STUDY

# Active Cyber Defense Operations

## Case Studies Cheat Sheets

Sven Herpig, Anushka Shah

July 10, 2024

Tech analysis and policy ideas for Europe

interface

# Stiftung Neue Verantwortung is now interface

Since 2014, our team has worked on building an independent think tank and publishing well-researched analysis for everyone who wants to understand or shape technology policy in Germany. If we have learned something over the last ten years, it is that the challenges posed by technology cannot be tackled by any country alone, especially when it comes to Europe. This is why our experts have not only focused on Germany during the past years, but also started working across Europe to provide expertise and policy ideas on AI, platform regulation, cyber security, government surveillance or semiconductor strategies.

For 2024 and beyond, we have set ourselves ambitious goals. We will further expand our research beyond Germany and develop SNV into a fully-fledged European Think Tank. We will also be tapping into new research areas and offering policy insights to a wider audience in Europe, recruiting new talent as well as building expert communities and networks in the process. Still, one of the most visible steps for this year is our new name that can be more easily pronounced by our growing international community.

Rest assured, our experts will still continue to engage with Germany's policy debates in a profound manner. Most importantly, we will remain independent, critical and focused on producing cutting-edge policy research and proposals in the public interest. With this new strategy, we just want to build a bigger house for a wider community.

Please reach out to us with questions and ideas at this stage.

# Table of Contents

# Introduction

The on-going debates around active cyber defense remain unresolved on both the international and national stages. Given its inherent risks, active cyber defense, 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" (Reference),

merits in-depth scrutiny. While determining if a measure aligns with an existing legal framework serves as a valuable initial consideration, the application of established legal frameworks to emerging domains like active cyber defense poses challenges, especially when specific legal structures are absent. In such cases, it becomes imperative to discern whether pursuing activities that may not necessarily be illegal is warranted. Hence, the question arises: How should we evaluate the implementation of active cyber defense measures? Notably, definitions of active cyber defense exhibit extensive variations, as do the technical strategies encompassed within these definitions. Towards this end, this is a comprehensive framework for the assessment of whether these measures should be put into practice.

The assessment is divided into two categories – Criteria and Safeguards. When calculating the risks for (unintended or cyber-physical, especially in critical infrastructure) damage, fundamental rights violations, violations of sovereignty, conflict escalation and success, there is not a common measurement for active cyber defense. Therefore, criteria include, but are not limited to, the who, against whom, where, to what effect and when of an active cyber defense operation. The criteria and their indicators offer an analytical framework for examining the operation's crucial elements to better assess the operation's risks, usefulness and potential costs. Although these steps are aimed at a case-by-case analysis, governments deciding to implement measures from the broad range that the definition offers also need to implement structural and procedural safeguards and apply them to all active cyber defense operations. These structural safeguards will inter alia guarantee privacy protections, ensure alignment with human rights and national and international law, maintain geopolitical stability and create a net-gain for national security.

- Hafnium-Web-Shell-Removal
- Operation LADYBIRD
- Operation MEDUSA
- Hive Takedown

## Criteria Legend

| Category | Criteria |
|---|---|
| Purpose | **Goal:**<br>An operation can aim to mitigate the impact, neutralize a malicious cyber operation or campaign and/or attribute it technically. |
| | **Success:**<br>To better estimate the success of the effect, it is crucial to look at the expected goal.<br>A possible categorization, connected to the frequency of operations, is whether the operation is likely to result in a tactical or strategic success.<br>In both cases, active cyber defense operations could be labeled during the ex-ante impact assessment. |
| Effect | **Type:**<br>If an active cyber defense operation leads to unintended consequences, reversibility ensures damage control.<br>Whereas the intrusiveness of measures is not necessarily binary but, on a spectrum, it may be useful to divide them into intrusive and non-intrusive methods for operationalizing the framework. |
| | **Space:**<br>Blue space is defined as the area within the jurisdiction of the government, including the private sector among others; Green space is defined as IT systems and infrastructure affected in a malicious cyber operation or campaign that are within the jurisdiction of an allied government; Red space is defined in this paper as IT systems and infrastructure used in a malicious cyber activity that is within the jurisdiction of the country in which the operation or campaign originates; Gray space is defined in this paper as IT systems and infrastructure used in a malicious cyber operation or campaign that are not located in blue, green, or red space. |
| | **Target:**<br>Concrete target systems and infrastructure must be considered and possibly treated differently, especially if the type of effect is intrusive and non-reversible.<br>It is crucial to make a clear distinction between critical infrastructure and non-critical infrastructure, treating targets of active cyber defense operations as critical infrastructure if they would be considered critical infrastructure in the nation conducting active defense.<br>Targeting critical infrastructure with active cyber defense operations should be assessed with utmost care, as it can lead to unintended and even cyber-physical effects and the subsequent risk of escalation. |
| Actors | **Government-led agency:**<br>Several types of government actors are considered for the lead agency in active cyber defense operations, including national cybersecurity agencies, federal and state law enforcement agencies, intelligence agencies, and the military.<br>Considering the effect space, cybersecurity and law enforcement agencies may be appropriate for measures in blue and green spaces, while intelligence agencies may be suitable for measures in gray and/or red spaces. |
| | **Cooperativeness:**<br>Cooperation is important, as the risks and resources involved in exchanging |

| | |
|---|---|
| | information with an affected third party may be lower than attempting to compromise that third party through an active cyber defense operation against their will. |
| Timing | **Attribution:**<br>Attribution through technical intelligence and other means plays a major role in implementing active cyber defense measures.<br>Confidence levels in technical attribution are crucial for several active cyber defense measures, and they may vary from uncertain to proven based on technical, intelligence, and geopolitical evidence.nce. |
| | **Time:**<br>The effectiveness of an active cyber defense operation may decrease as more time passes between the incident(s) and the operation.<br>The timing of active cyber defense operations is crucial to meet necessary self-defense criteria and avoid being seen as pure retribution. |
| Operations | **De and escalation:**<br>Several active cyber defense measures carry the potential risk for escalation.<br>The potential for de-escalation and escalation involves multiple parties and factors, including third parties and the controllability of the active cyber defense measures deployed.<br>Adhering to proportionality is important to maintain legality and avoid escalation.<br>Confidence-building measures, such as open communication channels, can help in avoiding escalation and promoting de-escalation. |
| | **Automation:**<br>Linked to the question of escalation is the level of automation of an active cyber defense operation.<br>Increased automation when targeting heterogeneous systems may lead to unintended consequences and loss of control.<br>How risky an automated process is depends on other criteria, such as the type of effect and the overall picture. |
| | **Frequency:**<br>The overall calculation should factor in the frequency of measures needed to achieve the goal.<br>A one-off operation may be preferable in terms of efficiency and effectiveness.<br>If repetitions of active cyber defense operations are necessary, other measures such as passive cyber defense or a combination may become more effective and efficient. |
| | **Cost:**<br>The costs for an active cyber defense operation in terms of resources are closely linked to the frequency criterion.<br>The cost of specialized staff time, procurement of tools and exploits, and third-party support from the private sector may make an operation prohibitively expensive or inefficient.<br>The overall operation may be rendered inefficient and ineffective in achieving the goal based on the costs criterion. |
| | **Collateral Consequences:**<br>Collateral consequences go beyond simply referring to accidental consequences, as actions may be taken while it is clear that collateral consequences could occur.<br>Planners of active cyber defense operations may either expect (and accept or not accept) or not expect collateral consequences. |

# Safeguard Legend

| Safeguard | Assessment |
|---|---|

| | |
|---|---|
| Define and limit the scope | • One safeguard to be clearly outlined in the warrant and impact assessment is the scope of the operation. It should be as narrow as possible and at least assess whether the targets of an operation are in blue, green, gray or red space.<br>• Ideally, the scope is limited to blue space, as this normally means directly supporting the targets of a malicious cyber operation or campaign within one's own jurisdiction, rather than going after the threat actor. |
| Establish a national legal framework | • Adherence to the rule of law is facilitated by having a clear and specific legal framework for active cyber defense operations.<br>• The legal framework and its elements should be regularly revisited, evaluated and refined in accordance with national sunset clause procedures. |
| Require impact assessments | • A formal ex ante impact assessment is essential to weigh the risks, impact, chances and possible consequences of an operation, develop additional options and back-up plans and define circuit breaker conditions.<br>• Impact assessment is a useful basis for decision-makers and oversight bodies and, therefore, should be a requirement for every active cyber defense operation. |
| Implement oversight | • Active defense operations most likely have extraterritorial implications, touching on sovereign issues of other states. Therefore, a high standard of oversight (ex-ante and ex post) is required and should be enshrined in the legal framework. |
| Create transparency and auditability | • Disseminating, if necessary sanitized, details of active cyber defense operations beyond executive, judicial and legislative authorities allow external experts to independently assess measures and suggest improvements.<br>• A key aspect of transparency as well as oversight is the auditability, performed by independent auditors, of the operations. |
| Set up guidelines for tools and services | • Some measures, however, especially the more intrusive ones, may require procurement of software including exploits and services.<br>• Tools and services must be acquired only from third parties, ideally as open source, that are transparently vetted, and which do not conduct any business with governments or other entities that have been reported to conduct unlawful activities and violate human rights with their tools and services.<br>• Whenever unknown vulnerabilities or exploits are required for an active cyber defense operation, a vulnerability assessment and management should be established beforehand as a safeguard to manage the risks of reverse engineering, exploitation of leaks and/or use against the infrastructure by the threat actors. |
| Apply international law | • The international law framework for countermeasures seems appropriate for active cyber defense operations in red space, although they may not always qualify as countermeasures or reprisals under international law.<br>• The framework includes aspects such as the measures should be proportional, have the right timing, be temporary, be non-retributive and require a certain level of attribution.<br>• Governments leading active cyber defense operations must also be aware that they are contributing to the development of binding customary law. |

| | |
|---|---|
| Consider public interest | • Active cyber defense operations should be conducted only if there is a clear public interest in doing so, such as threats to public safety and security (thresholds may include interference with critical infrastructure or a local declaration of state of emergency), meddling with democratic processes or significant economic harm.<br>• If there is no clear public interest, then other, more passive, measures could be taken to neutralize or mitigate an ongoing cyber operation or campaign. |
| Adapt confidence-building measures | • If states consider active cyber defense operations in green, gray and/or red space, the states should set up international confidence-building measures or adapt existing ones for this purpose.<br>• The minimum viable communication should make sure that the targets of the active cyber defense operation understand ex post that it was a response and, therefore, had defensive intent. |

# Hafnium-Web-Shell-Removal

## Criteria

| Criteria | Indicators |
|---|---|
| Goal | Mitigation |
| Success | Tactical |
| Type | Reversible |
| | Intrusive |
| Space | Blue Space |
| Target | Non-critical Infrastructure<br>(Critical Infrastructure) |
| Government lead agency | Law Enforcement |
| Cooperativeness | Unknown |
| Attribution | Not necessary |
| Time | During operations of the same campaign<br>In-between sequential campaigns |
| De and escalation | (Potential de-escalation)<br>No change in the escalation cycle |
| Automation | Semi-Automated |
| Frequency | One-off<br>(Periodic) |

| | |
|---|---|
| Cost | Low |
| Collateral Consequences | Not expected |

## Safeguards

| Safeguard | Assessment |
|---|---|
| Define and limit the scope | Clearly defined |
| National legal framework | Non-active-cyber-defense-specific legal framework exists |
| Impact assessment | Unclear; at least an independent technical expert was consulted |
| Oversight | Ex ante warrant and ex post notification of targets were required |
| Transparency and auditability | Public was informed with a sufficient amount of information; auditability is unknown |
| Guidelines for tools and services | Unknown process of development/ procurement and testing. Country has a Vulnerability Equities Process. |
| Apply international law | Not applicable (in the sense of between nations) |
| Consider public interest | Appears to have been in the public interest |
| Confidence-building measures | Not required (in the sense of international measures) |

### References

Sven Herpig (2021): Active Cyber Defense Operations – Assessment and Safeguards, Stiftung Neue Verantwortung

# Operation LADYBIRD

## Criteria

| Criteria | Indicators |
|---|---|
| Goal | Mitigation (Neutralization) |
| Success | Tactical (Strategic) |

| | |
|---|---|
| Type | (Reversible) |
| | Intrusive |
| Space | Blue Space |
| | Green Space |
| | (Gray Space) |
| Target | Non-critical Infrastructure<br>(Critical Infrastructure) |
| Government lead agency | Law Enforcement |
| Cooperativeness | Unknown |
| | Noncooperative |
| Attribution | Not necessary |
| Time | During operations of the same campaign<br>In-between sequential campaigns |
| De and escalation | No change in the escalation cycle |
| Automation | Semi-Automated |
| Frequency | One-off |
| Cost | Low |
| Collateral Consequences | Not expected |

# Safeguards

| Safeguard | Assessment |
|---|---|
| Define and limit the scope | Unclear |
| Establish a national legal framework | Applicability of existing non-specific legal framework questionable |
| Require impact assessments | Unclear |
| Implement oversight | Ex ante warrant and ex post notification of targets were required |
| Create transparency and auditability | Public was provided with a minimum of information; auditability unknown |
| Set up guidelines for tools and services | Likely implemented |
| Apply international law | Unknown |
| Consider public interest | Appears to have been in the public interest |

| | |
|---|---|
| Adapt confidence-building measures | Unknown |

## References

Sven Herpig (2021): Active Cyber Defense Operations – Assessment and Safeguards, Stiftung Neue Verantwortung

# Operation MEDUSA

## Criteria

| Criteria | Indicators |
|---|---|
| Goal | Mitigation |
| Success | Tactical |
| Type | Reversible |
| | Intrusive |
| Space | Blue Space |
| Target | Unclear whether critical infrastructure was included |
| Government lead agency | Law Enforcement |
| Cooperativeness | Cooperative<br>Unknown |
| Attribution | Proven |
| Time | In-between sequential campaigns |
| De and escalation | Potential de-escalation<br>No change in the escalation<br>Proportional |
| Automation | Semi-Automated |
| Frequency | One-off |
| Cost | High |
| Collateral Consequences | Not expected<br>Known Unknown |

## Safeguards

| Safeguard | Assessment |
|---|---|
| Define and limit the scope | Clearly defined |
| National legal framework | Non-active-cyber-defense-specific legal framework exists |
| Impact assessment | Unclear |
| Oversight | Ex ante warrant and ex post notification of targets were required |
| Transparency and auditability | Public was informed with a sufficient amount of information; auditability is unknown |
| Guidelines for tools and services | Unknown process of development/ procurement and testing. Country has a Vulnerability Equities Process. |
| Apply international law | Not applicable (in the sense of between nations) |
| Consider public interest | Appears to have been in the public interest |
| Confidence-building measures | Not required (in the sense of international measures). However, The FBI also alerted local authorities in other countries to take down Snake infections on compromised machines outside the United States.) and published an in-depth technical report .... To explain why they did it and that it was a response to longlasting, damaging operation. |

## References


- Anna Ribeiro (2023): DOJ executes court-authorized disruption of Snake malware network controlled by Russia's FSB
- Cybersecurity and Infrastructure Security Agency (2023): Hunting Russian Intelligence "Snake" Malware
- United States Department of Justice (2023): Justice Department Announces Court-Authorized Disruption of the Snake Malware Network Controlled by Russia's Federal Security Service
- Robert Legare (2023): FBI takes down Russia's sophisticated 20-year-old malware network known as "Snake"

# Hive Takedown

## Criteria

| Criteria | Indicators |
|---|---|
| Goal | Mitigation<br>Neutralization |
| Success | Strategic<br>Tactical |
| Type | Reversible |
| | Intrusive |
| Space | Blue Space<br>(unknown) |
| Target | Non-critical Infrastructure |
| Government lead agency | Law Enforcement |
| Cooperativeness | Cooperative |
| Attribution | Proven |
| Time | During operations of the same campaign |
| De and escalation | No change in the escalation |
| Automation | Manual |
| Frequency | One-off |
| Cost | High |
| Collateral Consequences | Not expected |

## Safeguards

| Safeguard | Assessment |
|---|---|
| Define and limit the scope | Clearly Defined |
| Establish a national legal framework | Non-active-cyber defense-specific frameworks exits |
| Require impact assessments | Unclear |

| | |
|---|---|
| Implement oversight | Ex ante warrant |
| Create transparency and auditability | Public was informed with a sufficient amount of information; auditability is unknown |
| Set up guidelines for tools and services | Unknown process of development |
| Apply international law | Not applicable<br><br>Law Enforcement Agencies in the United States, Germany and the Netherlands took action against infrastructure based in their countries. |
| Consider public interest | Appears to have been in the public interest |
| Adapt confidence-building measures | Coordination with EU Member States and Five Eye Countries |

## References


- United States Department of Justice (2023): U.S. Department of Justice Disrupts Hive Ransomware Variant
- Lawrence Abrams (2023): Hive ransomware disrupted after FBI hacks gang's systems

# Authors

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

Anushka Shah

# 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.
Ebertstraße 2
D-10117 Berlin



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