The Hague Centre for Strategic Studies

CyberPeace Institute

Cyber Transparency Value Chain

From Awareness and Understanding to Attribution, Monitoring and Sanctioning

Open Source Monitors and Observatories

The Hague Centre for Strategic Studies (HCSS)
CyberPeace Institute
January 2024

Cyber Transparency Value Chain

From Awareness and Understanding to Attribution, Monitoring and Sanctioning

Open Source Monitors and Observatories

Authors: The Hague Centre for Strategic Studies (HCSS), CyberPeace Institute

January 2024

Cover photo: Unsplash

This report is the result of independent research. The responsibility for the content lies entirely with the authors. The study was commissioned by CyberPeace Institute.

The Hague Centre for Strategic Studies (HCSS)
Website: www.hcss.nl

The CyberPeace Institute
Website: cyberpeaceinstitute.org

Table of Contents

About the Authors

The Hague Centre for Strategic Studies (HCSS)

HCSS conducts research on geopolitical, defence & security issues to governments, international institutions and businesses. Our research is characterised by a data driven, multidisciplinary approach, specialist expertise and a strategic orientation. We combine broad, conceptual knowledge with qualitative and quantitative methods and present our findings in the form of recommendations, strategic explorations and scenario analyses.

CyberPeace Institute

The CyberPeace Institute is an independent and neutral non governmental organisation (NGO), whose mission is to ensure the rights of people to security, dignity and equity in cyberspace. The Institute works in close collaboration with relevant partners to reduce the harms from cyberattacks on people's lives worldwide. By analysing cyberattacks, the Institute exposes their societal impact, how international laws and norms are being violated, and advances responsible behaviour to enforce cyberpeace.

Do you want to know more?

If you are interested in this in this initiative and want to know more about it? Please visit the website www.cybertransparency.org or www.hcss.nl and www.cyberpeaceinstitute.org

The Hague Centre for Strategic Studies

CyberPeace Institute

Executive Summary

Transparency in cyber space is widely seen as an important mechanism to build confidence, to prevent inadvertent incidents and damage and to increase accountability of those who seek to misuse cyberspace. The (im-)plausible deniability and lack of transparency on attribution of cyberattacks limits diplomatic efforts both track 1 (formal negotiations between states) and track 2 (non-governmental conflict resolution), as discussions lack a sound basis for common exchange.

The Hague Centre for Strategic Studies (HCSS) and the CyberPeace Institute are working together to increase cyber transparency, to inform policy processes and capacity building efforts, and contribute to accountability efforts. This Report provides an overview of the monitors and observatories developed to date by each organisation, which evolve as the threat and policy landscape evolves.

To support and enhance transparency specific online monitors and analysis are published on the cyber threat landscape to enable a broader understanding of threat actors, the impact and harms of cyber incidents, and accountability efforts of actors, including sanctioning of actors for breaches of laws and norms in cyberspace.

Based on our analysis we came to the following key points:

- Why transparency matters?
    - States should further engage in focused discussions with stakeholders regarding existing and potential threats that present systemic risks to sectors and services deemed essential. While recognising the need to share objective information on cyber threats in the context of international security, States should further tap into the potential information that can be provided by civil society, industry, and academia. As States have expressed concerns that *"a lack of awareness of existing and potential threats and a lack of adequate capacities to detect, defend against or respond to malicious ICT activities may make them more vulnerable"*, (OEWG, Second Annual Progress Report, para. 15[1]. There is an opportunity to tap into the work of private companies, research and civil society organisations, among others, which have a proven track record in analysing the threat landscape in a neutral and transparent way, creating repositories of cyber incidents, collecting and investigating such incidents, and mapping their impacts.
    - Civil society organisations play a key role in providing input on the cyber threat landscape, including on issues such as the impact of cyberattacks on human rights, safety and security of people, and implementation challenges of the agreed norms in practice. A multistakeholder approach is key for building a global culture of cybersecurity and sustainable operationalization of the framework.
    - States need to build data-driven understandings of the harm inflicted by cyberattacks. Thematic exchanges with stakeholders should be encouraged to foster context-aware approaches to tackling the malicious use of cyber. Civil society

[1] OEWG Second Annual Progress Report, Rev. 1 Draft of 12 July 2023 (hereafter referenced as the "Second Annual Progress Report"), para. 10. The APR is a part of the "Letter from the OEWG Chair," July 12, 2023, available at: https://docs-library.unoda.org/Open-Ended_Working_Group_on_Information_and_Communication_Technologies_-_(2021)/Letter_from_OEWG_Chair_12_July_2023_(technical_re-issue).pdf

organisations and academia have acquired relevant expertise both through their work with affected individuals and communities and conducting research into the differentiated harm stemming from cyber incidents that different groups of people may experience, e.g., based on their gender or factors of vulnerability.

- Some of the monitors fit in several categories of the Cyber Transparency Value Chain. Uptil now the decision was made to include the monitors where the most work has been carried out to date, e.g. Awareness and Understanding. On Monitoring and Sanctions until now little analysis is so far done. The ambition however is to also start working on these parts of the transparency value chain. Support to make this possible is welcomed. Collaboration and complementarity of the monitors of HCSS and the CyberPeace Institute is important as we can further identify where to add value. Both institutes look for further collaboration options inviting beneficiaries to enable us to continue doing so. E.g. on cross referencing work on HCSS Cyber Norms Observatory and CBMs policy work of the CyberPeace Institute.

- There are a number of key focus areas for potential collaboration and where we have specific asks: funding, data access, to address constraints identified in report, concrete feedback on how monitors are being used for Track 1 and Track 2 diplomacy (even if this cannot be made public) as this will strengthen our ability to continue the evolution of this work.

  - Cyber Arms Watch ambition - funding to sustain the platform, Sounding Board, yearly analysis/update report, data on 60 countries to cover blindspots, data on new group of countries, language capabilities.
  - Cyber Comparator ambitions - resources to sustain the platform, for updating and extension of the monitor adding new data categories, and extra analysis functions.
  - Cyber Norms Observatory - resources to sustain the platform, carry out analysis on various norms regimes, update and extend to Observatory with newly developed norms.
  - CyberPeace Watch ambition - funds and resources (expertise, data collection) to carry out continued analysis of laws and jurisprudence related to accountability measures against threat actors for the Platform which will launch in 2024. Collaboration on the development of the Harms Methodology to finalise the methodology, analysis of cases, indicators and metrics.
  - Cyber Attacks in Armed Conflict Platform #Ukraine - funding to sustain the platform and to analyse data for on harms and on threat actors for accountability processes, including potential submissions of cases to ICC or regional courts and for potentially monitoring cyber incidents subject to potential future ceasefire agreement.
  - Cyber Incident Tracer #Health - resources and data to resume data collection and analysis particularly of harm to people and society related to cyber incidents for the development of the Harms Methodology.

Potential for collaboration

If your organisation is interested in collaborating please contact us at info@hcss.nl and info@cyberpeaceinstitute.org.

Cyber Transparency, an International Initiative

What is cyberspace?

Cyberspace is the interdependent network of information technology infrastructures and includes the internet, telecommunication networks, computer systems and embedded processors and controllers in critical industries. In general, the term also refers to the virtual environment of information and interactions between people. The functioning of cyberspace is important for all human activity, be it business, education or social interactions. At the same time it is also used for cyber operations and attacks including by nation state actors and their proxies, hacktivists, collectives and cyber-criminal groups.

Digital security requires an ongoing and complex balancing act between diverse interests, digital threats and digital resilience. Many parties are making efforts to increase digital resilience against the malicious use of cyber and increase the security against cyber incidents.

The cyber and digital regulatory environment is continually evolving, and there are several ongoing negotiations in diplomatic and policy fora to develop and/or clarify laws and norms to regulate and govern behaviour in cyberspace. Governments and organisations are striving to make cyberspace more safe and secure. However, cyber incidents and threats, including from cyber influence operations, will continue to dominate the landscape at international, national and individual levels.

State and non-state actors threaten cyberspace infrastructure[2], data and users for political and ideological purposes, espionage, to disrupt, exploit and for criminal gain.

In view of the range of different actors operating in cyberspace and engaging in cyber operations and attacks, it is important to analyse and document the Tactics, Techniques and Procedures employed, as well as attribution efforts of and the harms caused to people and society. This is important as it is often not clear which actors are active, what their targets and intentions are, their *modus operandi* and the effects and damage they cause.

Transparency is widely seen as an important mechanism to build confidence, to prevent inadvertent incidents and damage and to increase accountability of those who seek to misuse cyberspace. The (im-)plausible deniability and lack of transparency on attribution of cyberattacks limits diplomatic efforts both track 1 (formal negotiations between states) and track 2 (non-governmental conflict resolution), as discussions lack a sound basis for common exchange.

[2] CyberGreen Institute's Internet Infrastructure Health Metrics Framework (IIHMF) is a set of models and metrics for measuring the "public health" of internet infrastructure. It allows nations to measure their overall risk, understand how it changes over time, and take steps to mitigate that risk for their citizens. CyberGreen scans the internet to detect vulnerabilities that exist within the cyber ecosystem in order to compile data and statistics. CyberGreen has devised and publishes its own scanning method for transparency and to maintain trust. https://cybergreen.net/datametrics/

Aim

HCSS and the CyberPeace Institute are working together to increase cyber transparency, to inform policy processes and capacity building efforts, and contribute to accountability efforts. This Report provides an overview of the monitors and observatories we have developed to date, which evolve as the threat and policy landscape evolves.

Objective

HCSS and the CyberPeace Institute are working together to increase cyber transparency, to inform policy processes and capacity building efforts, and contribute to accountability efforts. This Report provides an overview of the monitors and observatories we have developed to date, which evolve as the threat and policy landscape evolves.

Methodology

HCSS and the CyberPeace Institute are using the approach of the Cyber Value Chain, an architecture for collecting and communicating about Cyber Transparency which enables the identification of where the various monitors add value, and where more analysis is needed.

In the Cyber Value Chain the main focus is at country-level and the State as the entity that officially has the authority to adopt national laws and regulations, and to enter into relations with other states including to negotiate and adopt international laws and norms.

- The chain starts with **Awareness**, trying to build systemic knowledge and transparency about details of the cyber threat landscape, harms caused to people and society of cyber incidents, and to providing clarity on the offensive cyber capabilities of countries;

- The next link is **Understanding**, enhancing knowledge through the provision of evidence and facts to enable actionable engagement;

- **Attribution** is next in the chain, and this is an activity with multiple complexities.

    - A technical attribution is often carried out by cyber security companies through the publication of analytical reports determining who or what is responsible for an attack based on the analysis of technical artefacts (e.g. through forensics analysis). A core step in this process is associating the attack to specific software (e.g. malware strain), hardware (e.g. a server), code or modus operandi.
    - Political attribution: determining or disclosing who is the party(s) responsible for an attack (such as a nation state, State-sponsored group, or criminal group) by a State based on analysis, assessment and/or judgement. States have to balance the protection of national interests with maintaining diplomatic relations with other states, and showing that they are aware of attacks whilst not exposing their full operational capabilities. States are often reluctant to attribute attacks because of the corresponding obligations that it would impose on them, e.g. attribution under international law triggers State responsibility, may trigger the application of international humanitarian law, and/or a self-defence response. These dilemmas

make it useful for States to refer to independent sources of analysis that support transparency.

- Legal attribution: determining who is responsible for an attack based on technical means to identify the origin of the attack and legal criteria in order to ascribe legal consequences and/or other sanctions (e.g. through a court of law or through the application of sanctions). Attribution of a cyberattack under international law may trigger the application of IHL, State responsibility, and/or a response in self-defence.
- Self-attribution: some threat actors publicly disclose a cyberattack and attribute themselves as the actor behind the attack. Although not as formal a category of attribution as the other three, it remains one of the ways in which actors involved in a cyber incident are documented.

- The **Monitoring** link in the chain is about observing how states behave in line with their obligations[3] to respect and ensure respect for laws, whether they violate existing legal instruments and their respect of norms.

- The final link is **Sanctioning** - the economic, financial and legal measures taken against individuals or entities responsible for malicious cyber activities - takes place in various jurisdictions (be it national, regional or international) and thus is the most difficult aspect of the cyber transparency value chain. It entails compliance with international law and norms of state behaviour, including due diligence, a general international obligation for every State not to knowingly allow its territory to be used for internationally wrongful acts using cyber means.

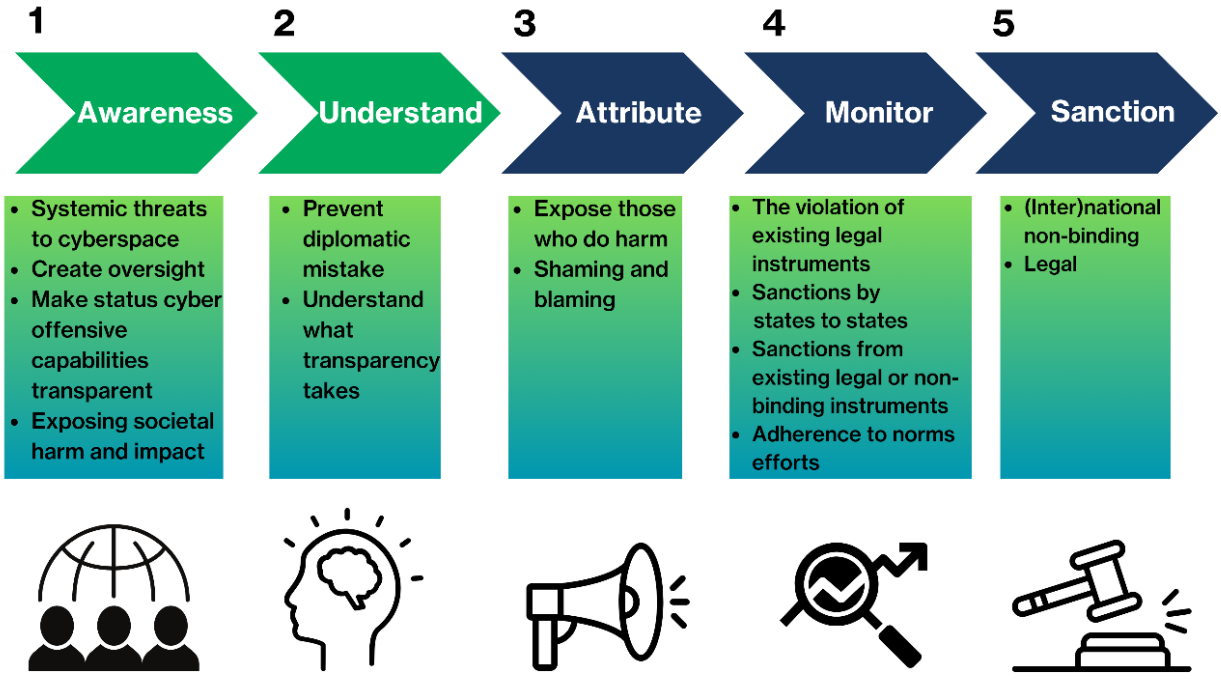


Figure 1:The Cyber Value Chain

[3] The United Nations Institute for Disarmament Research (UNIDIR) Cyber Policy Portal is an online reference tool - interactive map - of the global cyber policy landscape. It provides profiles of the cyber policies of all 193 UN Member States, in addition to various intergovernmental organisations and multilateral frameworks. This confidence building tool seeks to support informed participation by relevant stakeholders in all policy processes and promote trust, transparency, and cooperation in cyberspace. Most of the data on the Portal is compiled from official and publicly available sources, primarily traced back to the official documentation disseminated by the State or intergovernmental organisation.

Cyber Value Chain Transparency, In Practice

Building Awareness: Monitoring Cyber Threats

HCSS and the CyberPeace Institute leverage our expertise and capabilities to build platforms with carefully selected data collection, scope and associated indicators to build awareness. Collecting, analysing and publishing data and information about cyber capabilities and threat landscapes is challenging and requires resources to source, filter and analyse the range of available data and information systematically and consistently over time, provide relevant indicators/indices and collate them in a manner to enable patterns, trends and insights for audiences.

Several monitors are maintained to enable understanding of the cyber threat landscape.

Cyber Attacks in Times of Conflict Platform #Ukraine

Rationale: *better understand the harm and impact of cyberattacks and operations used in an international armed conflict.*

The CyberPeace Institute collects publicly available information on cyberattacks and operations in the context of the ongoing war between Ukraine and Russia in order to better understand the harm and impact of cyberattacks and operations, and who is responsible for them. The data driven Platform provides accessible and filterable data visualisations of observed trends and patterns and is complemented with a quarterly analysis report. Data on cyberattacks can be expanded to explore specific details relating to the incident itself, the threat actor, the targeted country and sector, and the harm related to the incident.

Data collection relates to cyber incidents in the context of the Russian-Ukrainian war including incidents in Ukraine, the Russian Federation and more than 50 other countries. Collection is concentrated on, but not limited to, incidents targeting and/or impacting civilians, civilian objects including private companies, and some 23 infrastructure sectors ensuring the delivery of essential services to civilians. The Platform also endeavours to document the hidden impacts that these attacks have on people and society.

Cyberattacks and operations are any incident conducted by a threat actor using a computer network or system with the intention to disrupt, disable, destroy, control, manipulate, surveil or extract a computing environment/infrastructure and/or data. Incidents documented to date include, but are not limited to, malware including wiper malware, distributed denial of service (DDoS), malspam, influence/information operations, hack and leak, account takeover and website defacements. As there is a reliance on publicly available data, the data on documented cyberattacks has been given a classification of certainty based on the reliability of the information source.

Figure 2: Monitoring the harm to civilians from cyberattacks, in the quest for cyberpeace.

The Institute does not conduct its own attribution of incidents to identify the actor(s) involved but documents the attribution efforts by others to link a particular individual, group or state to a specific incident. Self-attribution has been a particular feature of this conflict, with more than 110 threat actors publicly disclosing a cyberattack and attribute themselves as the actor behind the attack. They often do this by publishing data extracted as a result of an incident on dedicated websites.

The Platform also provides information on the legal and normative ecosystem related to cyberattacks and operations deployed during an international armed conflict, and specific challenges related to the application of laws and norms in cyberspace.

Our Ask

Funding to sustain the platform and to analyse data for on evidence harms and on threat actors for accountability processes, including potential submissions of cases to ICC or regional courts and for potentially monitoring cyber incidents subject to potential future ceasefire agreement.

Cyber Incident Tracer #HEALTH

Rationale: *provide visibility of the impact of cyberattacks against the healthcare sector (during the COVID-19 pandemic).*

The Cyber Incident Tracer (CIT) #HEALTH is a data platform that presents cyberattacks on the healthcare sector which disrupt the delivery of healthcare and subsequent impact on patients, healthcare professionals and facilities. This platform bridges the information gap about cyberattacks on the healthcare sector and their impact on people. The platform brings greater visibility to the scale of the problem and how such attacks impact people and the provision of care.

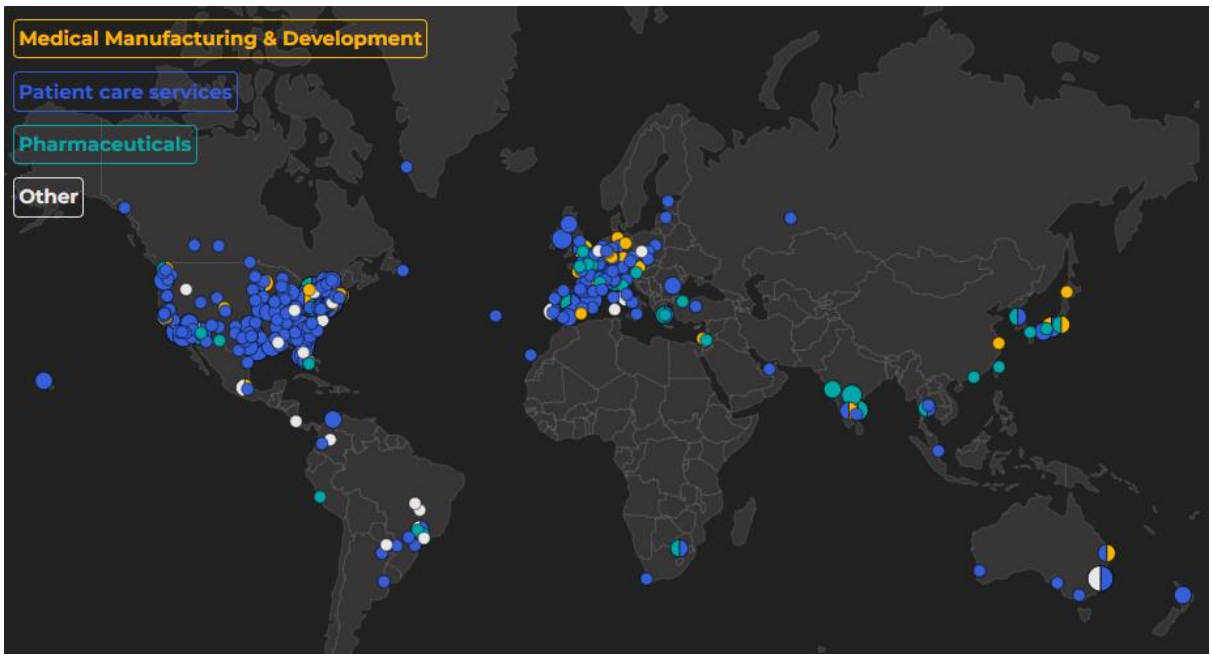


Figure 3: The Cyber Incident Tracer (CIT) #HEALTH screenshot

Data in the CIT #HEALTH is collected from publicly available sources. The CyberPeace Institute cross references information about a cyberattack to ensure accuracy. Data sources are categorised according to the proximity of the reporting source to the target organisation:

- Primary sources: firsthand information provided by the target organisation or reporting by an official public entity;
- Secondary sources: retrospective or ongoing reporting in the form of articles or blog posts;
- Tertiary sources: summarising or aggregating sources, such as data sets or reports.

The CyberPeace Institute developed CIT #HEALTH following its Strategic Analysis Report released in March 2021, entitled Playing with Lives: Cyberattacks on Healthcare are Attacks on People. This report made several key recommendations related to the existing data gap on cyberattacks against the healthcare sector, and the CIT #HEALTH platform is a step to remedy this issue by improving visibility of such attacks.

The data is displayed through filterable data visualisations and an aggregated searchable data table in which each incident can be expanded to explore specific details relating to the sources, incident itself, the targeted organisation, and the impact of the incident.

The CIT #HEALTH contains information on disruptive cyberattacks targeting the healthcare sector from June 2020 to September 2022. The last update was done on 30.09.2022 in order to concentrate resources on the Cyber Attacks in Times of Conflict Platform, and other tracers. Any support that can be provided to get a more thorough understanding of the frequency and harm of cyberattacks against the healthcare sector and to help protect patients against this threat is welcomed.

Our Ask

Resources and data to resume data collection, processing and analysis particularly of harm to people and society related to cyber incidents for the development of the Harms Methodology.

CyberPeace Watch

Rationale: *build understanding and share knowledge on the state of cyber peace and measures for greater accountability.*

The CyberPeace Watch (platform launching Q1, 2024) is an interactive online platform that provides a publicly accessible baseline of data to understand and share knowledge about cyberattacks, including threat analysis, societal harm, applicable laws and norms, and related paths for accountability.

Through this Platform, the CyberPeace Institute aims to provide a baseline of understanding and shared knowledge on the status of cyber peace and enable a mechanism to track this. The Platform's goal is to assess cyber peace based on evidence of the societal harm caused by cyberattacks and the actions taken by states and other relevant actors to strengthen responsible behaviour in cyberspace, and to empower victims of incidents in accountability processes. The Institute will publish a report on its developing Harm Methodology - a data and evidence driven methodology to measure the harms and impacts of cyberattacks on people, society and the environment, in December 2023.

In 2024, the CyberPeace Watch will focus on the development and use of the harms methodology and the monitoring of the cyberattacks and incidents carried out by more than 100 specific threat actors, and on the use of commercial spyware, in order to build awareness of harms and impact. The CyberPeace Watch will progressively provide a comprehensive analysis of systemic threats to cyber peace that arise in peacetime and wartime.

Our Ask

Funds and resources (expertise, data collection) to carry out continued analysis of laws and jurisprudence related to accountability measures against threat actors for the Platform which will launch in 2024. Collaboration on the development of the Harms Methodology to finalise the methodology, analysis of cases, indicators and metrics.

Cyber Incident Tracer #NGO

Rationale: *carry out research and analysis of cyber threats to vulnerable populations in order to build awareness, support capacity building and resilience.*

As the volume type and sophistication of cyberattacks are increasing year on year, it is important that vulnerable communities are able to access expert support to analyse threats against them and to receive data-driven insights to mitigate risks. The CyberPeace Institute is monitoring and analysing cyberattacks against non-governmental organisations in the humanitarian and development sectors. The Institute detects, investigates and analyses cyberattacks against NGOs and shares actionable threat intelligence with the NGO community. As well as providing free cybersecurity support to some 200 NGOs to build cyber preparedness and resilience, the Institute develops standards, fosters multi-stakeholder collaboration, and advocates for the protection of the humanitarian sector through its Humanitarian Cybersecurity Center.

The Institute provides independent and evidence-based insights on how vulnerable communities are targeted and harmed by malicious activities in cyberspace. r Assistance to NGOs and legal and policy contributions are intrinsically linked with the ability to deliver unique data-driven analysis. The

Institute makes its data and analysis freely accessible to raise awareness of the harm of cyberattacks and to be used in further research.

The Institute undertakes analysis across the three levels of cyber threat intelligence:

- Tactical analysis – micro-level data analysis to detect and monitor cyber threats to vulnerable communities.
- Operational analysis – investigations to identify how attacks have unfolded.
- Strategic analysis – macro-level research and analysis to identify who is behind cyberattacks targeting vulnerable communities, the motive(s) behind them and the harm they cause.

The CyberPeace Institute launched in November 2023 a unique data driven report regarding the cyber threat landscape of NGOs in "International Geneva", providing actionable recommendations. Further reports related to the cybersecurity posture of NGOs in other cities will be released in 2024. This will be complemented by a Cyber Incident Tracer #NGO in 2024 providing an accessible data platform regarding cyber threats affecting the humanitarian and development sectors.

Our Ask
Resources and data to resume data collection and analysis.

Cyber Arms Watch

Rationale: *counter the lack of transparency of cyber offensive capabilities.*

Conflict between states has taken on new forms, and cyber operations play a leading role in this increasingly volatile environment, earning them a top spot among states' most critical security concerns and fears in diplomatic circles that tensions in cyberspace are escalating. There is strategic competition in cyberspace, and States are increasingly using cyber defensively and offensively, generally with operations considered as below the threshold of an armed attack or armed conflict. States do so directly, through e.g. in-house intelligence services or military assets, or indirectly, through sponsoring affiliated groups performing a cyber operation. This contestation heightens the risk of miscalculation. Avoiding "inadvertent escalation" remains a significant challenge in cyberspace. *[According to the Council on Foreign Relations, 52 states are suspected of directly sponsoring cyber operations since 2005.[4]]*

A major contribution to the uncertainty in cyberspace is the lack of transparency of offensive cyber capabilities of states, as well as an understanding of the actual capabilities this requires (see PETIO[5] framework), and no common understanding of what is considered as "cyber force".

Unlike other military capabilities, offensive cyber capabilities are largely treated as dark secrets[6] from the espionage world. With offensive cyber capabilities, States are left guessing as to the overall capability of another state (albeit at varying degrees of detail) without, for the most part, being able to detail the exact order of battle, table of equipment, tactics, techniques, procedures or other basic

[4] Council of Foreign Relations, "Cyber Operations Tracker", last accessed in May 2023.
[5] People, Exploits, Toolset, Infrastructure and Organizational Structure (PETIO) from "No Shortcuts Why States Struggle to Develop a Military Cyber-Force", Max Smeets.
[6] Espionage is required for targeting and may be carried out for a long time before an attack happens. Espionage is part of state craft and cyber capabilities are both used for espionage and for attack.

information – unless the intelligence assessment is very complete.[7] This secrecy has implications not only for intelligence and national security assessments, but also for the lack of transparency which impedes intergovernmental and institutional dialogue and public discussion on international peace and security in cyberspace.[8] It is important to have a common language on offensive cyber capabilities and transparency to reduce the scope for misunderstanding, provide clarity of intent and predictability, and to help establish norms of restraint and communication – all essential ingredients for stability.

Objective

The Cyber Arms Watch offers a transparency index on the offensive cyber capabilities of 60 states. It compares the degree to which states are transparent about their declared cyber capabilities to outside perceptions of those capabilities. The Cyber Arms Watch aims to contribute to international peace and security by developing the first iteration of a "Cyber Transparency Index" that offers insights into stated and perceived offensive cyber capabilities of 60 states. The results are visualised as an interactive world map monitor[9], offering diplomats, policy makers, academics and researchers alike a one-stop full access to the data and insights into the current state of transparency in offensive cyber capabilities.

Academic research has shown time and time again that transparency about "new weapons" helps reduce the scope for misunderstandings, provides for clarity of intent and predictability, and helps establish norms of restraint and communication – all essential ingredients for stability. Finally, more transparency would enable public, media, and academic discussions to be closer to reality.

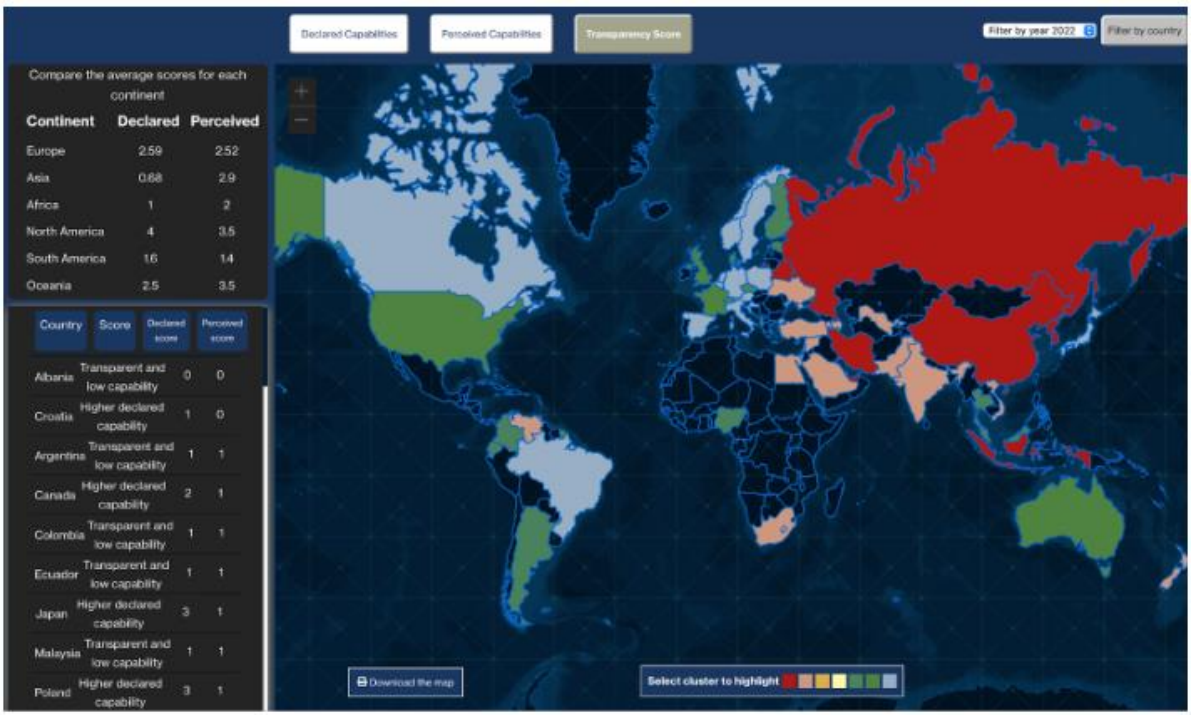


Figure 4 The Cyber Arms Watch monitor showing the results of the Cyber Transparency Index

[7] Alexander Klimburg and Louk Faesen, "Balance of Power in Cyberspace," in Dennis Broeders and Bibi van den Berg (ed.), "Governing Cyberspace: Behavior, Power, and Diplomacy" (2020).

[8] Track 1 (formal negotiations between States) and Track 2 (conflict resolution efforts by conflict resolution practitioners) require transparency and a common language. The absence of transparency frustrates meaningful progress for predictability, confidence-building measures (e.g. within regional organisations such as ASEAN and the OSCE), norms of responsible state behaviour (e.g. within the United Nations processes), and other stability measures. The lack of transparency also impacts and limits the wider public discussion: the absence of information means that much of the public, media, and academic discussion is not in sync with reality and risks becoming irrelevant.

[9] Inspired by the Freedom House Index.

Traditional arms control efforts have depended upon the ability to count weapon systems and/or to impose limitations through international treaties. In relation to control of cyber capabilities, the situation equates to that of the negotiations in the 1950s in relation to nuclear weapons.

The Cyber Arms Watch enables the determination of an overall "Cyber Transparency Index" by using two specific ratings:

- The **Declared Capabilities Rating (DCR)** indicates to what extent a state publicly officially discloses information about its offensive cyber capabilities through e.g. official communication on strategies, doctrines, and authorised media reports. This cumulatively indicates the level of declared capability using a seven-tiered labelling system (see Table 2). The classification ranges from no official indications of offensive cyber capabilities, to stated aspirations, disclosed tactical effects (low and high) and declared strategic effects (low and high).

- The **Perceived Capabilities Rating (PCR)** indicates the perceived offensive cyber capabilities of a state using open-source information and categorises them using a similar seven-level categorization system. This rating uses disclosures from external sources which indicate the offensive cyber capabilities of a state, including intelligence reports and assessments from other governments or non-state actors, indictments, sanctions, past operations, leaked documents, etc.

Figure 5: Result of a country comparison using the Cyber Arms Watch monitor

Level	Abstract	Declared Capabilities Rating (DCR)	Perceived Capabilities Rating (PCR)
0	Nothing	No official indications of Offensive Cyber Capability	No indications of actions to obtain offensive cyber capabilities
1	Aspirations	Stated aspiration for offensive cyber capability	Perceived to have intent to or be working on obtaining offensive cyber capabilities
2	Emerging capabilities	Some form of offensive capabilities referenced in national documents or authorised media reporting, but details are not disclosed.	Perceived to have developed some form of cyber capability, but details are unavailable. Furthermore, these capabilities cannot be linked to the military or intelligence structure of a state.
3	Low tactical effects	Disclosed the use of offensive cyber means at the tactical unit level, including CEMA, but does not provide insights into methods, missions or conditions of deployment.	Perceived to possess offensive cyber capabilities at the military tactical level, but no evidence of cyber effects is identifiable.
4	High tactical effects	Provides details on used methods and effects of offensive cyber at the tactical level, including CEMA, and transparency on military exercises at the tactical level.	Perceived to have integrated offensive cyber capabilities at the tactical level of its military structure and use these capabilities either in exercises or operations.
5	Low strategic effects	National documents mentioning a designated unit integrated within the military structure specifically tasked with cyber operations beyond the tactical level, e.g. a cybercommand. Specific details on size or capabilities are not provided.	Perceived to have integrated offensive cyber capabilities into their military structure at the strategic level, but no evidence of attribution to offensive cyber operations is displayed. Can also be perceived to have developed 'branched-in' spyware and used it against foreign actors for strategic purposes.
6	High strategic effects	National documents mention a designated unit integrated within the military structure specifically tasked with cyber operations beyond the tactical level, e.g. a cybercommand, with specific details on size and capabilities or conducted operations.	Perceived to have integrated offensive cyber capabilities into its military structure at the strategic level with specific details of strategic cyber operations linked to it.

Table 1: The seven-tiered labelling system for the Cyber Arms Watch

Constraints

Offensive cyber capabilities are very difficult to measure objectively due to the ambiguity, uncertainty, and duality inherent in cyberspace. Specifically, in relation to the HCSS methodology used, constraints are due to:

- Overview and definition overload: the methodology has been developed by HCSS with an effort to have distinct definitions although the cyberspace landscape is constantly evolving.
- Ambiguous language: terminology used in the various strategies and documents are subject to interpretation in order to categorise them in the methodology.
- Language limitations: HCSS is only analysing materials available in English.
- Data availability and bias: the number of countries analysed is subject to limitations on the availability of open source materials and the bias that this engenders. HCSS analyses materials from a liberal democratic standpoint and according to the rating methodology outlined above to ensure transparency and fairness.
- State and state-proxy relations: it is not always clear whether a state is deploying or directing a non-State threat actor as a State may obfuscate the extent of the relationship. To address this problem, the CAW only list those incidents in which a direct relationship between the state and the state-proxy is proven or considered highly likely. For every case, the CAW indicates the extent to which the source believes that a link between the two exists. If a link cannot be proven, but it is based on a suspicion, the CAW will log the incident, without attaching a score to it. The incident is added to provide context, but does not feed into the overall score of that country.

Our Ask

Funding to sustain the platform, setup and maintain an international Sounding Board enabling yearly analysis and update report, data on 60 countries to cover blindspots, data on a new group of countries, and extending the monitor doing analysis in more languages.

Sounding Board: If you are interested in helping funding to create more robustness of the monitor for example focusing on countries you are familiar with by e.g. helping review strategy documents using the methodology of the Cyber Arms Watch, or if you have other suggestions please contact info@hcss.nl

Cyber Comparator

Rationale: *Compare cyber capabilities of different countries. Users can easily assess the status of various aspects of cyber infrastructure and compare multiple countries.*

The HCSS Cyber Comparator enables comparison of different countries through a comprehensive overview of cyber infrastructure, strategy documents and other data on a country level. By selecting various data points, derived from open sources, users are able to rank countries across the globe on certain aspects, or build a detailed comparison on a selected number of countries across multiple variables in one overview.

Constraints

The Cyber Comparator Monitor is initially showing a limited set of data. More data could be integrated over time.

Our Ask

Resources to sustain the platform, for updating and extension of the monitor adding new data categories, and extra analysis functions.

If you have suggestions to enhance the monitor with extra datasets contact info@hcss.nl.

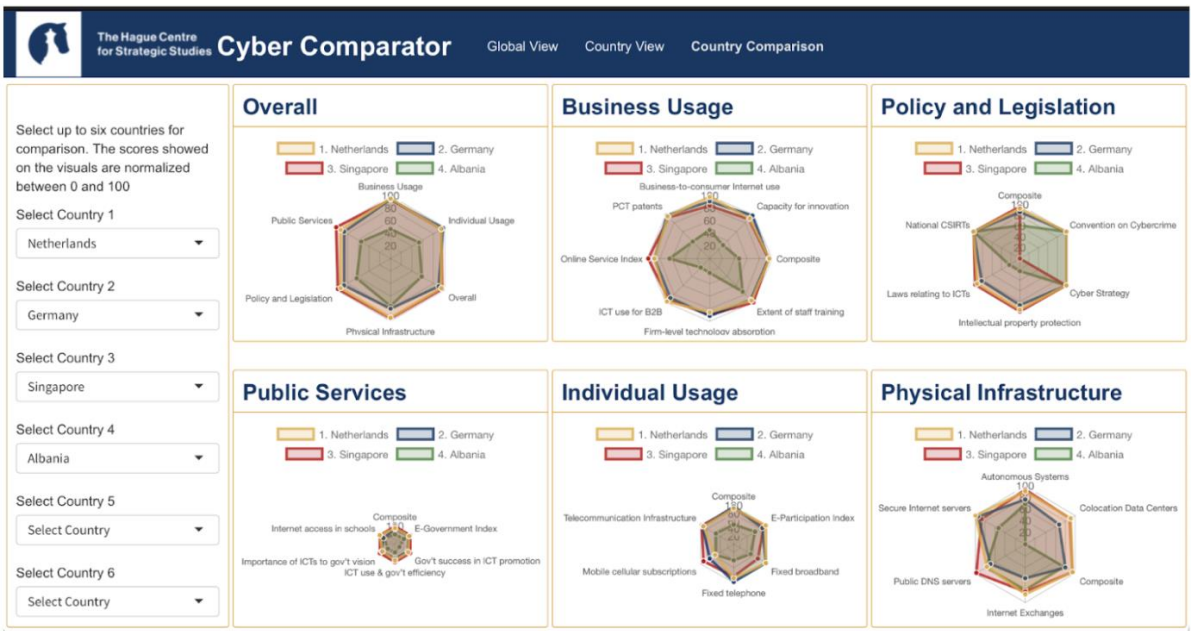


Figure 6: Cyber Comparator screenshot

Understanding: Enable Actionable Engagement

Understanding is about enhancing knowledge through the provision of evidence and fact-based comprehension to enable actionable engagement which can support mechanisms for confidence building, oversight and accountability.

Cyber Norms Observatory

Rationale: *what norms are there and is there overlap between different norm regimes or are there gaps in cyberspace norms regimes we should consider developing?*

Cyberspace is being shaped by many cyber norms or standards regimes designed and developed by and/or for civil society, governments and the private sector, each focusing on different themes or topics and for different purposes. Analysing the similarities and differences is useful in order to understand the interlinkages or lack thereof, to identify gaps and overlaps as well as to assess how these regimes propagate.

The Cyber Norms Observatory demonstrates how 1400 existing norms connect cyberspace. Taken from a large range of normative documents, the monitor shows which actors are trying to shape cyberspace, and how cyberspace depends on the connectivity between actors. The Observatory maps the intricate landscape of existing cyber norms and standards regimes, and thereby shows how connected norms are on different granularity levels. This includes measuring the propagation of these norms using advanced techniques such as text-mining and data-mining, employing Machine Learning and Natural Language Processing. Additionally, the Observatory aims to codify and identify documentation that is relevant to each norm.

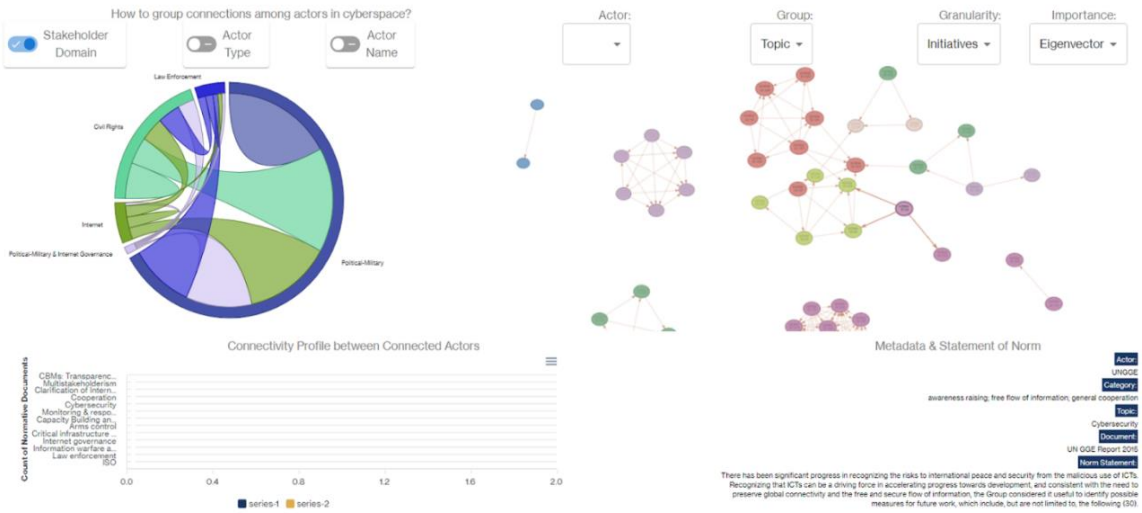


Figure 7: Norms Observatory screenshot

The Cyber Norms Observatory has a multi-faceted approach to understanding cyber norms. Firstly, it seeks to analytically separate each norm regime or initiative, focusing on their individual principles, such as those concerning confidence-building measures. Secondly, the observatory evaluates the relative importance of each norm, considering how they connect and relate to one another. Finally, it assesses the relationships between these norms, exploring what connects them and what differentiates them from each other.

A primary focus of the Cyber Norms Observatory is to facilitate the exchange of information. It invites "standing rapporteurs" to report or integrate relevant associations between different normative frameworks, with the aim to inform rather than consult. Furthermore, the Observatory frames the feasibility of establishing a "permanent dialogue" among various norms regimes. This dialogue is envisioned to promote cross-fertilization and collaborative growth among these regimes. The outcome of such analyses can also be used to raise awareness and understanding of different policy perspectives. This understanding, in turn, helps in identifying potential levers that could contribute to increased resilience, transparency, security, safety, and trust in cyberspace.

Constraints

The Cyber Norms Observatory is a monitor of more than 900 norms and standards built through the use of Natural Language Programming to source the information and interlinked commonalities between the wording used in normative documents. The terminology might be the same in the various norms and standards regimes, but these terms might have different meanings, the norms and standards regimes are continually evolving.

To exemplify the potential of a norms regime analysis the use case of Confidence Building Measures (CBMs) on critical infrastructure (CI) was analysed.

Cyber Confidence Building Measures (CBMs): Increasing transparency around designations of critical infrastructure under CBMs

Malicious cyber incidents targeting critical infrastructure violate the agreed-upon framework for State conduct in cyberspace, which prohibits attacking such infrastructure by cyber means.[10] This normative framework further mandates States to take appropriate measures to protect their critical infrastructure from cyber threats.[11]

Critical infrastructure is not defined. The CyberPeace Institute compared the designations of critical infrastructure by a number of States – indicating that definitions are generally overbroad, and only some States provide a clearer picture of the sectors they consider as critical in their national frameworks. While several countries provide guidance to their national approaches, there is no overarching direction on how to define critical infrastructure in a way that would allow an effective operationalization of the rules, norms and laws. Critical infrastructure could be defined as a "…universal set of core critical civilian infrastructure sectors that enable delivery of essential services to the population and which may be threatened by the use of ICTs".

The importance of the role of confidence building measures (CBMs) in preventing escalation and strengthening cooperation for international peace in cyberspace is recognized by States however, the operationalization of CBMs is a challenge. The CyberPeace Institute and HCSS have separately carried out specific workstreams related to CBMs.

The CyberPeace Institute has provided insights and recommendations on increasing transparency around designations of critical infrastructure (CI) to the United Nations Open-ended Working Group on security of and in the use of information and communications technologies 2021-2025 (OEWG). This contribution has been evidence-based and forward-looking, with action-oriented proposals.

The Institute highlighted that States should provide more clarity on what constitutes critical infrastructure under their national frameworks. The Institute underscored the importance for all States to form a common understanding by sharing, providing, and exchanging information that would facilitate establishing a baseline and increased transparency around what constitutes CI. Determining what constitutes CI is a matter of national security, but increased transparency in this regard is important to building common understanding and coordination on its protection. Transparency about how countries approach CI would reduce chances for misunderstandings and contribute to increase trust and confidence between and among States.

[10] United Nations, General Assembly, Report of the Group of Governmental Experts on Advancing responsible State behaviour in cyberspace in the context of international security, July 22, 2015, available from: https://documents-dds-ny.un.org/doc/UNDOC/GEN/N15/228/35/PDF/N1522835.pdf?OpenElement; The 2015 GGE Report states: 13(f) States should not conduct or knowingly support ICT activity contrary to its obligations under international law that intentionally damages critical infrastructure or otherwise impairs the use and operation of critical infrastructure to provide services to the public.

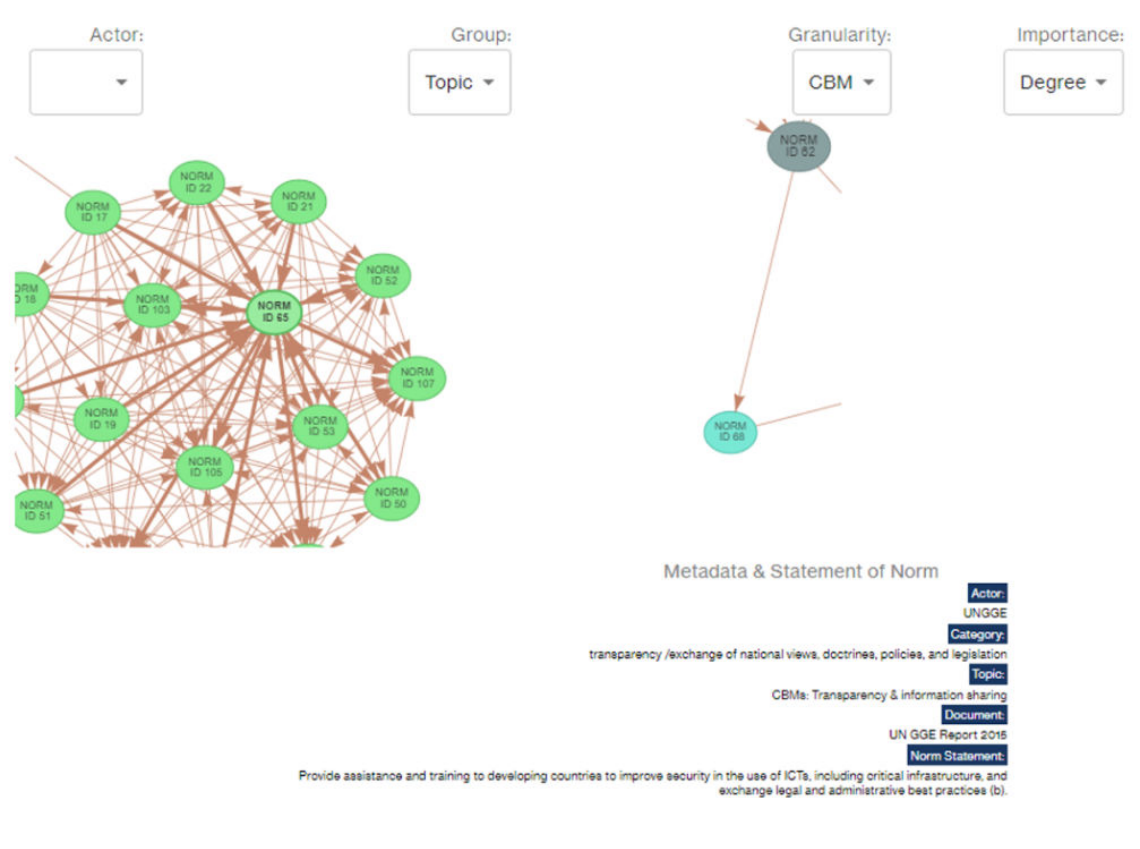
[11] United Nations, General Assembly, Report of the Group of Governmental Experts on Advancing responsible State behaviour in cyberspace in the context of international security, July 22, 2015, available from: https://documents-dds-ny.un.org/doc/UNDOC/GEN/N15/228/35/PDF/N1522835.pdf?OpenElement; The 2015 GGE Report states: 13(g) States should take appropriate measures to protect their critical infrastructure from ICT threats, taking into account General Assembly resolution 58/199 on the creation of a global culture of cybersecurity and the protection of critical information infrastructures, and other relevant resolutions.

Achieving secure, stable, and peaceful cyberspace requires building trust, and reducing tensions, and strengthening confidence between States. Voluntary communication, information sharing and transparency mechanisms related to what constitutes critical infrastructure for States is a necessary step. The 2015 UN GGE norms and the Organization for Security and Cooperation in Europe (OSCE) 16 ICT confidence building measures should be fully leveraged in this regard. Among other measures, it is also that all efforts are undertaken to refrain from targeting critical infrastructure.

The Norms Observatory looks distinctly at five granularity levels: CBMs, best practices, initiatives, principles, and proper norms. HCSS, in the Cyber Norms Observatory, identifies norms from various fora under thirteen topics, among which 'CBMs: Transparency & information sharing.'

Norms identified under this topics in the Observatory, that are related to CI, and that are on a CBM granularity level, include:

- [Norm ID 17] *'States should consider the development of practical confidence-building measures to help increase transparency, predictability and cooperation, including: (e) increased cooperation to address incidents that could affect ICT or critical infrastructure that rely upon ICT-enabled industrial control systems. This could include guidelines and best practices among States against disruptions perpetrated by non-State actors'* [UN Group of Governmental Experts (UN GGE) Report 2013].
- [Norm ID 105] *'Participating States will on a voluntary basis and at the appropriate level hold consultations in order to reduce the risks of misperception, and of possible emergence of political or military tension or conflict that may stem from the use of ICTs, and to protect critical national and international ICT infrastructures including their integrity'* [OSCE Permanent Council Decision No. 1106].
- [Norm ID 65] *'Provide assistance and training to developing countries to improve security in the use of ICTs, including critical infrastructure, and exchange legal and administrative best practices '* [UN GGE Report 2015].



The above figure portrayed the location of the four CI related norms identified in the Observatory dashboard among other norms, with norm ID 65 selected, showing its details at the right-bottom of the dashboard. Filtered by topic, norms categorised as 'CBMs: Transparency & information sharing' are shown in green[12], while the metadata is filtered to show norms with granularity setting 'CBM.' The Observatory shows the strong interrelations among CBM norms, and can be used to efficiently identify similar norm statements from different normative documents and actors.

In its policy submissions and recommendations the CyberPeace Institute can leverage the identification of norms related to CI from this range of different international fora to highlight the importance for States to provide more clarity on what constitutes critical infrastructure under their national frameworks. Increasing transparency about the designations of critical infrastructure would contribute to greater predictability, and enhance trust and confidence between and among states.

[12] The colour codes in the Observatory are random and not fixed, meaning they will change when refreshing the dashboard

Closing gaps in common baselines for designations of critical infrastructure is important in the implementation of the framework of responsible behaviour in cyberspace. This will also facilitate approaches to capacity building and resilience.[13] Streamlined and focused capacity building initiatives, among other measures, could be particularly beneficial for smaller countries with limited resources, to help them assess which infrastructure is critical and how to protect it.

The CyberPeace Institute compared the designations of critical infrastructure by a number of States – indicating that definitions are generally overbroad, and only some States provide a clearer picture of the sectors they consider as critical in their national frameworks. While several countries provide guidance to their national approaches, there is no overarching direction on how to define critical infrastructure in a way that would allow an effective operationalization of the rules, norms and laws.

The path from CBM to CI protection capacity can further utilise the HCSS Cyber Norms Observatory. Under the topic of 'Critical infrastructure protection,' the Cyber Norms Observatory collects norms on matters such as: how CI must be protected; appropriate measures; cross-border cooperation on IC protection must be facilitated; and withholding of pursuing or allowing cyber operations.

Using the example of CBMs on designations of CI, this case shows how the Observatory can be used by organisations such as the CyberPeace Institute to identify norm statements that show strong relations among one another, to enhance their insights and recommendations on cyber transparency, or to expand their research into related normative topics.

Our Ask

Resources to sustain the platform, carry out analysis on various norms regimes, update and extend to Observatory with newly developed norms. Do combined analysis on Norms and CBMs together with CPI.

[13] United Nations, General Assembly, Open-ended Working Group on developments in the field of information and telecommunications in the context of international security, Final Substantive Report, A/AC.290/2021/CRP.2, March 10, 2021, https://front.un-arm.org/wp-content/uploads/2021/03/Final-report-A-AC.290-2021-CRP.2.pdf; The 2021 OEWG consensus report states: *"Capacity-building aimed at enabling States to identify and protect national critical infrastructure and to cooperatively safeguard critical information infrastructure was deemed to be of particular importance. Capacity-building may also help States to deepen their understanding of how international law applies. Information sharing and coordination at the national, regional and international levels can make capacity-building activities more effective, strategic and aligned to national priorities."*

Cyber Attacks in Times of Conflict Platform #Ukraine

Rationale: *better understand the harm and impact of cyberattacks and operations used in an international armed conflict*

Since the start of the armed invasion of Ukraine in February 2022, the CyberPeace Institute has documented cyberattacks against critical infrastructure and civilian objects in Ukraine and the Russian Federation and cyberattacks against targets beyond the two belligerent countries. Between January 2022 and July 2023, the CyberPeace Institute has documented a total of 2,356 cyber incidents conducted by 112 different threat actors.

The data is publicly available through the Cyber Attacks in Times of Conflict Platform #Ukraine, and to facilitate understanding analytical graphs and maps of the data are provided and analysis reports published quarterly.

Attacks on infrastructure such as energy, water and sanitation facilities, healthcare, financial institutions, transport and communication services can have devastating consequences on the civilian population. Cyberattacks also sow distrust and limit access to accurate information or spread false information. On their own, and/or combined with kinetic attacks, cyberattacks can also be highly disruptive and create a sense of fear and uncertainty and accelerate violence and/or the displacement of people.

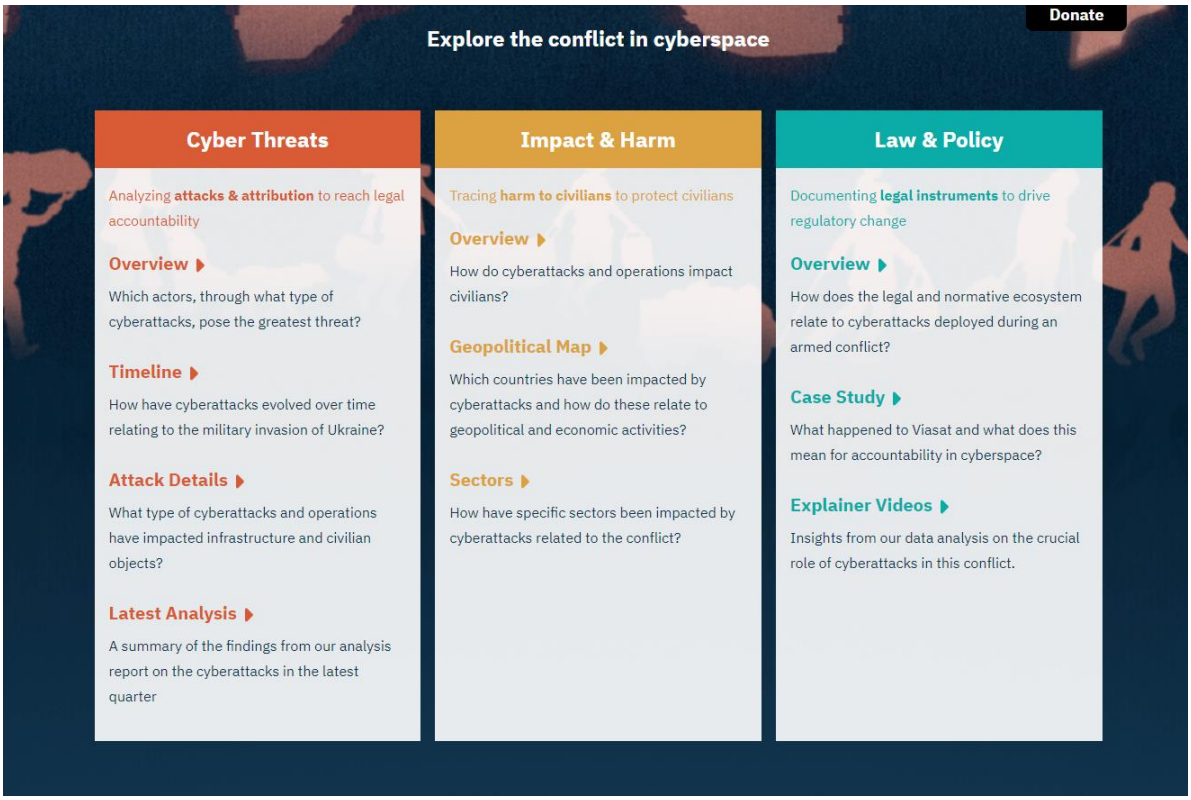


Figure 8: Cyber Attacks in Times of Conflict Platform #Ukraine screenshot

Objective

The Cyber Attacks in Times of Conflict Platform #Ukraine aggregates and publishes data on cyberattacks and operations against critical infrastructure to demonstrate different patterns, trends and emerging issues relating to cyber incidents taking place in Ukraine, the Russian Federation and other countries impacted by cyberattacks in the context of the armed conflict.

The documentation of these attacks not only provides an insight into the role that cyber has played in the Russian-Ukrainian war but also contributes to analysis of the use of cyber in other or future armed conflicts. It is important to document the harm and impact of cyberattacks affecting the civilian population, and to infrastructure essential for the survival of the civilian population, as well as the wider impact related to the destabilisation of cyberspace and international security. Aggregating data on cyberattacks against critical civilian infrastructure is important as these are protected under international humanitarian law during situations of armed conflict, and could be used for future criminal proceedings brought by international courts.

In the context of the war, there are eight dominant cyber threats that have at times served to destruct, disrupt, disinform, and weaponize data. The monitoring Platform and Analysis Reports provide further insights on a wide range of patterns and trends, for example:

- the types of cyberattacks and operations
- which actors pose the greatest threat
- how cyber incidents are evolving over time
- the types of attacks deployed against organisations in each country, etc.

The Institute collects publicly available (open source) information on cyberattacks through the monitoring of news/media outlets, government, CERTs, cybersecurity companies and civil society organisations' reports, advisories and blogs, and social media feeds, among other sources.

Data collection is concentrated on, but not limited to, incidents targeting and/or impacting civilians, civilian objects (including private companies), and infrastructure ensuring the delivery of essential services to civilians.

Constraints

Data collection: There is a reliance on publicly available data on documented cyberattacks, thus it is necessary to give the data a classification of certainty based on the reliability of the information source.[14] Every identified incident, and the associated content, is reviewed by at least two internal analysts and, wherever possible, the incident is linked to at least two separate sources of information. We continuously scan for information on previous incidents to update the data on societal harm and attribution which is often reported significantly after the actual incident. There is no data collection

[14] The classification levels are as follows:

- Confirmed: attacks in this category are based on official government reports / records, official press releases by the targeted organisation, official letters addressed to customers by the target organisation or the government, or social media communication by the targeted organisation. In cases where an incident has been self-attributed by a threat actor and a government entity has confirmed the attack, it will be classified as confirmed.
- Probable: attacks in this category are based on media reports of a press conference by the targeted organisation, social media communication by the targeted organisation or quotes from the targeted organisation's staff in media articles. In cases where an incident has been self-attributed by a threat actor, and the attack has been corroborated by a third-party through independent research or the analysis of stolen data, this is also classified as a probable incident. Incidents identified and reported on as a result of a technical/forensics investigation will also be classified as probable.
- Possible: attacks in this category are based on media reports with no direct reference to primary source information. This can be in the form of a news article that mentions a letter sent to patients or a blog post that references a statement published by the targeted organisation, but no direct record of this material is available. This category also includes data published by a threat actor online with no further corroborating information.

on cyber incidents against military objects, in as far as it is possible to differentiate these from incidents against civilian objects. It is important to note that there are particular challenges in verifying and/or confirming incidents, and if a specific cyberattack or operation has been committed with political, military, activist and/or strategic motives. The CyberPeace Institute does not publicly document data related to 'Hearsay' incidents which contain uncorroborated information originating from a third party, i.e. as a result of media reporting of the allegation by a third party.

Harm and impact of cyberattacks: Analysing the harm and impact of cyberattacks is at the heart of the CyberPeace Institute's work. There is not an existing standard methodology for measuring harm from cyber incidents, thus the Institute is currently developing indicators and a methodology to document and measure the harm and impact of cyberattacks on people, organisations, and society. As a first step, information is collected on the harm and impact of cyberattacks as reported by the source of the information. Insofar as it is possible, quantitative data is documented, such as the duration of a given impact or the number of individuals affected.

Attribution of attacks: The Institute does not conduct its own attribution of incidents to identify the actor(s) involved but documents the attribution efforts by others to link a particular individual, group or state to a specific incident. The challenges and complexity in the attribution of cyberattacks have been discussed by numerous experts.[15]

Our Ask

Funding to sustain the platform and to analyse data for on evidence of harms and on threat actors for accountability processes, including potential submissions of cases to ICC or regional courts and for potentially monitoring cyber incidents subject to potential future ceasefire agreement.

CyberPeace Watch

Rationale: *build understanding of the cyber threat landscape and paths for accountability.*

The malicious use of Information and Communication Technologies is escalating in sophistication, frequency and scale of attack, affecting people in countries throughout the world. Cyber tools are being used as weapons and methods of warfare; harmful content and disinformation campaigns are proliferating online, and essential critical infrastructure is being targeted on a regular basis. The harm to people is real and cyber threats pose significant risks to people and the enjoyment of their fundamental rights, to national security, economic stability, and global peace.

Cyber peace is possible when the world's digital ecosystems promote human security, dignity, and equity. Achieving peace and security in cyberspace requires a means to measure the current situation of threats and harm, the effectiveness of our responses, and a focus on accountability, which is so far absent. A baseline of understanding and shared knowledge is required to achieve this goal.

Objective

The CyberPeace Watch provides a publicly accessible baseline of data to understand and share knowledge about cyberattacks, including threat analysis, societal harm, applicable laws and norms, and related paths for accountability. The platform's goal is to assess cyber peace based on evidence of the societal harm caused by cyberattacks and the actions taken by states and other relevant actors

[15] See for example: Joseph S. Nye Jr., 'Deterrence and Dissuasion in Cyberspace', International Security 41, no. 3 (1 January 2017): 44–71, https://doi.org/10.1162/ISEC_a_00266.; Mariarosaria Taddeo, 'The Limits of Deterrence Theory in Cyberspace', Philosophy & Technology 31, no. 3 (1 September 2018): 339–55, https://doi.org/10.1007/s13347-017-0290-2.

to strengthen responsible behaviour in cyberspace. This will establish a level of transparency to build the understanding necessary for people to make an assessment of the state of cyber peace and empower victims by:

- Capturing and visualising the impact and harm caused by cyberattacks,
- Highlighting the need for redress and protection of victims,
- Giving a voice, backed by evidence, to those who wish to change the status quo.

Furthermore, this understanding can drive responsibility of State actors and non-state actors to take action by:

- Evaluating the actions of state and non-state actors to better understand if their behaviour in cyberspace is responsible or not,
- Monitoring and mapping the commitments, obligations and responsibilities they have undertaken,
- Documenting attribution of cyber incidents or threats, particularly the existence of political, legal, and technical attribution.

The absence of transparency arises from various domains, including because of the challenges associated with the attribution of cyberattacks, as the nature of cyberspace enables attackers to conceal their identity in a way that would not be possible offline. The impact of cyberattacks has been constantly rising in terms of both severity of the damages and number of victims, and the challenge of accountability needs to be prioritised by decision-makers at all levels.

Constraints

Data collection: There is a reliance on publicly available data on documented cyberattacks, thus it is necessary to give the data a classification of certainty based on the reliability of the information source. Every identified incident, and the associated content, is reviewed by at least two internal analysts and, wherever possible, the incident is linked to at least two separate sources of information.

Applicable rules and norms: There are many current challenges related to the interpretation and application of law and policy in relation to cyberattacks and operations due to their unique characteristics. It is important to identify and clarify these in order to ensure the protection afforded by the law, the limits imposed by existing rules and - where required - develop additional law and policy. Clarifications related to the interpretation of the rules are still required by States. Challenges regarding interpretation are obstacles to enforcing and developing the legal framework and to elaborating adequate and accurate policies at the international level.

Harm and impact of cyberattacks: One such important unknown to recognize is the true scale of the human impact of cyber operations. This is also true for kinetic operations, but cyber operations lend another layer of uncertainty as the impact on victims can materialise only after a time delay or may be indirect but cause harm. Analysing the harm and impact of cyberattacks is at the heart of the CyberPeace Institute's work. There is not an existing standard methodology for measuring harm from cyber incidents, thus the Institute is currently developing indicators and a methodology to document and measure the harm and impact of cyberattacks on people, organisations, and society. As a first step, information is collected on the harm and impact of cyberattacks as reported by the source of the information. Insofar as it is possible, quantitative data is documented, such as the duration of a given impact or the number of individuals affected.

Attribution of attacks: It is difficult to directly attribute impact to one cyberattack or operation, as sometimes these operations can take place over a long period of time such as espionage-related endeavours or disinformation campaigns, or they can be one in a series of operations that changes ever so slightly each time to avoid detection. There will also be a level of uncertainty in this regard, but the investigation of cyberattacks and operations is a developing field that needs time to mature. Attribution of a cyberattack or operation is significantly more difficult than for kinetic operations. Some reasons for this are:

- the origin of the attack can be misleading when the attacker uses compromised machines to launch their attack.
- attribution requires deep analysis of Tactics Techniques and Procedures (TTPs) used by threat actors, similarities of the code or the compilation of malware and data on the command & control infrastructure.
- the consequences or damages from cyberattacks may be difficult to quantify as they may be delayed (occurring several hours, days, months after the attack was launched), direct and indirect.

  The attribution of responsibility for a cyberattack to a certain attacker or group of attacks must be based on evidence, which may be of a technical and legal nature. The quality of an attribution is a function of available resources, time, evidence, data, verification means, etc. Speculating about or wrongly attributing an attack may lead to an escalation in hostilities. The Institute does not conduct its own attribution of incidents to identify the actor(s) involved but documents the attribution efforts by others to link a particular individual, group or state to a specific incident.

Information on accountability measures: There are different ways to prosecute individuals for violations of the law in wartime and peacetime, under domestic, regional or international instruments, e.g. for grave breaches of the Geneva Conventions and their Additional Protocols, at the International Criminal Court or a special tribunal to prosecute alleged war criminals for violations of international law. Collecting data on such prosecutions will be required and is a long term endeavour.

Our Ask

Resources and data to resume data collection and analysis particularly of harm to people and society related to cyber incidents for the development of the Harms Methodology.

Attribution: Technical, Legal, Political, Self-attribution

Attribution is a technical step in international law for attaching a given act or omission to a relevant actor, such as a state, for the purposes of determining who is responsible for a violation of international law and which is the appropriate legal framework establishing the rights and obligations of states affected by an incident and impose consequences, if appropriate. Thus, attribution is critical for accountability under the law, remedy, and redress for victims of cyberattacks.

Attribution of an attack under international law triggers State responsibility, and may trigger the application of IHL, and/or activate the right to respond in self-defence. It can be very difficult to attribute a cyberattack to a State among others because of the complexity of attribution and anonymity of attacks, and a reluctance to publicly reveal technical sources and methods of intelligence.

During armed conflict, attribution of acts to individuals is essential to be able to trace the actions of the belligerents and/or attackers, to assess their compliance with international humanitarian law and/or other legal regime they are bound by and the possible misconducts and violations, and for accountability for violations of the law. There is also an opportunity to identify good practices, gaps in the law or where clarification of the law is required.

The attribution of cyberattacks presents technical, legal, and political challenges, nonetheless it is a necessary prerequisite for holding actors accountable for malicious activity and ensuring peace in cyberspace.

- Technical challenges: the identification of the machine (s) used to carry out a cyberattack is complex due to the decentralised structure of the Internet and the multiple possible vectors of attack. Government agencies, law enforcement, and private companies have progressively refined their capacity to attribute cyberattacks due to the ability to collect, analyse, and match specific sets of Tactics, Techniques and Procedures (TTPs) with Advanced Persistent Threats or criminal groups. The identification phase is of crucial importance and should not be underemphasized, however, the key challenge for accountability revolves around what measures are actually taken against the identified actors. While further improvements in terms of identification capabilities are necessary, ensuring accountability in cyberspace ultimately remains a legal and political challenge.
- Legal Challenges: Both being able to ascribe to a perpetrator the legal responsibility of a cyber operation and enforce law or a norm of behaviour by punishing those found guilty. Domestic, regional, and international legal systems have evolved, adapted, and are interpreted to respond to the realities of cyber operations.[16] Nevertheless, the diverse nature of cyber operations poses thorny interpretive challenges on how the relevant rules apply. According to the existing norms on

[16] For example, as of January 2022, the vast majority of States have adopted specific cybercrime legislation, discussions for a new UN international convention on cybercrime are proceeding, and consensus has been reached on the applicability of international law to the use of Information and communication technologies (ICTs) by States.

responsibility outlined in the International Law Commission's Articles on the Responsibility of States for Internationally Wrongful Acts and various judgements[17] the legal attribution of cyber operations to a state can be quite challenging and the spectrum of state responsibility for malicious attacks hard to define.[18] This and other questions render even legal attributions particularly thorny to the point that both attribution and accountability, while sometimes supported and informed by international law argumentations, are oftentimes a political decision.

- Political Challenges: Political considerations are a major hindrance to ensuring accountability. Governments will take into account the consequences of publicly attributing a cyberattack to a state, perhaps considering the potential destabilisation of bilateral and multilateral relations, or even fearing further attacks. Similarly, attributing responsibility to a non-state actor operating within the borders of a certain state could also raise tensions between governments or provoke additional misbehaviour from the threat actor.

There is an urgent need to consider attribution not an end itself, but as a means to holding malicious actors accountable. The question of accountability remains a crucial one for ensuring stability in cyberspace and the international community must further address this challenge in a consistent way.

Cyber Attacks in Armed Conflict Platform #Ukraine

Rationale: *contribute to efforts to hold threat actors accountable for breaches of laws.*

The Cyber Attacks in Times of Conflict Platform #Ukraine aggregates and publishes data on cyberattacks, threat actors and attribution of attacks against critical infrastructure to demonstrate different patterns, trends and emerging issues in Ukraine, the Russian Federation and other countries impacted by cyberattacks in the context of the armed conflict. Throughout the war, different types of cyber threats have largely been contingent on the threat actors that perpetrate them and their respective motivations and capabilities. Both in times of peace and armed conflict, holding an actor accountable for a violation of the law is only possible when the action in question is attributed to that actor. The applicable law will change depending on the actor concerned, for instance, whether it is attributable to a State or an individual for instance.

Objectives

Attributing acts conducted during an armed conflict helps in the fight against impunity by highlighting and prosecuting actors involved in misconduct and violations. To prosecute and hold criminals accountable, there is a requirement to create evidence. Evidence can be hard to find or to gather in relation to attacks committed in cyberspace. This platform can contribute to this endeavour.

A notable phenomena in relation to cyberattacks being monitored in the armed conflict between Russia and Ukraine is that of "self-attribution". Some threat actors publicly disclose a cyberattack and attribute themselves as the actor behind the attack, often by publishing data extracted as a result of an incident on dedicated websites. Although not a formal category of attribution - like technical, legal and policy attribution - it remains one of the ways in which actors involved are documented.

[17] Of the International Court of Justice (ICJ) and the International Criminal Tribunal for the former Yugoslavia.

[18] For example, governments could back a group of cyber criminals to conduct cyber operations on their behalf, nonetheless, the degree of state control – thus the "interaction" – necessary to encourage and prompt non-state actors may be such that it does not trigger the responsibility of the state under international law norms.

Unlike conventional attacks, cyberattacks can be hard to accurately attribute to a particular actor, and plausible deniability exists because cyberattacks can be launched and a State can deny knowledge of the attack, responsibility for it and/or knowledge of the threat actor.

With self-attributed attacks, it is important to continue to monitor the threat actors, the harm they cause and future efforts to hold persons accountable for attacks which have breached laws and norms.

The attribution of acts can have different consequences. Attributing an act of violence against another State to a State can trigger the applicability of IHL related to an International Armed Conflict. Attributing acts conducted during an armed conflict helps in the fight against impunity by highlighting and prosecuting actors involved in misconduct and violations.

Constraints

Data and evidence: To prosecute and hold criminals accountable, there is a requirement to create evidence. Evidence can be hard to find or to gather in relation to attacks committed in cyberspace, especially in times of armed conflict and the so-called “fog of war”.

Attribution of an attack to a particular actor: It is difficult to directly attribute impact to one cyberattack or operation, as sometimes these operations can take place over a long period of time such as espionage-related endeavours or disinformation campaigns, or they can be one in a series of operations that changes ever so slightly each time to avoid detection. Attribution of a cyberattack or operation is significantly more difficult than for kinetic operations. Some reasons for this are:

- the origin of the attack can be misleading when the attacker uses compromised machines to launch their attack.
- attribution requires deep analysis of Tactics Techniques and Procedures (TTPs) used by threat actors, similarities of the code or the compilation of malware and data on the command & control infrastructure.
- the consequences or damages from cyberattacks may be difficult to quantify as they may be delayed (occurring several hours, days, months after the attack was launched), direct and indirect.

Proliferation of cyber threat actors: The armed conflict between Ukraine and Russia has shown that cyber operations can be an integral part of the way war is waged. It also means that anyone, from his or her computer or cell phone can have some ability to help or harm another device, another group of persons, another entity that might be on the other side of the planet. Today, we see individuals using their devices to participate in the hostilities, to act against one of the belligerents, or allies. The conflict has blurred the lines between state and non-state actors with hacktivists and collectives, and IT Armies of volunteers aligned with both parties to the conflict or “sides”. Cyberattacks by such groups, including those who are State proxy actors have fostered confusion about who is responsible, posing attribution challenges and enabling deniability for states to limit escalation and potential prosecution.

Direct participation in hostilities: Anyone, from his or her computer or cell phone can have some ability to help or harm another device, another group of persons, another entity that might be on the other side of the planet. Today, individuals use their devices to participate in the hostilities, to act against one of the belligerents, or allies. This raises questions about the notion of Direct Participation in Hostilities. A State supplying a belligerent State with weapons is usually not

considered as a belligerent unless it has some sort of control over the operations conducted or over the group it is supplying.

Our Ask

Funding to sustain the platform and to analyse data for on evidence of harms and on threat actors for accountability processes, including potential submissions of cases to ICC or regional courts and for potentially monitoring cyber incidents subject to potential future ceasefire agreement.

Monitoring and Sanctioning: Observing State Behaviour and Economic, Financial and Legal Measures

To better understand and monitor the status of state and non-state behaviour in cyberspace it is important to monitor the incidents showing violations of existing legal instruments and norms, sanctions and other accountability measures that are imposed. Tracking and analysis of conduct in cyberspace in order to provide information on measures for accountability and responsible behaviour is critical.

It is important to observe how states behave in line with their obligations to respect and ensure respect for laws, whether they violate existing legal instruments and their respect of norms.

Sanctioning is about the economic, financial and legal measures taken against individuals or entities responsible for malicious cyber activities. It takes place in various jurisdictions, be it national, regional or international, and thus is the most complex aspect of the cyber transparency value chain.

CyberPeace Watch

Rationale: *Monitor State observance of laws and norms to strengthen responsible state behaviour, and sanctions imposed.*

The CyberPeace Watch, an interactive online platform will provide a publicly accessible baseline of data to understand and share knowledge about cyberattacks, applicable laws and norms, and related paths for accountability.

Objectives

The platform's goal is to assess cyber peace based on evidence of the societal harm caused by cyberattacks and the actions taken by states and other relevant actors to strengthen responsible behaviour in cyberspace.

Data collection and analysis will focus on monitoring and mapping the commitments, obligations and responsibilities States and non-state actors have undertaken, evaluating the actions of state and non-state actors to better understand if their behaviour in cyberspace is responsible or not and documenting attribution of cyber incidents or threats, particularly the existence of political, legal, and technical attribution.

Our Ask

Funding to sustain the platform and to analyse data for on evidence of harms and on threat actors for accountability processes, including potential submissions of cases to ICC or regional courts and for potentially monitoring cyber incidents subject to potential future ceasefire agreement.

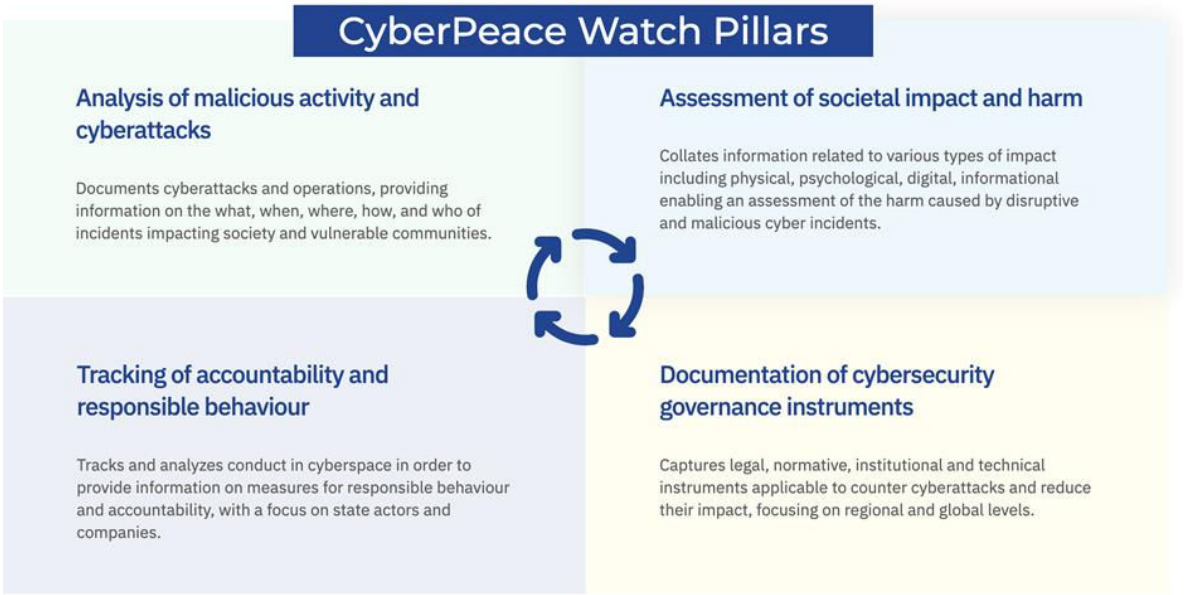


Figure 9: CyberPeace Watch Pillars

Cyber Transparency, an International Initiative

The Hague Centre for Strategic Studies (HCSS) and the CyberPeace Institute are working together to increase cyber transparency, to inform policy processes and capacity building efforts, and contribute to accountability efforts.

Potential for collaboration

If your organisation is interested in collaborating please contact us at info@hcss.nl and info@cyberpeaceinstitute.org