDR system, allowing customers to regularly monitor for suspicious behaviour. This proactive approach ensures continuous security checks and reduces the risk of re-victimization. We recommend performing these checks more often after an incident to prevent worst-case scenarios in case of being attacked again.

Key Takeaways

Threat hunting is an ongoing, iterative process that should be integrated into both the incident response plan and overall security strategy. Like the testing and evaluation of threat actors and their real-world behaviours, it requires continuous attention. Rather than treating it as a one-time compromise assessment following a forensic investigation, threat hunting should be a proactive method to prevent threat actors from exploiting vulnerabilities unnoticed.

This approach enables rapid improvement, helping to maximize, mature, and measure the success of security investments and overall security posture. A continuous threat hunting development plan can be as effective as continuous testing, and when combined, these efforts ensure a deeper understanding of your environment while naming defensive gaps. Knowing the adversary is one aspect, but truly countering and understanding their behaviour is essential for a resilient defense.

Emmanuelle Bernard
Mobile Network Security Expert
Orange

Stéphane Gorse
Senior Security Expert
Orange

Sébastien Roché
Corporate Internal Auditor / Senior Security Expert
Orange

Research: Mobile Security

Carriers, Networks and Security

Mobile phones are essential tools in modern society, thanks to fast, affordable data making internet access convenient beyond Wi-Fi. Mobile networks, a remarkable engineering feat, support reliable, simultaneous wireless communication for hundreds or thousands of devices, with interoperability allowing seamless network access when traveling abroad.

Behind this ease of use lies complex technology, and with complexity comes vulnerability. Intelligence agencies have long been aware of these weaknesses, and criminals are increasingly exploiting well-known flaws. We previously raised concerns about managing vulnerabilities in enterprise mobile phone estates, predicting that as mobile phones become central to enterprise security, criminals will adopt advanced hacking tactics to bypass controls like Multi-factor Authentication.

In previous Security Navigators, we predicted that mobile device attacks would increase as these devices become integral to personal, business, and cybersecurity infrastructures. While sophisticated, targeted attacks on high-profile individuals by private firms contracted to governments have intensified[182], we have not seen a significant rise in vulnerabilities or exploits affecting mainstream mobile platforms. However, there have been notable cases of mobile network infrastructure abuse—a topic we address for the first time in this report.

For example:

- In May 2024, UK police arrested two suspects for using a "homemade mobile antenna" to send phishing texts directly to mobile phones, bypassing network protections that typically block such messages[183].
- In early 2023, reports in Île-de-France described criminals driving with IMSI catchers to send fraudulent texts[184].
- In September 2023, a man was arrested and charged with espionage in Oslo for driving with an IMSI-catcher around the office of Norway's Prime Minister, the Defense Ministry and other government buildings[185].
- In January 2024, an attacker accessed Orange España's infrastructure by compromising an employee account lacking MFA, with credentials obtained through malware[186].
- In March 2024, SS7 and Diameter vulnerabilities were reportedly exploited to track individuals and intercept calls and texts, with potential abuse of the GSMA Global Title feature, previously linked to NSO Group and Intellexa[187][188].
- In August 2024, the UK National Crime Agency revealed that three men were sentenced for running an OTP-stealing service, "OTP Agency." This service phished One Time Pins (OTPs) by calling victims and warning of unauthorized account activity, prompting them to provide OTPs[189], which were then relayed to criminals.
- In September 2024, authorities arrested 17 suspects linked to an international network using the "iServer" phishing-as-a-service platform to unlock stolen or lost phones.
- In October 2024, reports emerged that "Salt Typhoon" breached several major US telecom providers, allegedly accessing systems related to lawful communication interception and other infrastructure areas[190][191].

In this chapter, we will pull the curtain back on the security risks associated with mobile phone networks. We'll discuss how mobile networks have evolved over the past two decades and how technology has adapted to address emerging threats. Note: we use a lot of acronyms in this chapter. You can find detailed explanations of these in the appendix on page 112.

The Mobile Telecommunication Ecosystem

Mobile networks like Orange are operated by telecommunications companies, but the underlying network functions are provided by network vendors like Ericsson, Nokia, and Huawei.

The secure deployment and operation of a mobile network depends largely on the operator's strategy, but is heavily influenced by each vendor's ability to meet these strategic requirements.

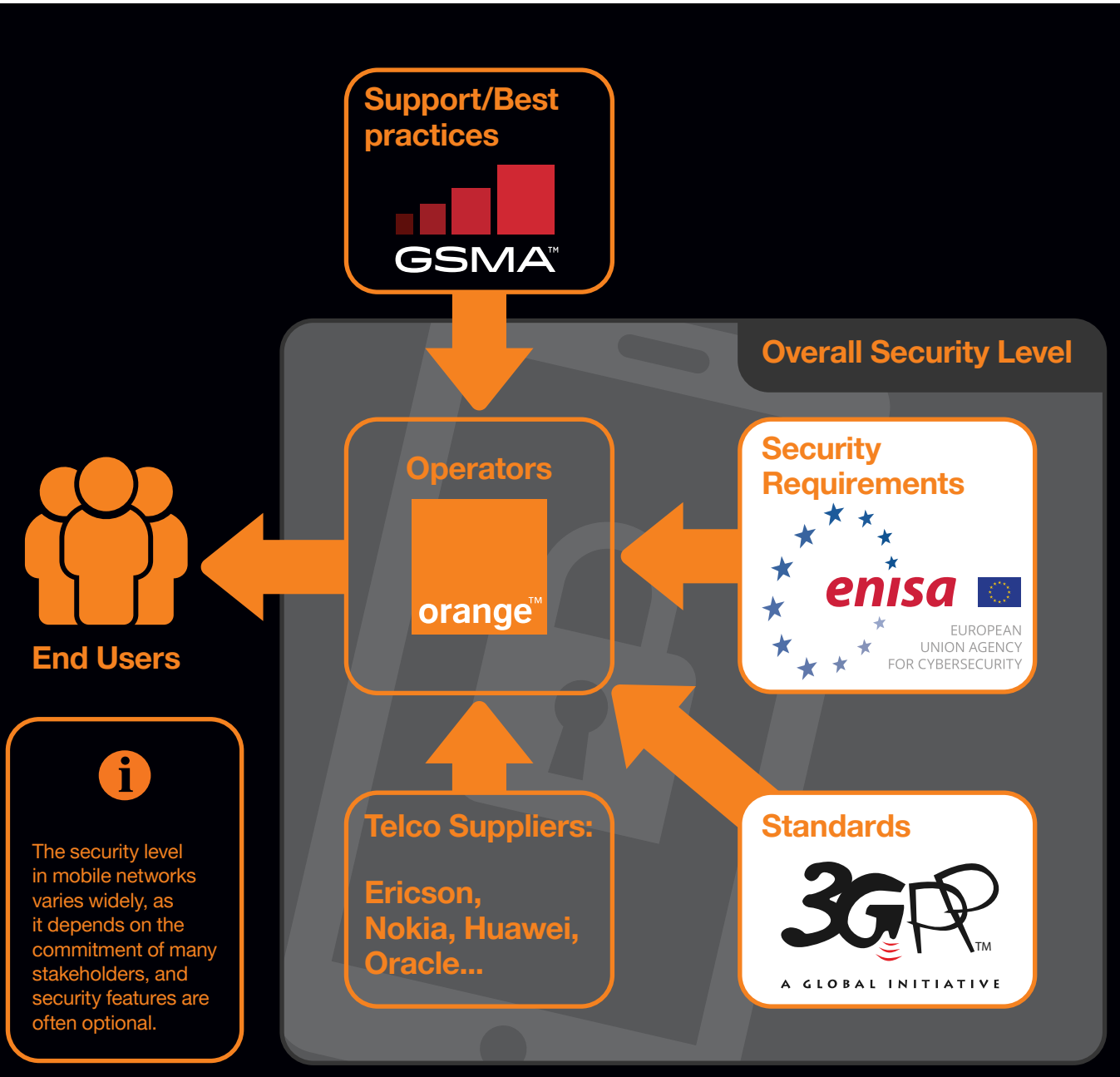
The **3rd Generation Partnership Project (3GPP)** is an organization that brings together several standards bodies to develop protocols for mobile telecommunications. 3GPP standards are designed to ensure interoperability between networks and network functions across different vendors. However, 3GPP does not specify all security mechanisms for a network; it only defines those required for interoperability, such as mobile authentication using SIM credentials. Security features available in network functions can vary significantly by vendor, which is a key differentiator in the market.

The **GSM Association (GSMA)** is a global organization representing the interests of mobile network operators and companies in the mobile ecosystem, including device manufacturers, software providers, equipment vendors, and internet companies.

Established to support the standardization and interoperability of mobile networks, GSMA develops industry guidelines, promotes collaboration, and advocates for policies that foster the growth and security of mobile communications.

It also develops key initiatives on security, IoT, 5G, and digital identity. The GSMA continually enhances the security support offered to the telco community as the threats targeting the mobile ecosystem evolve[192].

The **European Union Agency for Cybersecurity (ENISA)** is the EU's agency dedicated to improving cybersecurity across member states, including in mobile network security. ENISA provides strategic guidance, policy recommendations, and technical standards to enhance the resilience and security of critical infrastructure like mobile networks. Through collaboration with national cybersecurity authorities, mobile operators, and industry stakeholders, ENISA plays a pivotal role in strengthening defenses against threats within the mobile sector.



Mobile Telecommunications History

Launched in the 1990s, 2G or GSM (Global System for Mobile Communications), marked the transition from analog to digital telephony[193]. This technology introduced basic services such as voice calls and SMS. To support mobility of mobile users across networks and even international roaming, SS7[194][195] protocol called MAP was introduced. MAP operates within the SS7 framework, using SS7's signaling to enable mobile-specific functions across telecommunications networks.

3G – Universal Mobile Telecommunications System – was introduced in the early 2000s. It offered significantly higher data speeds and enabled mobile internet access[196]. SS7 was used again in 3G for core network signaling.

In 2010, 4G, or LTE (Long Term Evolution) was launched, revolutionizing mobile connectivity with significantly improved download and browsing speeds[197]. 4G introduced a new protocol called Diameter[198] for signaling exchange between core network functions.

Currently being deployed worldwide, 5G promises even faster speeds, reduced latency, and the ability to connect a much larger number of devices simultaneously. 5G uses advanced technologies like Massive MIMO (Multiple Input Multiple Output) and beamforming. In the core network, HTTP/2 replaces Diameter, and network functions now expose to other network functions via API – whether in the same network or in a partner network for roaming[199][200][201].

New Tech, New Threats

The mobile operator ecosystem has evolved significantly over the last 30 years – from 2G to 5G - and the attack surface has evolved with it. As new generations of mobile technology are deployed on top of older generations, not in place of them, the risk continues to accumulate.

In 2G, most reported attacks resulted from weak encryption algorithms (known as A5/1) on air interfaces, leading to possible "Attacker in The Middle" attacks. Tools like "IMSI Catchers" (or fake base stations) were used to mimic cell towers, allowing attackers to capture communications from unsuspecting users, or send them SMS.

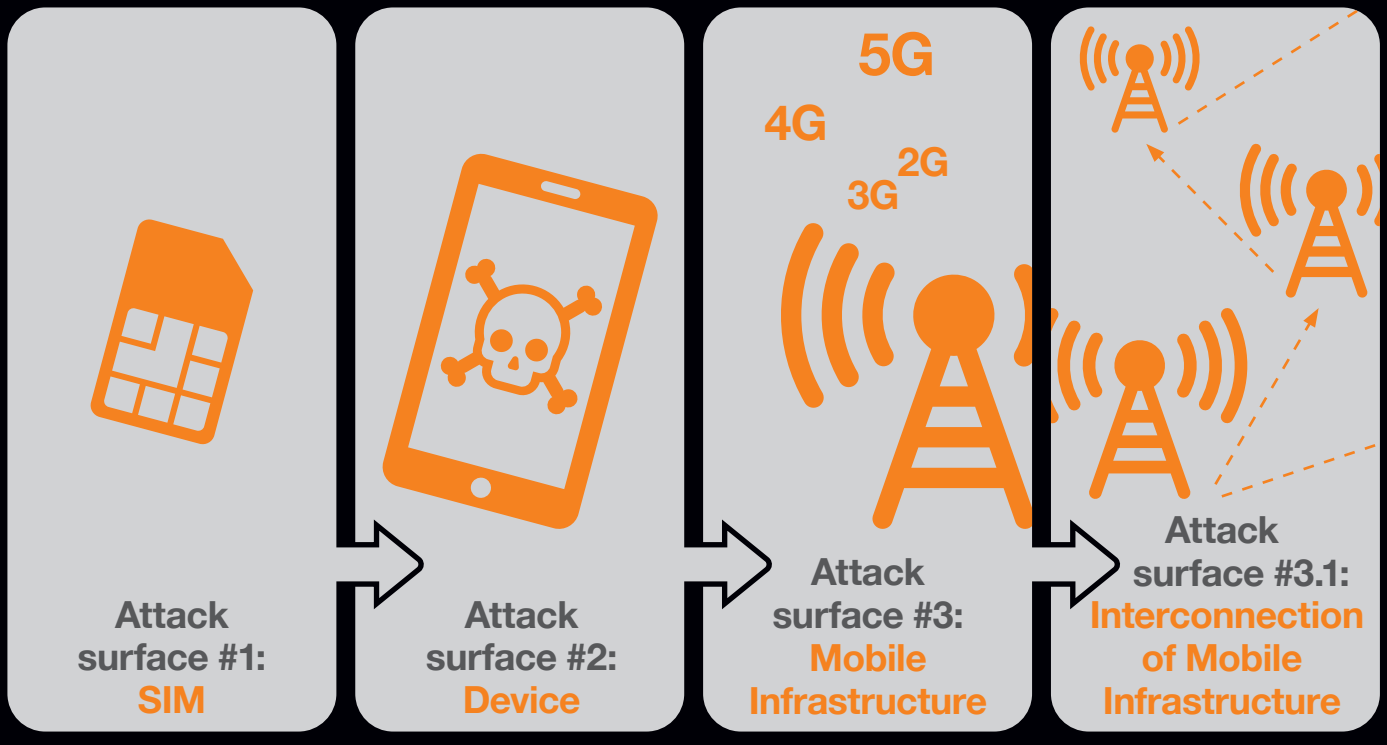
In 2G and 3G, SS7/MAP was unauthenticated and unencrypted on interfaces between operators, allowing for data theft and denial-of-service attacks. As roaming was initially designed within a "trust" relationship between operators, security was not considered in the SS7 protocol.

Later, as 4G networks began to roll out, vulnerabilities in the Diameter protocol were exploited[202]. Attackers could manipulate signaling messages to gain unauthorized access to user data or to disrupt services.

The 5G core network is virtualized and API-based, so the attack surface is also increasing and 5G networks still rely on 4G infra when 5GSA is not deployed. Threats such as software supply chain attacks (e.g. via 3rd party dependencies), attacks targeting critical infrastructure, and distributed denial of service attacks via IoT device vulnerabilities (like Mirai), all exacerbate this threat.

In one 2020 report, for example, researchers from Positive Technology cautioned that "Vulnerabilities in the GPRS Tunnelling Protocol (GTP) expose 4G and 5G cellular networks to a variety of attacks, including denial-of-service, user impersonation, and fraud"[203].

High Level View of the Attack Surface for Mobile Telecommunication



The Mobile Attack Surface

The mobile network attack surface emerges across 3 distinct domains:

1. Universal Integrated Circuit Card (UICC)/SIM
2. Device
3. Infrastructure

SIM

SIM cards are vulnerable to various threats. For example, a fraudster can take over a bank customer's telecom subscription by misappropriating their SIM card. By doing so, the fraudster gains control of the "possession" authentication factor, enabling access to the victim's accounts when combined with stolen personal data. This technique can be applied not only to banking applications but also to any other applications on the mobile phone, such as social media. Three primary methods are commonly used:

SIM Swap

A SIM swap occurs when a fraudster requests the operator to produce and activate a new SIM card. Once activated, the new SIM renders the original SIM inactive, causing the legitimate subscriber to lose access to the mobile network and their online services.

Portability

In this method, the fraudster uses the subscriber's Number Transfer PIN (NTP) to request outbound portability with a different operator. The new operator then issues a new SIM card for the transferred number.

Cloning

Cloning involves physically replicating a SIM card. Although technically complex and rarely used for fraud today, research has shown that it is possible to extract secret credentials from a SIM card via side-channel attacks, even with physical security modules in place[204][205].

Not so eSIMple

eSIM technology is also susceptible to fraud. Although the provisioning process is generally secure, the user controls activation, creating an opportunity for phishing or smishing attacks. Through these tactics, fraudsters can obtain credentials or one-time passwords (OTPs) used in the enrolment process.

Case Study

In April 2024, a significant eSIM fraud incident was detected by one of our European operators, involving multiple unauthorized eSIM swaps.

The fraud was initially flagged due to an unusual activity involving eSIM swaps. Specifically, multiple swaps were performed using the same device IMEI, which raised a red flag.

The fraudsters employed social engineering techniques to deceive victims. They contacted the victims, pretending to be representatives from the mobile service provider.

During the call, they generated an OTP (One-Time Password) for the provider's app and convinced the victims to share this code.

With the OTP, the fraudsters logged into the app and initiated SIM swaps on the victims' phone numbers. These swaps were primarily executed outside of working hours to avoid detection.

This specific incident affected at least 14 different phone numbers. Customers experienced unauthorized SIM swaps, leading to potential breaches of personal information and disruption of mobile services. Some customers even terminated their contracts out of fear of being hacked again.

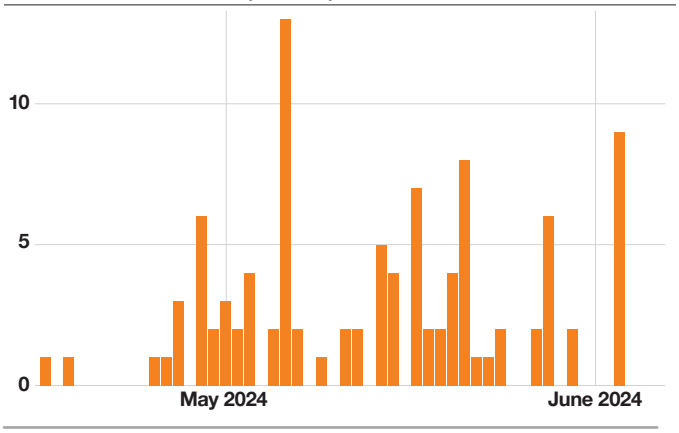
The situation prompted quick decisions and actions by the operator's security and fraud teams. Including:

- Informing the authorities.
- Blocking eSIM functionality via the provider's app.
- Improved Know Your Customer (KYC) measures to prevent further incidents.

Additionally, there were discussions about updating message templates to include warnings that the provider would never ask for the OTP code.

The incident outlined above is not an isolated case. As the chart below illustrates, over 30 days during May 2024, one European operator recorded 110 fraudulent eSIM swaps and 337,000 fraudulent SMS messages.

Fraudulent eSIM swaps
over time at one European operator



The shift to eSIMs introduces new fraud risks rather than eliminating them.

A recent University of Aalto study outlines over 12 root causes of eSIM fraud and hacking, highlighting common vulnerabilities like phishing, smishing, caller ID (CLI) spoofing, and brute-force attacks. Once hackers obtain a victim's identity data, they can exploit provisioning weaknesses to execute eSIM swaps. The study recommends implementing Know Your Customer (KYC) solutions or temporarily blocking features until new anti-fraud tools are available. Vendors like Thales, Nokia, and Ericsson are developing solutions, but current monitoring capabilities are limited, requiring further investment and regular updates.

For end-users, a compromised eSIM profile gives hackers control over the victim's phone number, facilitating financial fraud and bypassing two-factor authentication (2FA) for banking and other sensitive accounts. Hackers can also duplicate eSIM profiles for identity theft. For operators, eSIM fraud threatens revenue and reputation, increasing the need for audits of eSIM and IoT B2B services and heightening GDPR compliance risks. Operators face recurring costs to enhance fraud protection as technology evolves.

The Device Itself

Modern mobile phones operate like powerful computers, running operating systems and applications while connecting via mobile networks, Wi-Fi, Bluetooth, NFC, and even satellite networks. As we reported in 2021, 547 vulnerabilities were identified in Android and 357 in iOS, with 18 Android vulnerabilities rated critical, compared to 45 for iOS. This suggests Android has more vulnerabilities but fewer severe ones, while iOS is harder to exploit yet offers greater reward. Android exploits are widely used across devices, whereas iOS exploits are often associated with sophisticated mobile surveillance actors.

Apple's consistent ecosystem means iPhone users are more vulnerable when flaws are disclosed, though updates are quicker, with 70% upgrading within 51 days. Android's fragmented system leaves older devices exposed to older vulnerabilities while somewhat protected from newer exploits. However, malware remains the most pressing threat for everyday users.

Both Apple and Android use dedicated marketplaces—the App Store and Google Play—with security measures like app reviews and sandboxing to limit exposure to malicious apps. In 2022, Google Play had 781% more malicious apps than the App Store, likely due to higher malicious submissions, Android's low-complexity vulnerabilities, and ready-made exploits.

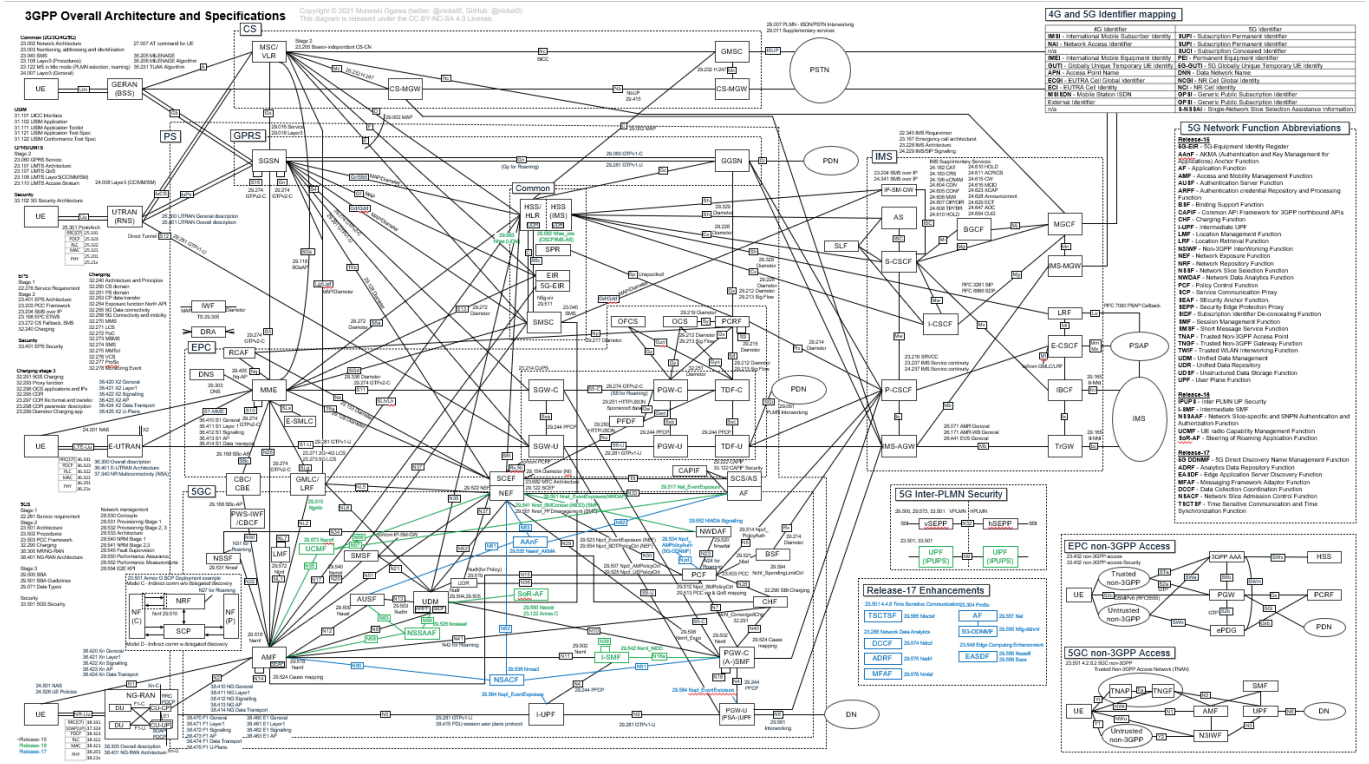
Google's review processes may also be less strict than Apple's, and unofficial Android app stores adding further risks. Android users can also sideload apps. This feature, often exploited by trojans, poses a major risk, particularly as alternative app stores usually lack robust security.

While currently Android-only, iOS is expected to allow sideloading in the EU by 2024 (starting with iOS 17) to meet EU regulations, potentially introducing new security challenges for Apple users.

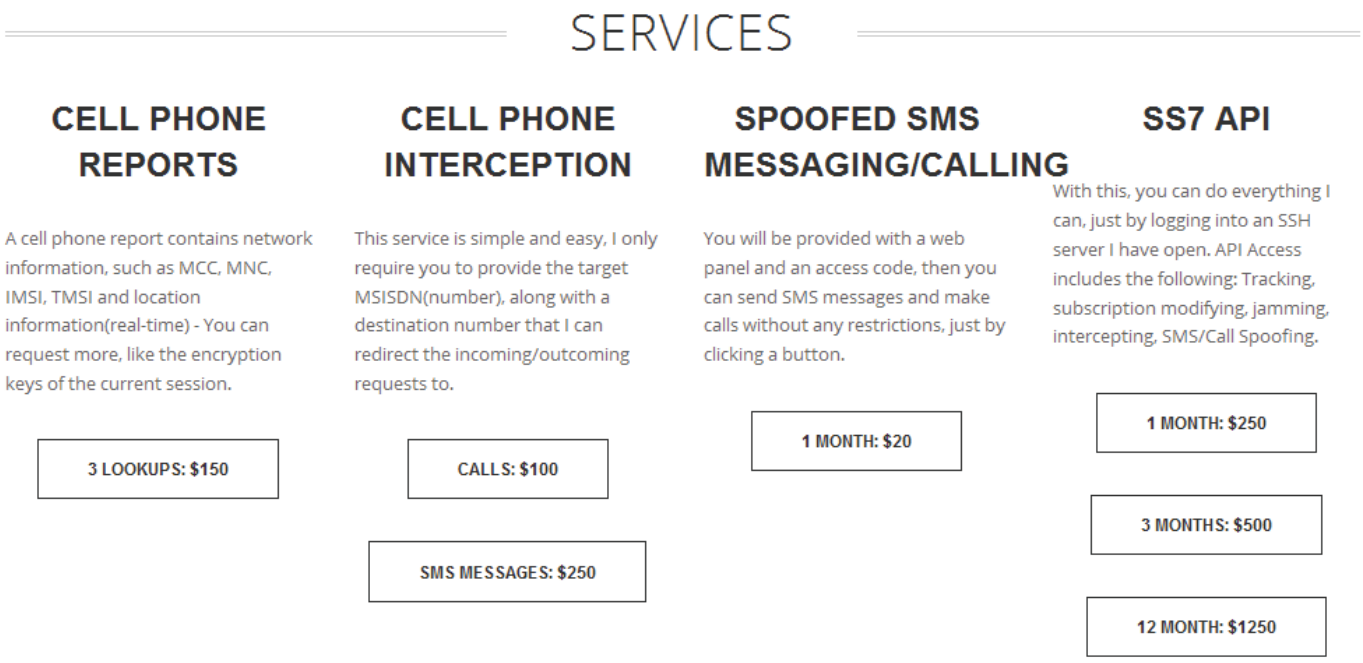
Infrastructure

The attack surface in mobile infrastructure has expanded significantly with the advancement of mobile technology. A quick overview of this complexity is provided below[206]:

The GSMA's "Security Landscape 2024" report[207] highlights several critical areas of concern for the mobile telecommunications industry. Key points include the increasing frequency and sophistication of attacks on virtualized infrastructure, such as virtual machines and container solutions. The report also emphasizes the vulnerabilities within supply chains, and the growing issue of spyware.



Detailed view of the mobile network attack surface at infrastructure level



- **Services identified on Dark Net related to signalling aspect**

Pushing MFA

Given the inherent weaknesses in mobile network technologies, single-use passwords (OTP) sent via SMS, were deemed insecure by NIST as early as 2016. Fraudsters have adapted to bypass SMS OTP by using techniques like caller ID spoofing to impersonate banks, tricking customers into authorizing fraudulent transactions.

Effective Multi-factor authentication (MFA) today therefore largely leverages two main methods for a second authentication factor beyond device possession:

Third-party Authentication:

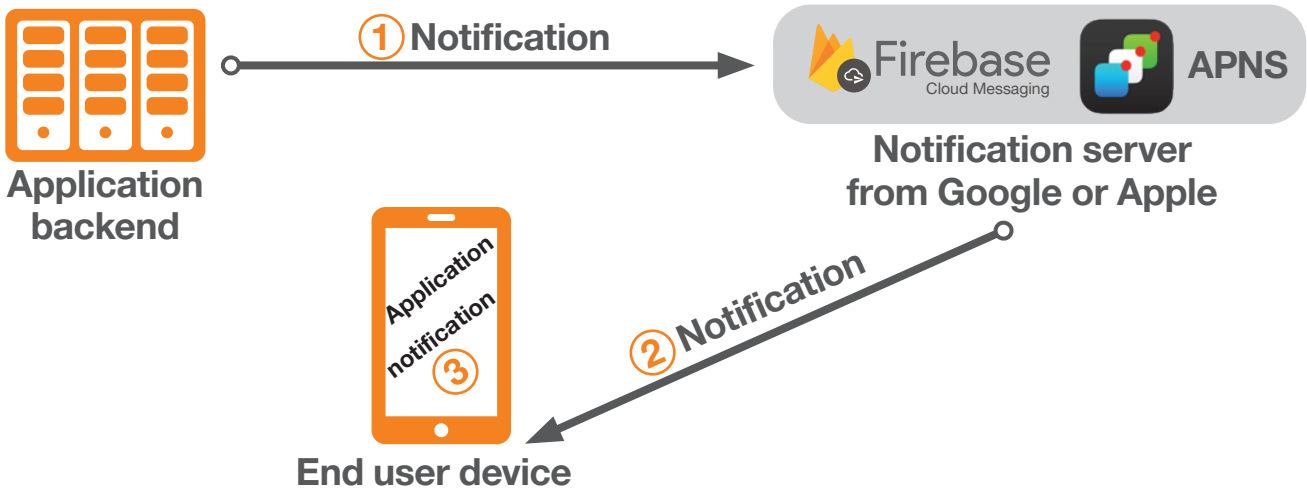
High value accounts such as banking often use their mobile banking apps to provide a second factor via a PIN code (knowledge) or smartphone biometrics (inherence).

Operator-Based Authentication:

An alternative method, standardized by GSMA as "Mobile Connect", uses SIM-based authentication and requires a PIN code (knowledge). Sending OTPs via SMS is discouraged due to vulnerability to SS7 rerouting attacks (as noted in NIST-800-63B)[208].

Also increasingly popular, a dedicated MFA mobile app uses push notifications provided by Google or Apple, a system natively supported by Android and iOS [209][210][211]. In this model, telecom operators are excluded from the process, which poses potential privacy risks for users, as Google and Apple retain the ability to collect usage data from backend interactions, even if exchanges are encrypted.

More generally, third-party instant messaging and VOIP applications use the application layer to manage traffic according to their own standards, with security and data protection measures thus depending on the efforts and success of the software vendor.



- **Push notification channel**

A Brief History Of Mobile Network Hacking

Intelligence Agency Exploits

Mobile infrastructure has been a target for intelligence agencies, with several incidents highlighting this fact since the 2000s. For example:

Mark Klein, a former AT&T technician, revealed his role in exposing the U.S. National Security Agency's (NSA) use of AT&T's infrastructure for mass surveillance. Klein revealed that the NSA had installed splitters to divert Internet traffic, allowing them to monitor communications without warrants[212].

The Greek wiretapping case of 2004–05 involved the illegal surveillance of over 100 mobile phones belonging to high-ranking Greek officials, including the Prime Minister. The surveillance was conducted through the exploitation of vulnerabilities in Vodafone Greece's mobile network infrastructure. The attackers, suspected to be state-supported cyber threat actors, installed rogue software that intercepted calls and messages. This software exploited lawful interception capabilities meant for legal wiretaps, redirecting the data to unknown recipients[213].

The Thales Group's investigation[214] into the alleged hacking of Gemalto's SIM card encryption keys revealed significant vulnerabilities in mobile network. The breach, reportedly conducted by the NSA and GCHQ, involved the theft of encryption keys, which allowed these agencies to intercept and decrypt mobile communications without the need for cooperation from telecom companies or legal warrants. Exploitation of these encryption keys enabled the attackers to bypass traditional security measures, gaining unauthorized access to voice and data communications on a global scale.

Vulnerabilities Exposed and abused

In 2016, researcher Karsten Nohl demonstrated[215] how to intercept a voice call from a U.S. senator, following his 2014 presentation at the Chaos Computer Club conference with researcher Tobias Engel, where they exposed vulnerabilities in the SS7 protocol. Then in 2017, operator O2 confirmed that hackers targeted its network by exploiting SS7/SMS protocol weaknesses used in two-factor authentication. Combined with phishing attacks, attackers managed to trigger money transfers and redirect two-factor verification codes via SMS, resulting in customer losses totaling approximately €200,000.

Mitigations

Defending the SIM

Mitigating SIM card vulnerabilities requires multiple strategies. Operators should deploy GSMA-certified SIM cards with a generic protection profile, and embedding a firewalling Java applet within the SIM can block unexpected external interactions.

For SIM swaps, telecom operators like Orange have updated customer processes with stricter controls. But SIM swap attacks often rely on social engineering, making customer awareness essential. Operators have also introduced APIs allowing service providers to check if a SIM card was recently renewed.

Device manufacturers are also strengthening mobile security, implementing stricter controls in app stores and limiting API access for application developers to improve security.

Solutions providing dynamic application analysis to detect threats are now common, and mobile device management (MDM) systems are highly recommended for organizations to address major security risks.

Defending the Infrastructure

Since security is not built into SS7/MAP and Diameter protocols, operators like Orange have implemented specialized protection solutions known as Signaling Firewalls. These solutions provide key functions like Traffic Filtering, Anomaly Detection, Protocol Validation, Access Control, Logging and Reporting.

One valuable feature for network security is "velocity checks," which prevent attacks by verifying that user mobility aligns with realistic speeds (e.g., not exceeding airplane travel). This rule helps detect and block attempts to impersonate a visited network identity.

Defending the Device

Securing mobile devices against threats is challenging, as these devices are high-performance computers running complex operating systems. Like any computer, they require monitoring for malicious activity and malware.

For individual users, solutions like antivirus software with added services (e.g., personal fraud investigation) are available. In the business sector, Mobile Device Management (MDM) systems like Checkpoint and Pradeo Mobile Threat Defense help protect entire device fleets by collecting device data and enabling rapid mitigation.

Attacks exploiting radio channels are harder to counter, as they require access to the modem baseband, which is not available in standard consumer devices, necessitating specialized, hardened devices.

A good start for businesses may be to standardize on a mobile device platform that can be trusted to be up to date and monitored using a reliable MDM system.

Defending MFA

In Europe, the Payment Services Directive 2 (PSD2), enacted in 2018, mandates strong customer authentication (SCA) for digital transactions by financial institutions, particularly banks, to enhance security.

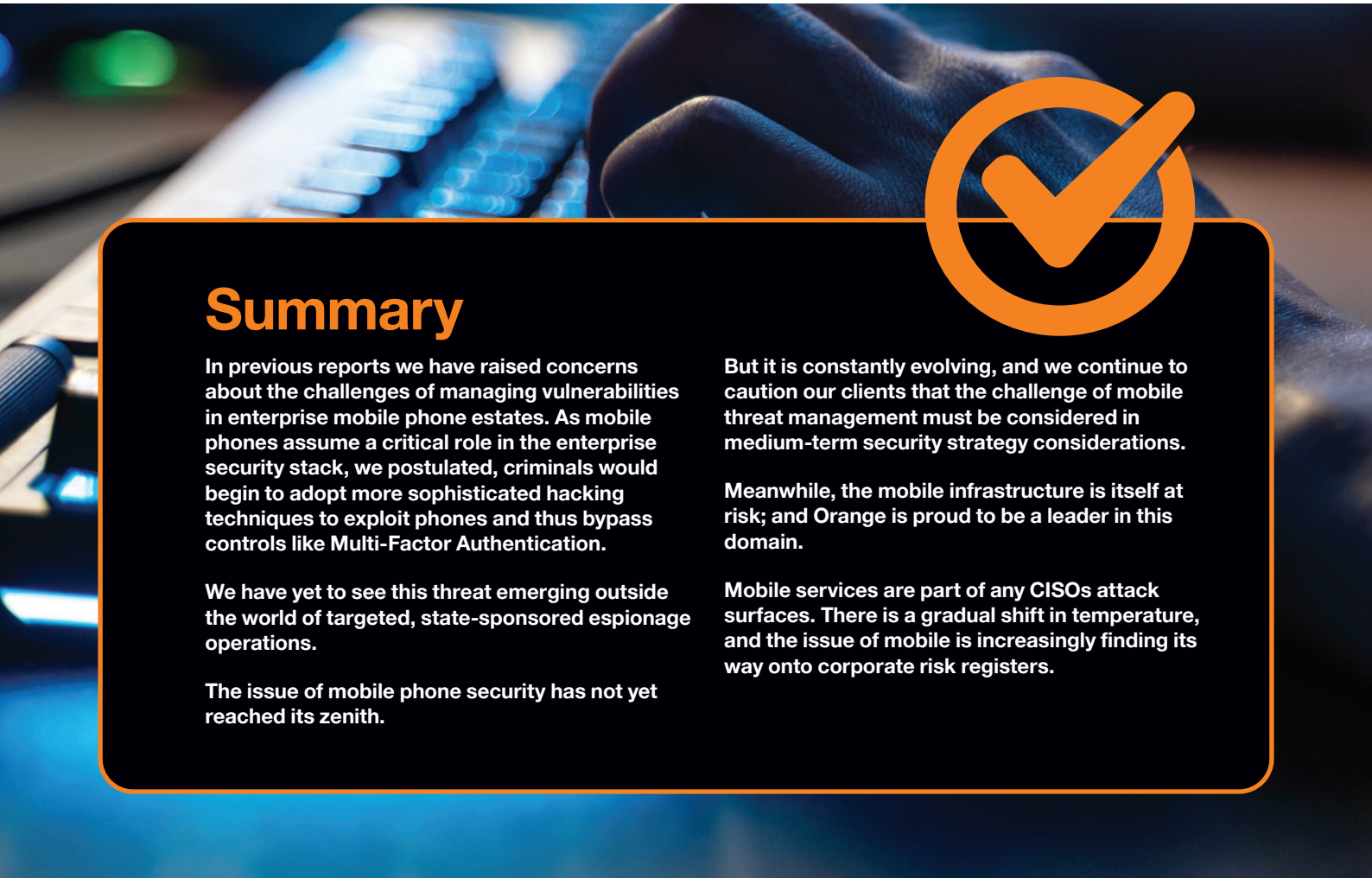
By implementing proprietary applications, banks comply with PSD2 and can legally reject customer claims in commercial disputes, excluding fraud cases. The directive's revision presents an opportunity for the European Commission to reinforce banks' financial accountability in fraud cases, even when strong authentication has been applied.

The revised directive also introduces implicit responsibilities for telecom operators if a spoofed call is involved in fraud, including caller ID spoofing (fake calls), sender ID spoofing (fake SMS), or SIM-based actions (SIM swap, number portability, or cloning).

Passkeys are a replacement for passwords, always strong and phishing resistant[216][217][218]. The Fast Identity Online (FIDO) alliance has published a specification that is based on public-key cryptography where each passkey contains a unique public/private key-pair. The passkey can be stored on a dedicated hardware token or be integrated into a device that supports the specification. Mobile devices such as Apple's iPhone and Google's Pixel mobile phones are examples. Passkeys use the trusted relationship of the hardware and the tightly bound identity of the user to facilitate authentication. The user uses the device to relay a cryptographically verifiable value that cannot be faked.

FiGHT or Flight

The MITRE FiGHT (5G Hierarchy of Threats) project is designed to identify and categorize potential security threats specific to 5G networks and related technologies. FiGHT provides a structured framework for understanding the unique risks within 5G environments by mapping out threat scenarios across various layers of the 5G infrastructure[219].



Summary

In previous reports we have raised concerns about the challenges of managing vulnerabilities in enterprise mobile phone estates. As mobile phones assume a critical role in the enterprise security stack, we postulated, criminals would begin to adopt more sophisticated hacking techniques to exploit phones and thus bypass controls like Multi-Factor Authentication.

We have yet to see this threat emerging outside the world of targeted, state-sponsored espionage operations.

The issue of mobile phone security has not yet reached its zenith.

But it is constantly evolving, and we continue to caution our clients that the challenge of mobile threat management must be considered in medium-term security strategy considerations.

Meanwhile, the mobile infrastructure is itself at risk; and Orange is proud to be a leader in this domain.

Mobile services are part of any CISOs attack surfaces. There is a gradual shift in temperature, and the issue of mobile is increasingly finding its way onto corporate risk registers.

A Hierarchy of Needs

Incident response readiness: Where to begin

It is quite the experience to sit across from a person – usually a CISO or an IT Security Manager – that has not slept in days, and to be asked: "how can we stop this from ever happening again?". Perhaps we had spent the last two days containing a threat actor that had exploited an unpatched VPN appliance and penetrated deep into the infrastructure. Or maybe it was the worst-case scenario: this person's entire infrastructure had been ransomwared, and there were no backups to rely upon. In all these cases, my answer would be: "We can't – but here's how we can react better next time".

The reality is that a cybersecurity incident is a matter of when, and not if. Organizations unwilling to face this reality will be caught unprepared again. Incident response readiness is a complex beast, with various areas of concern demanding attention. So where do we start?

Saskia Kuschke, Senior CSIRT Analyst, **Orange Cyberdefense**

Your Hierarchy of Needs

A simple starting point may be the "Incident Response Hierarchy of Needs" model, from the mind of Matt Swann[220].

Similarly to Maslow's hierarchy, the model depicts several needs in the original diagram – inventory, telemetry, detection, triage, threats, behaviours, hunt, track, act, all coming to a point in collaboration. Each tier depicts a deceptively straightforward question that - depending on the organization's policy, budget, risk appetite and culture - likely has a complicated answer. However, this way of ranking "needs" may present a simple and practical way to prioritize your efforts. Each tier builds upon the previous: for example, with a better view on your inventory position, you gain a better understanding of your coverage needs in terms of telemetry and visibility – and better telemetry leads to increased detection opportunities (and so on). One of the criticisms of this model is that one can still perform incident response even if all the tiers do not have adequate controls – as such, it useful to recognize that while you do not need to finish off the entire tier before moving to the next, the activities described higher up in the pyramid become significantly smoother if you have invested in a solid prior foundation.

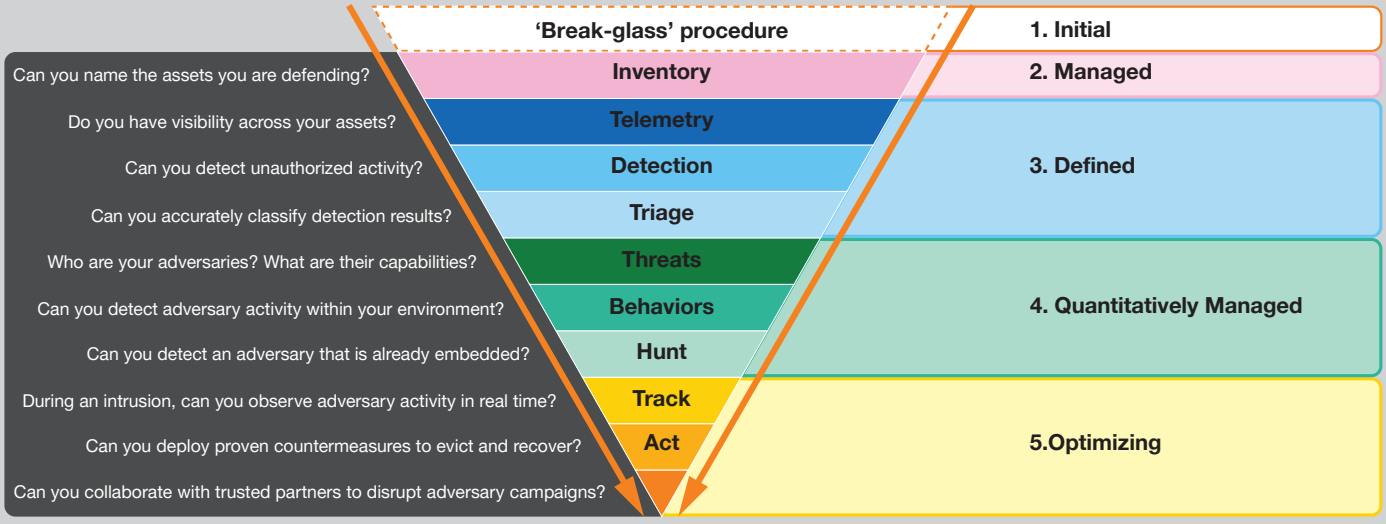
As far as models go, this simple breakdown of identifying "needs" can be an effective starting point in the journey towards being incident response-ready in time for the next attack. But if it is this simple, then why does our CSIRT still encounter multiple organizations that seem to struggle with even the foundational tiers?

A Roadmap for Response Readiness

Part of the difficulty of organizing effective incident response is because the building blocks are a mix of organizational, people, process and technological considerations. To make the original Hierarchy model more concrete, and taking these building blocks into account, our CSIRT has developed a "roadmap" of where to start on the journey to incident response readiness. For illustrative purposes, a simplified representation of this roadmap is in the diagram below.

The reasoning behind the roadmap is simple: be practical. Complexity is the enemy during an incident, and many of the preparatory activities on the road to IR readiness involves reducing ambiguity as far as possible in the decision-making process. To measure the distance travelled in the IR readiness journey, we make use of the capability maturity model integration (CMMI) model as a guideline[221]. At each phase, one must consider the people, process and technology needed to achieve the goal.

This is a good time to quote famous statistician George E.P Box: "All models are wrong, but some are useful". In the spirit of this sentiment, take this diagram as a suggestion on how to structure and measure the journey of maturity, rather than a concrete mapping of phases and absolute truths. In practice, you will likely find yourself moving back and forth between the various tiers and associated activities, rather than having the luxury to complete all in sequential order.



1. Ready Your Fire Extinguisher

In the Initial phase, it is about making sure you have a fire extinguisher that you know how to operate – in other words, addressing the very basics of incident response. Do you know who is responsible for what during an incident, and who needs to be called and informed? Further still, do your operational teams understand how to collect data from endpoints, how to perform emergency firewall rule changes? And can all of this be recalled and performed under pressure? Examples of elements to have in place are:

- **Process, People:** An IR plan clearly listing roles & responsibilities assigned to specific individuals.
- **Process:** Communication plan during an incident.
- **Process:** Playbooks for containment and data collection (e.g. emergency firewall changes, endpoint isolation, running forensic collector software on affected systems).

2. Map Your Environment

With the essentials in place, you can now tackle the challenge of mapping your environment to progress towards the managed level, which aligns with the Inventory tier: where are your assets, and what are your critical systems and data? Which systems are vital for your business, and how are they configured? Are any internet-facing? Many clients struggle with mapping and maintaining infrastructure knowledge as environments grow in complexity. However, a thorough understanding of your setup is crucial for a stronger incident response. Consider:

- **Process, Technology:** Creating & maintaining asset lists (automated, where possible).
- **Process, Technology:** Creating & maintaining IT architecture documentation (e.g. network diagrams, cloud architecture diagrams, Active Directory topology).
- **Process, People:** Documenting system owners and how to contact them (especially out of hours).
- **Technology:** Understanding and mapping software and configuration vulnerabilities.

3. Tune Your Smoke Detectors

Once you understand your key systems and pressure points, ensure you have the telemetry, detection, and triage capabilities to assess activity on them. This defines the "defined" level: knowing the completeness, accessibility, accuracy, and retention of your data. First responders, analysts, and decision-makers need information to identify threats, enact containment, or even shut down the network if necessary. Logs, SIEM, and EDR/XDR data are vital here, and knowing what data is retrievable (even under pressure) is crucial for mastering incident response. Considerations may include:

- **Technology:** available log sources & forwarding to centralized repository (e.g. to a SIEM).
- **Technology:** EDR/XDR coverage and capability.
- **Technology:** detection engineering and monitoring use cases configured for your available telemetry.
- **Process:** tuning your event & incident classification frameworks to better suit your organization.
- **People:** personnel trained to monitor, triage and analyze data, events and alerts using security tooling.
- **Technology:** data quality in terms of accuracy, completeness, coverage, accessibility & retention timeframes.

4. Conduct Your Fire Drills

Reaching the quantitatively managed level means that you have a good grasp of your break-glass processes, environment and information position available to you during an incident. This is an ideal point to focus efforts into conducting “fire drills” and measuring the efficacy of your IR capability in the vein of tabletops and assessments. While continuous testing can (and should) be used to measure your response throughout the IR readiness journey, these activities will likely start revealing less “obvious” improvements to be made at this point. In this maturity phase, your capability should also be controlled enough to dive into the more proactive parts of your “needs” – such as the incorporation of strategic and operational cyber threat intelligence (CTI) and proactive threat hunting to identify threats and malicious behaviors directly relevant to your organization. Consider:

- **People, Process:** tabletop exercises to test specific elements of the IR process.
- **Technology:** assessments of configurations of security tooling and related systems.
- **People, Process, Technology:** proactive, continuous CTI-driven threat hunting.
- **People:** additional training for personnel where gaps are identified.

5. Iteration And Continuous Improvement

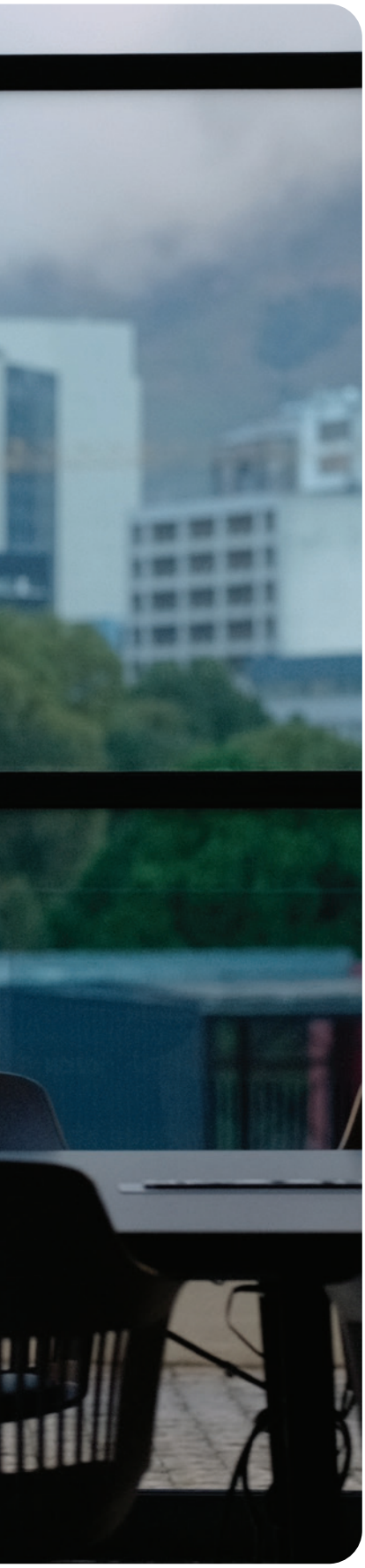
And finally, the coveted Optimizing phase, where your processes, people and technology are well-oiled enough so that any improvements are essentially incremental instead of instrumental. Tuning policies, designs and tooling to ensure that tracking, acting and collaborating during an incident can run as unhindered as possible by preventable issues and poor starts. Here your focus may include:

- **Process:** Maintaining a robust lessons-learned cycle.
- **Process:** Tuning and creating IT security policies.
- **Process, Technology:** Improving the design of your IT infrastructure.

Key Takeaways

In our CSIRT’s experience, incident response is something that improves via iteration – every incident you survive makes you better equipped for the next one, provided you put in the effort to learn from the encounter. Preparing as best you can prior to an incident puts you in an optimal position to fully leverage this experience to identify what your organization needs the most, at whatever maturity level you find yourself occupying.

Break your problem areas down into people, processes and technology, and prioritize your solutions and mitigations in a way that supports the work to come. Above all: be practical and be prepared, so that when we reach the end of an incident, you are the one telling me what your organization will do better next time.



Tatiana Chamis-Brown
SVP Global Strategic Marketing
Orange Cyberdefense

Vivien Mura
Global CTO
Orange Cyberdefense

Security predictions

A story of Convergence, Intelligence And Resilience

Join us once more as we take a step back and try to predict how the big picture presents itself and where the trends are going.

What will shape the digital world in the year to come? Which threats should we prepare to face and how should we go about it? What will be the major trends and tendencies our industry and others?

This year we will focus on five key trends we believe are going to be relevant in the field of cybersecurity and associated risks.

Apts Will Not Leave Room for Ransomware

The landscape of cyber threats is becoming increasingly complex, with a notable rise in extortion victims, often compromised and subsequently threatened with a data leak multiple times, often with the same set of stolen data. This escalation is not merely a trend; it reflects a broader shift in the tactics employed by cybercriminals, who exploit sophisticated methods to achieve their goals, increase their resilience, and impose fewer moral or geographical limits on themselves. Disinformation on the web is integrated into destabilization methods to amplify pressure on victims, and drastically improved impersonation capabilities through generative AI allow for deception of even the most discerning individuals.

In this already concerning context, the conduct of more discreet attacks, involving the infiltration of information systems for espionage or to prepare for future aggressions, must remain on the defender's radar. In 2024, the accidental discovery of a backdoor methodically introduced over several years into a component of Linux systems (XZ utils, openssh) highlights the determination of major powers to occupy strategic positions in cyberspace without being detected.

Critical vulnerabilities discovered in security equipment are indeed exploited for this purpose. Advances in quantum computing pose an additional risk to data encrypted with current algorithms. The migration to quantum-resistant cryptographic systems will take time and must begin as soon as possible to account for the retroactive effect of a future quantum threat on today's encrypted communications.

Furthermore, global outages triggered in 2024 by a faulty update of Crowdstrike's Falcon solution remind us of the fragility of the digital space in the face of systemic crisis risks, which could be caused by attacks on software maintenance chains. This type of attack is not new; numerous cases have been reported in the press (NotPetya in 2017, SolarWinds in 2020, Kaseya in 2021), and the hyperconnectivity of physical assets (OT, IoT) only increases the attack surface.

Generative AI Boosts Automation

A Matter of Time

The distance between the attacker and the defender is often temporal: the attacker has the advantage of surprise, forcing the defender to equip themselves and prepare to react as quickly as possible whenever a vulnerability appears, or a security event arises. In these circumstances, automating detection, alerting, and response mechanisms (CyberSOC, SOC, CERT, and VOC) allows for time savings that can make a difference in remediating critical vulnerabilities and resolving incidents. This is why the significant advancements in artificial intelligence algorithms present an opportunity to support the automation of services, thereby increasing the speed and quality of our cyber defense.

Maintaining Control of Security

The widespread use of generative AI solutions to assist humans in handling increasingly complex tasks also expands the attack surface across a new value chain: training databases, consultation data, prompts and responses, LLM hosting infrastructures, RAG systems, generative AI models, etc.

In the future, we can expect generative AI systems to become more interconnected with the rest of the digital landscape, with increasingly elevated action privileges (bank transactions, control of physical systems, etc.). Securing this chain often involves implementing traditional and proven security measures and solutions. However, certain characteristics unique to AI systems require adaptations of existing security products and specifically trained expertise. Nonetheless, the conveniences offered by new AI technologies should not lead us to neglect data protection aspects. Typically, no software code generated by a virtual assistant should escape secure development practices, and no ChatBot solution should be deployed without risk analysis and security measures.

Finally, social engineering techniques are greatly facilitated by generative AI, allowing criminals of all levels to perfectly imitate a person's style, voice, or appearance. Therefore, we can expect a surge in fraud and scams in the coming months and years, which will require an adaptation of digital offerings to better protect society.

Regulations: The More Compliant, the Better Prepared

The excellent preparation of the stakeholders involved in the 2024 Olympic Games has paid off: despite numerous security events, the overall increase in security levels and operational rigor have helped avoid a crisis. This outcome proves that security can be successful and that sufficient investment can protect against the worst. This is why regulations regarding the protection of digital assets are strengthening.

From Theory to Practice

The year 2024 is pivotal in the European regulatory landscape. First, the implementation of the NIS 2 directive in member states expands the regulated scope to many entities, categorized by their criticality into important and essential entities. The directive aims to better protect small and medium-sized enterprises, which are particularly affected by cybercrime (as evidenced by the Security Navigator figures). In effect since 2023, the DORA directive complements NIS2 by specifically targeting the financial sector to enhance the resilience of operators against threats.

Finally, the recently adopted Cyber Resilience Act by the Council of the European Union aims to raise the security level of many digital products marketed in the European market, based on their criticality.

Indeed, products with digital components can introduce vulnerabilities into uses or information systems that pose cyber risks with economic and societal impacts.

For example, these vulnerabilities can be exploited to orchestrate massive denial-of-service attacks or to steal valuable data, whether personal, strategic, or characteristic of intellectual property.

Lernaean Hydra

Additionally, there has been an increase in arrests and dismantling of cybercriminal members and networks, thanks to effective international collaboration, as recently seen with the LockBit group. While law enforcement interventions are commendable as they hinder the activities of mafia groups and sometimes recover seized data, the organizational model of cybercrime makes it particularly resilient, and we should expect it to continue growing.

Resilience 2.0

With only a few days to the opening of the Paris 2024 Olympics, it was not a cyber attack that caused significant disruption across the world. The CrowdStrike update of Friday, 19. July highlighted the perils of concentration and supply chain risks, and the importance of a robust back-up and recovery plan.

In parallel, the Security Navigator 2025 report highlights an increase of 15.29% in cyber extortion victims, notably SMB victims increased by 62%. This is especially concerning as many larger organizations depend on SMBs in their supply chain. These smaller organizations often lack advanced cybersecurity practices, and due diligence of third party risks is not infallible.

Resilience for larger organizations requires third-party contingency and incident response plans. Moreover, larger organizations can increase their own resilience by sharing best practices across their supply chain to lift their capabilities, especially to SMBs.

Effective risk management has for some time involved more than investment in prevention and protection – it also needs deliberate investment on back-up, response and recovery for resilience.

We see this shift accelerating in the year ahead given events of 2024, with increased investment on crisis management training and drills, recovery strategies and solutions, third-party risk management and best practice sharing.

And though automation boosted by AI is here to stay, IT systems are not fully autonomous. The capability to confirm an anomaly, declare a crisis, implement an incident response plan and manage impacts across the direct scope of the organization are all powered by people. The human element remains a central element of the resilience equation.

Security ROI in Focus

Many organizations suffer from so-called technology bloat. The problem stems not only from the number of cybersecurity solutions adopted, but that these do not always streamline. Consequently, in-house security teams are stretched, spending significant time managing disparate tools that are not integrated, instead of deriving value from this investment. This is aggravated by the fact that the cybersecurity vendor ecosystem is characterized by a plethora of tools and technologies and a scarcity of skilled personnel to manage them effectively, according to Forrester[222].

As security architecture matures, security leaders are increasingly undertaking a critical review of existing solutions, identifying redundancies, gaps and under-utilisation and pruning solutions that are not yielding value. In fact, Gartner[223] estimates that 70% of organizations use 20% of the functionality of security products. Improved security ROI may come from better utilising and integrating existing tools.

While Gen-AI may be leveraged to augment existing tools and bridge the resource gap, many organizations are wary of further bloating their stack. Consolidation may be a solution for some, though it does not necessarily entail adopting one single platform and neglecting innovation. Partnering with a leading MSSP to bridge the gap is an option to both derive further security ROI – via fusion of solutions, threat intelligence enrichment and access to security experts that can deliver outcome-based services – and to future-proof the security technology stack.

We believe ROI from security investments will be increasingly under scrutiny. Security leaders will need to identify improvements and potential gains to secure buy-in for further investments.

Report Summary

What Have We Learned?

Sara Puigvert
EVP Global Operations
Orange Cyberdefense

Essential Insights For CISOs, CTOs, and Security Managers

Security Navigator 2025 highlights critical cybersecurity trends, providing insights and strategic guidance tailored to address the challenges faced by today's CISOs, CTOs, and Security Managers. This year's findings underscore how organizations are increasingly exposed to aggressive cyber extortion (Cy-X), sophisticated hacktivism, targeted Operational Technology (OT) threats, and the evolving demands of integrated threat and risk management.

Cyber Extortion (Cy-X): Growing Aggression and Targeted Attacks

Cyber extortion remains a pervasive threat, impacting organizations of all sizes and sectors, especially small and medium-sized enterprises (SMEs). SMEs this year faced a 53% rise in ransomware incidents, and this year marks the biggest ever ransom obtained by a ransomware group: 75 million dollars were paid to Dark Angels. With the emergence of AI tools designed specifically for fraud, extortion, and impersonation, AI has enabled an increase in the volume and sophistication of extortion incidents across sectors. The impact of these attacks reaches beyond the immediate target, with disruptions cascading through supply chains and posing risks to larger companies. We observe a growing cynicism as criminals no longer avoid critical services like healthcare.

We need resilience-building strategies to counter these risks. This includes the implementation of robust recovery protocols and reliable backup systems to reduce downtime and data loss after an attack. Our previous report[224] offers detailed guidance for CISOs.

Hacktivism and Cognitive Attacks: A Rising Threat to Public Trust

Hacktivism is still evolving from activism into destabilizing campaigns, often aligned with geopolitical conflicts like the war against Ukraine, with a particular impact in Europe. In the Nordics, through a combination of distributed denial-of-service (DDoS) attacks and disinformation tactics, pro-Russian hacktivists have launched extensive attacks targeting government services, critical infrastructure and other "symbolic" entities[225][226]. AI can be used to create fake news and digitally altered images as part of campaigns targeting elections and eroding trust in democratic institutions.

Attackers increasingly target perception and trust through these "cognitive" attacks. These attacks aren't technical disruptions. They aim to manipulate public opinion, undermine trust in institutions, and destabilize societal confidence.

To limit the spread of disinformation and safeguard institutional credibility, the report recommends organizations prepare to counter these "cognitive attacks".

This involves equipping cybersecurity teams with monitoring tools to identify disinformation early and implementing rapid-response protocols to counter false narratives effectively. It is paramount to protect high-visibility assets like public-facing websites and social media accounts, which Orange Cyberdefense anti cybercrime teams work toward daily. By managing public perception and maintaining a trusted information environment, organizations can mitigate the reputational damage that often accompanies these attacks.

Operational Technology Security (OT): Unique Risks for Critical Infrastructure

Operational Technology (OT) environments, which control essential physical processes, are now vulnerable to cyber extortion and hacktivism, with attackers frequently using techniques that specifically target OT systems. Unlike information technology (IT) systems, OT environments have specialized requirements that make conventional cybersecurity approaches inadequate.

We highlight direct threats called "Category 2 attacks", which target OT directly and aim to interfere with physical processes. The techniques tend to leverage existing, legitimate OT functionality, and are therefore very hard to detect or block. We can't simply copy the defenses we have for IT in an OT environment. Basic controls like network segmentation remain essential, while more advanced practices like penetration testing need to be carefully examined to ensure they add value to OT.

Evolving Threat and Risk Management: A Shift Beyond "Vulnerability Management"

With over 264,000 vulnerabilities cataloged globally, the load is impossible to manage. Moreover, threats like zero-day vulnerabilities in widely used products like Ivanti, Palo Alto, and Cisco, continue to be exploited by actors reportedly backed by states like China[227]. 2024 has demonstrated that traditional "vulnerability management" must evolve toward a dual strategy of threat-informed prioritization for publicly exposed assets, combined with systemic risk reduction for internal environments.

For large internal environments, we need to conceive architectures that are immune to compromise via an individual system. This requires three strategies: firstly, minimizing attack surfaces by removing unnecessary systems. Secondly, limiting attack impact through robust segmentation and Zero Trust architecture. Thirdly, defining and implementing appropriate configurations, recorded in an asset inventory, and enrolled in software management systems.

Conclusion

As cybersecurity threats become more sophisticated and unpredictable, today's CISOs, CTOs, and Security Managers stand at a pivotal crossroads. The cyber landscape demands more than just defenses; it requires a proactive, intelligence-driven approach that anticipates and mitigates risks before they materialize. Cyber extortion, hacktivism, zero-day exploits and OT-specific threats are no longer isolated issues but interconnected challenges that call for a cohesive and adaptable strategy.

The path forward lies in building resilient organizations equipped to protect, recover, and evolve in response to shifting tactics and emerging vulnerabilities.

This means embracing not only technical solutions but also cognitive defenses to safeguard public trust and prioritizing risk-informed management over sheer volume in vulnerability tracking. By adopting these approaches, security leaders can transform challenges into opportunities for stronger, more resilient infrastructures.

A strong security strategy requires adaptation and readiness to address constantly evolving threats, supported by tools and an organization that can swiftly adjust to new circumstances.

» The path forward lies in building resilient organizations equipped to protect, recover, and evolve in response to shifting tactics and emerging vulnerabilities. «

Sara Puigvert, EVP Global Operations Orange Cyberdefense

Terminology we use in the report

Glossary

Organizational Teams

CERT – Computer Emergency Response Team – produce threat intelligence and coordinate our response to critical threats and vulnerabilieis

VOC – Vulnerability Operations Centers – deliver managed vulnerability scanning services for clients

CSOC – CyberSOC Operations Centers – deliver managed threat detection services for clients

SOC - Security Operations Centers – manage client security equipment like firewalls and VPN.

VERIS 4A Categories [p13]

Actors are entities that cause or contribute to an incident.

Actions describes what the threat actor(s) did to cause or contribute to the incident.

Asset describes the information assets that were compromised during the incident.

Attribute describes which security attributtes (CIA) were compromised during the incident.

Threat Actions [p13]

The Threat Action categories used in the VERIS framework consist of the following 7 primary categories:

Malware is any malicious software, script, or code run on a device that alters its state or function without the owner's informed consent. Examples include viruses, worms, spyware, keyloggers, backdoors, etc.

Hacking is defined within VERIS as all attempts to intentionally access or harm information assets without (or exceeding) authorization by circumventing or thwarting logical security mechanisms. Includes brute force, SQL injection, cryptanalysis, denial of service attacks, etc.

Social tactics employ deception, manipulation, intimidation, etc to exploit the human element, or users, of information assets. Includes pretexting, phishing, blackmail, threats, scams, etc.

Misuse is defined as the use of entrusted organizational resources or privileges for any purpose or manner contrary to that which was intended. Includes administrative abuse, use policy violations, use of non-approved assets, etc. These actions can be malicious or non-malicious in nature. Misuse is exclusive to parties that enjoy a degree of trust from the organization, such as insiders and partners.

Physical actions encompass deliberate threats that involve proximity, possession, or force. Includes theft, tampering, snooping, sabotage, local device access, assault, etc.

Error broadly encompasses anything done (or left undone) incorrectly or inadvertently. Includes omissions, misconfigurations, programming errors, trips and spills, malfunctions, etc.

Environmental not only includes natural events such as earthquakes and floods, but also hazards associated with the immediate environment or infrastructure in which assets are located. The latter encompasses power failures, electrical interference, pipe leaks, and atmospheric conditions.

Mobile Networks Acronyms [p96]

2G: The second generation of mobile networks, providing digital voice and basic data services with low-speed data transmission.

GSM: Global System for Mobile Communications, a standard developed to ensure compatibility between mobile networks worldwide, widely used in 2G networks.

3G: The third generation of mobile networks, enabling faster data speeds and improved multimedia services over 2G networks.

SMS: Short Message Service, a text messaging protocol that allows brief text communication over mobile networks.

Air interface refers to the radio-based communication link between a mobile device (like a smartphone) and the cell tower (base station).

SS7 (Signaling System No. 7) is a global telecommunications protocol standard used to enable communication between mobile and fixed network carriers.

MAP (Mobile Application Part) is a key protocol within the SS7 suite, specifically responsible for handling mobile-related services, like roaming, SMS, and subscriber data management.

A5/1 is an encryption algorithm used to secure voice and data communications over 2G GSM (Global System for Mobile Communications) networks.

Diameter: A protocol that succeeded Radius to support authentication, authorization, and accounting in mobile networks, mainly used in 4G and 5G.

MIMO: Multiple Input Multiple Output, a technology that uses multiple antennas at both transmitter and receiver to improve data throughput and reliability.

HTTP/2: The second major version of the HTTP protocol, offering enhanced security and performance for web applications over mobile networks.

IMSI: International Mobile Subscriber Identity, a unique identifier assigned to each mobile user, crucial for authenticating on mobile networks.

3GPP: The 3rd Generation Partnership Project, a collaborative organization that creates technical standards for mobile communications, including 3G, 4G, and 5G.

UICC: Universal Integrated Circuit Card, a smart card used in mobile devices to secure user identity, network access, and data.

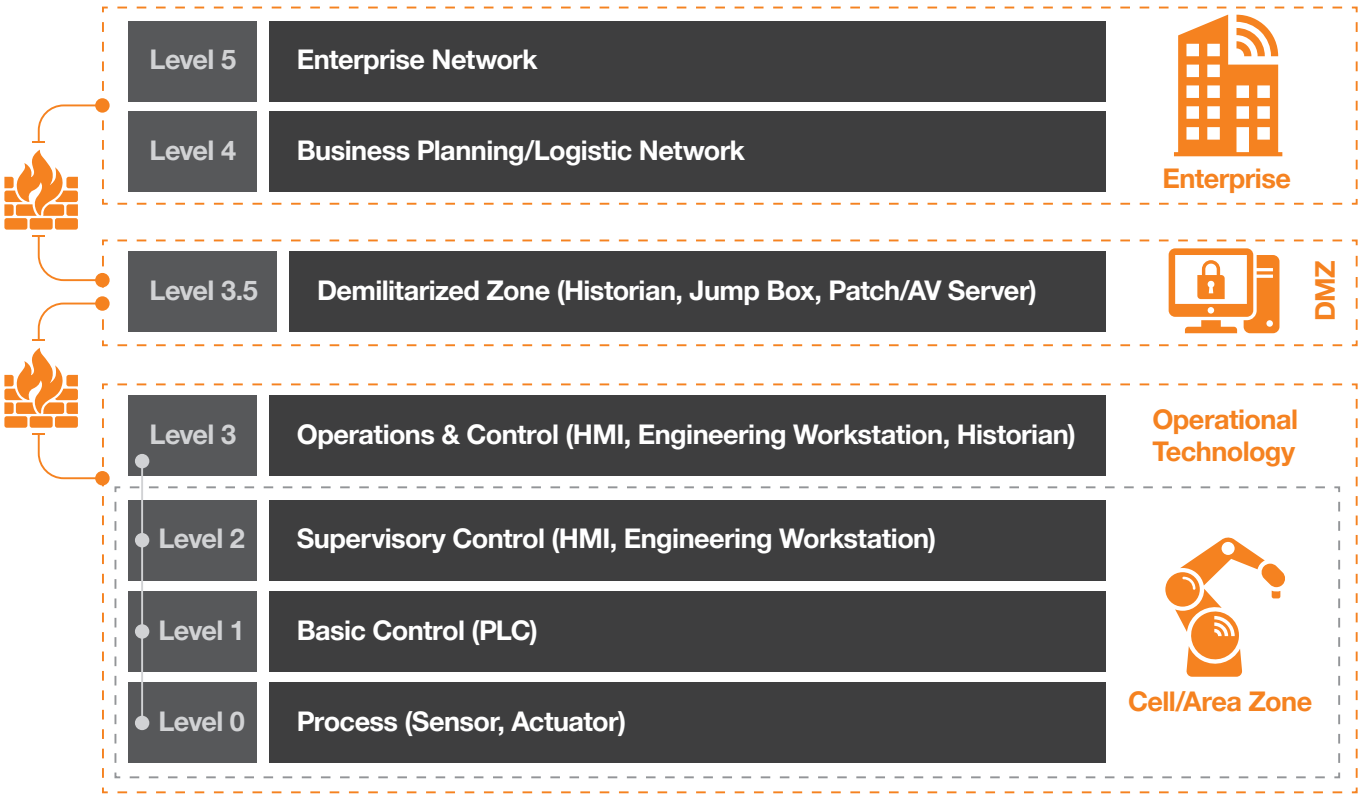
SIM: Subscriber Identity Module, a card that securely stores information, like IMSI, to authenticate users on mobile networks.

eSIM: Embedded SIM, a digital version of a SIM card that is embedded in the device and can be reprogrammed remotely by operators.

SIP: Session Initiation Protocol, a protocol for establishing and managing voice and video calls over IP networks, used in VoIP and mobile network applications.

The Purdue Model [p80]

The Purdue Enterprise Reference Architecture



Contributors, Sources & Links

Sources

This report could not have been created without the hard work of many researchers, journalists and organizations around the world. We've gratefully used their online publications for reference or context.

Sources/links

[1] https://www.bbc.com/news/articles/cz04m913m49o

[2] https://www.reuters.com/world/middle-east/israel-planted-explosives-hezbollahs-taiwan-made-pagers-say-sources-2024-09-18/

[3] https://therecord.media/south-africa-national-health-laboratory-service-ransomware-recovery

[4] https://www.techtarget.com/searchSecurity/news/366614476/Fortinet-discloses-critical-zero-day-flaw-in-FortiManager

[5] https://blogs.microsoft.com/on-the-issues/2024/07/30/protecting-the-public-from-abusive-ai-generated-content/

[6] https://www.dhs.gov/sites/default/files/2023-09/23_0913_ia_23-333-ia_u_homeland-threat-assessment-2024_508C_V6_13Sep23.pdf, page 14

[7] https://www.orangecyberdefense.com/global/blog/research/from-cyber-aware-to-cyber-judgement-how-cisos-can-use-the-aida-marketing-model-to-drive-change

[8] https://www.gartner.com/en/newsroom/press-releases/2020-09-14-gartner-security---risk-management-summit--day-1-high

[9] https://www.cyentia.com/why-your-mttr-is-probably-bogus/

[10] https://www.cisa.gov/securebydesign

[11] https://www.cybersecuritydive.com/news/microsoft-security-debt-crashing-down/714685/

[12] https://www.ivanti.com/blog/our-commitment-to-security-an-open-letter-from-ivanti-ceo-jeff-abbott

[13] https://cwe.mitre.org/data/definitions/1000.html

[14] https://cwe.mitre.org/data/definitions/707.html

[15] https://cwe.mitre.org/data/definitions/664.html

[16] https://www.nationalcrimeagency.gov.uk/the-nca-announces-the-disruption-of-lockbit-with-operation-cronos

[17] https://slcyber.io/a-timeline-of-events-operation-cronos-and-lockbit/

[18] https://www.europol.europa.eu/media-press/newsroom/news/lockbit-power-cut-four-new-arrests-and-financial-sanctions-against-affiliates

[19] https://www.orangecyberdefense.com/global/offering/managed-services/threat-and-risk-management/world-watch

[20] https://cloud.google.com/blog/topics/threat-intelligence/information-operations-surrounding-ukraine

[21] https://therecord.media/polish-anti-doping-agency-polada-hack-leak

[22] https://dfrlab.org/2024/08/01/russia-linked-operations-target-paris-2024-olympics/

[23] https://www.newsguardtech.com/special-reports/2024-paris-olympics-misinformation-tracking-center/

[24] https://harfanglab.io/insidethelab/doppelganger-operations-europe-us/

[25] We choose not to name these groups as we believe they benefit from excessive publicity.

[26] https://socradar.io/what-is-ddosia-project/

[27] https://news.liga.net/ua/politics/news/sait-liganet-bulo-zlamano-nevidomi-opublikuvaly-rosiisku-dezinformatsiiu-pro-avdiivku

https://www.welivesecurity.com/en/eset-research/operation-texonto-information-operation-targeting-ukrainian-speakers-context-war/

https://informnapalm.org/en/website-networks-in-europe-used-as-tools-for-russian-information-warfare-osint-investigation-informnapalm-insight-news/

https://blogs.microsoft.com/on-the-issues/2024/04/17/russia-us-election-interference-deepfakes-ai/

[28] https://www.reuters.com/world/europe/russian-hackers-were-inside-ukraine-telecoms-giant-months-cyber-spy-chief-2024-01-04/
https://portal.cert.orangecyberdefense.com/worldwatch/advisory/1478
[29] https://portal.cert.orangecyberdefense.com/worldwatch/advisory/1478
https://www.kyivpost.com/post/36471
https://www.kyivpost.com/post/36570
https://www.epravda.com.ua/news/2024/07/24/717061/
[30] https://blog.cloudflare.com/cyber-attacks-in-the-israel-hamas-war/
https://www.microsoft.com/en-us/security/security-insider/microsoft-digital-defense-report-2023
https://securelist.com/a-hack-in-hand-is-worth-two-in-the-bush/110794/
https://www.bloomberg.com/news/articles/2023-10-26/israel-taps-blacklisted-pegasus-maker-nso-to-track-gaza-hostages-and-hamas?
[31] https://www.bellingcat.com/news/2023/10/11/hamas-attacks-israel-bombs-gaza-and-misinformation-surges-online/
https://www.zerofox.com/blog/navigating-the-mis-and-disinformation-minefield-in-the-current-israel-hamas-war/
https://twitter.com/JohnHultquist/status/1711605715888955747?s=20
[32] https://blog.cloudflare.com/malicious-redalert-rocket-alerts-application-targets-israeli-phone-calls-sms-and-user-information/
[33] https://cybernews.com/cyber-war/israel-redalert-breached-anonghost-hamas/
[34] https://www.malwation.com/blog/new-muddywater-campaigns-after-operation-swords-of-iron
https://portal.cert.orangecyberdefense.com/worldwatch/advisory/1482
https://research.checkpoint.com/2024/new-bugsleep-backdoor-deployed-in-recent-muddywater-campaigns/
[35] https://www.cisa.gov/news-events/cybersecurity-advisories/aa23-335a
https://www.westernpeople.ie/news/hackers-hit-erris-water-in-stance-over-israel_arid-4982.html
https://portal.cert.orangecyberdefense.com/worldwatch/advisory/1674
[36] https://www.gov.il/en/pages/ziv181223
https://intezer.com/blog/research/stealth-wiper-israeli-infrastructure/
https://www.securityjoes.com/post/bibi-linux-a-new-wiper-dropped-by-pro-hamas-hacktivist-group
[37] https://www.cbc.ca/news/world/hezbollah-pagers-explosions-1.7326969
[38] MTTR is “Mean Time To Resolve”. Once an alert is raised by a security technology and a case is created, MTTR measures the time it takes for the case to be analyzed and then reported to the client, who must investigate, take action, and confirm the finding.
[39] https://darktrace.com/resources/darktrace-ai-combining-supervised-and-unsupervised-machine-learning
[40] https://www.proofpoint.com/us/solutions/nexusai
[41] https://www.microsoft.com/en-us/security/business/siem-and-xdr/microsoft-defender-office-365
[42] https://www.crowdstrike.com/falcon-platform/artificial-intelligence-and-machine-learning/
[43] https://www.microsoft.com/en-us/security/business/ai-machine-learning/microsoft-copilot-security
[44] https://dspace.mit.edu/bitstream/handle/1721.1/147544/Mihretie-yosefmih-meng-eecs-2022-thesis.pdf?sequence=1
[45] https://securityintelligence.com/deeplocker-how-ai-can-power-a-stealthy-new-breed-of-malware/
[46] https://www.rsaconference.com/Library/presentation/USA/2019/the-rise-of-the-machines-ai-and-mlbased-attacks-demonstrated
[47] https://securityaffairs.com/169253/malware/rhadamanthys-information-stealer-uses-ai.html
[48] https://www.ft.com/content/b977e8d4-664c-4ae4-8a8e-eb93bdf785ea
[49] https:/www.microsoft.com/en-us/security/blog/2024/02/14/staying-ahead-of-threat-actors-in-the-age-of-ai
[50] https://openai.com/index/disrupting-malicious-uses-of-ai-by-state-affiliated-threat-actors/
[51] https://openai.com/global-affairs/an-update-on-disrupting-deceptive-uses-of-ai/
[52] https://www.microsoft.com/en-us/security/security-insider/intelligence-reports/microsoft-digital-defense-report-2024
[53] https:/cloud.google.com/blog/topics/threat-intelligence/ai-powered-voice-spoofing-vishing-attacks/
[54] https://www.theatlantic.com/technology/archive/2024/09/microsoft-ai-oil-contracts/679804/
[55] https://www.trendmicro.com/vinfo/us/security/news/cybercrime-and-digital-threats/exploiting-ai-how-cybercriminals-misuse-abuse-ai-and-ml
[56] https://arstechnica.com/science/2024/10/the-more-sophisticated-ai-models-get-the-more-likely-they-are-to-lie/

[57] https://genai.owasp.org/llm-top-10/
[58] https://www.malwarebytes.com/blog/news/2024/10/ai-girlfriend-site-breached-user-fantasies-stolen
[59] https://www.schneier.com/blog/archives/2012/12/feudal_sec.html
[60] https://www.pnas.org/content/early/2020/03/17/1915768117
[61] https://techcrunch.com/2024/04/14/generative-ai-is-coming-for-healthcare-and-not-everyones-thrilled/
[62] https://www.techpolicy.press/shining-a-light-on-shadow-prompting/
[63] https://www.techpolicy.press/author/eryk-salvaggio
[64] https://www.trendmicro.com/en_us/research/24/j/rogue-ai-part-4.html
[65] https://www.cisa.gov/news-events/news/dhs-cisa-and-uk-ncsc-release-joint-guidelines-secure-ai-system-development
[66] https://www.coalitionforsecureai.org
[67] https://www.cnil.fr/fr/definition/modele-ia
[68] https://fr.wikipedia.org/wiki/Alignement_des_intelligences_artificielles
[69] https://www.cyberark.com/resources/threat-research-blog/operation-grandma-a-tale-of-llm-chatbot-vulnerability
[70] https://josephthacker.com/ai/2023/05/19/prompt-injection-poc.html
[71] https://x.com/LeGuideDuSecOps/status/1841180286836441499
[72] https://mistral.ai/fr/
[73] https://huggingface.co/blog/alonsosilva/nexttokenprediction
[74] https://medium.com/@munnangisravya/ascii-smuggler-the-invisible-prompt-injection-d4188d2ff951
[75] https://arxiv.org/pdf/2402.11753
[76] https://promptengineering.org/system-prompts-in-large-language-models/
[77] https://x.com/LeGuideDuSecOps/status/1844298679655727618
[78] https://x.com/literallydenis/status/1708283962399846459
[79] https://www.gladia.io/blog/prompt-injection-in-speech-recognition-explained
[80] https://ai.google.dev/gemma
[81] https://x.com/LeGuideDuSecOps/status/1844298679655727618
[82] https://x.com/literallydenis/status/1708283962399846459
[83] https://www.gladia.io/blog/prompt-injection-in-speech-recognition-explained
[84] https://www.phoronix.com/news/Linux-CVSS-9.9-Rating
[85] https://www.bleepingcomputer.com/news/security/automattic-blocks-wp-engines-access-to-wordpress-resources/
[86] https://therecord.media/vulnerability-database-backlog-nist-support
[87] https://cyberscoop.com/plan-to-resuscitate-beleaguered-vulnerability-database-draws-criticism/
[88] https://www.cnnvd.org.cn/home/childHome
[89] https://www.sentinelone.com/labs/labscon-replay-is-cnvd-%E2%89%A5-cve-a-look-at-chinese-vulnerability-discovery-and-disclosure/
[90] https://www.atlanticcouncil.org/in-depth-research-reports/report/sleight-of-hand-how-china-weaponizes-software-vulnerability/
[91] https://storage.googleapis.com/gweb-uniblog-publish-prod/documents/Year_in_Review_of_ZeroDays.pdf
[92] https://storage.googleapis.com/gweb-uniblog-publish-prod/documents/Year_in_Review_of_ZeroDays.pdf
[93] Cyentia Institute and Kenna Security. 2022. Prioritization to Prediction Vol 8. (2022). https://www.kennasecurity.com/resources/prioritization-to-prediction-reports/
[94] https://www.first.org/cvss/
[95] https://www.cisa.gov/resources-tools/resources/kev-catalog
[96] https://www.orangecyberdefense.com/global/offering/managed-services/threat-and-risk-management/managed-vulnerability-intelligence-watch
[97] https://www.orangecyberdefense.com/global/blog/research/exploring-the-exploit-prediction-scoring-system
[98] https://www.thoughtco.com/complement-rule-example-3126549
[99] https://www.first.org/epss/user-guide
[100] https://www.first.org/epss/user-guide#3-EPSS-Can-Scale-to-Produce-System-Network-and-Enterprise-level-Exploit-Predictions

[101] https://github.com/JohnLaTwC/Shared/blob/master/Defenders%20think%20in%20lists.%20Attackers%20think%20in%20graphs.%20As%20long%20as%20this%20is%20true%2C%20attackers%20win.md
[102] https://attack.mitre.org/
[103] https://sensepost.com/blog/2024/dumping-lsa-secrets-a-story-about-task-decorrelation/
[104] https://math.stackexchange.com/questions/4624889/what-is-the-name-of-this-formula-1-1-pn-x
[105] https://www.mathsisfun.com/data/binomial-distribution.html
[106] https://www.theregister.com/2024/09/20/cisa_software_cybercrime_villains/
[107] https://security.googleblog.com/2024/10/pixel-proactive-security-cellular-modems.html
[108] https://security.googleblog.com/2024/09/eliminating-memory-safety-vulnerabilities-Android.html
[109] https://www.cybersecuritydive.com/news/microsoft-security-debt-crashing-down/714685/
[110] https://www.ivanti.com/blog/our-commitment-to-security-an-open-letter-from-ivanti-ceo-jeff-abbott
[111] https://www.fastly.com/blog/the-dept-of-know-live-sounil-yu-on-why-embracing-the-die-security-model-means-faster-innovation/
[112] https://www.cisa.gov/securebydesign
[113] https://www.cisa.gov/resources-tools/resources/secure-demand-guide
[114] https://www.cisa.gov/resources-tools/resources/secure-design-alert-eliminating-cross-site-scripting-vulnerabilities
[115] https://www.cisa.gov/news-events/cybersecurity-advisories/aa20-107a
[116] https://www.cisa.gov/news-events/directives/supplemental-direction-v1-ed-24-01-mitigate-ivanti-connect-secure-and-ivanti-policy-secure
[117] https://cloud.google.com/blog/topics/threat-intelligence/hacktivists-targeting-ot-systems/
[118] Kushner, D., 2013. The real story of stuxnet. ieee Spectrum, 50(3), pp.48-53.
[119] https://www.dragos.com/blog/protect-against-frostygoop-ics-malware-targeting-operational-technology/
[120] https://www.mitre.org/sites/default/files/pdf/08_1145.pdf
[121] https://www.theregister.com/2023/12/08/polish_trains_geofenced_allegation/
[122] https://www.bbc.co.uk/news/technology-62072480
[123] https://icscsi.org/library/Documents/White_Papers/SANS%20-%20ICS%20Cyber%20Kill%20Chain.pdf
[124] https://www.cyberphysicalsecurity.info/
[125] https://attack.mitre.org/matrices/ics/
[126] https://www.isa.org/standards-and-publications/isa-standards/isa-iec-62443-series-of-standards
[127] https://csrc.nist.gov/News/2023/nist-publishes-sp-800-82-revision-3
[128] https://www.ccn-cert.cni.es/publico/InfraestructurasCriticaspublico/CPNI-Guia-SCI.pdf
[129] Smith, P., 2021. Pentesting Industrial Control Systems: An ethical hacker's guide to analyzing, compromising, mitigating, and securing industrial processes. Packt Publishing Ltd.
[130] Ackerman, P., 2017. Industrial Cybersecurity: Efficiently secure critical infrastructure systems. Packt Publishing Ltd.
[131] Knapp, E.D., 2024. Industrial Network Security: Securing critical infrastructure networks for smart grid, SCADA, and other Industrial Control Systems. Elsevier.
[132] Staves, A., Gouglidis, A., Maesschalck, S. and Hutchison, D., 2024. Risk-based safety scoping of adversary-centric security testing on operational technology. Safety science, 174, p.106481.
[133] Castellanos, J.H., Ochoa, M. and Zhou, J., 2018, December. Finding dependencies between cyber-physical domains for security testing of industrial control systems. In Proceedings of the 34th Annual Computer Security Applications Conference (pp. 582-594).
[134] https://cloud.google.com/blog/topics/threat-intelligence/global-revival-of-hacktivism?hl=en
[135] https://www.trendmicro.com/vinfo/us/security/news/cyber-attacks/rising-from-the-underground-hacktivism-in-2024
[136] https://radar.cloudflare.com/reports/ddos-2024-q1
[137] https://www.weforum.org/agenda/2023/12/2024-elections-around-world/
[138] https://www.cbsnews.com/news/2-sudanese-nationals-charged-cyber-attack-for-hire-gang/
[139] https://www.radware.com/h1-2024-global-threat-analysis-report-lpc-39853846/
[140] https://www.enisa.europa.eu/publications/enisa-threat-landscape-for-dos-attacks
[141] https://www.radware.com/security/threat-advisories-and-attack-reports/hacktivism-unveiled-april-2023/

[142] https://ir.netscout.com/investors/press-releases/press-release-details/2024/DDoS-Attacks-Skyrocket-and-Hacktivist-Activity-Surges-Threatening-Critical-Global-Infrastructure-According-to-NETSCOUTs-1H2024-Threat-Intelligence-Report/default.aspx

[143] https://www.ccc.de/en/hackerethik

[144] https://www.pewresearch.org/internet/2014/03/11/world-wide-web-timeline/

[145] https://www.statista.com/forecasts/1137817/household-internet-penetration-forecast-in-europe

[146] https://www.wired.com/1999/06/coming-soon-back-orifice-2000/

[147] https://www.reuters.com/investigates/special-report/usa-politics-beto-orourke/

[148] https://www.ccc.de/en/hackerethik

[149] http://www.cultdeadcow.com/news/statement19990107.html

[150] https://www.congress.gov/bill/99th-congress/house-bill/4718

[151] https://www.legislation.gov.uk/ukpga/1990/18/contents

[152] Nihilism - The belief things are inherently meaningless.

[153] First activities like date back to the Kosovo war in 1999 where cyber actors targeted the North Atlantic Threat Organization (NATO) and other government websites to protest NATO's bombing of Yugoslavia; by the mid-2000s activities like this became much more prominent. D. (2000) Activism, Hacktivism, and Cyberterrorism: The Internet as a Tool for Influencing Foreign Policy.

[154] https://securityintelligence.com/posts/the-decline-of-hacktivism-attacks-drop-95-percent-since-2015/ https://www.fbi.gov/contact-us/field-offices/newyork/news/press-releases/leading-member-of-the-international-cyber-criminal-group-lulzsec-sentenced-in-manhattan-federal-court

[155] Smith, M. (2023). The Irregulars: Third-Party Cyber Actors and Digital Resistance. CyCon 2023 Proceedings. DOI: 10.23919/CyCon58705.2023.10182061 https://ccdcoe.org/uploads/2018/10/Ottis2008_AnalysisOf2007FromTheInformationWarfarePerspective.pdf

[156] https://www.atlanticcouncil.org/blogs/ukrainealert/the-2008-russo-georgian-war-putins-green-light/

[157] https://www.washingtonpost.com/lifestyle/style/the-hacktivists-of-telecomix-lend-a-hand-to-the-arab-spring/2011/12/05/gIQAAosraO_story.html

[158] Kostiantyn Korsun, former Head of Ukrainian CERT and former Deputy Head of Computer Crime Division at the Security Service of Ukraine posted a request on LinkedIn asking for help on the cyber front. https://docslib.org/doc/8087108/cyber-proxies-and-the-crisis-in-ukraine

[159] Maurer, T. (2018). Cyber mercenaries: The state, hackers, and power. Cambridge University Press.

[160] https://cyberforumkyiv.org/A_Decade_in_the_Trenches_of_Cyberwarfare.pdf

[161] https://therecord.media/ukraine-monobank-ddos-attack-donations

[162] The total number of requests sent to overwhelm a service

[163] https://blog.cloudflare.com/how-cloudflare-auto-mitigated-world-record-3-8-tbps-ddos-attack/

[164] The total data volume (in bits) sent per second

[165] Bandwidth-based attacks aim to saturate the network and can be more challenging to mitigate.

[166] https://www.radware.com/security/threat-advisories-and-attack-reports/project-ddosia-russias-answer-to-disbalancer/

[167] https://media.defense.gov/2024/May/01/2003454817/-1/-1/0/DEFENDING-OT-OPERATIONS-AGAINST-ONGOING-PRO-RUSSIA-HACKTIVIST-ACTIVITY.PDF

[168] https://www.bleepingcomputer.com/news/security/us-govt-warns-of-pro-russian-hacktivists-targeting-water-facilities/

[169] https://www.lawfaremedia.org/article/what-impact-if-any-does-killnet-have

[170] Nissen, T. E. (2015). "The Weaponization of Social Media: Information Operations in the Context of 21st Century Warfare." Royal Danish Defense College.

[171] https://bindinghook.com/articles-hooked-on-trends/russias-strategic-culture-drives-its-foreign-hacking/

[172] https://en.wiktionary.org/wiki/Russophobic

[173] https://www.reuters.com/world/americas/canadian-pm-apologises-after-parliamentary-speaker-publicly-praised-nazi-2023-09-27/

[174] https://en.wiktionary.org/wiki/Russophobic

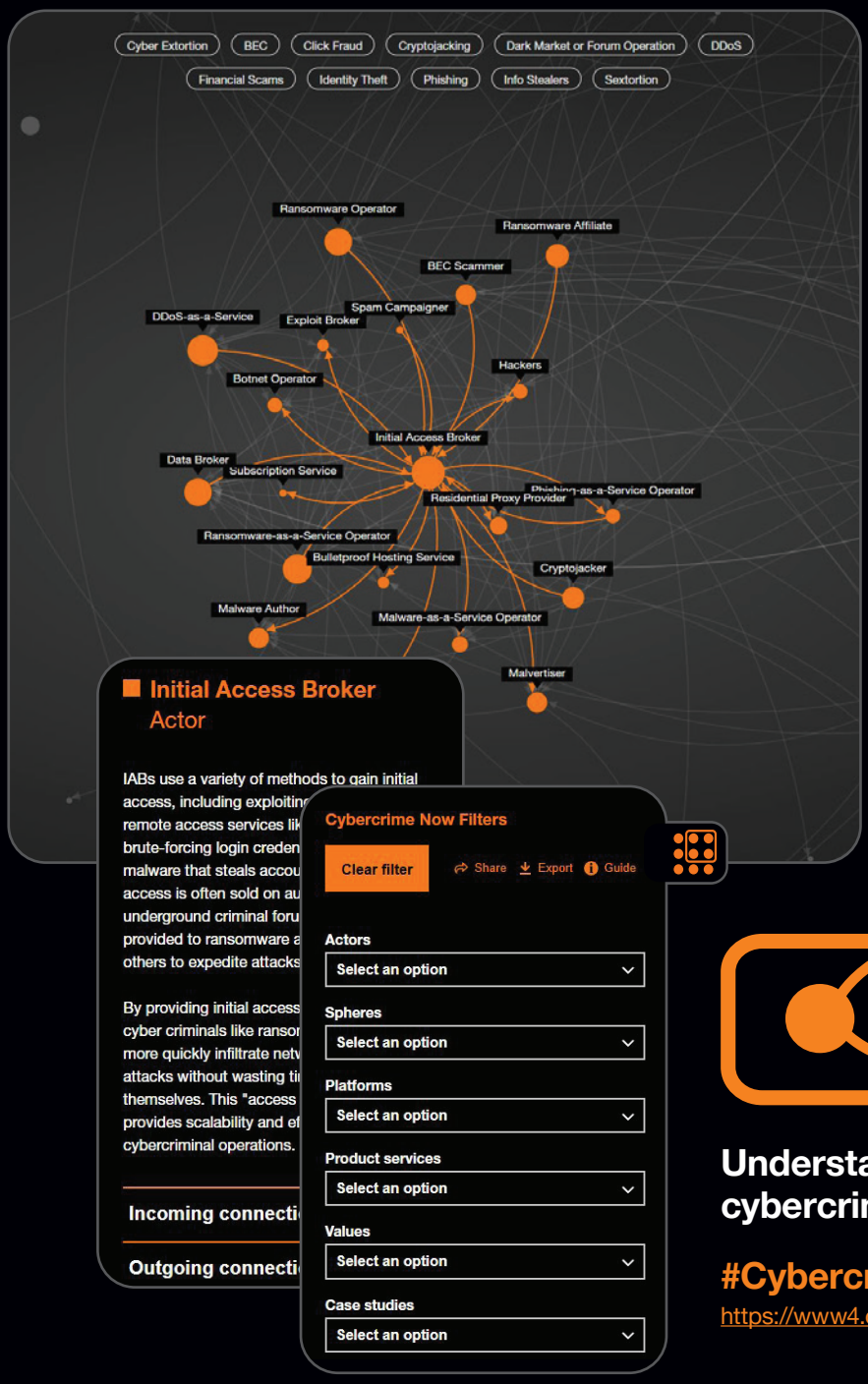
[175] https://www.pravda.com.ua/eng/news/2023/09/9/7419101/

[176] https://mitre-engenuity.org/cybersecurity/center-for-threat-informed-defense/

[177] https://www.mitre.org/sites/default/files/2021-11/prs-19-3892-ttp-based-hunting.pdf
[178] https://www.splunk.com/en_us/blog/security/peak-threat-hunting-framework.html
[179] https://center-for-threat-informed-defense.github.io/summiting-the-pyramid/
[180] https://medium.com/detect-fyi/akira-in-the-chang-way-server-ecosystem-re-vicitimization-a9011fbc6dff
[181] https://mitre-engenuity.org/cybersecurity/center-for-threat-informed-defense/our-work/measure-maximize-and-mature-threat-informed-defense-m3tid/
[182] https://citizenlab.ca/tag/nso-group/
https://blog.sekoia.io/active-lycantrox-infrastructure-illumination/
https://www.amnesty.org/en/documents/act10/7245/2023/en/
https://gijn.org/stories/the-rapid-rise-of-phone-surveillance/
https://www.amnesty.org/en/latest/press-release/2021/07/world-leaders-potential-targets-of-nso-group-pegasus-spyware/
https://www.business-humanrights.org/en/latest-news/nso-group-spyware-sold-to-governments-used-to-target-activists-politicians-journalists-according-to-pegasus-project-investigation-company-denies-allegations/
https://rm.coe.int/pegasus-and-similar-spyware-and-secret-state-surveillance/1680ac7f68
https://apnews.com/article/poland-spyware-pegasus-nso-group-israel-413bb3cb27daac011d52b524c6d16160
https://www.reuters.com/technology/cybersecurity/spain-reopens-israeli-spyware-probe-sharing-information-with-france-2024-04-23/
[183] https://therecord.media/sms-blasting-arrests-uk-homemade-antenna
[184] https://www.francetvinfo.fr/faits-divers/escroquerie-aux-sms-de-l-assurance-maladie-les-suspects-volaient-les-numeros-de-telephone-depuis-leur-voiture_5665943.html
[185] https://commsrisk.com/oslo-imsi-catcher-arrest-suspected-malaysian-spy-now-investigated-for-fraud-with-international-ramifications/
[186] https://therecord.media/orange-espana-outage-hacker-internet-ripe-bgp-rpki
[187] https://www.gsma.com/solutions-and-impact/technologies/security/gtleasing/
[188] https://www.lighthousereports.com/investigation/ghost-in-the-network/
[189] https://krebsonsecurity.com/2021/09/the-rise-of-one-time-password-interception-bots/
[190] https://www.wsj.com/tech/cybersecurity/u-s-wiretap-systems-targeted-in-china-linked-hack-327fc63b
[191] https://www.rcrwireless.com/20241008/telecom-software/verizon-att-lumen-among-telcos-hacked-by-chinese-group-reports
[192] https://www.gsma.com/solutions-and-impact/technologies/security/
[193] https://networkencyclopedia.com/global-system-for-mobile-communications-gsm/
[194] https://ss7.info/
[195] https://en.wikipedia.org/wiki/Signalling_System_No._7
[196] https://www.umtsworld.com/umts/faq.htm
[197] https://www.thalesgroup.com/en/markets/digital-identity-and-security/technology/lte
[198] https://ss7.info/ss7-vs-diameter/
[199] https://www.gsma.com/solutions-and-impact/gsma-open-gateway/gsma-open-gateway-api-descriptions/
[200] https://www.gsma.com/solutions-and-impact/gsma-open-gateway/
[201] https://www.rcrwireless.com/20240625/5g/philippine-telcos-join-gsma-open-gateway-initiative
[202] https://www.enisa.europa.eu/publications/signalling-security-in-telecom-ss7-diameter-5g/@@download/fullReport
[203] https://www.securityweek.com/gtp-vulnerabilities-expose-4g5g-networks-high-impact-attacks/#:~:text=Positive%20Technologies%20performed%20security%20assessments%20on%20behalf%20of,it%20does%20not%20check%20the%20user's%20actual%20location.
[204] https://www.blackhat.com/docs/us-15/materials/us-15-Yu-Cloning-3G-4G-SIM-Cards-With-A-PC-And-An-Oscilloscope-Lessons-Learned-In-Physical-Security.pdf
[205] https://www.kaspersky.co.za/blog/sim-card-history-clone-wars/11091/
[206] https://github.com/nickel0/3GPP-Overall-Architecture/blob/master/diagram/3GPP_Overall_Architecture_and_Specifications.pptx
[207] https://www.gsma.com/solutions-and-impact/technologies/security/wp-content/uploads/2024/07/Security-Landscape-2024-Issue-intro-contents.pdf

[208] https://pages.nist.gov/800-63-3/sp800-63b.html
[209] https://www.magicbell.com/blog/expert-guide-to-push-notifications
[210] https://www.airship.com/resources/explainer/ios-push-notifications-explained/
[211] https://medium.com/@KaushalVasava/push-notification-in-android-how-its-work-2679d0bc0720
[212] https://en.wikipedia.org/wiki/Mark_Klein
[213] https://en.wikipedia.org/wiki/Greek_wiretapping_case_2004–05
[214] https://www.thalesgroup.com/en/markets/digital-identity-and-security/press-release/gemalto-presents-the-findings-of-its-investigations-into-the-alleged-hacking-of-sim-card-encryption-keys
[215] https://www.cbsnews.com/news/60-minutes-hacking-your-phone/
[216] https://fidoalliance.org/passkeys/
[217] https://support.google.com/accounts/answer/13548313?hl=en-EN
[218] https://support.apple.com/en-za/guide/iphone/iphf538ea8d0/ios
[219] https://fight.mitre.org/
[220] https://github.com/swannman/ircapabilities
[221] https://cmmiinstitute.com/learning/appraisals/levels
[222] https://www.forrester.com/blogs/2025-security-risk-budget-planning-guide/
[223] https://open.spotify.com/episode/7dNpU6mxd7UUou2pz2mxIN
[224] https://www.orangecyberdefense.com/global/cyber-crisis-management
[225] https://www.politico.eu/article/meet-killnet-russias-hacking-patriots-plaguing-europe/
[226] https://www.thalesgroup.com/en/worldwide/security/press_release/ukraine-whole-europecyber-conflict-reaches-turning-point
[227] https://www.securityweek.com/volexity-catches-chinese-hackers-exploiting-ivanti-vpn-zero-days/

What are the criminals doing?



Defenders think in lists. Attackers think in graphs. As long as this is true, attackers win.

John Lambert,
Microsoft

Understanding cybercrime

#CybercrimeNow

https://www4.orangecyberdefense.com/cybercrime-now

Disclaimer

A very special thanks
to all our experts including cyber hunters, researchers, analysts, engineers, ethical hackers and incident responders.

Why Orange Cyberdefense?

Orange Cyberdefense is the expert cybersecurity business unit of the Orange Group, providing managed security, managed threat detection & response services to organizations around the globe. As Europe's go-to security provider, we strive to build a safer digital society.

We are a threat research and intelligence-driven security provider offering unparalleled access to current and emerging threats.

Our organization retains a 25+ year track record in information security, 250+ researchers and analysts 17 SOCs, 15 CyberSOCs and CERTs distributed across 11 location in the world and sales and services support in 160 countries. We are proud to say we can offer global protection with local expertise and support our customers throughout the entire threat lifecycle.

Orange Cyberdefense has built close partnerships with numerous industry-leading technology vendors.

We wrap elite cybersecurity talent, unique technologies and robust processes into an easy-to-consume, end-to-end managed services portfolio.

At Orange Cyberdefense we embed security into Orange Business solutions for multinationals worldwide. We believe strongly that technology alone is not a solution. It is the expertise and experience of our people that enable our deep understanding of the landscape in which we operate. Their competence, passion and motivation to progress and develop in an industry that is evolving so rapidly.

We are proud of our in-house research team and proprietary threat intelligence thanks to which we enable our customers to focus on what matters most, and actively contribute to the cybersecurity community. Our experts regularly publish white papers, articles and tools on cybersecurity which are widely recognized and used throughout the industry and featured at global conferences, including Infosec, RSA, 44Con, BlackHat and DefCon.

www.orangecyberdefense.com