STUDY

Assessing Irresponsibility in Cyber Operations

A Guide for Operators and Decision-Makers in Times of Strategic Competition

Sven Herpig

December 01, 2025

Tech analysis and policy ideas for Europe

interface

Table of Contents

Executive Summary

State-backed cyber operations are increasingly shaping the contours of modern strategic competition between rival powers. These operations target critical infrastructures, democratic institutions, and military assets, turning cyberspace into a domain in which states compete below the threshold of armed conflict. Yet, current frameworks, such as the *United Nations norms of responsible state behaviour in cyberspace*, that are meant to define "responsible" conduct, whether political, legal, or operational, often fail to provide the clarity, operational guidance, or enforcement mechanisms needed to prevent irresponsible conduct.

Moreover, the extent to which states with advanced offensive cyber programs pursue compliance with existing international law provisions and treaties remains unclear. This paper argues that existing notions of *responsibility* need to be more concrete and attuned to operational and geopolitical realities to effectively guide state behavior.

To this end, this paper proposes a pragmatic approach: seven "red flags" that signal when state-backed cyber operations cross a threshold into irresponsibility. These include inflicting physical injuries or widespread psychological harm, interfering with political processes, prepositioning for or causing physical disruptions, shaping the future battlefield, and losing operational control.

By formulating thresholds and selectively referencing past incidents for context, this paper offers practical guidance for state-backed cyber operators conducting activities and for the executives of affected states responding to them. This guidance helps both sides better navigate the area of conflict below the threshold of armed conflict. This framework not only helps identify irresponsible state behavior in the cyber domain but also underscores the strategic necessity of firm responses to operations that defy the bounds of such conduct.

Responsible Cyber Behavior: A Difficult Distinction

Geopolitical tensions have been on the rise in the last few years, marking the beginning of a new era of strategic competition between states. In this environment, states and their proxies employ a spectrum of instruments to contest, influence, and secure strategic advantage over their rivals. Among these tools are state-backed cyber operations. For the purpose of this analysis, *state-backed cyber operations* are defined as any action ranging from state encouraged to state integrated on the *Spectrum of State Responsibility.*[1]

Several factors reflect the increasing role of cyber operations in interstate conflicts, especially below the threshold of armed conflict[2] —in the "contested arena somewhere between routine statecraft and open warfare—the *gray zone*."[3] Publicly available data reflect a substantial increase in cyber operations carried out by states in recent years,[4] chief among them the People's Republic of China.[5] Concurrently, government officials from the US and the UK have publicly said that they are going to increasingly engage in offensive cyber operations.[6] Several Western governments have also begun publishing more information about how they are conducting cyber operations,[7] and states are building a multistakeholder network for offensive cyber operations. While the UK only recently launched its UK Cyber Effects Network, the People's Republic of China has been developing its robust ecosystem for over a decade.[8] Finally, both theory and empirical evidence suggest that cyber operations are rather seen as "pressure relief," which are limited actions intended to defuse tensions without altering the underlying dispute; rather than serving as means of escalation, they offer states a versatile tool for managing international strategic power competition.[9]

Cyber espionage has become a consistent feature of strategic competition, reflecting the view that it is "merely another form of espionage"[10] and that "espionage for national security purposes is an implicitly, if begrudgingly, accepted state practice."[11] Furthermore, the International Group of Experts advising the Tallinn

[1] Levels 4 (*state encouraged*) to 10 (*state integrated*) on the *Spectrum of State Responsibility* in Jason Healey (2011): Beyond Attribution: Seeking National Responsibility for Cyber Attacks

[2] NATO Cooperative Cyber Defence Centre of Excellence (2025): International armed conflict

[3] Center for Strategic & International Studies (2025): Gray Zone Project

[4] See European Repository of Cyber Incidents (2025): Detailed Table View for incidents between January 1, 2020, and June 30, 2025, with direct or indirect state responsibility indicator.

[5] Between January 1, 2020, and June 30, 2025 the People's Republic of China was the originator of almost 30% (78/268) of cyber incidents with indirect or direct state responsibility outside of the Russia–Ukraine War according to the European Repository of Cyber Incidents (2025): Detailed Table View.

[6] Tom Uren (2025): Why America Needs Its Own Salt Typhoon and Tom Uren (2025): It's Like Signal, but Dumb and Patrick Gray (2025): BONUS INTERVIEW: Senator Mark Warner on Signalgate, Volt Typhoon and tariffs and Kevin Townsend (2025): The UK Brings Cyberwarfare Out of the Closet

[7] See, for example, Royal Danish Defence College (2019): Joint Doctrine for Military Cyberspace Operations and UK Government (2023): The National Cyber Force: Responsible Cyber Power in Practice and ABS News In-depth (2023): How Intelligence agencies catch criminals and U.S. Air Force (2023): AIR FORCE DOCTRINE PUBLICATION 3-12 CYBERSPACE OPERATIONS and Bundeswehr (2025): Cyber- und Informationsraum „Hacking" bei der Bundeswehr Auf unsichtbarer Mission: Wie offensive Cyberoperationen ablaufen

[8] Royal United Services Institute for Defence and Security Studies (2025): UK Cyber Effects Network and Eugenio Benincasa (2025): Defense-Through-Offense Mindset: From a Taiwanese Hacker to the Engine of China's Cybersecurity Industry and Winnona DeSombre Bernsen (2025): Crash (exploit) and burn: Securing the offensive cyber supply chain to counter China in cyberspace and Dakota Cary and Eugenio Benincasa (2024): Capture the (red) flag: An inside look into China's hacking contest ecosystem and Dakota Cary and Kristin Del Rosso (2023): Sleight of hand: How China weaponizes software vulnerabilities

[9] Jasin Healey and Robert Jervis, in The Escalation Inversion and Other Oddities of Situational Cyber Stability, conclude that it may only be a matter of time and geopolitical changes before "the tipping point [is] reached" and cyber operations become escalatory.

[10] David Weissbrodt (2013): Cyber-Conflict, Cyber-Crime, and Cyber-Espionage

[11] Erica Lonergan and Michael Poznansky (2025): A Tale of Two Typhoons: Properly Diagnosing Chinese Cyber Threats

Manual 2.0, a nonlegally binding scholarly work on how international law applies in the cyber context, agreed that customary international law does not prohibit espionage *per se.*[12] As the experts further note, "if an aspect of a cyber espionage operation is unlawful under international law, it renders the cyber espionage unlawful."[13]

However, operations in the cyber domain can be conducted to achieve a variety of objectives beyond espionage and may be connected with or integrated into other forms of activities, such as cyber-enabled information operations[14] : "Whether massive military and commercial espionage campaigns or international extortion rings and theft, the cyber domain offers an outlet for states to advance their interests."[15] Indeed, the offensive cyber operations that states, such as the People's Republic of China or the US, are conducting now extend beyond espionage into more coercive forms,[16] which "are hostile acts, not just part of the cost of doing business."[17]

This distinction between traditional espionage and more expansive offensive cyber operations underscores a crucial challenge: While peacetime political cyber espionage may be largely accepted by affected states, the broader spectrum of cyber activities, particularly those that cause disruption or coercion, raises a set of complex questions, such as how to time and calibrate responses. Many current operations, such as the Volt Typhoon campaign against US critical infrastructures,[18] are more aggressive than political espionage while still staying below the threshold of armed conflict. However, states may still be unwilling, or even not permitted under international law, to use coercive countermeasures in response.[19]

This tension is especially apparent in ongoing international efforts to define responsible state behavior in cyberspace. States have been debating and advancing the framework for more than 20 years.[20]

[12] Michael N. Schmitt (2017): Tallinn Manual 2.0 on the International Law Applicable to Cyber Operations and NATO Cooperative Cyber Defence Centre of Excellence (2025): Peacetime cyber espionage

[13] Michael N. Schmitt (2017): Tallinn Manual 2.0 on the International Law Applicable to Cyber Operations

[14] See, for example, Chris Kremidas-Courtney (2025): Hybrid storm rising: Russia and China's axis against democracy

[15] Benjamin Jensen and Brandon Valeriano (2019): What Do We Know About Cyber Escalation? Observations From Simulations And Surveys

[16] See Annex 5.1. Examples

[17] Emily Harding, Julia Dickson, and Aosheng Pusztaszeri (2025): A Playbook for Winning the Cyber War

[18] Cybersecurity & Infrastructure Security Agency (2024): PRC State-Sponsored Actors Compromise and Maintain Persistent Access to U.S. Critical Infrastructure

[19] See, for example, Michael N. Schmitt (2017): Tallinn Manual 2.0 on the International Law Applicable to Cyber Operations: "Finally, it must be understood that 'use of force' and 'armed attack' (Rule 71) are standards that serve different normative purposes. The 'use of force' standard is employed to determine whether a State has violated Article 2(4) of the UN Charter and its related customary international law prohibition. By contrast, the notion of 'armed attack' has to do with whether the target State may respond to an act with a use of force without itself violating the prohibition of using force. This distinction is critical in that the mere fact that a use of force has occurred does not alone justify a use of force in response. States facing a use of force not amounting to an armed attack will, in the view of the International Group of Experts, have to resort to other measures if they wish to respond lawfully, such as countermeasures (Rule 20) or actions consistent with the plea of necessity (Rule 26)." See also Henning Lahmann (2020): Unilateral Remedies to Cyber Operations – Self-Defence, Countermeasures, Necessity, and the Question of Attribution

Progress has been made in agreeing upon a consensus framework for responsible state behavior, especially through *voluntary nonbinding* norms at the United Nations level.[21] UN Member States have agreed that international law applies to the "ICT environment,"[22] have published national views on how international law applies to cyber activities since,[23] and have outlined operational considerations, such as specific rules of engagement.[24] **Current norms and legal or operational frameworks therefore serve as intentionally broad principles for responsible state behavior in the cyber domain.**

The most operationally applicable existing norms for irresponsible state behavior during times of strategic competition are the *Norm to Avoid Tampering* and the *Norm Against Commandeering of ICT Devices into Botnets* in the Singapore Norm Package,[25] the United Nations norm (k) to *not harm the information systems of the authorized emergency response teams,*[26] and explanations from *Rule 32 – Peacetime cyber espionage* of the Tallinn Manual 2.0.[27]

Although this breadth grants interpretive leeway to policymakers, from a technical and operational perspective, this preexisting corpus of norms can be prohibitively ambiguous.[28] **What is needed is guidance for policymakers that is concrete, cyber specific, and attuned to operational and geopolitical realities. Otherwise, norms risk remaining abstract principles rather than actionable standards.**[29]

For example, the UN norm against damaging or impairing critical infrastructure states, "a State should not conduct or knowingly support ICT activity contrary to its obligations under international law that intentionally damages critical infrastructure or otherwise impairs the use and operation of critical infrastructure to provide

[20] United Nations Office for Disarmament (2025): Developments in the field of information and telecommunications in the context of international security, and for background, see also Bart Hogeveen (2022): The UN norms of responsible state behaviour in cyberspace – Guidance on implementation for Member States of ASEAN.

[21] United Nations (2015): Group of Governmental Experts on Developments in the Field of Information and Telecommunications in the Context of International Security (A/70/174)

[22] United Nations (2013): Group of Governmental Experts on Developments in the Field of Information and Telecommunications in the Context of International Security (A/68/98)

[23] NATO Cooperative Cyber Defence Centre of Excellence (2025): Applicability of international law and Kubo Mačák, Talita Dias and Ágnes Kasper (2025): Handbook on Developing a National Position on International Law and Cyber Activities – A Practical Guide for States

[24] For an overview, see Appendix 5.2.

[25] Global Commission on the Stability of Cyberspace (2018): Norm Package Singapore

[26] United Nations (2015): Group of Governmental Experts on Developments in the Field of Information and Telecommunications in the Context of International Security (A/70/174)

[27] Michael N. Schmitt (2017): Tallinn Manual 2.0 on the International Law Applicable to Cyber Operations

[28] They require additional guidance and operationalization for effective implementation; see, for example, Bart Hogeveen (2022): The UN norms of responsible state behaviour in cyberspace – Guidance on implementation for Member States of ASEAN and Alexandra Paulus and Christina Rupp (2023): Government's Role in Increasing Software Supply Chain Security: A Toolbox for Policy Makers and Sven Herpig (2024): Vulnerability Disclosure: Guiding Governments from Norm to Action – How to Implement Norm J of the United Nations Norms of Responsible State Behaviour in Cyberspace.

[29] See Louise Marie Hurel (2025): New Ways to Frame Responsible Cyber Behaviour Beyond the UN or the chapter "Cyber Diplomacy Meets Vulnerability Realpolitik" in Sven Herpig (2024): Vulnerability Disclosure: Guiding Governments from Norm to Action – How to Implement Norm J of the United Nations Norms of Responsible State Behaviour in Cyberspace. For additional examples of the relevant frameworks, see the Appendix.

services to the public."[30] While the norm is sound in theory, it is vague in practice when applied to national policies, given that there is no universally agreed-upon definition of critical infrastructure. States consider their designation a national prerogative,[31] which in practice varies significantly across UN Member States, and, in many cases, is not even publicly accessible.[32] Germany, for example, like other European Union Member States, has legally defined a subset of hospitals as a critical infrastructure in the context of cybersecurity.[33] Simply going by this norm and adhering to the black letter of German law would mean that impairing smaller hospitals in Germany, which are not designated as critical infrastructure, may not be irresponsible, although it could lead to injury and death.

The Tallinn Manual 2.0's formulation on violations of sovereignty[34] is another illustrative example.[35] It states that "if an agent of one State [physically present on another State's territory] uses a USB flash drive to introduce malware into cyber infrastructure located in another State, a violation of sovereignty has taken place."[36] At first glance, this establishes a clear red line, allowing the affected state to respond in accordance with international law. However, it offers no guidance on whether the state *should* respond. Suppose, for instance, that the malware introduced by an agent of one state using a USB flash drive is used to exfiltrate the medical records of the target state's leader. Although such an operation is undoubtedly intrusive and politically sensitive, it may not be considered an inherently irresponsible act in the broader context of interstate cyber operations, that necessarily requires a firm response.

Finally, consider the complexity of the US military's public operational framework; planning a cyber operation requires navigating a vast number of documents and pages of guidance. The US Air Force Doctrine on Cyber Operations[37] (around 40

[30] See Norm F of the United Nations Norms of Responsible State Behaviour in Cyberspace in United Nations (2015): Group of Governmental Experts on Developments in the Field of Information and Telecommunications in the Context of International Security (A/70/174).

[31] United Nations (2025): Open-ended working group on security of and in the use of information and communications technologies 2021-2025 (A/AC.292/2025/CRP.1)

[32] Valentin Weber, Maria Pericàs Riera, and Emma Laumann (2023): Mapping the World's Critical Infrastructure Sectors

[33] Bundesministerium der Justiz und für Verbraucherschutz (2025): Verordnung zur Bestimmung Kritischer Infrastrukturen nach dem BSI-Gesetz (BSI-Kritisverordnung - BSI-KritisV) Anhang 5 (zu § 1 Nummer 4 und 5, § 6 Absatz 6 Nummer 1 und 2) Anlagenkategorien und Schwellenwerte im Sektor Gesundheit

[34] NATO Cooperative Cyber Defence Centre of Excellence (2025): Sovereignty

[35] See Rule 4 in Michael N. Schmitt (2017): Tallinn Manual 2.0 on the International Law Applicable to Cyber Operations.

[36] Michael N. Schmitt (2017): Tallinn Manual 2.0 on the International Law Applicable to Cyber Operations states, "6. Based on its internal sovereignty, a State may control access to its territory and to the superjacent national airspace. A State's territory includes the land territory, internal waters, territorial sea (including its bed and subsoil), and archipelagic waters (where applicable). The Experts agreed that a violation of sovereignty occurs whenever one State physically crosses into the territory or national airspace of another State without either its consent or another justification in international law (see Rule 19 on circumstances precluding wrongfulness). As an example, the nonconsensual exercise of enforcement jurisdiction in another State's territory (Rule 11) is a violation of that State's sovereignty.22 In the cyber context, therefore, it is a violation of territorial sovereignty for an organ of a State, or others whose conduct may be attributed to the State, to conduct cyber operations while physically present on another State's territory against that State or entities or persons located there. For example, if an agent of one State uses a USB flash drive to introduce malware into cyber infrastructure located in another State, a violation of sovereignty has taken place."

[37] U.S. Air Force (2023): AIR FORCE DOCTRINE PUBLICATION 3-12 - CYBERSPACE OPERATIONS

pages) apparently needs to be read in conjunction with 100 pages of the US joint doctrine on cyber operations[38] and another almost 100 pages on targeting.[39] If it is part of a NATO operation, there will be another 50 pages on the allied joint doctrine on cyber operations[40] to be observed, bringing it to roughly 300 pages of guidance. That is just one branch of the US military. In addition to the US Air Force, US Army, US Navy, and US Marine Corps, which carry out cyber operations in conjunction with US Cyber Command, cyber operations in the intelligence sector are conducted by the National Security Agency and the Central Intelligence Agency, each with its own set of doctrines. With reference to such complex frameworks, government officials have asked whether "lawyers can lose wars"[41] and, even more specifically, whether "lawyers [can] lose wars by stifling cyber capabilities".[42]

This tension between extensive legal and operational frameworks and the practical realities of strategic competition highlights a core challenge: Even with guidance in place, states often navigate a landscape in which norms are difficult to interpret or enforce, creating gaps between codified responsibility and actual behavior.

In practice, *Realpolitik* means that cyber operations deemed "not responsible" under broad norms, or even "wrongful" under international law, are often carried out without consequences, even though states have responses[43] to such operations at their disposal. Decision makers on both the conducting and affected sides may view existing norms and frameworks as too vague or impractical to guide action. This results in uncertainty over which incursions merit a response and increases the risk of unintended overreach.

To address this gap, this paper—drawing on insights from researchers, practitioners, and existing legal, normative, and operational frameworks—proposes a set of criteria, or red flags. **As opposed to the red lines of international law,[44] these red flags help identify which cyber operations, specifically those conducted below the threshold of armed conflict, should be considered irresponsible by the "affected state," prompting a firm response[45] and, consequently, avoided by the "conducting state" to reduce the risk of unintended overreach.**

[38] Joint Chiefs of Staff (2018): Joint Publication 3-12 - Cyberspace Operations

[39] U.S. Air Force (2021): AIR FORCE DOCTRINE PUBLICATION 3-60 - TARGETING

[40] UK Ministry of Defence (2020): Allied Joint Publication-3.20 - Allied Joint Doctrine for Cyberspace Operations

[41] Stewart Baker (2011): Denial of Service

[42] Anonymous European Intelligence Official (2024): Can lawyers lose wars by stifling cyber capabilities?

[43] Sven Herpig (2021): Die Beantwortung von staatlich verantworteten Cyberoperationen and Talita Dias (2024): Countermeasures in international law and their role in cyberspace

[44] Denise Tennant, Louis Nolan and Deanna House (2024): CYBER RED LINES—Government Responses to Cyberattacks on Critical Infrastructure

[45] Fully aware that those responses may not deter current or future actions of the adversary, especially in a geopolitical context where the adversary is very powerful, see, for example, Carley Welch (2024): NSA's China specialist: US at a loss to deter Chinese hackers and Erica Lonergan and Michael Poznansky (2025): A Tale of Two Typhoons: Properly Diagnosing Chinese Cyber Threats and Tom Uren (2024): FCC to Demand Telcos Improve Security.

The purpose of these red flags is to clarify permissible conduct and its consequences while ensuring that compliance is achievable and deviations can be readily addressed. **The overall damage resulting from cyber operations can be mitigated by establishing these thresholds.**

Red Flags for Cyber Operations

Cyber operators should strictly adhere to a broad set of normative and operational frameworks when conducting operations outside the context of armed conflict. **These operators rarely seek to fully align with normative standards of responsible behavior, but they do carefully weigh the political costs and risks of retaliation, often shaping their actions around this calculus.**

No framework can eliminate the inherent ambiguity of cyber operations, such as those arising from unpredictable effects, normative gray zones, and plausible deniability, but the red flags proposed here are designed to reduce them in practice. They cover the full spectrum of potential impacts, ranging from counterforce effects against military facilities, such as delaying operations, to countervalue effects targeting civilian populations with the aim of destabilizing society.[46] The red flags are not a definitive checklist but a set of guardrails that help cyber operators and decision makers identify actions that, individually or linked as campaigns,[47] are likely to—and, in the author's view, should—trigger a strong adversarial response by the affected state. **Simply put, red flags serve as yardsticks for identifying when an operation crosses the line into irresponsibility.**[48]

For operators, the red flags sharpen awareness that certain effects, if pursued, are almost certain to cross a line and risk escalation; for decision makers of the state conducting the operation, they provide guidance to assess planned activities and decide on the rules of engagement. For the decision makers of an affected state, the red flags provide a structured basis for determining when adversary behavior has breached acceptable bounds and requires a response. In this way, the framework does not seek to resolve all uncertainty but instead narrows discretion and curbs plausible deniability, which is the ability to maintain credible doubt about a state's involvement. It also ensures that those engaged in offensive activities are better able to recognize when they are operating irresponsibly, with the ultimate goal of decreasing the overall damage resulting from cyber operations.

[46] Erica Lonergan and Michael Poznansky (2025): A Tale of Two Typhoons: Properly Diagnosing Chinese Cyber Threats

[47] As Harding, Dickson, and Pusztaszeri put it, "pinpricks add up to an intolerable chorus of pain," Emily Harding, Julia Dickson, and Aosheng Pusztaszeri (2025): A Playbook for Winning the Cyber War.

[48] An operation that is not assessed as irresponsible based on the red flags is not automatically responsible.

Each red flag carries equal weight within this framework and is therefore presented without implying any hierarchy or sequence of importance.

Red Flag 1: Causing Physical Harm, Injury, or Death

A state has a fundamental duty to protect those within its jurisdiction from harm.[49] In the context of cyber operations, the clearest red flag arises when an operation results in physical harm, injury, or loss of life, regardless of the operator's intent.

In such cases, irresponsibility stems from effects, not intentions. Whether the harm is immediate and direct, such as a hospital system failure leading to critical care delays, or arises indirectly through second- or third-order effects, such as a prolonged infrastructure outage triggering medical complications or fatalities, any resulting injury or death should be treated as a red flag. These effects may not always be easy to trace and evaluate. Fatalities caused by heating outages during a winter blackout or by a lack of cooling during a tropical heatwave might unfold over days and involve multiple interdependencies. However, this complexity does not relieve cyber operators or planners of their responsibility and, ultimately, their accountability.

Operators must assess downstream risks proactively. Likewise, decision makers responding to cyber operations need to trace second- and third-order effects to make informed judgments.

While the presence of injury or loss of life is what ultimately defines the red flag in this case, **context still plays a critical role in anticipating the likelihood and severity of these effects.** A temporary power outage, for example, may be manageable in Rome on a mild autumn afternoon but far more dangerous in Kyiv during sub-zero winter temperatures or in Manila amid a heatwave and widespread reliance on cooling systems. In this sense, operators must not only consider what is being targeted but also **when and where**—and what that means for the safety of civilian populations.

Operators and decision makers need to know that such operations are not only red flags but may, under certain conditions, constitute a use of force or even an armed attack under international law.[50]

[49] For reference, see the "Cyber Harm Framework," Global Cyber Security Capacity Centre (2025): The Cyber Harm Framework.
[50] Michael N. Schmitt (2017): Tallinn Manual 2.0 on the International Law Applicable to Cyber Operations

Several known state-backed operations, such as the attempted poisoning of Israel's water system,[51] BlackEnergy's disruption of Ukraine's winter power supply,[52] and the Industroyer attack on Ukrainian substations during winter,[53] approached this threshold. However, these operations took place in conflict situations (short of armed conflict), and there is still no publicly confirmed case that has demonstrably crossed the threshold of physical harm, injury, or death during peacetime operations.

Red Flag 2: Inflicting Widespread Psychological Harm

While the deliberate infliction of physical harm through cyber means is an unmistakable threshold violation, the intentional imposition of psychological harm, especially when directed at civilians or broader segments of society, likewise constitutes a serious red flag. Such actions amount to deliberate interference in a state's domestic stability, creating direct political pressure on decision makers to act, since passivity risks signaling weakness or acceptance, inviting further aggression. To determine when psychological harm crosses the threshold into a red flag, this assessment can draw on two interrelated criteria: the scope and visibility of the effects and the operational logic behind the operation. A red flag may be triggered if one or both are sufficiently evident.

Regarding **scope and visibility**, operations must substantially affect segments of the population, either through direct interaction (e.g., halting public transport and disabling critical services), through highly publicized data exposure (e.g., health records), or through data breaches (e.g., personal files). Effects that remain covert or limited to a small number of individuals generally fall outside this category, as the extent of the psychological impact is comparatively minimal. The red flag depends on whether the operation is likely to be noticed and resonates with the affected population. Effects are inherently context dependent: The same technical action might evoke fear depending on media coverage, societal sensitivities, and public awareness.

Regarding **operational logic**, the operation instils fear rather than achieving any other clearly defined operational or strategic objective. Accidental or tangential effects do not automatically trigger the red flag; what matters is whether the operation is designed, or carried out with reckless disregard to cause psychological impact, for example, by leveraging cyber-enabled information operations.

[51] TOI Staff and Agencies (2020): Iran cyberattack on Israel's water supply could have sickened hundreds – report
[52] SANS ICS and E-ISAC (2016): Analysis of the Cyber Attack on the Ukrainian Power Grid – Defense Use Case
[53] ESET Research (2022): Industroyer2: Industroyer reloaded

Operations designed to maximize visibility and psychological impact clearly cross the threshold into irresponsibility.

Several known state-backed operations have triggered this red flag. They include the disruption of the Iranian railway system[54] and petrol station infrastructure,[55] as well as the data breach of SingHealth.[56]

Red Flag 3: Intervening in Domestic Political Processes

Domestic political processes, such as elections, leadership transitions, or succession mechanisms, are foundational to the internal legitimacy and external recognition of any state. While not every cyber operation targeting political stakeholders qualifies as a red flag, as they can be acts of political espionage, certain forms of interference cross a critical threshold. Most cyber intrusions aimed at political actors constitute political espionage, which, while unfriendly, falls within the longstanding practice of interstate behavior. Operations that seek to directly alter, disrupt, or delegitimize a state's political structure or leadership pose a grave threat to national sovereignty. Failing to counter them risks normalizing a dangerous precedent, inviting more brazen campaigns, emboldening foreign actors, and eroding both leadership confidence and public trust in the state's ability to safeguard its political integrity.

First, interference with the core mechanisms of a country's leadership selection **challenges the very essence of statehood.**[57] Undermining this system is tantamount to challenging the state's identity and internal cohesion. A salient example is the manipulation of election infrastructure in an electoral democracy. Through cyber-enabled vote tampering, threat actors may alter the composition of parliament or enable a government that proceeds illegitimately to dismantle key constitutional norms.

Second, the "public" nature of many such operations elevates their strategic impact and the need for deliberate counteraction. Operations designed to manipulate or discredit key political figures, such as so-called "hack and leak" operations,[58] are often intended to sway public opinion, fracture ruling coalitions, or influence

[54] JD Work (2021): Balancing on the rail – considering responsibility and restraint in the July 2021 Iran railways incident

[55] Hamid Kashfi (2024): The Curious Case of Predatory Sparrow – Reconstructing the Attack from a 4th Party Collector's Point of View

[56] Government of Singapore (2019): Government's response to the report of the COI into the cyber attack on SingHealth

[57] Bundeszentrale für Politische Bildung (2025): Drei-Elemente-Lehre

[58] Sven Herpig, Julia Schuetze and Jonathan Jones (2018): Securing Democracy in Cyberspace – An Approach to Protecting Data-Driven Elections

succession outcomes. For example, in a tribal governance system, single-party state, or theocratic regime, the release of compromising information about potential successors may derail leadership transitions and generate internal instability. Even if domestic responses are partially effective, **the mere visibility of such interference can signal vulnerability**.

Operators and decision makers need to know that such operations are not only a red flag but also a potential breach of international law.[59]

Several known state-backed operations have triggered this red flag. They include the operations against the US presidential elections in 2016,[60] the French elections in 2017,[61] the Ukrainian Central Election Commission in 2014,[62] and the UK Electoral Commission in 2021/2022.[63]

Red Flag 4: Triggering Physical Disruption or Destruction

Cyber operations that trigger physical destruction or major physical disruption are a red flag for several reasons, each of which independently raises the stakes. Together, they create a clear threshold of irresponsibility.

First, **physical destruction is not easily reversible**. Damaged or destroyed infrastructure, whether turbines, substations, or pipelines, cannot be rebooted or patched like (most) software. Repairs are costly and may take weeks or months, and supply chain limitations can delay restoration even further. The resulting outages can have unpredictable ripple effects on connected systems and entire sectors of society.

Second, such operations **raise the risk of unintended human harm**. Explosions, fires, flooding, or cascading mechanical failures can endanger the lives of bystanders, workers, or first responders. Even if no casualties or sustained environmental impact occur, the potential for harm is often enough to escalate political tensions or prompt preemptive defensive measures.

Third, the **visibility and physicality** of such effects make them politically untenable to ignore, and the psychological effects on society may be severe. Unlike covert

[59] Michael N. Schmitt (2017): Tallinn Manual 2.0 on the International Law Applicable to Cyber Operations

[60] U.S. Department of Justice (2018): Grand Jury Indicts 12 Russian Intelligence Officers for Hacking Offenses Related to the 2016 Election

[61] Clea Caulcutt (2025): Notorious Russian hackers behind 2017 'Macron leaks,' France says

[62] Mark Clayton (2014): Ukraine election narrowly avoided 'wanton destruction' from hackers

[63] UK Government (2024): UK holds China state-affiliated organisations and individuals responsible for malicious cyber activity

intrusions or subtle disruptions, a visible explosion or industrial malfunction is difficult to downplay or deny to citizens and businesses. It can generate public fear, media attention, and immediate pressure on decision makers to respond, whether diplomatically, economically, or even militarily. The psychological and political weight of such events amplifies their significance well beyond the actual damage done.

In sum, the **clear, calculable effects** of such operations, such as public visibility, political pressure, and irreversible physical damage, are sufficient to mark them as irresponsible. **Unpredictable downstream effects**, such as economic disruption, injuries, or second- and third-order failures, further underscore their irresponsibility.

Operators and decision makers need to know that such operations are not only a red flag but also a potential use of force or, depending on other factors, even an armed attack under international law.[64]

Several known state-backed operations have triggered this red flag. They include BlackEnergy's disruption of Ukraine's power supply,[65] the interference with a German steel mill,[66] and Operation Olympic Games, which disrupted nuclear enrichment facilities in Iran.[67]

Red Flag 5: Prepositioning for Civilian Disruption

Some cyber operations targeting civilian critical infrastructure providers may be tolerated under limited conditions, particularly those linked to traditional political espionage. However, the intentional placement of malware or persistent access within such systems for the purpose of future disruption marks a dangerous escalation. Realistically, distinguishing whether disruption is a goal is often impossible, especially since that goal can shift quickly from nondisruptive to disruptive during an active operation. Therefore, this behavior constitutes a red flag that warrants a timely and decisive response by the affected state.

First, the **nature of the targeted infrastructure** provides an initial basis for assessing the severity of the operation. Cyber intrusions into the networks of, for example, water utilities indicate a strategic focus on infrastructures providing essential

[64] Michael N. Schmitt (2017): Tallinn Manual 2.0 on the International Law Applicable to Cyber Operations
[65] SANS ICS and E-ISAC (2016): Analysis of the Cyber Attack on the Ukrainian Power Grid – Defense Use Case
[66] Robert M. Lee, Michael J. Assante and Tim Conway (2014): ICS CP/PE (Cyber-to-Physical or Process Effects) case study paper – German Steel Mill Cyber Attack
[67] Sven Herpig (2016): Anti-War and the Cyber Triangle – Strategic Implications of Cyber Operations and Cyber Security for the State

services to society. Moreover, cyber operations targeting systems such as electrical grid networks are inherently irresponsible,[68] as their compromise would almost certainly have immediate and cascading effects on civilian life and national stability. The presence of foreign access, even if inactive, raises legitimate concerns that the operator is preparing for a scenario in which these systems could be disabled to maximize chaos or weaken societal resilience.

Second, the specific capabilities and technical behavior of the tools deployed are crucial indicators of **operational intent**. While cyber actors may initially pursue espionage objectives, technical elements, such as a modular design that allows adding data-wiping functions, mass activation command-and-control structures, and malware tailored for industrial systems, clearly signal disruptive intent. For instance, the discovery of sophisticated malware with such functions in a hospital network or a power grid operator signals a willingness to endanger civilian populations and public safety for strategic advantage. Such prepositioning is not proportional and risks wide-ranging second- and third-order effects. Moreover, cyber operations that preposition capabilities to disable, degrade, or deny the functionality of civilian critical infrastructures, such as the power grid or telecommunication networks, may be **designed for use in the event of armed conflict** or preparation for a major geopolitical confrontation. As a result, these operations may raise multiple red flags, including signaling preparations for the battleground (Red Flag 6).

Finally, prolonged unauthorized access to civilian critical infrastructures introduces additional risks beyond the intentions of the original operator. The longer such access remains unmitigated, the greater the likelihood that **other malicious actors, whether state sponsored, criminal, or opportunistic, may discover and exploit these footholds**. Moreover, even if the original actor intended to use such access only in extreme scenarios, the mere existence of these capabilities creates ambiguity and tension in peacetime international relations.

Operators and decision makers need to know that such operations are not only a red flag but also a potential violation of international law, amounting to a threat of use of force or even an armed attack.[69]

Several known state-backed operations have triggered this red flag. They include all deployments of Industroyer malware,[70] Havex malware,[71] and HatMan malware,[72] US

[68] Valentin Weber (2023): How German (Cyber)diplomacy Can Strengthen Norms in a World of Rule-Breakers
[69] Michael N. Schmitt (2017): Tallinn Manual 2.0 on the International Law Applicable to Cyber Operations
[70] Anton Cherepanov (2017): WIN32/INDUSTROYER – A new threat for industrial control systems
[71] Daavid (2014): Havex Hunts For ICS/SCADA Systems

cyber operations against the Russian power grid in 2018,[73] and Chinese cyber operations against US critical infrastructure.[74]

Red Flag 6: Preparing the Military Battleground

Even during periods of intense strategic competition, states share one fundamental goal: avoiding open hostilities and armed conflict. Cyber operations that actively shape the conditions for future military confrontations,[75] commonly referred to as cyberspace operational preparation of the environment,[76] cross a critical line when they enable further offensive actions; such operations warrant a firm response from affected states, especially as there are "[enormous] incentives to start[ing] any military conflict with a significant attack in cyberspace[...]."

First, the nature of the targeted assets **where the effects are taking place** is pivotal in determining whether the operational behavior should be considered a red flag. When cyber operations focus on military assets,[77] such as air defense systems, naval vessels, or nuclear command and control infrastructures, the threshold is unequivocally breached. The rationale is straightforward: Cyber operators who cause or preposition effects on these military targets are clearly seeking to shape the battleground and gain operational advantage in preparation for armed conflict. These operations must be distinguished from intrusions targeting defense contractors, military academies, individual armed forces personnel's smartphones,[78] or research institutions. While still significant, such intrusions usually serve broader objectives, such as political espionage or technological intelligence gathering, rather than direct preparation for conflict.

Second, the **nature and intended effects** of the cyber activities themselves provide crucial indicators reveal whether they are designed for intelligence collection or for shaping conditions for conflict. Although operators with system access can often switch between different operational goals, ranging from data exfiltration to data degradation, certain behaviors or hardcoded functionalities within deployed malware offer clearer evidence of aggressive intent. For instance, malware

[72] U.S. Department of Homeland Security (2018): Analysis Report – MAR-17-352-01 HatMan – Safety System Targeted Malware (Update A)

[73] David E. Sanger and Nicole Perlroth (2019): U.S. Escalates Online Attacks on Russia's Power Grid

[74] U.S. Cybersecurity & Infrastructure Security Agency (2024): PRC State-Sponsored Actors Compromise and Maintain Persistent Access to U.S. Critical Infrastructure

[75] See, for example, Michael P. Fischerkeller, Emily O. Goldman, and Richard J. Harknett (2025): Setting the Stage: Cyber Contingency Campaigning

[76] U.S. Air Force (2023): AIR FORCE DOCTRINE PUBLICATION 3-12 – CYBERSPACE OPERATIONS

[77] Lukasz Olejnik (2021): The Dire Possibility of Cyberattacks on Weapons Systems

[78] Matthias Schulze (2025): Cyber-Operationen in den Kriegen in der Ukraine und im Gazastreifen: Noch keine Revolution der Kriegsführung

implanted within a defense contractor's network that contains a hardcoded capability to wipe the entire network, or features sufficient modularity to easily add such a function, strongly signals preparations to shape the operational environment in a hostile manner. Likewise, if such access is leveraged to manipulate critical software updates, such as those governing fighter jets, with triggers designed to disable systems when specific conditions are met (e.g., crossing a designated latitude), this clearly indicates cyberspace operational preparation for conflict.

Operators and decision makers need to know that such operations are not only a red flag but also a potential violation of international law, amounting to a threat of use of force or even an armed attack.[79]

Several known state-backed operations have triggered this red flag. This includes the disabling of Syrian air defense and radar systems during Operation Orchard[80] and Volt Typhoon's operations targeting US infrastructure in Guam.[81]

Red Flag 7: Lacking or Losing Operational Control

As states increasingly employ cyber operations to advance their strategic interests, maintaining effective operational control is essential. When that control is deliberately abandoned or unintentionally lost, the risks grow significantly. Such operations often produce effects that exceed what is necessary to achieve their objectives, resulting in disproportionate and avoidable harm. In these cases, the absence or breakdown of control warrants a response from those affected.

Organizational lack of control can take several forms, all of which raise the risk of irresponsible outcomes. One example is when government agencies are given excessive autonomy to plan and carry out cyber operations without clear oversight or accountability. In such cases, objectives can drift, and operations may extend beyond intended or lawful boundaries. Another concern is the outsourcing of operations to nonstate actors, such as private companies, criminal networks, or universities, without sufficient supervision. Delegating this level of authority while lacking control mechanisms increases the risk of escalation, misuse, or unintended consequences. In some instances, governments publicly signal strategic goals and

[79] Michael N. Schmitt (2017): Tallinn Manual 2.0 on the International Law Applicable to Cyber Operations

[80] ETH Zürich (2017): The use of cybertools in an internationalized civil war context: Cyber activities in the Syrian conflict

[81] U.S. Cybersecurity and Infrastructure Security Agency (2024): PRC State-Sponsored Actors Compromise and Maintain Persistent Access to U.S. Critical Infrastructure and Antonio Pequeño IV (2023): Microsoft Says China Hackers Targeted 'Critical' U.S. Infrastructure In Guam – A Key Military Asset In The Pacific

allow external actors, ranging from individuals and collectives ("patriotic hackers") to private contractors, to conduct operations on their own initiative and share the outcomes afterward. This fragmented approach undermines centralized control and governance, as third parties may not observe the same standards of discipline, restraint, or accountability expected of official state operators. Moreover, outsourcing can amplify risks beyond individual operations. Competition among contractors may drive a surge in overall activity, and the sheer volume of operations can itself be highly irresponsible, affecting how aggressive the sponsoring state is perceived to be.

Technical loss of control arises when operational decisions compromise containment, oversight, or predictability. A prominent example is the use of self-propagating malware, such as worms, without effective safeguards, such as reliable kill switches. Without these controls, malware can spread far beyond its intended targets, disproportionately increasing damage beyond operational necessity. Another risk stems from an insufficient understanding of the tools being deployed. Using externally sourced malware or code developed with the help of generative AI ("AI vibe coding"), without thorough review, can introduce unpredictable behaviors. In such cases, operators may not fully grasp the malware's behavior, its persistence, or how easily it could be exploited by others. Finally, failing to adequately remove access points, such as unsecured web shells, vulnerable implanted malware, or easily exploitable credentials, intentionally transferring them among threat actor clusters, or leaving behind advanced malware and exploits that can be repurposed can lead to proliferation. These tools may be discovered and reused by third parties, including criminal groups, causing extensive damage.

Sometimes, poor control during the planning phase overlaps with loss of control during execution. This is particularly evident when operators engage in large-scale, indiscriminate exploitation of vulnerable devices—well beyond what is operationally necessary. Common examples include the creation of botnets, Operational Relay Box (ORB)[82] networks, or compromised infrastructure as part of supply chain intrusions. These activities often result in the compromise of thousands of systems with no realistic prospect of maintaining oversight or containment. The result is a disproportionate level of collateral damage, including widespread economic costs, such as incident response and remediation, and a heightened risk of third-party hijacking. Such behavior reflects poor operational discipline and violates the principle of minimizing harm beyond the intended scope of the operation.

[82] Dutch Ministry of Justice and Security (2025): ORB networks and their impact on digital security in the Netherlands

Several known state-backed operations have triggered this red flag. They include operations conducted by Chinese companies i-Soon[83] and Sichuan Silence Information Technology[84] and the University of Electronic Science and Technology of China,[85] WannaCry[86] and NotPetya[87] worms causing disruptions worldwide, the Stuxnet worm[88] evading safeguards and causing incident response costs far beyond the operational target, and weakly secured web shells left behind by Hafnium[89] operators and exploited by third parties.

Responding to Red Flag Operations

The framework outlined in this paper enables both offensive and defensive stakeholders to better anticipate and assess the strategic consequences of cyber operations. For operators, recognizing red flags during the planning and preparation phase is essential. Ensuring a shared perception of these red flags within the team should inform internal escalation and oversight processes, particularly when potential national or international repercussions could outweigh the perceived operational benefit. For the affected states, the question is not whether to respond when red flags are raised but how to do so proportionally and when.[90] As stated by a number of mainly Western states in 2019, "there must be consequences for bad behavior in cyberspace."[91] Otherwise, "the reputation [...] may suffer if an adversary appears to cross a red line without generating an appropriate or implied response."[92]

Where technical attribution reaches a sufficient level of confidence,[93] red flag operations warrant a firm response. Initial assessments should be followed by reevaluations, as cyber operations can swiftly evolve from not irresponsible to irresponsible. For example, an operator might pivot from leveraging access to a system for espionage purposes to using it to disrupt civilian critical infrastructures. Additionally, while individual activities by the same operator may not cross the

[83] Bundesamt für Verfassungsschutz (2024): BfV CYBER INSIGHT – Die i-Soon-Leaks: Industrialisierung von Cyberspionage

[84] Ross McKerchar and Andrew Brandt (2024): Pacific Rim timeline: Information for defenders from a braid of interlocking attack campaigns

[85] Ross McKerchar and Andrew Brandt (2024): Pacific Rim timeline: Information for defenders from a braid of interlocking attack campaigns

[86] Counter Threat Unit Research Team (2017): WCry Ransomware Analysis

[87] Karan Sood and Shaun Hurley (2017): NotPetya Technical Analysis - A Triple Threat: File Encryption, MFT Encryption, Credential Theft

[88] Nicolas Falliere, Liam O Murchu, and Eric Chien (2011): W32.Stuxnet Dossier

[89] Patrick Howell O'Neill (2021): How China's attack on Microsoft escalated into a "reckless" hacking spree

[90] Benjamin Jensen and Brandon Valeriano (2019): What Do We Know About Cyber Escalation? Observations From Simulations And Surveys and Sven Herpig (2021): Die Beantwortung von staatlich verantworteten Cyberoperationen

[91] U.S. Department of Justice (2019): Joint Statement on Advancing Responsible State Behavior in Cyberspace

[92] Denise Tennant, Louis Nolan and Deanna House (2024): CYBER RED LINES – Government Responses to Cyberattacks on Critical Infrastructure quoting Dan Altman and Kathleen E. Powers (2022): When Redlines Fail – The Promise and Peril of Public Threats.

[93] Timo Steffens (2021): Attribution of Advanced Persistent Threats: How to Identify the Actors Behind Cyber-Espionage

threshold into red flags, a combination of cyber operations might.

Given the potential severity of effects, a proportionate response may extend beyond the cyber domain. States have access to a broad spectrum of instruments under national security policy and international relations and tend to "proportionally respond to a threat to maximize their position short of escalation."[94] This response toolbox includes the following:

- **Cyber defense and operational countermeasures** from threat hunting and system hardening[95] to counter-cyber operations aimed at stopping the adversary's operation and disrupting adversary infrastructure
- **Intelligence missions**, including covert collection and information operations
- **Diplomatic measures**, such as public statements in international fora, summoning foreign ambassadors, recalling one's own diplomatic personnel, delivering formal démarches, or ending diplomatic relations
- **Public attributions** through technical reporting or official public political attribution[96]
- **Criminal justice actions**, including domestic indictments or the issuance of international arrest warrants
- **Sanctions regimes**, such as targeted financial measures (individual and institutional listings) or sector-specific trade restrictions
- **Military posturing or operations**, ranging from increased readiness to strategic troop deployments or, in extreme cases, the use of force in accordance with international law

The appropriate choice or combination of responses depends on the red flags triggered and the overall resulting harm. Contextual factors, including the status of diplomatic relations, economic interdependence, geopolitical priorities, and the balance of power, also influence the response. As such, **responses to red flag operations may not be immediate**; strategic patience is sometimes required to identify the right moment for a credible and effective countermeasure. Responses should be timely to remain effective.

While deliberate nonresponse may, in rare cases, be a calculated decision, it should not become the norm because **inaction risks normalizing irresponsible state behavior** and lowering the threshold for unacceptable conduct.

On the other hand, **responsive action may also be taken before a red-flagged cyber operation reaches its full effect**. For instance, if reconnaissance reveals systems containing malware designed to disrupt industrial control systems (Red Flag 5) or

[94] Benjamin Jensen and Brandon Valeriano (2019): What Do We Know About Cyber Escalation? Observations From Simulations And Surveys
[95] Erica Lonergan and Michael Poznansky (2025): A Tale of Two Typhoons: Properly Diagnosing Chinese Cyber Threats
[96] Christina Rupp and Alexandra Paulus (2023): Official Public Political Attribution of Cyber Operations – State of Play and Policy Options

inadequately controlled, wormable malware (Red Flag 7), preemptively degrading these systems through counter-cyber operations may be permissible.

By identifying red flags, **this analysis seeks to support more consistent national and international responses, particularly from actors hesitant to characterize certain operations as irresponsible.** A clearer understanding of where strategic thresholds lie can help prevent inadvertent conflict, strengthen accountability, and uphold a rules-based international order in the cyber domain.

This paper is intended as a contribution to the current broader debate on operational cyber responsibility[97] by proposing criteria for irresponsibility. Such guidelines must meet at least two of three criteria—*specific*, *binding*, and *global*[98] —to provide an effective framework for cyber stability. The red flags presented here are intended to be specific and global and may become effectively politically binding through consistent state practices in responding to violations.

As more states expand their cyber forces and adopt increasingly assertive operational postures, the need for clear guardrails becomes ever more pressing. With governments also developing greater capacity for offensive cyber operations outside armed conflict, there is a risk of institutionalizing destabilizing patterns and increasing the potential for broader harm. The guardrails help mitigate risk by defining the boundary between operational advantage and irresponsible conduct.

Appendices

Examples

The following table shows sample operations that raise red flags, including explanations of why they did.

Red Flag	Operation	Explanation
1: Causing Physical Harm, Injury, or Death	None	No operation has been publicly and directly linked to this red flag as of this writing.
2: Inflicting Widespread	Disruption of the Iranian railway	It publicly disrupted a core public service, displayed provocative messages on station boards telling

[97] Louise Marie Hurel (2025): New Ways to Frame Responsible Cyber Behaviour Beyond the UN and Benjamin Jensen and Brandon Valeriano (2019): What Do We Know About Cyber Escalation? Observations From Simulations And Surveys

[98] Jason Healey and Robert Jervis (2020): The Escalation Inversion and Other Oddities of Situational Cyber Stability

Psychological Harm	system	passengers to call the Supreme Leader's office, and created nationwide confusion and anxiety.
	Disruption of Iranian petrol station infrastructure[99]	It was a broad disruption of public infrastructure (petrol stations), designed to create societal distress and put political pressure on decision makers by directly affecting citizens' daily lives.
	Data breach of SingHealth	Exfiltrating the sensitive health records of 1.5 million citizens is an act designed to create widespread public anxiety and erode trust in the state's ability to protect its people.
3: Intervening in Domestic Political Processes	Hack and leak operation against the US presidential elections	It deliberately undermined the integrity of US democratic institutions to influence the outcome of the 2016 election by publishing politically sensitive documents to influence the public.
	Hack and leak operation against the French elections	It aimed at compromising Emmanuel Macron's 2017 presidential campaign and publishing sensitive internal documents to influence the election outcome.
	Interference with the Ukrainian Central Election Commission	Infiltrating Ukraine's Central Election Commission, attempting to display false election results, and delaying the vote tally all aimed at undermining the legitimacy of the electoral process itself.
	Breach of the UK Electoral Commission	It involved targeting the UK's Electoral Commission and parliamentary figures to undermine the integrity of democratic institutions and influence political stability.
4: Triggering Physical Disruption or Destruction	Disruption of Ukrainian power supply	It deliberately caused major physical disruption to Ukraine's electrical grid by remotely manipulating Supervisory Control and Data Acquisition systems, resulting in widespread power outages affecting roughly 225,000 customers.
	Disruption of a German steel mill	It led to the improper shutdown of a blast furnace and caused significant physical damage to the German steel mill.
	Disruption of Iranian nuclear enrichment facilities	It caused physical damage to nuclear centrifuges and operational disruption of Iran's nuclear enrichment process.
5: Prepositioning for Civilian Disruption	Deployment of Industroyer malware	The malware was specifically designed to disrupt industrial control systems, particularly targeting electrical substations.
	Deployment of the Havex malware	The malware exfiltrates information for the purpose of taking control of and disrupting industrial control systems.
	Deployments of the HatMan malware	The malware disables safety controls to cause physical harm or production disruption.
	Deployment of disruptive malware in the Russian power	Malware was deployed in parts of the Russian electrical power grid with the intent to disrupt it, if certain conditions in the context of election interference are met.

	grid	
	Compromising a range of civilian critical infrastructures in the US	Threat actors compromised systems in infrastructure, including the energy, water, and wastewater sectors, in which the only operational goal can be subsequent disruption.
6: Preparing the Military Battleground	Disabling of Syrian air defense and radar systems	It neutralized Syrian radar systems, thereby facilitating a covert airstrike on a suspected nuclear reactor site.
	Operation compromising US infrastructure in Guam	It infiltrated key components of the US military’s logistical and operational framework in the Pacific.
7: Lacking or Losing Operational Control	Operations carried out by i-Soon	A nonstate actor conducted operations under broad direction, partially without government awareness or oversight.
	Operations carried out by Sichuan Silence Information Technology	A nonstate actor conducted operations under broad direction, partially without government awareness or oversight.
	Operations carried out by the University of Electronic Science and Technology of China	A nonstate actor conducted operations under broad direction, partially without government awareness or oversight.
	Deployment of the WannaCry malware	The self-propagating malware spread rapidly and indiscriminately across networks worldwide, including systems that were not the intended primary targets.
	Deployment of the NotPetya malware	The self-propagating malware spread rapidly and indiscriminately across networks worldwide, including systems that were not the intended primary targets.
	Deployment of the Stuxnet malware	The self-propagating malware spread rapidly and indiscriminately across networks worldwide, including systems that were not the intended primary targets.
	Proliferation of web shell access used in Hafnium operations	Following initial detection, vulnerable systems were indiscriminately backdoored at scale, enabling follow-on exploitation by a wide range of threat actors, including potential criminal actors, and demonstrating intentional loss of operational control.

[99] Hamid Kashfi (2024): The Curious Case of Predatory Sparrow – Reconstructing the Attack from a 4th party collector’s point of view

Frameworks

Below is a non-exhaustive list of existing operational, legal, and normative frameworks for cyber operations that have been considered when formulating the red flags for cyber operations.

International Humanitarian Law	*Overview by the International Committee of the Red Cross*	Since 1864	Armed Conflict	Legal
International Cyber Law in Practice: Interactive Toolkit	NATO Cooperative Cyber Defence Centre of Excellence	Since 1907	Strategic Competition and Armed Conflict	Legal
United Nations Charter	United Nations	1945	Strategic Competition and Armed Conflict	Legal
International Human Rights Law Treaties and Instruments	*Overview by the United Nations*	Since 1948	Strategic Competition and Armed Conflict	Legal
Norms, rules, and principles for the responsible behaviour of States	United Nations	2015	Strategic Competition	Normative
Tallinn Manual 2.0 on the International Law Applicable to Cyber Operations	International Group of Experts (NATO Cooperative Cyber Defence Centre of Excellence)	2017	Strategic Competition and Armed Conflict	Legal
Joint Publication 3-12 Cyber Operations	US Joint Chiefs of Staff	2018	Strategic Competition and Armed Conflict	Operational
Nine Common Principles to Secure Cyberspace	Paris Call for Trust and Security in Cyberspace	2018	Strategic Competition	Normative
Singapore Norm Package	Global Commission on the Stability of Cyberspace	2018	Strategic Competition	Operational
Advancing Cyberstability	Global Commission on the Stability of Cyberspace	2019	Strategic Competition	Normative
Joint Doctrine for Military Cyberspace Operations	Royal Danish Defence College	2019	Armed Conflict	Operational
Allied Joint Publication-3.20 – Allied Joint Doctrine for Cyberspace Operations	NATO Standardization Office	2020	Armed Conflict	Operational

FM 3-12 CYBERSPACE OPERATIONS AND ELECTROMAGNETIC WARFARE	US Department of the Army	2021	Strategic Competition and Armed Conflict	Operational
8 rules for “civilian hackers” during war, and 4 obligations for states to restrain them	Tilman Rodenhäuser and Mauro Vignati (International Committee of the Red Cross)	2023	Armed Conflict	Operational
Active Cyber Defense – Toward Operational Norms	Sven Herpig (interface – Tech analysis and policy ideas for Europe)	2023	Strategic Competition	Operational
AIR FORCE DOCTRINE PUBLICATION 3-12 CYBERSPACE OPERATIONS	US Air Force	2023	Strategic Competition and Armed Conflict	Operational
The National Cyber Force: Responsible Cyber Power in Practice	UK National Cyber Force	2023	Strategic Competition and Armed Conflict	Operational
The Pall Mall Process: tackling the proliferation and irresponsible use of commercial cyber intrusion capabilities	Participant Representatives of the Pall Mall Process	2024	Strategic Competition	Normative
Responsible cyber behaviour in the Indo-Pacific: Views from Cambodia, Fiji, India, Indonesia, Japan, Pakistan and Taiwan	Gatra Priyandita and Louise Marie Hurel (Australian Strategic Policy Institute)	2025	Strategic Competition	Normative

Methodology

In January 2025, interface started to set up an international and interdisciplinary working group made up of 38 practitioners and researchers working on various topics within the field of cybersecurity. Members were recruited from 150 alumni of the Transatlantic Cyber Forum (TCF), founded in 2017, and beyond. They formed the TCF working group “Irresponsible Behavior During Cyber Operations,” with a particular emphasis on state-backed Chinese campaigns.

In parallel with desk research, 34 semi-structured interviews on irresponsible operational cyber behavior were conducted with working group members between January 2025 and March 2025.

In March 2025, an expert survey on irresponsible behavior in cyber operations by state-backed threat actors was conducted, with 29 respondents.

On June 3–4, 2025, 15 members of the working group came together for a two-day interactive workshop in Berlin to discuss irresponsible state behavior during cyber operations in times of strategic competition.

The draft policy paper, developed based on knowledge and insights from desk research, interviews, survey responses, and workshop interactions between July 2025 and August 2025, was reviewed in September and October 2025 by 26 members of the working group and additional reviewers.

Acknowledgments

This analysis was supported by the TCF working group "Irresponsible Behavior During Cyber Operations." The views and opinions expressed in this paper are those of the author and do not reflect the official policy or position of the working group members or that of their respective employer(s). In alphabetical order, the author acknowledges the essential contributions of the following individuals:

1. Charles-Pierre Astolfi, Seedfence
2. Eugenio Benincasa, Center for Security Studies at ETH Zurich
3. Mei Danowski, Natto Thoughts
4. Yorck Hesselbarth, Orcrist Technologies
5. Sherry Huang
6. Christoph Lobmeyer, Deutsche Cyber-Sicherheitsorganisation (DCSO)
7. Igor Mikolic-Torreira, Center for Security and Emerging Technology (CSET)
8. Aleksandar Milenkoski, SentinelLABS (SentinelOne)
9. Lukasz Olejnik, Department of War Studies, King's College London
10. Alexandra Paulus, German Institute for International and Security Affairs (SWP)
11. Pavlina Pavlova
12. Daniel Plohmann, Fraunhofer FKIE
13. Thomas Reinhold, Peace Research Institute Frankfurt (PRIF)
14. Matthias Schulze, Institute for Peace Research and Security Policy at the University of Hamburg (IFSH)
15. Adam Segal, Council on Foreign Relations
16. Fernando Sol III
17. Timo Steffens, German Federal Office for Information Security (BSI)
18. Julian-Ferdinand Vögele, Recorded Future
19. Valentin Weber, German Council on Foreign Relations (DGAP)

The author also extends special gratitude to the entire Cybersecurity Policy and Resilience team at interface, as well as to Luisa Seeling, Alina Siebert, Scribendi editor Maria (EM643) and ChatGPT for their assistance with the editing and layout

of this publication.

This project was made possible by the generous support of the German Federal Foreign Office. The views expressed in this paper do not necessarily represent the official position of the ministry.

Author

Sven Herpig
Lead Cybersecurity Policy and Resilience
sherpig@interface-eu.org
+49 (0)30 81 45 03 78 91

Imprint

interface – Tech analysis and policy ideas for Europe
(formerly Stiftung Neue Verantwortung)

W www.interface-eu.org
E info@interface-eu.org
T +49 (0) 30 81 45 03 78 80
F +49 (0) 30 81 45 03 78 97

interface – Tech analysis and policy ideas for Europe e.V.
c/o Publix
Hermannstraße 90
D-12051 Berlin

Design by Make Studio
www.make.studio
Code by Convoy
www.convoyinteractive.com