SWP Comment
NO. 46 OCTOBER 2024

The Attribution Dividend: Protecting Critical Infrastructure from Cyber Attacks

Annegret Bendiek, Jakob Bund, and Mika Kerttunen

International law and voluntary norms have not effectively prevented state, proxy, or other criminals from malicious and harmful behaviour in cyberspace. Geopolitical confrontation and tension beyond cyberspace with major threat actors have largely exhausted sanctions. Faced with threats that prove difficult to deter, the European Union (EU), its Member States, and international partners need to refocus their attention on creating friction for malicious activity and increasing the costs of adversary operations. Through their contributions to resilience, forensic capabilities and international cooperation on technical investigations offer practical opportunities to blunt the tools of adversaries. By coordinating technical, political, and legal attribution at the EU level, Member States could reinforce a victim-oriented approach to cyber diplomacy.

Pipeline ruptures in the Baltic Sea, cut data cables, and disruptions of satellite communications are raising questions in European capitals about how to respond to suspected attacks against critical infrastructure. Addressing incidents affecting Germany and at least six other member states, NATO allies in early May called attention to the scope of detected Russian efforts to destabilize supporters of Ukraine. Moscow's documented hybrid activities have ranged from influence campaigns, air traffic signal jamming, and violent intimidation and assassination attempts against regime critics to acts of sabotage, including through cyber-enabled means. To strengthen capabilities to prepare and assist Member States in countering these threats, in late May 2024 the Council of the EU agreed on a first framework to set up EU Hybrid Rapid Response Teams. Deploying such mechanisms effectively in the EU, however, depends on a shared understanding among Member States of threat actors and their capabilities. For cyber-enabled sabotage, the practices for analyzing the tactics, techniques, and procedures of threat actors as well as the trade-offs involved in publicly attributing malicious activity have been tested and evolved over more than a decade against the backdrop of Russia's attempts to subvert Ukraine's election systems and electric grid following Russia's invasion of Crimea in 2014.

Stiftung Wissenschaft und Politik
German Institute for International and Security Affairs

In comparison to other types of cyber operations, sabotage missions designed to destroy data or cause operational disruptions have been a focus of public attribution efforts. Practical differences, such as the visible effects of a successful sabotage, may contribute to public pressure to address these operations and a higher attribution ratio. By their design, sabotage attempts seek to manipulate targeted systems to cause a specific disruptive effect. Concerns in this regard have focused, for example, on the opening of circuit breakers in substations to cut off electricity or the altering of chemicals used in water treatment plants to levels dangerous to health. Such physical consequences can make it easier for investigations to identify initial effects and likely strategic intentions that attribution statements can address.

Espionage operations, in contrast, tend to develop their value through cumulative effects. This can be accomplished either through depth, by maintaining access to target environments for prolonged periods of time, or breadth, through intrusions across certain sectors. This campaigning approach requires a similarly sustained effort to attribution, that assembles and connects the pieces of this puzzle. Attributing sabotage operations targeting critical infrastructure is part of a larger effort to assert UN-level agreements about responsible behaviour in cyberspace. The credibility of the red line these agreements established for the protection of critical national infrastructure in times of peace depends on whether norm violations are identified and perpetrators are held accountable. A notable trend has been the extension of the attribution pattern from state-sponsored disruptive operations to espionage operations that are conducted to prepare for sabotage. Attribution efforts have also expanded to criminal groups that seek to hold critical infrastructure at risk for financial gain. In line with shifts in the threat landscape, this has led to a surge in the attribution of malicious cyber activity.

Flashpoint critical infrastructure

To assess threats against critical infrastructure, a fundamental distinction is necessary. Network environments of many essential service providers, such as power plants or water treatment facilities, contain both IT and operational technology (OT) components. While IT networks are used for administrative tasks, such as to send emails or issue invoices, OT sustains the actual operations of critical infrastructure. Central within OT are industrial control systems (ICS), which manage processes that, for example, clean water, distribute electricity, or direct traffic. The vast majority of intrusions affecting critical infrastructure do not reach OT components. However, as IT and OT networks converge, disruptions of IT appliances can develop cascading effects. For instance, Colonial Pipeline shut down its distribution of gas and jet fuel after the Russian ransomware group DarkSide encrypted its billing system in May 2021. This decision by the largest United States (US) transport pipeline operator out of an abundance of caution led to temporary supply shortages at US East Coast gas stations.

In contrast, ICS specific malware that directly targets OT remains rare. To date, only nine cases have been publicly tracked. Instances of destructive effects against OT through cyber capabilities are rarer yet. Increasingly, OT systems that were developed for operations in isolated networks are brought online, as part of digitalisation efforts to assist with remote maintenance. These connectivity initiatives change the threat model of critical infrastructure operators and can lead to new vulnerabilities. OT components and protocols that are legacies from air-gapped operations were not designed for the potential exposure introduced with deployment adjacent to IT networks. Digital transformation of industries needs to take into account how such increases in OT connectivity can be secured.

Driver 1: Russian efforts to erode support for Ukraine

Russia's use of cyber operations has become a central component of its broader strategy to weaken international support for Ukraine. These threats, however, are not solely orchestrated by the Kremlin. Various hacktivists with suspected links to state entities in Russia, such as the Cyber Army of Russia Reborn (CARR), operate with varying degrees of autonomy. The involvement of these groups poses a significant risk of inadvertent escalation, as highlighted by repeated warnings from the United Kingdom's (UK) Government Communications Headquarters in 2023 and 2024.

The sophistication of these threat actors remains limited, as demonstrated by the US imposition of sanctions on July 19th, 2024, against two prominent CARR members. While these actors may not possess advanced capabilities, the US government has designated their activities as a "significant threat to the national security, foreign policy, or economic health or financial stability" under Executive Order 13694. This classification is less a judgment about the technical prowess of the hackers and more about the vulnerabilities within the targeted organisations, particularly those in critical infrastructure sectors. National Security Advisor Jake Sullivan and the Environmental Protection Agency have both called for enhanced support for local utilities, emphasizing that strengthening resilience must be a key component of the defence strategy.

Driver 2: Chinese preparations for a geopolitical contingency

The strategic intent behind cyber operations targeting sensitive networks is difficult to assess through technical means alone. This is particularly true when the objective is to establish access for potential future disruptions without observable actions in the present to that end. In May 2023, US intelligence disclosed a series of stealthy intrusions affecting sensitive telecommunication networks that it linked to Volt Typhoon, a group with suspected links to the Chinese state. US intelligence assessments have concluded that the detected activities differ fundamentally from previous espionage operations. This distinction is based on the nature of the intrusions, which is aligned more with preparations for potential offensive actions rather than mere intelligence gathering.

A beachhead for sabotage?

Volt Typhoon is a new but highly sophisticated Advanced Persistent Threat actor, attributed to China, focusing on espionage and network discovery. It has been active since mid-2021, primarily targeting critical infrastructure. Microsoft first disclosed Volt Typhoon's activity on May 24th, 2023, in a report published alongside a joint advisory by the US and the cybersecurity agencies of Five Eyes partner countries. The observed behaviour suggests that the threat actor is attempting to maintain access for as long as possible without being detected, using living-off-the-land techniques. Microsoft states with moderate confidence that the analysed activity is already at a form of developing the capabilities to disrupt infrastructure in the future (prepositioning). Notable activities include the successful infiltration of US communication infrastructure in the Pacific region in 2021 that would be critical to US logistics in an armed confrontation with China.
Author: *Erik Kellenter*, SWP.

Driver 3: Cybercrime as emerging national security threat

The rise of ransomware attacks targeting critical infrastructure with low disruption tolerance, particularly in the healthcare sector, represents a new dimension of national security threats posed by criminal actors. Over the past 12 months, ransomware attacks on hospitals have surged, placing the healthcare sector among the most targeted industries. To speed up recovery and prevent the leak of sensitive health data, the US-based drug distribution company Cencora, for example, agreed to pay a ransomware group $75 million, the highest publicly recorded sum to date. This trend marks a significant shift. Ransomware

groups previously had been reluctant to target organisations in the healthcare sector out of concern of attracting law enforcement attention. Another ransomware attack in June caused delays in blood processing in London, forcing the city to limit blood transfusions. Such incidents underscore the potentially disastrous impact of ransomware on essential services, including the delivery of care.

The threat posed by criminal cyber actors has been recognized at the highest levels of government. The European Commission's new Political Guidelines include a commitment to develop an action plan within the first 100 days in office to address this escalating threat. Additionally, the EU's decision to sanction criminal actors for the first time in June 2024 reflects an increasing concern over the impact of these threats on critical national infrastructure and the limitations of traditional law enforcement tools in mitigating these risks.

Returns from technical attribution

While the immediate purpose of attribution is to establish responsibility, its underlying objective, as Jason Healey elaborated, is to enable an understanding of malicious activity that can put a stop to it. Efforts in this direction are typically categorized according to three lines of action: technical, political, and legal. At the foundation, forensic investigations aim to determine the tools, tactics, techniques, and procedures (TTPs) employed in an operation and trace the systems and networks through which it was conducted. Such technical assessments link this information to an intrusion set to compare and find matches across other operations. This allows observations of how threat actor behaviour evolves over time. Although the initial focus is on reconstructing the breach, these insights can also inform targeted defences against the detected intrusion patterns. Hardening the entry points used by threat actors to break into networks and monitoring for their instruments impose operational costs on threat actors, forcing them to adjust their approach to gain access or avoid detection.

Findings from this technical investigation may be leveraged for political attribution, which assigns responsibility to a state. Judgments of state responsibility can cover a spectrum, ranging from direct authorship of an operation and, support for proxies to failing to meet due diligence obligations in prohibiting criminal activity originating from its territory. Public and private communications of these conclusions seek to weigh in on adversaries' strategic cost calculus to deter them from continuing their operation. A key lesson in this regard has been the interference in the 2016 US election. The delay in a US government response to expose the leaks that were directed by the Russian government allowed for influence material that was stolen from the Clinton campaign to circulate without official challenge. When in August 2024, several US media organisations were approached with internal documents from the Trump campaign, the outlets decided against reporting on details from the hacked material on the suspicion that the files had been obtained by a foreign power. Early warnings by Microsoft and supporting statements by the US intelligence community that linked the leak to a hostile power vindicated this approach, denying momentum to the influence attempt.

Building on such efforts to establish responsibility, legal attribution aims to hold perpetrators and their sponsors accountable with a longer-term view — beyond an individual operation — of strengthening the rule of international law and the respect of norms of responsible behaviour. Calling out violations strengthens these frameworks as the basis for response measures.

While political and legal responses rely on the conclusions of forensic investigations, technical attribution, including those by government actors, can proceed independent of decisions to assign responsibility and pursue accountability. In its public response to the compromise of the defence company RUAG, the Swiss government

notably focused on technical attribution in describing the breach as conducted by the espionage group Turla. Attribution speed is of essence if the goal is to protect other potential victims by sharing knowledge of TTPs and reducing their effectiveness. This positions technical attribution as the fast track to disseminating insights with the direct potential to influence adversary behaviour.

In the current threat environment critical infrastructure faces, political and legal measures face an uphill challenge in diminishing the strategic motivations of states at war or preparing for conflict and the financial motivations of criminals sheltered in these countries. Technical attribution provides a channel to introduce friction into the capabilities of adversaries by creating awareness about attacker TTPs that allows for the development of mitigation measures.

In a best-case scenario, early detection can thwart an operation. In 2022, the OT cybersecurity company Dragos in collaboration with the US government discovered a toolkit capable of causing physical disruption and destruction in electrical systems, oil and gas pipelines, water systems, manufacturing plants, and military control systems. As the tool uses the inherent functionalities of target systems, its use cannot be prevented with a software fix. Detection requires continuous vigilance. Noting the capability's robustness and applications across a variety of sectors, the analysts called it PIPEDREAM. The name, however, also captures the success of detecting the tool before its deployment — "left of boom".

Plans like the announcement of the UK National Cyber Security Centre (NCSC) in August to upgrade its Active Cyber Defence (ACD) program seek to expand early detection capabilities to disrupt adversary activity. In particular, the NCSC is preparing to add deceptive measures to the ACD portfolio. Part of this consideration are tripwires and honeypots that are designed for threat actors to inadvertently reveal their presence and tradecraft. Making these services available at scale, especially to small- and medium-sized enterprises that otherwise lack the resources for these capabilities, can further extend the visibility into threat activity that supports technical attribution.

The addition of deceptive measures also points to a psychological component of technical attribution capabilities. Knowledge about the potential presence of deception tools in target networks may in itself have an effect on threat actor behaviour — an area of influence that the NCSC has identified for further study.

Maintaining the attribution dividend

In a new development, China's National Computer Virus Emergency Response Center and the 360 Digital Security Group made ham-fisted attempts in two reports to misconstrue Western threat intelligence reporting to link Volt Typhoon to a criminal ransomware group. Without any corroboration, these reports refer to updates to industry reporting as signs of US government pressure on industry to hide the purported ransomware identity of Volt Typhoon. The incentive structure of ransomware groups to openly claim responsibility for their breaches to exert pressure on victims to respond to demands makes these claims doubtful. Despite their lack of supporting evidence, Chinese officials repeatedly cited these reports, in an attempt to frame the official disclosures on Volt Typhoon as a "disinformation campaign" conceived by the US intelligence community.

Despite these steps to deflect responsibilities, political attribution has achieved no perceptible change in behaviour. In early August, Sherrod DeGrippo, Director of Threat Intelligence Strategy for Microsoft — the company that first publicly reported on Volt Typhoon's activities — assessed that the group's operations continue unabated. DeGrippo identified a pattern of "consistency and persistence" in its targeting. A central concern, as the former Chief Security Officer of Yahoo and Facebook Alex Stamos put it, is that publicity "does not deter" actors and their sponsors.

As Volt Typhoon shows, despite coordinated efforts to root out the group from targeted networks with a dedicated technical attribution campaign, for well-resourced groups with long-term strategic objectives, these effects have a short half-life. While technical attribution raises the costs of operations for threat actors, it remains in constant competition with adversary efforts to retool. For advanced actors that are unlikely to be priced out, successes remain temporary. Defender advantages developed through technical attribution are therefore part of a cycle: Knowledge about TTPs and tools close the gap and can deprive threat actors of their foothold, requiring them to identify new entry vectors. Threat actors regain momentum as they pivot to new intrusion techniques — a gap that renewed detection and attribution efforts need to close. However, not all attack infrastructure is equally easy to replace. Takedowns enabled through technical attribution, such as in the case of the KV botnet of hijacked small office/home office routers in January 2024 in the US that allowed Volt Typhoon to obfuscate its operations, can set adversaries back.

An FBI-led disruption of a second China-nexus botnet in September 2024 underscores this contribution of continuous technical attribution. Flax Typhoon, the group who controlled the botnet, seeks to maintain a low profile by minimizing malware use and avoiding easily identifiable intrusion patterns. This approach allowed the group to remain undetected during a multi-year espionage campaign against critical infrastructure targets, including in Taiwan and the US. To communicate findings beyond jurisdictions cooperating in the takedown, the US together with its Five Eyes partners published an advisory documenting more than 60 vulnerabilities that Flax Typhoon had exploited to grow its botnet. Assembling a cumulative picture of these compromises maps the attack infrastructure and can inform the focus of preventive efforts. Consistent tracking of actor groups that enables substantive degradation of their infrastructure raises the bar for the efforts of threat actors to stay undetected over time.

In a new espionage campaign targeting US Internet service providers (ISPs) and digital services that cater to government and military users publicly reported in August, Volt Typhoon was observed using a zero-day vulnerability in Versa Director, a software that allows for the central management and monitoring of network devices. Such previously unreported software flaws are a rare commodity. Volt Typhoon's use of a zero-day exploit highlights that the group needs to resort to valuable means to regain access. It also demonstrates its efforts to avoid early detection and technical attribution as it continues its operations.

The EU sanction-attribution nexus

The EU's response to malicious cyber activities has evolved significantly over the past few years. The EU's cyber sanctions regime was formally established with Council Regulation (EU) 2019/796 and Council Decision (CFSP) 2019/797 of 17 May 2019. This legal framework laid the foundation for imposing targeted restrictive measures on individuals, entities, and bodies responsible for cyberattacks threatening the EU or its Member States. However, it also made a clear distinction between sanctions and the attribution of responsibility to a third state. The Decision emphasized that attribution is a sovereign political decision that each Member State can make independently on a case-by-case basis.

The initial rounds of sanctions under this regime, imposed on 30 July 2020 and 22 October 2020, followed a more cautious approach. These measures were imposed with a significant time difference between the events they are intended to sanction and relied heavily on public third-party reporting that required minimal information sharing between Member States. For the first round of sanctions, the EU clearly stated that the imposition of restrictive measures was to have a deterrent effect and should be distinguished from attributing responsibility to any state actor.

Over time, the EU's approach to attribution in the context of cyber sanctions has evolved, particularly as the threat landscape has grown more complex. The Council Conclusions from 21 May 2024 on the Future of Cybersecurity highlighted the rising threat level as a key reason to review the 2020 EU Cybersecurity Strategy. The Conclusions explicitly pointed to the increasing frequency and severity of cyberattacks, particularly those targeting critical infrastructure and essential services, often using disruptive techniques such as ransomware and wiper malware.

These changes in the threat environment were a direct impetus for the third round of EU sanctions imposed on 24 June 2024. This latest round marked a significant shift in the EU's strategy by more actively linking attribution to the sanctions process. The EU relied on attribution not only to justify sanctions but also as a tool to directly connect individuals to specific malicious activities. Notably, two individuals were sanctioned by the EU who had not previously been indicted or sanctioned by other countries, despite being identified by the Ukrainian Security Service as officers of the Russian Federal Security Service (FSB). This marked the first time that the EU had sanctioned individuals based on attributions that had not been publicly established by other major powers, signalling a new willingness of the EU to lead.

However, despite this progress, there are still areas where the EU's attribution, coordination, and sanctions processes have room to catch up to the efforts of partner countries. For example, the UK and US have often documented and sanctioned threat actors based on intelligence that was not publicly available at the time, demonstrating a higher level of agility and coordination in their response. While the Council's press release accompanying the third round of sanctions expressed an ambition to enhance cooperation with the UK and US in disrupting and responding to cybercrime, the EU's current attribution coordination has yet to reach this proactive level. The influence of the ongoing conflict in Ukraine has influenced the EU's approach to attribution. For instance, in response to the disruptions of modems linked to the KA-SAT satellite network, the EU's High Representative issued a statement in solidarity with Ukraine and attributed the attack to Russian state actors. This quick response is part of a broader trend of shrinking attribution timelines that has accelerated the exchange of threat information.

The revised implementing guidelines of the EU Cyber Diplomacy Toolbox (CDT) adopted in 2023 further underscore the EU's commitment to improving attribution processes. The guidelines include detailed recommendations on how to use attribution strategically in communications while maintaining that political attribution remains the sovereign decision of Member States. The guidelines also recognize the importance of shared situational awareness and information sharing to facilitate coordinated political attribution at the EU level. Whereas the guidelines acknowledge the impact of technical attribution on decision-making, they do not explicitly address the mechanisms through which technical attribution contributes to the overall deterrent effect.

Leveraging technical attribution: Recommendations for the EU

While the technical dimension of attribution remains under-leveraged by EU institutions due to limited analytical capabilities provided by Member States, the EU can significantly increase its engagement through strategic partnerships, support for international mechanisms, and alignment with victims of cyber operations.

A victim-oriented approach to cyber diplomacy

One of the most effective ways for how the EU can strengthen its attribution efforts is by enhancing the cyber capabilities of its partner countries. Capacity building support is crucial in enabling these nations to establish and improve their own attribution

capabilities, which can serve as a force multiplier for collective security. By focusing on partnerships, the EU can support a network of countries equipped to conduct credible and timely attribution.

The EU should prioritize coordination with countries that have been victims of cyber operations, particularly in the use of measures provided by the CDT. The European External Action Service can play a pivotal role in this effort by building on its initial consultations, e.g., with Ukraine. Ukraine's experience and ongoing defence efforts offer valuable insights that can inform a more victim-oriented approach to cyber diplomacy. Law enforcement cooperation has taken steps in this direction on a case-by-case basis, documenting the behaviour of ransomware groups to prevent further compromises. Diplomatic initiatives can systematically scale this exchange between victims and potential targets at the international level.

Recent exchanges as part of the EU-Ukraine Cyber Dialogue signal a shift towards a target-centred approach that recognizes the experience of frontline states in their value for identifying targeted responses. Such dialogues offer a platform not only to discuss strategic priorities to advance local cybersecurity capabilities, but to ascertain how coordinated disclosures can contribute to the degradation of adversary capabilities and thereby enhance the defensive posture of the victim's state.

Supporting international rule of law through best practices

In addition to capacity building, the EU has a critical role to play in the development and dissemination of best practices for attribution on a global scale. It is imperative that the international community establishes clear, consistent, and legally sound standards for attribution. Such shared standards may facilitate joint reporting by Member States and international partners. Merging insights into the operations of North Korean threat actors, Germany and South Korea, for instance, developed comprehensive mitigation guidance.

The EU should support efforts for the creation of a United Nations accountability mechanism specifically for cyber operations. Such a mechanism would provide a platform for the international community to collectively assess and respond to cyber incidents, ensuring that perpetrators are held accountable under international law.

The EU's commitment to supporting international legal frameworks can be further demonstrated by backing the International Criminal Court (ICC) in its investigations into potential cyber war crimes. In June 2024, ICC officials confirmed an ongoing probe into four cyberattacks against Ukraine, investigating them as possible war crimes. The EU should actively support these investigations, which test the application of international humanitarian law in cyberspace. The war context in Ukraine has put international views on how international humanitarian law (IHL) applies to cyber operations to the test. Cyberattacks that target civilian infrastructure, disrupt essential services, or contribute to broader war crimes challenge traditional interpretations of IHL. The EU could support legal and factual assessments that affirm the application of IHL in cyberspace.

Dr Annegret Bendiek is a Senior Fellow at the EU/Europe Research Division and Co-Head of the Research Cluster Cybersecurity and Digital Policy at SWP.
Jakob Bund is an Associate at the EU/Europe Research Division at SWP and Senior Researcher at the European Cyber Conflict Research Initiative (ECCRI).
Mika Kerttunen is the Director of the Cyber Policy Institute.
All authors are members of the research consortium managing the European Repository of Cyber Incidents (EuRepoC, www.eurepoc.eu), a project funded by the German Federal Foreign Office.

SWP
Stiftung Wissenschaft und Politik
German Institute for International and Security Affairs

Ludwigkirchplatz 3—4
10719 Berlin
Telephone +49 30 880 07-0
Fax +49 30 880 07-100
www.swp-berlin.org
swp@swp-berlin.org

ISSN (Print) 1861-1761
ISSN (Online) 2747-5107
DOI: 10.18449/2024C46