November 2023 · Dr. Sven Herpig

Active Cyber Defense

Toward Operational Norms

Stiftung Neue Verantwortung

Think Tank at the Intersection of Technology and Society

Acknowledgments

This analysis was supported by the Transatlantic Cyber Forum working group on "Active Cyber Defense" through online collaboration and a joint workshop in Berlin. The views and opinions expressed in this paper are those of the author and do not necessarily 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. Dave Aitel
2. Charles-Pierre Astolfi, Polygones
3. Tabea Breternitz
4. Maximilian Heinemeyer, Darktrace
5. Sherry Huang, William and Flora Hewlett Foundation
6. Carolin Kemper, German Research Institute for Public Administration (FÖV)
7. Andreas Kuehn, Observer Research Foundation America
8. James Andrew Lewis, Center for Strategic and International Studies (CSIS)
9. Christoph Lobmeyer
10. Lennart Maschmeyer, ETH Zurich
11. Igor Mikolic-Torreira, CSET, Georgetown University
12. Lukasz Olejnik, Independent Researcher
13. Pavlina Pavlova, CyberPeace Institute
14. Thomas Reinhold, Peace Research Institute Frankfurt (PRIF)
15. Christine Runnegar, Internet Society
16. Christina Rupp, Stiftung Neue Verantwortung
17. Janine Schmoldt, University of Erfurt
18. Guido Schulte, German Armed Forces
19. Aleksandra Sowa, PET, German Informatics Society
20. Jasmin Stadler, European Cyber Conflict Research Initiative
21. Timo Steffens, Federal Office for Information Security
22. Eric MSP Veith, Carl von Ossietzky University Oldenburg
23. Julian-Ferdinand Vögele, Recorded Future

Finally, the author would like to thank Rebecca Beigel, Frederic Dutke, and Christina Rupp for their support in designing and implementing the workshop, as well as Luisa Seeling, Alina Siebert, and Scribendi Editor EM1733 for their support with this publication.

Executive Summary

An increasing number of governments are engaging with active cyber defense — either through policy debate or in practice. However, the public discourse on what a good policy framework should look like and what operational norms should be followed lags significantly behind, especially when it comes to concrete operational norms (for intrusive measures) undertaken at home and against IT systems abroad. In simple terms, active cyber defense is the practical response by technical means of a government to malicious cyber activity targeting organizations within its country or allied states and partner states. This paper outlines nine operational norms that may decrease the risk of collateral damages and diplomatic escalation stemming from active cyber defense operations:

1. **Respond, don't retribute:** Active cyber defense operations should always be a response to a malicious cyber operation or campaign, thus neutralizing, mitigating, or attributing a malicious cyber activity.

2. **Prioritize operational spaces:** Governments should focus their measures on their own jurisdictions, communicate with allies before engaging in their jurisdictions, and try to avoid the jurisdictions of uninvolved third parties.

3. **Don't just do it — explain it:** Governments should set up political, legal, and oversight frameworks for active cyber defense operations and put an emphasis on impact assessment and transparency.

4. **Shape the international discourse:** Governments should be aware of their role in shaping international law and should engage in confidence-building measures.

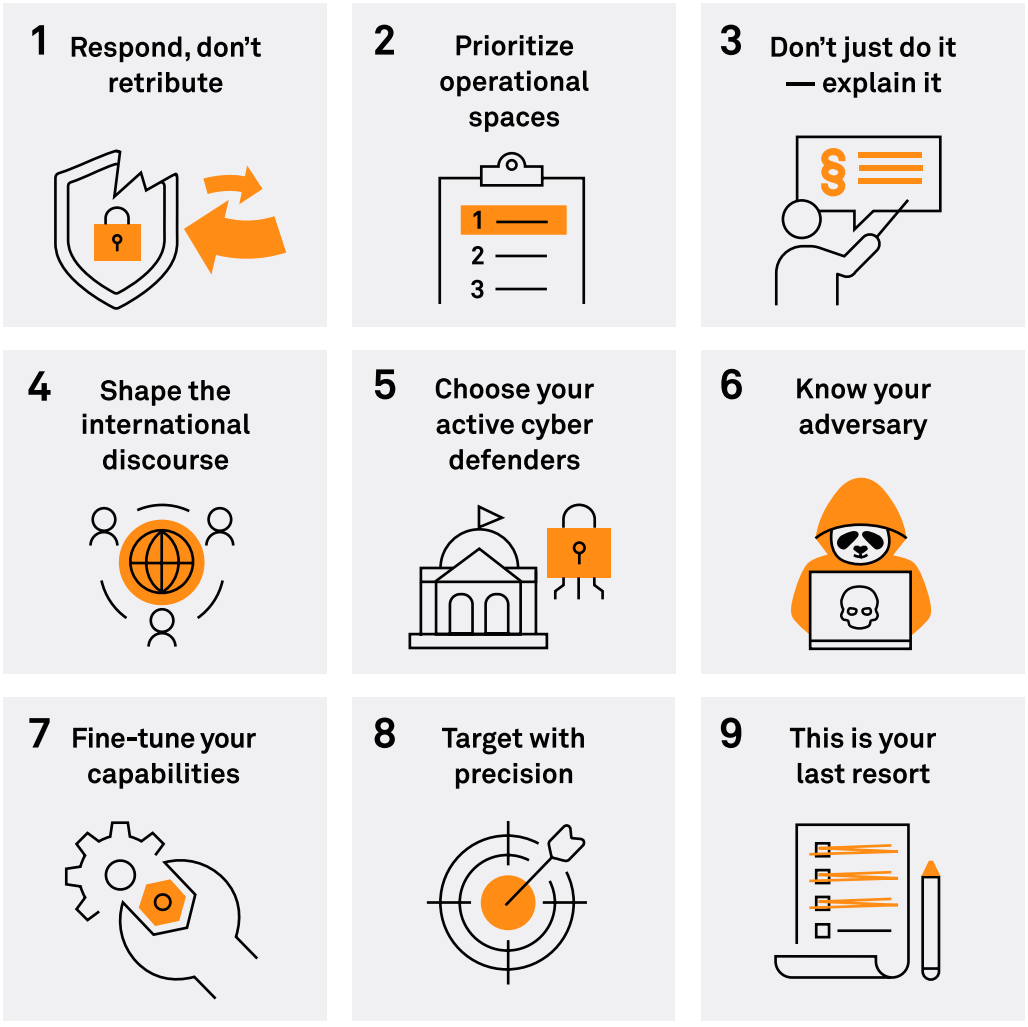
5. **Choose your active cyber defenders:** Technical excellence, operational expertise, and the willingness to subject itself to strict frameworks under a central authority should be key requirements for the primary operational agency.

6. **Know your adversary:** A deep level of technical understanding about the adversary's cyber-operational environment is crucial for an active cyber defense operation.

7. **Fine-tune your capabilities:** The procuring, designing, and testing processes of capabilities need to be meticulous in order to guarantee the efficiency, effectiveness, and proportionality of the measures.

8. **Target with precision:** Independent from the operational space, measures should be as limited as possible and avoid targeting third parties, especially supply chains and critical infrastructures.

9. **This is your last resort:** Governments should be aware that every intrusive active cyber defense operation is likely a resource-intensive, one-off activity that does not sustainably improve the overall level of national cybersecurity or resilience.



These operational norms are meant to serve not only as a contribution to the ongoing debate but also as a starting point for governments that are looking for advice on how to develop their active cyber defense policies. Additionally, these norms may also contribute to increasing convergence among like-minded states regarding active cyber defense policies that reflect shared values.

Table of Content

International Development of Active Cyber Defense Policy and Practice

Despite governments' efforts to improve cybersecurity through regulations, policies, technical facilities, training, and resource allocation, the threat level seems to remain elevated.[1] Well-resourced threat actors, such as organized crime, groups offering their services for hire, security agencies, and militaries, continue to be a bane for government bodies, the industry, and society-at-large. This is likely one of the main drivers behind government attempts to find additional ways to make cyberspace safer and more secure. Active cyber defense operations are one set of activities that has been discussed for years[2] but only now appears to be making fast-paced headway in the policy world.

> Active cyber defense operations are "one or more technical measures implemented by an individual state or collectively, carried out or mandated by a government entity with the goal to neutralize and/or mitigate the impact of and/or attribute technically a specific ongoing malicious cyber operation or campaign."[3]

Examples of measures taken in active cyber defense operations range from mandating Internet Service Providers to block or reroute malicious traffic to taking over a command-and-control infrastructure used in malicious cyber campaigns to uninstall or neutralize malware on the victims' systems and/or deploy patches. Active cyber defense is therefore different from passive measures, such as running anti-malware software or firewalls. As active cyber defense has the goal of supporting cybersecurity efforts, it differs from (offensive) cyber operations, which are, for example, aiming to collect intelligence on targets or to preposition in adversarial IT systems for future military missions.[4] **Thus, active cyber defense is an extension of passive cyber defense, while its — sometimes intrusive — operations are a defensive subset of cyber operations in general.**

In April 2022, the **United Kingdom**'s National Cyber Force published a primer on its activities, stating that "[c]ountering threats which undermine the confidentiality, integrity and availability of data, and effective use of systems by users

1 E.g., compare annual threat assessments of Germany's Federal Office for Information Security, see Bundesamt für Sicherheit in der Informationstechnik (2021): Archiv Lageberichte and Bundesamt für Sicherheit in der Informationstechnik (2022): Die Lage der IT-Sicherheit in Deutschland 2022
2 E.g., public discourse in Germany started around 2017, see Sven Herpig et al (2020): Aktive Cyberabwehr/ Hackback in Deutschland
3 Sven Herpig (2021): Active Cyber Defense Operations – Assessment and Safeguards
4 For a taxonomy of cyber defense, see for example, Tanya Gärtner (2023): Towards a Taxonomy of Cyber Defence in International Law and for a categorization of active cyber defense, see for example, Sven Herpig (2018): Aktive Cyber-Abwehr/ Hackback

[...] ca involve conducting cyber operations [...]."[5] Although not legally binding, the **Council of the European Union** "recall[ed] its encouragement to Member States to further develop their own capabilities to conduct cyber defence operations, including when appropriate proactive defensive measures to protect, detect, defend and deter against cyberattack"[6] in its May 2023 Council Conclusions on the EU Policy on Cyber Defence. In June 2023, **Germany** published its first-ever national security strategy. There, the federal government stated that it "[...] will examine what capabilities and legal authority it requires to defend against threats in cyberspace — including for an ongoing or imminent cyberattack [...]."[7] This would be a profound step, as it would require a change in the country's Basic Law (*Grundgesetz*).[8] Later in June 2023, the head of the cyber division of the Romanian Intelligence Service announced that they "will hack back the command and control servers of foreign APT [Advanced Persistent Threat] groups targeting the country" as they have done already to counter cybercrime operations.[9]

In parallel to debates about active cyber defense in Europe, similar discussions also take place globally. In November 2022, the Australian government announced that "the Australian Federal Police and the Australian Signals Directorate will initiate an ongoing, joint standing operation to investigate, target and disrupt cyber criminal syndicates with a priority on ransomware threat groups."[10] In its discussion paper on the 2023–2030 Australian Cyber Security Strategy, it is stated that "[w]hen a cyber incident does occur, we have an agile and rapid response to mitigate its harm, recover quickly, and disrupt further malicious acts."[11] In its National Security Strategy, Japan announced in December 2022 that it will "introduce active cyber defense for eliminating in advance the possibility of serious cyberattacks [inter alia by] penetrat[ing] and neutraliz[ing] attacker's servers and others in advance to the extent possible."[12] In April 2023, the People's Republic of China amended its counter-espionage law, which now states that "[...] state security organs may employ technical investigative measures [...]." Additionally, it mandates that "relevant departments are to [...] order the telecommunications operators or internet service providers to promptly employ measures such as repairing vulnerabilities, solidifying network protections, stopping transmission, deleting programs or content, suspending related services, removing related applications, or closing relevant websites,

5 National Cyber Force (2023): The National Cyber Force: Responsible Cyber Power in Practice
6 Council of the European Union (2023): Council Conclusions on the EU Policy on Cyber Defence
7 The Federal Government (2023): Robust. Resilient. Sustainable. Integrated Security for Germany – National Security Strategy
8 Currently, the authority to implement, especially the intrusive end of, "emergency response" [Gefahrenabwehr] measures in cyberspace is, with few exceptions (e.g., responses to terrorist threats), held by state law enforcement. The federal government plans to grant that authority to federal law enforcement, which would require a change in the Basic Law.
9 Catalin Cimpanu (2023): Risky Biz News: Romania to hack-back foreign APTs
10 Mark Dreyfus (2022): Joint standing operation against cyber criminal syndicates
11 Australian Government (2022): 2023 - 2030 Australian Cyber Security Strategy Discussion Paper
12 Cabinet Secretariat (2022): National Security Strategy of Japan [Provisional Translation]

and store the related records" as possible responses to "risks such as information content or network attacks involving acts of espionage."[13] Active cyber defense measures in the United States are being actively pursued. After the removal of the Hafnium web shells in 2021,[14] the United States announced the takedown of the Hive Network in January 2023[15] and the removal of the Snake malware in *Operation MEDUSA*[16] in May 2023. All three can be considered active cyber defense operations. Additionally, the United States has been engaged in "defensive hunt forward" operations in Albania,[17] Lithuania,[18] and elsewhere.[19] The detection of "malicious cyber activity on host nation networks"[20] by hunt forward operations may have set the stage for partner states to conduct active cyber defense operations, as judged appropriate by that nation. In July 2023, the head of NATO's cyber and hybrid policy section hinted at the military alliance debating a more proactive posture in cyberspace to be "more decisive in imposing costs for many of these malicious cyber activities," for example, through "defensive cyberspace operations."[21]

With multiple countries having set up related policy frameworks and/or having conducted actual active defense cyber operations, more countries will likely move toward conducting active cyber defense operations in the future. At the same time, policy debates fall short of providing clarity as to what constitutes responsible behavior at the operational level that lessens the risk of collateral damage and diplomatic escalation. With regard to defining criteria for the responsible conduct of active cyber defense operations, the 2021 predecessor of this analysis stated that "[...] the different indicators cannot be hard coded with risk levels" and that "[...] decisions about deployment of active cyber defense operations must always be made on a case-by-case basis."[22] While these conclusions still hold true, the present analysis attempts — based on the latest developments in the field — to provide concrete advice to governments as to which operational norms may contribute to more responsible state behavior.[23]

There is a need to do so — for example, the outgoing director of the UK Government Communications Headquarters articulated that the primer on National Cyber Force activities "provides a benchmark for the UK's approach and a basis for like-minded

13 China Law Translate (2023): Counter-espionage Law of the P.R.C. (2023 ed.)
14 Sven Herpig (2021): Active Cyber Defense Operations – Assessment and Safeguards
15 U.S. Department of Justice (2023): U.S. Department of Justice Disrupts Hive Ransomware Variant
16 U.S. Department of Justice (2023): Justice Department Announces Court-Authorized Disruption of Snake Malware Network Controlled by Russia's Federal Security Service
17 U.S. Embassy in Albania (2023): "Committed Partners in Cyberspace": U.S. Concludes First Defensive Hunt Operation in Albania
18 U.S. Cyber Command (2023): "Building Resilience": U.S. returns from second defensive Hunt Operation in Lithuania
19 Julia Schuetze and Eglė Daukšienė (2023): Cybersecurity Support Deployments: An emerging cooperative approach
20 U.S. Cyber Command (2022): CYBER 101: Hunt Forward Operations
21 Alexander Martin (2023): NATO's Christian-Marc Liflander on how the alliance can take a 'proactive' cyber stance
22 Sven Herpig (2021): Active Cyber Defense Operations – Assessment and Safeguards
23 The paper therefore aims to contribute an interdisciplinary, intersectoral, Western perspective to the growing body of literature on the responsible conduct of cyber operations with focus on active cyber defense.

governments to come together internationally to establish a shared vision and values for the responsible use of cyber operations."[24] Similarly, practitioners and academics already argued in August 2021 that "the United States needs to build international support for drawing lines between responsible and irresponsible operations in cyberspace."[25] Additionally, a former practitioner and current academic argued in September 2021 that "it has become increasingly important to understand how this instrument [offensive cyber operations capabilities] ought to be used — and not merely how it has been abused in the past by malicious actors."[26] While most of these comments refer to cyber operations in general and not active cyber defense operations in particular, Germany specifically emphasizes active cyber defense operations. **In its national security strategy, Germany states that it "will also develop standards for [the use of active cyber defense], while [...] respecting the norms of responsible state behaviour in cyberspace."**[27]

While substantial research on the topic of norms for active cyber defense operations does not yet exist, there are norms and contributions on responsible state behavior in the use of information and communication technologies (ICTs) that may also influence the conduct of active cyber defense operations by states.[28] Together with the analytical framework and safeguards of the predecessor study,[29] these aspects form the basis of the main part of the analysis. **This paper presents nine non-binding operational norms to adhere to for governments that wish to conduct active cyber defense operations responsibly. These norms are especially designed to be applicable to the intrusive end of active cyber defense operations measures — those that bypass security mechanisms, exploit vulnerabilities, and interfere with the integrity of IT systems.**

24 Alexander Martin (2023): UK says its offensive cyber operations are 'accountable, precise, and calibrated'
25 Perri Adams, Dave Aitel, George Perkovich, and JD Work (2021): Responsible Cyber Offense
26 JD Work (2021): Balancing on the rail – considering responsibility and restraint in the July 2021 Iran railways incident
27 The Federal Government (2023): Robust. Resilient. Sustainable. Integrated Security for Germany – National Security Strategy
28 General norms can be found at the United Nations (2015): Report of the Group of Governmental Experts on Developments in the Field of Information and Telecommunications in the Context of International Security and Global Commission on the Stability of Cyberspace (2019): Advancing Cyberstability: Final Report. More specific, operational, norms can be found in expert analyses, for example, in Perri Adams, Dave Aitel, George Perkovich, and JD Work (2021): Responsible Cyber Offense.
29 Sven Herpig (2021): Active Cyber Defense Operations – Assessment and Safeguards; a summary can be found at Andreas Kuehn and Sven Herpig (2022): The EU and Responsible Active Cyber Defence. A number of sentences, paragraphs and conclusions have been adopted from a prior study by the same author, see Sven Herpig (2021): Active Cyber Defense Operations – Assessment and Safeguards. To improve readability, lengthier parts adopted from that study have not been marked with quotation marks.

A First Glimpse of Operational Norms for Cyber Operations

Cyber operations are an instrument in the toolset of states. That instrument includes offensive military cyber operations[30] as well as cyber operations for intelligence collection[31] or for subversion[32] and active cyber defense operations[33]. As with other potentially harmful activities that states conduct in cyberspace, the question arises of how to engage appropriately. Or, in language agreed upon by all United Nations Member States,[34] what characterizes "responsible state behavior"[35] in the conduct of cyber operations? While there are some general guidelines — in the form of voluntary, non-binding norms — on the use of ICTs by states from the United Nations in this respect, there were few discussions on concrete norms on the operational level until recently.

Therefore, it was remarkable when the UK government decided to publish its primer on "The National Cyber Force: Responsible Cyber Power in Practice."[36] Almost at the same time, ABC News (Australia) aired "Breaking the Code: Cyber Secrets Revealed", a show featuring the work of the Australian Signals Directorate (ASD).[37] By outlining why they are conducting cyber operations the way they do, both the primer and the interviews with ASD staff on the show can be interpreted as steering the international debate in the direction of operational norms that Australia and the UK already regard as responsible.

In its primer, the UK government has set out three high-level operational principles for the responsible conduct of cyber operations: **accountability**, **precision**, and **calibration**.[38] Accountability includes a "robust legal framework,"[39] "robust authorisation and oversight procedures."[40] Additionally, the UK primer includes rather broad expressions of cyber operations needing to be in accordance with national

30 See for example Matthias Schulze (2020): Militärische Cyber-Operationen – Nutzen, Limitierungen und Lehren für Deutschland
31 Michael Warner (2017): Intelligence in Cyber—and Cyber in Intelligence
32 See for example Lennart Maschmeyer (2023): Subversion, cyber operations, and reverse structural power in world politics
33 See for example Sven Herpig (2021): Active Cyber Defense Operations – Assessment and Safeguards
34 United Nations General Assembly (2015): Group of Governmental Experts on Developments in the Field of Information and Telecommunications in the Context of International Security and United Nations General Assembly (2021): Group of Governmental Experts on Advancing Responsible State Behaviour in Cyberspace in the Context of International Security
35 United Nations General Assembly (2015): Group of Governmental Experts on Developments in the Field of Information and Telecommunications in the Context of International Security and United Nations General Assembly (2021): Group of Governmental Experts on Advancing Responsible State Behaviour in Cyberspace in the Context of International Security
36 National Cyber Force (2023): The National Cyber Force: Responsible Cyber Power in Practice
37 ABC News (2023): How Intelligence agencies catch criminals
38 National Cyber Force (2023): The National Cyber Force: Responsible Cyber Power in Practice
39 Conrad Prince (2023): Government shines a light on UK cyber operations
40 Tom Uren (2023): UK's National Cyber Force: A Bunch of Mindf-ckers

and international law, "national values," and to be conducted in an ethical manner.[41] Precision refers to a good understanding of the specific operational environment, as well as accurate timing and targeting of the operations.[42] Lastly, calibration references impact assessment, geopolitical context, and live responses to technical and political changes.[43] The Australian perspective, on the other hand, at least what has been publicly communicated through official channels, remains vague on what it regards as responsible. The former director of the Australian Signals Directorate simply referred to Australian cyber operations during *Operation Valley Wolf* as "generating effects through cyberspace in a very **clever, precise,** [and] **timed** way [...]."[44] Operators that were part of it described how much work went into **reconnaissance** of the target environment, as well as **custom-tailoring the capabilities**.[45] Outside of official channels, Australia's take on responsible cyber operations has been described in a short analysis by the Australian Strategic Policy Institute as following the principles of **necessity**, **specificity**, **proportionality,** and **"considering whether an act causes greater harm than is required to achieve the legitimate military objective."**[46] This aligns with what the UK government describes as precision and calibration. Accountability is covered by the Australian implementing agency's existing legislative and oversight framework, with approval for operations going up to the Minister of Defence.[47] Therefore, there appears to be some convergence in this area between the two countries.

While the operational norms highlighted by the two governments are still not overly specific, they are a far cry from "we observe national and international law." Still, there remains a lack of internally agreed-upon, unified standard frameworks for cyber operations.

Academic and former practitioner *JD Work*, while going into much more detail during his analysis, describes responsible cyber operations as "not merely the result of good intentions. Rather, it requires deliberate planning, engineering, operational, management, and oversight efforts throughout the lifecycle of a campaign to ensure that access and actions on objectives are **properly aligned**, adequately tailored, and **appropriately balance potential harms** to competing equities whilst accomplishing mission objectives. Responsible conduct requires programmatic maturity, individual professionalism, and organizational focus to achieve."[48]

41 National Cyber Force (2023): The National Cyber Force: Responsible Cyber Power in Practice
42 National Cyber Force (2023): The National Cyber Force: Responsible Cyber Power in Practice
43 National Cyber Force (2023): The National Cyber Force: Responsible Cyber Power in Practice
44 ABC News (2023): How Intelligence agencies catch criminals
45 ABC News (2023): How Intelligence agencies catch criminals
46 Fergus Hanson and Tom Uren (2018): Australia's Offensive Cyber Capability
47 Fergus Hanson and Tom Uren (2018): Australia's Offensive Cyber Capability
48 JD Work (2021): Balancing on the rail – considering responsibility and restraint in the July 2021 Iran railways incident

Academics and practitioners *Perri Adams, Dave Aitel, George Perkovich,* and *JD Work* advise technical operators of cyber operations to be **cautious**, **precise**, and to **minimize unintended harm.**[49] They suggest "technical operational norms [that] should address irresponsible actions that cause adverse effects, such as collateral damage. Consideration should also be given to **verifiability.**"[50] The operational norms discussed include testing tools, avoiding indiscriminate targeting, carefully selecting which devices to target as pivoting points, constraining automation, and preventing third-party access to backdoors.

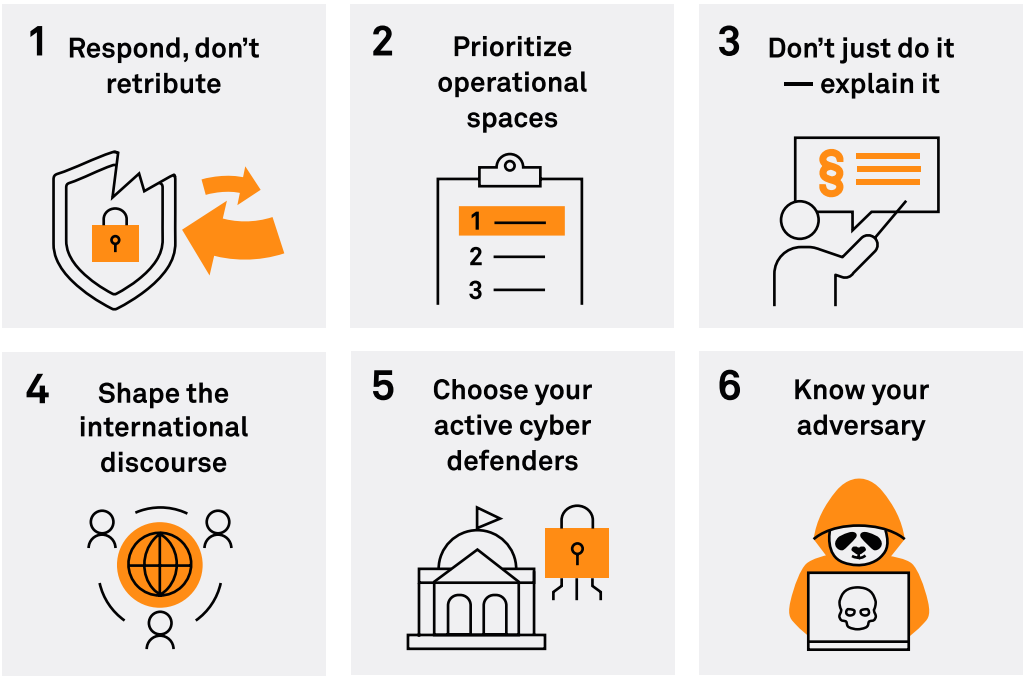
While going into more detail, the operational norms suggested by selected researchers and practitioners align well with the broader strokes outlined by the British and Australian governments.

The mentioned operational norms offer a first glimpse of the responsible conduct of cyber operations, yet they are not custom-tailored to active cyber defense operations. In addition, they are based on a very small body of government publications and expert analyses. Despite these limitations, they do provide a basic understanding regarding operational norms, which is leveraged in the following sections.

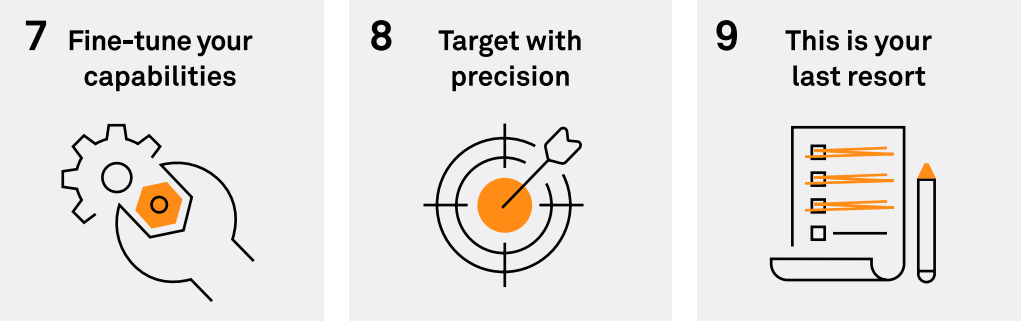
49 Perri Adams, Dave Aitel, George Perkovich, and JD Work (2021): Responsible Cyber Offense
50 Perri Adams, Dave Aitel, George Perkovich, and JD Work (2021): Responsible Cyber Offense

Developing Operational Norms for Active Cyber Defense

Although active cyber defense operations are a subset of cyber operations that can be intrusive — similar to those operations aiming to collect intelligence or preposition for disruptive effects — they serve an inherently defensive purpose. At first glance, this defensive purpose would give them more flexibility in terms of responsible conduct, as they are a non-retaliatory means to defend. At the same time, active cyber defense operations, especially if they are intrusive, need to be conducted with a high level of caution to avoid further victimizing affected agencies, companies, and citizens. Based on the analytical framework developed in the predecessor study[51] and the basic understanding of operational norms for cyber operations discussed in the last segment, this section offers concrete operational norms that governments should abide by if they want to ensure that they are conducting active cyber defense operations in a more responsible manner. Recounting the definition, active cyber defense operations are defined as **"one or more technical measures implemented by an individual state or collectively, carried out or mandated by a government entity with the goal to neutralize and/or mitigate the impact of and/or attribute technically a specific ongoing malicious cyber operation or campaign."**[52]



51 Sven Herpig (2021): Active Cyber Defense Operations – Assessment and Safeguards
52 Sven Herpig (2021): Active Cyber Defense Operations – Assessment and Safeguards



1. Respond, don't retribute

While active cyber defense operations and other cyber operations may bear some similarities, it is important to clearly communicate which of the two actions is being discussed. If implemented efficiently, effectively, and proportionately[53], active cyber defense operations have, for example, a low risk of diplomatic escalation due to the fact that they are responses to a hostile action. This may not be the case with other cyber operations, which may bear escalation risks both from the effect or intended effect of the operation itself and the manner in which the operation was conducted.[54] These differences lead to important nuances in the operational norms of those actions. However, they hold true only if active cyber defense operations are indeed a response and not retribution. Under international law, any active cyber defense operation conducted in response to an internationally wrongful act by another state may not amount to retribution.[55]

Therefore, active cyber defense operations need to be conducted with the right level of technical attribution, in a timely fashion, aimed at systems and infrastructure that are involved in an active or ongoing malicious cyber operation or campaign against one's own state or, by extension and through cooperation,[56] an allied or partner state.

53 James Andrew Lewis (2022): Creating Accountability for Global Cyber Norms on proportionality in response to malicious cyber activities.

54 In James Andrew Lewis (2021): Toward a More Coercive Cyber Strategy the author however notes that "[...] in two decades of malicious cyber action, there has never been an incident that has led to escalation. While there have been a few instances of unintended consequences and collateral damage, these did not lead to an escalation of conflict".

55 Michael N. Schmitt (2022): Lieber Institute White Paper: Responding to Malicious or Hostile Actions under International Law
For a general take on the permissibility of active cyber defense operations under International Law, see for example Tanya Gärtner (2023): Towards a Taxonomy of Cyber Defence in International Law

56 Ashley Deeks (2020): Defend Forward and Cyber Countermeasures

2. Prioritize operational spaces

Recapping from the prior analysis,[57] there are four operational spaces in which active cyber defense operations can have effects:[58]

Blue space
IT systems and infrastructure under the jurisdiction of the state

Green space
IT systems and infrastructure leveraged in the malicious cyber operation/campaign that are under the jurisdiction of an allied or partner state

Red space
IT systems and infrastructure leveraged in the malicious cyber operation/campaign that are under the jurisdiction of the originating state

Grey space
All IT systems and infrastructure that do not belong to any of the other spaces

I. Blue space

Active cyber defense operations should aim to achieve their goals within blue space. Governments have full jurisdiction over that space, meaning that they can rely on robust legal frameworks (if they are in place — and they should be) and use additional measures such as search and seizures to improve visibility for and effectiveness of active cyber defense operations. Additionally, safeguarding implementation, risk minimization, and escalation prevention are easier to achieve in the blue space.

II. Green space

Existing and future treaties, joint task forces, communication channels, and other instruments enable various cooperative ways among allies to counter malicious

57 Sven Herpig (2021): Active Cyber Defense Operations – Assessment and Safeguards
58 The categorization assumes correct, confident technical attribution. Thus, the assessment may change over the course of planning and implementing an active cyber defense operation (e.g., when a false flag operation is encountered) and, therefore, adapt to the new information. At the same time, due to the distribution of bots in a botnet or division of labor among various actors in a cybercrime campaign, an active cyber defense operation may target IT systems in a number of different spaces with different types of effects. Thus, the space criterion is abstract and may be challenging to operationalize.

cyber activities, highlighting the need for close cooperation on the issue. These options must be seriously considered before unilateral action is taken against the systems and infrastructures of an allied or partner state, even when the systems and infrastructures are (unknowingly) involved in malicious cyber activity. However, the risk of third-party collection[59] and the need for swift action may call for unilateral actions at times.

Operations carried out in green space may also have less potential for backfiring if a partner state or ally benefits from them. The German Bundeskriminalamt's *Operation Ladybird,* for example, resulted in an intrusion in computer systems across the globe by using the existing access of the malware operations in order to remove the malware.[60] However, none of the affected actors' governments complained publicly about this operation.

From an international law perspective, the only truly feasible argument for a unilateral, noncooperative active cyber defense operation in allied or partner states' jurisdictions would be their (repeated) failure to comply with due diligence obligations.[61] Exceptions would, for example, be cases of grave and imminent peril. **Active cyber defense operations in green space should be carried out only after all options in the blue space have been exhausted or in cooperation with the respective ally or partner state. Unilateral, noncooperative active cyber defense operations in green space should be avoided in principle.**

III. Red space

Although there is always a risk of conflict escalation, active cyber defense operations in red space have — if implemented efficiently, effectively, and proportionately — a low risk of diplomatic escalation due to the fact that they are responses to a hostile action. Moreover, a broader geopolitical context should be taken into consideration when discussing potential conflict escalation due to an active cyber defense operation: "It is only when stepping back to view the whole of the ongoing clandestine conflict that the most significant aspects of responsibility in the present case may be understood."[62] Additionally, unilateral and noncooperative active

59 Security agencies or other organizations that are not involved in a cyber operation but gather data from its victim for various purposes fall into the category of third-party collectors. For instance, if a Russian cyber threat actor compromises a German government agency and utilizes IT systems situated in France to store the stolen data, and French security agencies access this data for their own intelligence needs, it would qualify as third-party collection. In this scenario, France is not directly affected by the cyber operation, nor is it the perpetrator, yet French agencies are acquiring data from a German government agency.

60 Joseph Cox (2016): The FBI's ‚Unprecedented' Hacking Campaign Targeted Over a Thousand Computers and Andre Meister (2021): Schadsoftware-Bereinigung: BKA nutzt Emotet-Takedown als Türöffner für mehr Befugnisse und neue Gesetze

61 International cyber law: interactive toolkit contributors (2023): Due diligence. Another possible legal basis would be *Necessity* as outlined in Article 25 of International Law Commission (2001): Responsibility of States for Internationally Wrongful Acts. See also James Andrew Lewis (2022): Creating Accountability for Global Cyber Norms on political responsibility of states for wrongful acts originated from their territories.

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

cyber defense operations in red space may be more in line with international law requirements (e.g., for countermeasures[63]) than in green or gray space.[64] **If there is solid technical attribution, proportionality of action, and clear communication, the risk should be negligible, and therefore active cyber defense operations in red space should be on the table. However, it is advisable to first exhaust all options in blue space.**

IV. Gray space

Measures that can be taken in gray space are potentially more limited than in the other operational spaces. Again, from an international law perspective, the only truly feasible argument for a noncooperative active cyber defense operation in gray spaces may be a state's (repeated) failure to comply with due diligence obligations.[65] Exceptions would, for example, be cases of grave and imminent peril. Cooperative active cyber defense operations in gray space may have a substantial risk of third-party collection. This needs to be taken into consideration in operational planning. **Ideally, active cyber defense operations in gray space are conducted only when measures in all other operational spaces have been exhausted.**

63 Ashley Deeks (2020): Defend Forward and Cyber Countermeasures

64 Of course, there are different levels of national responsibility for malicious cyber operations or campaigns that are carried out from a country's territory. On the Spectrum of State Responsibility — see Jason Healey (2011): Beyond Attribution: Seeking National Responsibility for Cyber Attacks — the *state-prohibited and state-prohibited-but-inadequate* points could — depending on diplomatic relations and evidence of efforts to stop the malicious activity — lead to the space being considered green or gray rather than red.

65 International cyber law: interactive toolkit contributors (2023): Due diligence. Another possible legal basis would be *Necessity* as outlined in Article 25 of International Law Commission (2001): Responsibility of States for Internationally Wrongful Acts.
See James Andrew Lewis (2022): Creating Accountability for Global Cyber Norms on political responsibility of states for wrongful acts originated from their territories.

3. Don't just do it — explain it

I. Political framework

Governments that want to responsibly engage in active cyber defense operations should have a high-level, public concept that outlines their respective **goals, procedures, and safeguards** toward such operations as comprehensively as possible.[66] Ideally, these aspects are translated into policies and technical and non-technical operational norms, as well as into a robust legal framework.

II. Legal framework

A legal framework for active cyber defense operations is crucial for several reasons, especially due to the potential privacy invasiveness and security risks that may be inherent in active cyber defense operations and because it is vital to clarify the rules applying to everyone involved. Adherence to the rule of law may be facilitated by having **a clear and possibly specific legal framework for active cyber defense operations**. A legal framework can and should include a number of safeguards and clear assignments of liability that apply to active cyber defense operations. Only then can government agencies be certain that their measures are lawful and that citizens and other parties are appropriately protected from the impacts of such operations that are conducted by states. The legal framework and its elements **should be regularly revisited, evaluated, and refined** in accordance with national sunset clause procedures. The legal framework must also include oversight mechanisms to foster accountability.

III. Oversight framework

Active cyber defense operations should require an **ex ante warrant for intrusive measures**, including specific parameters of the planned operation, for example, in the form of an impact assessment. Although intrusive active cyber defense operations take time to prepare, there may be edge cases in which imminent danger can be responded to quickly by an active cyber defense operation. In these cases, immediate measures followed by a timely retroactive judicial or political review should be possible. Specialized courts or legislative bodies should be in charge and provided with technical capacity building and the option to bring in independent technical expertise to enable the decision-makers to understand the possible implications of the measures stated in the warrant.

Active cyber defense operations outside national jurisdiction (non-blue space) should require approval from the highest echelons of government. The Australian Signals Directorate, for example, must seek approval for active cyber defense

66 Compare this to the UK government stating that "[t]he application of operational cyber capabilities in a responsible way is governed by a defined strategy and doctrine, so that there is clarity about what they are used for and a well-understood set of principles governing their operational application. We have developed a robust framework and while most of this must remain secret, this document has set out a number of the most significant principles." See National Cyber Force (2023): The National Cyber Force: Responsible Cyber Power in Practice.

operations from the Minister of Defence.[67] Additionally, after-action reports of all intrusive and non-blue space active cyber defense operations should be provided to a legislative oversight committee for additional scrutiny, where feasible. To facilitate remedial actions, active cyber defense operations should include a **notification to the non-adversarial targets (non-red spaces)**. The notification needs to include the particularity requirements laid out in the warrant application if a warrant is involved. Due to operational concerns, the notice does not have to be provided before or during an ongoing operation but within a limit, for example, of 90 days after the end of the operation.

Additionally, agencies carrying out active cyber defense operations should have constant **internal oversight**, from the initial impact assessment to the last action on the target and beyond. Internal legal counsel should be present during this entire time.

To enable robust oversight and review processes, capability design, procurement, and testing, as well as operational actions, must be auditable. Regarding offensive cyber operations, *Perri Adams, Dave Aitel, George Perkovich,* and *JD Work* state that "[e]nsuring responsible activity requires instituting processes that require suitable political authorization and oversight over technical quality control to reduce risk and collateral damage."[68] The same holds true for active cyber defense operations. Event logging and written statements from operators should form the basis for follow-on actions, such as after-action reports, and be accessible to the oversight bodies. The technical side of the audit, therefore, should be as unchangeable as possible through a **secured audit trail**.

IV. Impact assessment

In addition to the political, legal, and oversight frameworks, and post-operation transparency, active cyber defense operations should require an impact assessment. A formal **ex ante impact assessment** is essential to weigh the risks, impact, chances, and possible consequences of an operation, define ex post actions on the targets, develop additional options and backup plans, and define circuit breaker conditions. As part of the design process, capabilities should be evaluated for their risks, including human rights and potential collateral damage. This assessment may exclude some capabilities from ever qualifying for responsible use.

Operational norms play a major role when drafting the impact assessment. **The assessment should also speak to why the envisioned active cyber defense operation is the most efficient, effective, and proportionate option on the table. Therefore, the impact assessment should also include a discussion of the impact of taking no action.** Legal advice should be present throughout the entire planning

67 Fergus Hanson and Tom Uren (2018): Australia's Offensive Cyber Capability
68 Perri Adams, Dave Aitel, George Perkovich, and JD Work (2021): Responsible Cyber Offense

and implementation stage, for example, in the form of **embedded lawyers**. Ideally, the impact assessment additionally includes independent technical, legal, economic, and policy expertise whenever feasible.

Nevertheless, it is clear that any ex ante impact assessment often has to work with limited information and therefore provides only a narrow picture. In the case of so-called "black box" assessments, where little information is available in advance, a set of predefined general criteria is suitable for performing a self-assessment. **Additionally, if necessary to achieve the intended goals and adhere to operational norms, adaptations and course corrections may and should occur during the operation.** However, the impact assessment is a useful basis for decision-makers and oversight bodies and, therefore, should be a requirement for every active cyber defense operation.

V. Post-operation transparency

While it is understandable that governments would have an interest in keeping most of the operational details classified, *Healey* and *Jervis* point out that "[t]he national security community must declassify and break down compartments to combat cognitive bias. The current situation—yelping about the adversary's punches but classifying one's own—is not tenable, leading to a biased view of cyber conflict that is poisonous in an open democracy."[69]

Post-operation transparency can serve several purposes. Proactive communication about what happened and why (in the shape of press releases, warrants, etc.), accompanied, for example, by technical analysis of the threat actor infrastructure that has been targeted,[70] improves understanding of the actions taken across the board and signals intention to adversaries and allies. **Thereby, declassification might enable governments to pursue other tools and goals, such as naming and shaming a threat actor or de-escalating diplomatic tensions. Additionally, more transparency may also help build a common understanding of the implementation of operational norms.** Moreover, transparency reports could enable limited independent external audits.

The minimum viable communication should make sure that the targets of the active cyber defense operation — independent of the operational space — understand ex post that it was a response and, therefore, had defensive intent.

However, due to operational and diplomatic concerns, among other reasons, as well as the active involvement of allies, the right level of transparency will have to be decided on a case-by-case basis. **For the above-mentioned reasons, governments may want to err on the side of transparency.**

69 Jason Healey and, Robert Jervis (2020): The Escalation Inversion and Other Oddities of Situational Cyber Stability

70 Cybersecurity and Infrastructure Security Agency et al (2023): Joint Cybersecurity Advisory – Hunting Russian Intelligence "Snake" Malware

4. Shape the international discourse

Governments leading active cyber defense operations must also be aware that their respective practices, if widespread and consistent, may be contributing to the development of binding customary international law[71] when also "recogniz[ing] an obligation or a right to act in this way."[72] **Therefore, states must conduct their operations in the most responsible manner, so that any possibly emerging binding customary law can follow.**

Apart from shaping international law, confidence-building measures are useful to advance the discourse and lower the risk of unintended escalation involving third parties across operational spaces. Focusing these measures on active cyber defense operations countering cybercrime activities may be the lowest common denominator to bring various parties to the table. Confidence-building measures with allies and partner states could, for example, include Track 1 dialogues, joint exercises,[73] and best practice exchange for operational norms and insights in past operations. **However, the emphasis on confidence-building measures should primarily focus on activities involving non-allied, non-partner states.** Activities with those states, including strategic rivals, could feature Track 2 expert exchanges, the sponsoring of international academic conferences on the topic, and "Red Phones."[74] Wherever possible, existing confidence building measure initiatives — such as the OSCE CYBER/ICT SECURITY CBMs,[75] the OAS ADDITIONAL CONFIDENCE BUILDING MEASURES (CBMs) TO PROMOTE COOPERATION AND TRUST IN CYBERSPACE,[76] or the ASEAN CBMs[77] — should be leveraged.

An active cyber defense operation in an allied or partner state's jurisdiction (green space) should be communicated to and approved by the respective government, ex ante if possible, as such an operation certainly "risks friction with allies."[78] However, there may be an occasional need for ex post notification, for example, to avoid losing a window of opportunity for action or to avoid third-party collection. Especially in these cases, it is crucial to avoid misunderstandings, disgruntled allies, and unnecessary escalation. **Ideally, allies would agree bilaterally or in multilateral fora, such as ASEAN, OSCE, EU, or NATO, which circumstances could justify non-cooperative action in an allied or partner state's jurisdiction.**

71 Sven Herpig (2021): Active Cyber Defense Operations – Assessment and Safeguards
72 United Nations (2018): Draft conclusions on identification of customary international law, with commentaries
73 Rebecca Beigel and Julia Schuetze (2021): Cybersecurity Exercises for Policy Work – Exploring the Potential of Cybersecurity Exercises as an Instrument for Cybersecurity Policy Work
74 Tim Maurer (2013): Creating „Red Phones" for Cyberspace
75 Organization for Security and Co-operation in Europe (2023): Cyber/ICT Security
76 Inter-American Committee Against Terrorism (2022): Additional Confidence Building Measures (CBMs) To Promote Cooperation And Trust In Cyberspace
77 Association of Southeast Asian Nations (2022): ASEAN Cybersecurity Cooperation Strategy (2021 – 2025)
78 Max Smeets (2019): Cyber Command's Strategy Risks Friction with Allies

5. Choose your active cyber defenders

The prior study argued that there must be **a central focal point at the strategic level** overseeing all active cyber defense operations and responsible for adjusting the overall active cyber defense policy.[79] This focal point should oversee legal structure and authorities — including the operational and non-operational norms mentioned here — independent from the operational agency that is implementing active cyber defense measures.

On the operational level, one or more government stakeholders may be engaged in active cyber defense measures. Due to the broad range of measures that fall under the definition of active cyber defense, it is **impractical to prescribe one agency** category — such as law enforcement or military — as the primary operational stakeholder. Additionally, security cultures and legal frameworks vary across countries and jurisdictions, making a one-size-fits-all approach rather difficult.

Moreover, and most importantly, many countries may only have a very limited number of agencies with the technical capacity needed to carry out active cyber defense operations. **Technical capacity** is therefore one of the primary attributes that governments should look for if there is a need to designate a primary operational agency for active cyber defense measures.

Whichever agency, or set of agencies, is legally and politically entrusted with carrying out active cyber defense operations should follow the operational norms mentioned in this analysis. This means that while a national cybersecurity agency may not normally be required to apply for a warrant to conduct its operational tasks, decision-makers might want to reconsider this setup for active cyber defense operations, as outlined in the oversight framework.

As ideally most active cyber defense operations take place in national jurisdiction (blue space), at least one agency equipped with an active cyber defense authority should have the legal framework and operational expertise to **operate inside their own country**.

Perception by the targets of active cyber defense operations plays another, secondary, role. As mentioned, the low likelihood of diplomatic fallout from active cyber defense operations is closely linked to it being a response to a hostile action. Intelligence agencies carrying out active cyber defense operations outside blue space may look, at first glimpse, like espionage, while military branches doing it may appear as prepositioning. Whereas this can certainly be cleared up diplomatically later on, avoiding that risk through a different choice of primary operational agency could be considered. Alternatively, these agencies should use distinctly

79 Sven Herpig (2021): Active Cyber Defense Operations – Assessment and Safeguards

different sets of capabilities, or toolchains, for active cyber defense operations compared to their other intelligence or military tasks.[80]

While in the case of many nations, the answer to which agency should take the lead may be **(federal) law enforcement**, it is critical that the primary operational agency meets those criteria. Whether one or more agencies are implementing active cyber defense measures, it is vital that **they all operate within the same framework** outlined by the central focal point.

6. Know your adversary

In order to behave responsibly with regard to any of the following operational norms, it is crucial to gather as much technical information about the threat actor's tactics, techniques, and procedures and its operation or campaign — the *cyber operational environment*[81] — as possible. Activities include technical reconnaissance and intelligence gathering through other active and passive means. The collected data will improve the effectiveness of planning and capabilities, enable more accurate impact assessments, and reduce associated risks such as accidental collateral damage. **Thus, active cyber defense operations require accurate, actionable, high-quality information about the cyber operational environment.**

Examples of extensive intelligence gathering prior to a cyber operation are Stuxnet[82] and the operation against the Iranian railway system, where capabilities "were well tailored, through substantial intelligence support including extensive reconnaissance within target networks."[83]

7. Fine-tune your capabilities

A crucial operational norm is the design, procurement, and testing of the capabilities to be used. The UK government states that "[a] core part of responsible cyber operations is the design and use of capabilities in a way that is predictable and controllable, and where the risks are proportionate to the outcome required."[84] This applies to all kinds of tools and services used in active cyber defense operations, especially if they are intrusive. Apart from meticulous design, needed capabilities must be procured from law- and ethics-abiding vendors and thoroughly evaluated

80 However, there is, of course, always the risk of false flag operations and deception when pretending to carry out an active cyber defense operation while actually having a different agenda.
81 National Cyber Force (2023): The National Cyber Force: Responsible Cyber Power in Practice
82 E.g., Jon R. Lindsay (2013): Stuxnet and the Limits of Cyber Warfare and Kim Zetter (2015): Countdown to Zero Day: Stuxnet and the Launch of the World's First Digital Weapon
83 JD Work (2021): Balancing on the rail – considering responsibility and restraint in the July 2021 Iran railways incident
84 National Cyber Force (2023): The National Cyber Force: Responsible Cyber Power in Practice

in realistic test scenarios before use. This is necessary even if the latter may be challenging when situations change and new capabilities are urgently needed due to grave and imminent peril.[85]

I. Design

First, in order to design capabilities, governments need extensive intelligence on the target systems and networks. This allows the government to procure, design, and tailor capabilities efficiently, effectively, and proportionately. **Increased visibility decreases the chances of capabilities triggering unforeseen events leading to collateral damage, allows for finding the least intrusive attack vector, and enables avoiding critical infrastructures.**[86]

Second, the **effects on the target should always be as limited as possible** while still achieving the goal of the operation.[87] An example of this would be the Hafnium web shell removal, in which the FBI only removed the shells without patching the affected systems afterwards.[88] One way to combine responsible design without excessively degrading effectiveness could be to develop adaptive implants that have multiple effect options, ranging from non-intrusive and reversible to intrusive and non-reversible. This was in part demonstrated by the Australian Signals Directorate, which designed three stages ranging from reversible (*Rickrolling*) to non-reversible (*Dark Wall*) effects.[89]

Third, **capabilities should not propagate without restrictions, and have automation only where consistent with responsible control.**[90] A higher degree of automation needs to be safeguarded with additional controls, such as validation controls, including "formal principles of target discrimination, in both the technical and legal senses of the term, be incorporated into command-and-control functionality [..., and have a] human on the loop [...]."[91] An example of a capability that has apparently been designed with controls in mind is *Meteor.* Though "the deployment of the Meteor destructive payload and its associated components was highly scripted, this automation did not permit indiscriminate autonomous behavior. The detailed automation features specified, to a high degree of control, elements of the target to be serviced for effects and did not allow for non-specified elements to be struck."[92]

85 ABC News (2023): How Intelligence Agencies Catch Criminals
86 JD Work (2021): Balancing on the rail – considering responsibility and restraint in the July 2021 Iran railways incident
87 Sven Herpig (2021): Active Cyber Defense Operations – Assessment and Safeguards
88 Sven Herpig (2021): Active Cyber Defense Operations – Assessment and Safeguards
89 ABC News (2023): How Intelligence agencies catch criminals
90 Sven Herpig (2021): Active Cyber Defense Operations – Assessment and Safeguards
91 JD Work (2021): Balancing on the rail – considering responsibility and restraint in the July 2021 Iran railways incident
92 JD Work (2021): Balancing on the rail – considering responsibility and restraint in the July 2021 Iran railways incident

Fourth, **active cyber defense operations should leverage, whenever possible, technical concepts, capabilities, and techniques that are considered known, reliable, trusted, and appropriate.** *JD Work* describes this as "an additional hallmark of responsible operations, in which more sophisticated planners will consider the potential proliferation implications of deploying code which may both be directly repurposed if recovered from a target system or network, as well as the potential knowledge transfer to both the immediate adversary based on weapons technical intelligence derived from reverse engineering and behavioral analysis of observed implants."[93] In addition to limiting proliferation, designing capabilities in this way makes the effects more predictable and decreases erroneous and unexpected behavior.

Fifth, **active cyber defense operations should feature technical characteristics to ensure ex post operational transparency and accountability.** Including this feature would help targets distinguish operations for active cyber defense from other cyber operations aiming, for example, for espionage or prepositioning. There are various ways of doing so, including the publication of YARA rules,[94] offensive toolkit tokens,[95] and contact details in the comments of the code of implants[96] or as a text file on an affected IT system, which will be discovered during forensic analysis. In order "to enable an operationally relevant window of time for mission completion,"[97] these auditable features may, for example, be obfuscated or temporarily encrypted.[98] An example of this aspect is the publishing of YARA rules for detecting the modified binary file in the Emotet takedown.[99]

II. Procurement

While governments themselves ideally have certain capabilities, some active cyber defense operations may be so specific that they require additional expertise and tooling. Those missing capabilities, from exploits to implants and even complete services, will need to be procured from third parties. **These third parties should be transparently vetted and should not conduct any business with governments or other entities that have been reported to engage in unlawful activities or**

93 JD Work (2021): Balancing on the rail – considering responsibility and restraint in the July 2021 Iran railways incident
94 Victor Alvarez (2023): YARA
95 See for example Dave Aitel (2021): Technical Measures for Signaling. For a description of cryptographic markers in the context of marking systems as off-limits, see Samuel Charap and Reinhard Krumm (2023): Proposals to Address Political Interference. Outcomes of a Trilateral Dialogue
96 For an interesting discussion about analyzing the purpose of malware through its code, see Thomas Reinhold and Christian Reuter (2020): Towards a Cyber Weapons Assessment Model – Assessment of the Technical Features of Malicious Software
97 JD Work (2021): Balancing on the rail – considering responsibility and restraint in the July 2021 Iran railways incident
98 JD Work (2021): Balancing on the rail – considering responsibility and restraint in the July 2021 Iran railways incident
99 Bundeskriminalamt (2021): YARA-Signatur zur Identifizierung der Emotet-Malware

violate human rights with their tools and services.[100] An example of a company being added to such an entity list[101] is the Israeli firm NSO. NSO was added to the list by the US government because "the company knowingly supplied spyware that has been used by foreign governments to "maliciously target" the phones of dissidents, human rights activists, journalists and others."[102] An addition to such an entity list could be requiring **Know Your Vendor laws** that "would provide government clients with the ability to check where their prospective supply chain might include firms on restricted entity lists before awarding contracts."[103]

Before procuring capabilities that rely on zero-day vulnerabilities, **a vulnerability assessment and management**[104] **should be established as a safeguard to manage the risks of reverse engineering, exploitation of leaks, and/or use against the infrastructure by the threat actors.**[105] An example of such a process is the American Vulnerabilities Equities Process (VEP).[106]

Governments should require source code access from the third party in order to ensure compliance of the capability with the operational norms. Additionally, this access is needed to analyze whether the capability fulfills the operational requirements and what kind of insights the third party has into the government's operation, especially when the capability is a complete service — such as *Access-as-a-Service* (AaaS) — from third parties.[107]

III. Testing

Capabilities to be used in an active cyber defense operation require thorough testing before they can become operational because "poorly written malware can run awry and cause unintended issues."[108] This is especially true because, in line with their design requirements, active cyber defense capabilities are likely to be "one trick

100 See for example Winnona DeSombre, James Shires, JD Work, Robert Morgus, Patrick Howell O'Neill, Luca Allodi, and Trey Herr (2021): Countering Cyber Proliferation: Zeroing in on Access-as-a-Service and Tim Cushing (2023): NSO Competitor QuaDream Shutting Down After Finding It Can't Make Money If It Can't Sell To Human Rights Abusers and Aleksandra Sowa and Jan Mönikes (2012): Programmier- und Exportverbote für Software?. Inspiration for such a list of governments can be drawn from — among others — arms export restrictions and the Universal Human Rights Index.

101 Bureau of Industry and Security (2020): Entity List

102 David E. Sanger, Nicole Perlroth, Ana Swanson, and Ronen Bergman (2021): U.S. Blacklists Israeli Firm NSO Group Over Spyware

103 Winnona DeSombre, James Shires, JD Work, Robert Morgus, Patrick Howell O'Neill, Luca Allodi, and Trey Herr (2021): Countering Cyber Proliferation: Zeroing in on Access-as-a-Service

104 Sven Herpig (2018): Governmental Vulnerability Assessment and Management – Weighing Temporary Retention versus Immediate Disclosure of 0-Day Vulnerabilities

105 Hotcobalt is an interesting example of an unknown vulnerability that may have been exploited in active cyber defense operations to temporarily deny service to malicious cyber operation or campaign infrastructure and may not have constituted a huge risk when not immediately disclosed to the vendor; see Gal Kristal (2021): Hotcobalt – New Cobalt Strike DoS Vulnerability That Lets You Halt Operations and compare with Florian Roth (2021) on Twitter; for a tool leveraging a different method for the same software, see Mario Henkel (2021): CobaltSpam.

106 White House (2017): Vulnerabilities Equities Policy and Process for the United States Government

107 See for example Winnona DeSombre, James Shires, JD Work, Robert Morgus, Patrick Howell O'Neill, Luca Allodi, and Trey Herr (2021): Countering Cyber Proliferation: Zeroing in on Access-as-a-Service

108 Perri Adams, Dave Aitel, George Perkovich, and JD Work (2021): Responsible Cyber Offense

ponies" and not mature implant platforms. The emulated infrastructure for testing should be as close to identical as possible to the target infrastructure to improve effectiveness and safety[109] but especially to reduce the risk of collateral damage.[110] An extreme level of care must be taken when critical infrastructures may be affected. Due to the heterogeneity of the potentially affected IT systems, creating identical testing infrastructures is a challenge in botnet takedown operations where a wide range of end users is directly affected by the capability. **The capability, therefore, should be as stable as possible, even under changing conditions. An additional step of care would be validation of the capability on the target infrastructure.**

While not necessarily without errors, Stuxnet is known to have been extensively tested before its use.[111] Another example is the malware deployed against the Iranian rail system, where "the payloads appear to suggest substantial quality assurance engineering. Multiple measures to ensure redundancy of key guardrail functions seem to have been deliberately introduced, and the developers apparently sought to provide continuing state of health status so that operators would have positive control throughout delivery and execution."[112]

8. Target with precision

Any active cyber defense operation should be based on careful planning and targeting. The UK government, for example, states with respect to preconditions for its cyber operations that "[t]here are multiple approval stages that consider the feasibility, operational plan, benefits and risks of an operation before it can be authorized."[113] *Perri Adams, Dave Aitel, George Perkovich,* and *JD Work* state that "[r]esponsible actors should carefully select targets, identify any risk of collateral damage, and plan accordingly."[114] Both statements hold true for active cyber defense operations as well.

Active cyber defense operations are, by definition, direct responses to the malicious activities of threat actors. However, it is worth highlighting that proportionality is a key requirement for cyber operations adhering to international law.[115] **Therefore, any active cyber defense operation outside the government's own jurisdiction must be a proportionate response to the malicious activity it aims to mitigate, neutralize, and/or technically attribute.**

109 Perri Adams, Dave Aitel, George Perkovich, and JD Work (2021): Responsible Cyber Offense
110 Regarding challenges of insufficient testing for offensive capabilities, see JD Work (2020): Who Hath Measured the (Proving) Ground: Variation in Offensive Capabilities Test and Evaluation
111 E.g., Jon R. Lindsay (2013): Stuxnet and the Limits of Cyber Warfare
112 JD Work (2021): Balancing on the rail – considering responsibility and restraint in the July 2021 Iran railways incident
113 National Cyber Force (2023): The National Cyber Force: Responsible Cyber Power in Practice
114 Perri Adams, Dave Aitel, George Perkovich, and JD Work (2021): Responsible Cyber Offense
115 Sven Herpig (2021): Active Cyber Defense Operations – Assessment and Safeguards

Active cyber defense operations may target the whole range of threat actors, including but not limited to cyber criminals, intelligence agencies, and military branches. However, the effect of the active cyber defense operation may also take place on the IT systems of the initial victims — see, for example, the Emotet takedown.[116] While — for several reasons outlined earlier — technical reconnaissance and intelligence prior to the operation will hopefully have revealed the nature of the threat, it is not relevant in terms of responsible state behavior based on the operational norms outlined in this analysis.

On the more operational level, target selection must focus on pivot points as close to malicious systems and infrastructures as possible, avoiding uninvolved third-party systems — including supply chains. Performing a collateral damage assessment beforehand and installing effective containment measures are important safeguards in this regard. If third parties are involved, ex post notification should be provided formally or informally — e.g., by deliberately triggering their anti-malware software as a hint. **Additionally, targeting should make sure to avoid infrastructure whose safety is crucial to the functioning of society** — critical infrastructure such as power plants, waterworks, and hospitals. Moreover, in cases where active cyber defense operations target the systems of victims, for example, to remove malware, operators need to avoid further victimizing those targets.

9. This is your last resort

The previous analysis[117] concluded that active cyber defense operations will be a minor part of the broader government cybersecurity efforts, with IT security and resilience taking precedence. The UK government's approach appears to — at least partially — agree with this conclusion. The government states that it is relying on a wide range of activities to counter security threats, including "measures such as cyber resilience, law enforcement action, sanctions, [and] diplomatic intervention," and further stating that "[w]here traditional responses are best placed to deal with the challenge effectively, NCF [National Cyber Force] would rarely if ever get involved."[118] *Tom Uren,* therefore, concludes that "[t]his implicitly recognises the limits of disruptive cyber operations."[119]

Considering these limitations, governments should limit intrusive active cyber defense operations to the necessary minimum and focus on non-intrusive, threat agnostic, scalable, and sustainable measures in blue space. If, however, in specific

116 Andre Meister (2021): Schadsoftware-Bereinigung: BKA nutzt Emotet-Takedown als Türöffner für mehr Befugnisse und neue Gesetze
117 Sven Herpig (2021): Active Cyber Defense Operations – Assessment and Safeguards
118 National Cyber Force (2023): The National Cyber Force: Responsible Cyber Power in Practice
119 Tom Uren (2023): UK's National Cyber Force: A Bunch of Mindf-ckers

cases, this does not suffice, conducting active cyber defense operations responsibly will involve choosing the least intrusive and escalatory measures possible. **These countermeasures must be the most efficient, effective, and proportionate ones available. Additionally, the economic, political, security, and social cost of doing nothing must be greater than the cost of conducting active cyber defense operations.**

Actively Cyber Defend Responsibly

While active cyber defense operations will only be a small fraction of the activities that increase overall cybersecurity, countries around the world are starting to develop policies and experiment with corresponding measures due to the worsening threat landscape. That is the reality, and therefore it is important to get active cyber defense operations right — conducting them as responsibly as possible. Similar to the discourse around offensive cyber operations, it is a choice "[...] whether and how to engage in them responsibly and minimize cost to societies. For there are better and worse ways for governments (and their explicit or de facto contractors) to operate in cyberspace."[120] And while the risk of escalation may be minimal in most cases — as responsible operators are likely to consider global stability as a whole in the best interests of everyone — formalizing a set of operational norms is appropriate.

Operational norms contribute to reflecting responsible state behavior and established values toward other actors in the space. Robust operational frameworks, adherence to international law, and effective communication with partner states — allies and strategic rivals alike — are what set responsible active cyber defense operations apart from others on a general level. On a technical level, operational norms should be centered on capabilities and how they are developed, tested, and used. The appropriate capabilities need to be designed and deployed precisely to achieve exactly what they ought to do and where they ought to do it. Only then can active cyber defense operations contribute to a net improvement in the state of cybersecurity and overall national security. To this end, this paper has presented nine operational norms on both the broader and technical levels.

While several countries and multinational organizations are pivoting toward active cyber defense operations, hackbacks, or counter-cyber operations in policy and practice, it is unclear how many of them already have a robust set of operational norms. A nuanced discourse among those stakeholders, as well as a public debate among researchers and practitioners, will contribute toward better national, regional, and global understandings of active cyber defense. This, in turn, will be a precondition for more responsible state practice.

A government aspiring to conduct active cyber defense operations responsibly, with the goal of improving global cyberspace stability, must establish an active cyber defense policy aligned with the operational norms outlined in this analysis.

120 Perri Adams, Dave Aitel, George Perkovich, and JD Work (2021): Responsible Cyber Offense

About Stiftung Neue Verantwortung

Stiftung Neue Verantwortung (SNV) is a non-profit think tank at the intersection of technology and society. At SNV's core is a methodology of collaborative development of policy proposals and analyses. SNV experts do not work alone – they develop and test ideas together with representatives from politics and public administration, technology companies, civil society and academia. Our experts work independently of interest groups and political parties. We guarantee our independence through diversified financing, comprised of contributions from different foundations, state and corporate actors.

About Transatlantic Cyber Forum

To further policy research in the area of international cybersecurity and provide concrete recommendations, the Stiftung Neue Verantwortung (SNV) established The Transatlantic Cyber Forum (TCF) in January 2017. TCF is an intersectoral network and currently consists of more than 150 practitioners and researchers from civil society, academia and private sector working in various areas of transatlantic cybersecurity policy.

About the Author

Dr. Sven Herpig is Director for Cybersecurity Policy and Resilience. He analyses the roles of government in vulnerability management, law enforcement hacking, responses to cyber operations, securing machine learning, and fostering open source security.

Contact the author

Dr. Sven Herpig
Director for Cybersecurity Policy and Resilience
sherpig@stiftung-nv.de
+49 (0) 30 81 45 03 78 91

Imprint

Stiftung Neue Verantwortung e.V.
Ebertstraße 2
10117 Berlin

T: +49 (0) 30 81 45 03 78 80
F: +49 (0) 30 81 45 03 78 97
https://www.stiftung-nv.de/en
info@stiftung-nv.de

Design:
Make Studio
www.make-studio.net

Layout:
me+Gestaltung