CCDCOE
NATO COOPERATIVE
CYBER DEFENCE
CENTRE OF EXCELLENCE

Autonomous cyber capabilities and unilateral measures of self-help against malicious cyber operations

Marta Stroppa
NATO Cooperative Cyber Defence Centre of Excellence

About the authors

Marta Stroppa is a PhD Candidate at the Sant'Anna School of Advanced Studies in Pisa, Italy. During her PhD, she was a Visiting Scholar at the North Atlantic Treaty Organisation (NATO) Cooperative Cyber Defence Centre of Excellence and the University of Westminster in London, United Kingdom. Marta has previously worked in the Legal Affairs Office of the Permanent Mission of Italy to the United Nations in New York and the Global Maritime Crime Programme of the United Nations Office on Drugs and Crime. She holds a Bachelor's and Master's Degree in International Relations from the University of Milan, Italy and a Master of Laws in International and European Law from Tilburg University, Netherlands.

CCDCOE

The NATO Cooperative Cyber Defence Centre of Excellence (CCDCOE) is a NATO-accredited knowledge hub offering a unique interdisciplinary approach to the most relevant issues in cyber defence. The heart of the CCDCOE is a diverse group of international experts from military, government, academia and industry, currently representing 39 nations.

The CCDCOE maintains its position as an internationally recognised cyber defence hub, a premier source of subject-matter expertise and a fundamental resource in the strategic, legal, operational and technical aspects of cyber defence. The Centre offers thought leadership on the cutting edge of all aspects of cyber defence and provides a 360-degree view of the sector. The Centre encourages and supports the process of mainstreaming cybersecurity into NATO and national governance and capability, within its closely connected focus areas of technology, strategy, operations and law.

The Tallinn Manual, prepared at the invitation of the CCDCOE, is the most comprehensive guide for policy advisers and legal experts on how international law applies to cyber operations carried out between and against state and non-state actors. Since 2010, the Centre has organised Locked Shields, the biggest and most complex technical live-fire cyber defence challenge in the world. Each year, Locked Shields allows cybersecurity experts to enhance their skills in defending national IT systems and critical infrastructure under real-time attacks. The focus is on realistic scenarios, cutting-edge technologies and simulating the entire complexity of a massive cyber incident, including strategic decision-making and legal and communication aspects.

The CCDCOE hosts the International Conference on Cyber Conflict, CyCon, a unique annual event in Tallinn, bringing together key experts and decision-makers from the global cyber defence community. The conference, which has taken place in Tallinn since 2009, attracts more than 600 participants each spring.

The CCDCOE is responsible for identifying and coordinating education and training solutions in the field of cyber defence operations for all NATO bodies across the Alliance. NATO-accredited centres of excellence are not part of the NATO Command Structure.

www.ccdcoe.org
publications@ccdcoe.org

Disclaimer

Acknowledgements

The author would like to extend a heartfelt thank you to Cdr Davide Giovannelli and Ann Väljataga for their inestimable support during the drafting process of the paper. Without their help, this paper would not exist. A sincere thank you also goes to Aleksi Kajander for his valuable and insightful comments, and to Karine Veersalu for her editorial assistance. Finally, I am particularly grateful to the Law Branch of the CCDCOE for their hospitality and the time spent together

Table of Contents

1. Abstract 5
2. Introduction 6
3. Autonomous cyber capabilities: A definition 7
4. Towards an increasing automation of cyber defence 8
5. Autonomous measures of self-help against malicious cyber operations 10
   - 5.1 Self-defence 10
   - 5.2 Countermeasures 17
   - 5.3 Plea of necessity 21
   - 5.4 Retorsions 23
6. Conclusions 25
7. References 27

1. Abstract

Autonomous cyber capabilities – that is, cyber capabilities able to operate without real-time human intervention – are currently being researched and developed by several states as a result of the increasing use of artificial intelligence and autonomy in the military domain. Whereas these capabilities are expected to be mainly employed to respond to malicious cyber operations, their use for defensive purposes raises some legal challenges that deserve to be explored. This paper seeks to analyse whether autonomous cyber capabilities can be used in compliance with international law to respond to malicious cyber operations using unilateral measures of self-help. After briefly introducing the notion of autonomous cyber capabilities and their current state of technological development, this paper will consider whether autonomous cyber capabilities can be used in compliance with the law regulating self-defence, countermeasures, plea of necessity and retorsions. As will be shown, their potential use in circumstances precluding wrongfulness (i.e., self-defence, countermeasures and plea of necessity) is highly problematic, as autonomous cyber capabilities seem to be currently unable to identify the objective and subjective element of the malicious cyber operation (whether it amounts to an internationally wrongful act and whether it is attributable to a state), and to calibrate their response in the light of the principles of necessity and proportionality. Yet, it will be suggested that states may still cautiously use autonomous cyber capabilities to carry out acts of retorsion.

2. Introduction

Artificial intelligence (AI) and autonomy are increasingly playing a significant role in all military domains, including cyberspace.[1] The prospect of using autonomous and adaptable cyber capabilities for defensive purposes is especially appealing for states, as it would allow them to overcome the shortfalls of traditional cybersecurity tools, by further strengthening their networks' robustness, resilience and response against malicious cyber operations.[2] Not only those types of cyber capabilities may be used by states to increase their systems' ability to discover and withstand external intrusions, but they may also be employed to further advance a system's capacity to autonomously defeat incoming attacks and mitigate their effects, allowing a greater speed, scale and precision of response that exceed human capabilities.[3]

While the current state of technological development is still far from reaching full autonomy in cyberspace, both states and the private sector are consistently researching and developing autonomous solutions for cyber defence.[4] Thus, autonomous cyber capabilities will likely shape future cyber warfare. At the same time, however, the potential use of autonomous cyber capabilities for defensive purposes raises important concerns *vis-à-vis* international law that need to be addressed. Yet, States have remained silent on the matter, often overlooking the potential implications that may arise from the overlaps of autonomy and cyberspace.[5]

This paper seeks to analyse whether autonomous cyber capabilities can be used in compliance with international law to respond to malicious cyber operations outside of armed conflicts. It will be structured as follows. After providing an overview of the current state of technological development in the automation of cyber defence, it will consider in which instances autonomous cyber capabilities can be used by states to carry out unilateral measures of self-help against malicious cyber operations. In particular, it will explore whether and to what extent autonomous cyber capabilities can be used to respond to a malicious cyber operation by resorting to self-defence, countermeasures, plea of necessity or retorsions. It will be suggested that, at the current stage of technological development, it is unlikely that autonomous cyber capabilities will be able to act in compliance with international law in circumstances precluding wrongfulness (i.e. self-defence, countermeasures and plea of necessity). Yet, states may still be able to cautiously use autonomous cyber capabilities to carry out acts of retorsion.

---

[1] Jacopo Bellasio and Erik Silfversten, 'The Impact of New and Emerging Technologies on the Cyber Threat Landscape and Their Implications for NATO', in *Cyber Threats and NATO 2030: Horizon Scanning and Analysis*, by Amy Ertan et al. (Tallinn: NATO CCDCOE Publications, 2020), pages 90-91; Maggie Gray and Amy Ertan, 'Artificial Intelligence and Autonomy in the Military: An Overview of NATO Member States' Strategies and Deployment' (Tallinn: NATO CCDCOE, 2021).
[2] Salvador Llopis Sanchez, 'Artificial Intelligence (AI) Enabled Cyber Defence', *European Defence Matters*, no. 14 (2017): 18. See also Mariarosaria Taddeo, Tom McCutcheon and Luciano Floridi, 'Trusting Artificial Intelligence in Cybersecurity Is a Double-Edged Sword,' *Nature Machine Intelligence* 1, no. 12 (2019): 557–60, page 557.
[3] Tim McFarland, 'The Concept of Autonomy', in *Autonomous Cyber Capabilities under International Law*, by Rain Liivoja and Ann Väljataga (Tallinn: NATO CCDCOE Publications, 2021): 12–35.
[4] Tanel Tammet, 'Autonomous Cyber Defence Capabilities,' in *Autonomous Cyber Capabilities under International Law*, by Ann Väljataga and Rain Liivoja (Tallinn: NATO CCDCOE Publications, 2021): 36–50, page 37.
[5] On the political debate over autonomous cyber capabilities, see Louis Perez, 'Is Stuxnet the next Skynet? Autonomous Cyber Capabilities as Lethal Autonomous Weapons Systems', in *Artificial Intelligence and International Conflict in Cyberspace*, by Fabio Cristiano et al., Routledge Studies in Conflict, Security and Technology (Oxon; New York: Routledge, 2023): 186–222.

3. Autonomous cyber capabilities: A definition

Autonomous cyber capabilities have been defined as cyber capabilities that are able 'to perform some task without requiring real-time interaction with a human operator'.[6] In other words, they are tools designed and programmed by human operators to follow a set of human-written instructions to achieve a pre-defined goal, but they do not need real-time human control or guidance while executing their tasks.

The more complex the tasks or the environment in which they operate are, the more advanced autonomous cyber capabilities will be. Autonomous cyber capabilities with very simple and limited tasks are often driven by programmes with encoded instructions in the form of: 'if < X happens > then < do action A > else < do action B >'. More advanced autonomous cyber capabilities are instead driven by complex self-adaptive (or even self-learning) programmes that allow them to adapt to evolving circumstances to achieve their pre-determined goal, without the help of human operators.[7]

Depending on the goal they pursue, autonomous cyber capabilities can be classified as offensive or defensive. Offensive autonomous cyber capabilities are tools designed and programmed to launch a cyber operation against a target without real-time human intervention. A famous example is Stuxnet, the worm used to target the Natanz nuclear enrichment facility in 2009, which was designed to operate in air-gapped networks disconnected from the internet without any form of interaction with a human operator.[8] Defensive autonomous cyber capabilities are tools designed and programmed to defend a system against a malicious cyber operation without requiring real-time human intervention. A widely discussed example is the Mayhem Cyber Reasoning System, developed by For All Secure and winner of the Defence Advanced Research Projects Agency's 2016 Cyber Grand Challenge. Although it was a prototype expected to operate in a simplified operating system, it was designed to protect the network from external intrusions, to stop cyber operations while occurring and identify their originator, and to exploit the vulnerabilities in the adversaries' network to harm the system where the intrusions originated from.[9]

Defensive autonomous cyber capabilities can also be distinguished by passive and active defensive measures. Passive cyber defence includes all those measures of detection and mitigation of intrusions that are aimed at making the defended network and system more resilient, as in the case of firewalls or anti-virus software. They solely operate within the system they are defending.[10] Active cyber defence involves proactive measures carried out outside the defended infrastructure to prevent, block or respond to a malicious cyber operation.[11] An example of active cyber defence is the 'hack-back', namely taking actions against an identified source of a malicious cyber operation to mitigate its effects or to gather technical evidence to be used for attribution.[12] When autonomous cyber capabilities are deployed, both passive and active defensive measures can be carried out without real-time human intervention.

---

[6] Rain Liivoja, Maarja Naagel and Ann Väljataga, 'Autonomous Cyber Capabilities under International Law' (Tallinn: NATO CCDCOE, 2019), page 10. François Delerue has defined autonomous cyber operations as 'cyber operations that, once activated, can select and engage targets without further intervention by a human operator' – François Delerue, *Cyber Operations and International Law*, (Cambridge: Cambridge University Press, 2020), page 158.
[7] McFarland, 'The Concept of Autonomy', pages 18-26.
[8] Paul Scharre, *Army of None: Autonomous Weapons and the Future of War* (New York: W. W. Norton & Company, 2018), pages 214-215.
[9] Liivoja, Naagel and Väljataga, 'Autonomous Cyber Capabilities under International Law', page 12.
[10] Michael N. Schmitt, *Tallinn Manual 2.0 on the International Law Applicable to Cyber Operations*, 2nd ed. (Cambridge: Cambridge University Press, 2017), Glossary, page 566.
[11] *Ibid*, page 563.
[12] *Ibid*, page 565.

4. Towards an increasing automation of cyber defence

In recent years, the international community has shown particular interest in the automation of cyber defence, as it would allow states to carry out defensive activities with greater speed, accuracy, precision and over a longer period, even when large quantities of data and information are involved.[13]

Currently, autonomous solutions for defensive purposes in cyberspace are being researched and developed by several states including Australia,[14] China,[15] France,[16] Germany,[17] Japan,[18] the Russian Federation,[19] the United Kingdom (UK),[20] and the United States (US).[21] NATO[22] and the European Union (EU)[23] are also particularly interested in the automation of cyber defence.

While current technological developments are still far from replacing human specialists with fully autonomous cyber capabilities, there are no doubts that automation will play a central role in the development of cyber defence.[24] Virus defence systems and firewalls, anomaly detection systems and network mappers, just to mention a few examples, are already routinely automated.[25] Current efforts are focusing on the development of defence systems able to automatically detect external intrusions, isolate

---

[13] McFarland, 'The Concept of Autonomy', page 18.
[14] The Cyber Warfare Operations branch of Australia's Cyber and Electronic Warfare Division is in charge of researching and developing 'autonomous, resilient and effective cyber capabilities'. See the website of the Cyberwarfare Operations Branch, available at: https://www.dst.defence.gov.au/capability/cyberwarfare-operations.
[15] US Department of Defence, 'Military and Security Developments Involving the People's Republic of China', Annual Report to Congress, 2021.
[16] Delerue, *Cyber Operations and International Law*, page 159.
[17] Germany Federal Ministry of the Interior and Community, Cyber Security Strategy for Germany 2021, 05 October 2021, page 47, available at: https://www.bmi.bund.de/SharedDocs/downloads/EN/themen/it-digital-policy/cyber-security-strategy-for-germany2021.html.
[18] 'Japan's New AI-Based Cyber Defence System', Cyber Security Intelligence (blog), 15 April 2020, available at: https://www.cybersecurityintelligence.com/blog/japans-new-ai-based-cyber-defence-system--4907.html; and John Leyden, 'Japan Tasks Fujitsu with Creating Search-and-Destroy Cyber-Weapon', The Register, 3 January 2012, available at: https://www.theregister.com/2012/01/03/japan_cyber_weapon_research/.
[19] Samuel Bendett et al., 'Advanced Military Technology in Russia: Capabilities and Implications', Research Paper, Russia and Eurasia Programme (London: Chatham House, September 2021), pages 71-72.
[20] United Kingdom, 'National Cyber Strategy', 15 December 2022, available at: https://assets.publishing.service.gov.uk/government/uploads/system/uploads/attachment_data/file/1053023/national-cyber-strategy-amend.pdf; and in particular the Autonomous Resilient Cyber Defence – Intelligent Agents project, available at: https://www.gov.uk/government/news/autonomous-resilient-cyber-defence-intelligent-agents.
[21] Scharre, *Army of None,* page 95; US Cyberspace Solarium Commission, 'Final Report', March 2020, page 16, available at: https://www.solarium.gov/report; and US National Security Commission on Artificial Intelligence, 'Final Report', 2021, page 79, available at: https://www.nscai.gov/2021-final-report/.
[22] In 2016, the NATO created the IST-152 Research and Technology Group 'Intelligent Autonomous Agents for Cyber Defense and Resilience' with the precise objective to help accelerate the development and transition to practice of autonomous intelligent cyber-defence agents, capable of autonomously monitor the networks, detect the enemy cyber activities while remaining concealed and then destroy or degrade the enemy malware. Alexander Kott et al., 'Autonomous Intelligent Cyber-Defense Agent (AICA) Reference Architecture, Release 2.0' (Devcom Army Research Laboratory, September 2019).
[23] In 2022, the EU stated in the Strategic Compass for Security and Defence its commitment to 'develop and make intensive use of new technologies, notably quantum computing, Artificial Intelligence and Big Data, to achieve comparative advantages, including in terms of cyber responsive operations and information superiority'. European Union, 'Strategic Compass for Security and Defence', 2022, page 45 available at: https://www.eeas.europa.eu/sites/default/files/documents/strategic_compass_en3_web.pdf. On this point, see also Sanchez, 'Artificial Intelligence (AI) Enabled Cyber Defence'.
[24] Tanel Tammet, 'Autonomous Cyber Defence Capabilities,' in *Autonomous Cyber Capabilities under International Law*, by Ann Väljataga and Rain Liivoja (Tallinn: NATO CCDCOE Publications, 2021), 36–50, page 50.
[25] *Ibid*, page 40.

the targeted system and recuperate after incidents.[26] More complex AI-based tools with higher levels of autonomy are also slowly emerging from the early experimental stage.[27]

Soon, both states and the private sector are expected to develop cyber capabilities with different degrees of autonomy that can cooperate with human operators when necessary and operate on their own when human control is unfeasible or undesirable. The increasing automation of cyber defence, however, brings with it new legal hurdles, especially when autonomous cyber capabilities will fully operate without real-time human intervention, that need to be addressed.

[26] *Ibid*, page 44.
[27] *Ibid*, page 37.

5. Autonomous measures of self-help against malicious cyber operations

Under international law, a state that is the victim of a malicious cyber operation may resort to a number of extrajudicial measures to compel the responsible state to fulfil its international obligations that take the form of unilateral measures of self-help.[28] Such extrajudicial remedies may include acts of retorsion or measures justified by circumstances precluding wrongfulness. While acts of retorsion are always available to the injured state, the other measures of self-help can be used only in specific circumstances which preclude the wrongfulness of a conduct that would be otherwise unlawful.[29]

If a state decides to deploy an autonomous cyber capability for defensive purposes outside of armed conflict, it will delegate the decision to respond to a malicious cyber operation targeting its network to the computer. Thus, it is important to understand whether and to what extent autonomous cyber capabilities can act in compliance with the provisions of international law regulating measures of self-help. In particular, this part will consider the four measures of self-help that are more significant when it comes to defensive measures against malicious cyber operations outside of armed conflict, namely self-defence, countermeasures, plea of necessity and retorsions.

5.1 Self-defence

The right of states to self-defence is long-established in customary international law and is one of those exceptions to the prohibition of threatening or using force under Article 2 paragraph 4 of the UN Charter. It is codified in Article 51 of the UN Charter, which recognises that all Member States of the United Nations have the 'inherent right of individual or collective self-defence'.[30]

Whereas the UN Charter (or any other treaty) does not explicitly mention whether such right applies also in cyberspace, in its Advisory Opinion on the Legality of the Threat or Use of Nuclear Weapons, the International Court of Justice (ICJ) has made clear that Article 51 applies to 'any use of force, regardless of the weapons employed'.[31] Thus, it has been argued that the right to self-defence also applies to cyber operations,[32] including those carried out using autonomous cyber capabilities.[33] Given

---

[28] Math Noortmann, *Enforcing International Law: From Self-Help to Self-Contained Regimes* (Aldershot: Ashgate, 2005), page 3; Henning Lahmann, *Unilateral Remedies to Cyber Operations: Self-Defence, Countermeasures, Necessity and the Question of Attribution* (Cambridge ; New York: Cambridge University Press, 2020).
[29] International Law Commission, 'Draft Articles on Responsibility of States for Internationally Wrongful Acts,' 2001, Chapter V, Articles 20-25.
[30] Although Article 51 only refers to the members of the United Nations, it is now generally accepted that it reflects rules of customary international law which apply to all states, irrespective of their membership to the United Nations. Yoram Dinstein, *War, Aggression and Self-Defence*, Sixth Edition (Cambridge ; New York: Cambridge University Press, 2017), page 200.
[31] Legality of the Threat or Use of Nuclear Weapons (Advisory Opinion), ICJ Reports 1996 226 (International Court of Justice 1996), para. 39.
[32] United Nations General Assembly, 'Report of the Group of Governmental Experts on Developments in the Field of Information and Telecommunications in the Context of International Security' (A/68/98*, 24 June 2013), para. 19. The same was reiterated by the GGE on Developments in the Field of Information and Telecommunications in the Context of International Security in its 2015 Report (A/70/174, 22 July 2015), paras. 24-25; and by the UN GGE on Advancing Responsible State Behaviour in Cyberspace in the Context of International Security in its 2021 Report (A/76/135, 14 July 2021), para. 69.
[33] Liivoja, Naagel and Väljataga, 'Autonomous Cyber Capabilities under International Law', pages 22-25; Michael N. Schmitt, 'Autonomous Cyber Capabilities and the International Law of Sovereignty and Intervention', in

the exceptionality of the right to self-defence, however, Article 51 poses important restrictions to its exercise, which raises important challenges when it comes to autonomy.

(i) Armed attack

Article 51 restricts the right to self-defence only to those situations involving an 'armed attack' – that is, 'the most grave forms of the use of force', to be distinguished from other less grave forms of use of force, such as 'mere frontier incidents' on the basis of their 'scale and effects'.[34] While this threshold is particularly ambiguous, especially in cyberspace, the International Group of Experts working on the Tallinn Manual identified some instances of standalone cyber operations that undoubtedly amount to armed attacks due to their scale and effects.[35] Cyber operations that seriously injure or kill a number of people or that cause significant damage to or destruction of property, for instance, clearly satisfy the scale and effects requirement.[36] The same holds for those cyber operations that individually fall below the threshold of an armed attack, but that taken together have the same scale and effects of kinetic armed attacks causing extensive death, injuries or physical damage or destruction.[37] However, the experts found that acts of cyber intelligence gathering and cyber theft, cyber operations involving brief or periodic interruption of non-essential cyber services do not qualify as armed attacks.[38] Nor does a low-intensity use of cyber force, such as the destruction of a single smartphone.[39]

In some other circumstances, however, it might be more difficult to determine whether a cyber operation reaches the threshold of armed attack. Consider, for instance, the case of cyber operations that do not result in injury, death, damage or destruction, but that still have extensive negative effects, for instance on the industrial or economic resources of the targeted state. While some experts did not consider them as armed attacks since they do not cause any harm or physical damage, others were of the view that 'it is not the nature (injurious or destructive) of the consequences that matters, but rather the extent of the ensuing effects'.[40] States are also divided on the matter: while a number of them have declared that only cyber operations causing injury or death to persons, damage or destruction amount to an armed attack,[41] others have a less restrictive position. The US, for instance, believes that 'the inherent right of self-

---

Autonomous Cyber Capabilities under International Law, by Rain Liivoja and Ann Väljataga (Tallinn: NATO CCDCOE Publications, 2021), pages 147-150.

[34] Military and Paramilitary Activities in and against Nicaragua (Nicaragua v. United States of America) (Judgement), ICJ Reports 1986 14 (International Court of Justice 1986), paras. 191 and 195.

[35] Schmitt, *Tallinn Manual 2.0 on the International Law Applicable to Cyber Operations,* Rule 71. The scale and effects standard was endorsed also by several states in their national positions, including Brazil (2021), Estonia (2021), Finland (2020), France (2019), Germany (2021), Iran (2020), Italy (2021), the Netherlands (2019), New Zealand (2020), Norway (2021), Sweden (2022), Switzerland (2021) and the United Kingdom (2021), available at: https://cyberlaw.ccdcoe.org/wiki/Self-defence.

[36] Schmitt, *Tallinn Manual 2.0 on the International Law Applicable to Cyber Operations,* Rule 71, Commentary, para. 8.

[37] According to the Experts, when different cyber incidents have been launched by the same originator, are related to each other and taken together meet the required scale and effects, they amount to a composite armed attack. *Ibid*, para 11. The same view was explicitly shared also by France (2019) and Singapore (2021) in their national positions, available at: https://cyberlaw.ccdcoe.org/wiki/Self-defence.

[38] Schmitt, *Tallinn Manual 2.0 on the International Law Applicable to Cyber Operations,* Rule 71, Commentary, para. 8.

[39] Delerue, *Cyber Operations and International Law,* page 330.

[40] Schmitt, *Tallinn Manual 2.0 on the International Law Applicable to Cyber Operations,* Rule 71, Commentary, para. 12.

[41] See, for instance, the national positions adopted by Estonia (2021), Germany (2021), the Netherlands (2019) and the UK (2021), available at: https://cyberlaw.ccdcoe.org/wiki/Self-defence.

defence potentially applies *any illegal use of force*', regardless of its scale or effects.[42] Other states have claimed that, in certain circumstances, even malicious cyber operations severely affecting the national critical infrastructures may be considered 'armed attack'.[43]

The lack of agreement on the definition of armed attack in cyberspace limits the possible use of autonomous cyber capabilities for defensive purposes as it is not clear when they are entitled to respond to a malicious cyber operation by resorting to self-defence. Even if there were consensus on the threshold of 'armed attack' in cyberspace, its indicators – the scale and effects of an attack – are qualitative in nature and therefore cannot be encoded in a computer program.[44] The assessment of the scale and effects of a malicious cyber operation must be carried out on a case-by-case basis, taking into account not only technical information but also other factors such as the strategic context and the effects of cyber operations beyond cyberspace, which currently goes beyond autonomous cyber capabilities' understanding.[45] This is highly problematic since there is a real risk that, if deployed in complex and unpredictable environments, autonomous cyber capabilities will wrongly qualify a malicious cyber operation as an armed attack and forcefully respond in self-defence.[46] Of course, the possibility of making a mistake of law and/or facts is not excluded even when the human is in the loop, but the fact that such assessment is delegated to autonomous cyber capabilities only complicates matters. Should this happen, it would qualify as a breach of the prohibition of threat or use of force and may lead to an unintended escalation of hostilities.[47]

(ii) Necessity, proportionality and immediacy

The use of force in self-defence is not unrestrained, but rather subjected to the principles of necessity, proportionality and immediacy.[48] The use of force in self-defence is deemed necessary only when it is a means of last resort to repel or defeat an armed attack. The determination of whether a forceful response is necessary must be taken on a case-by-case basis by the victim state and all other non-forceful measures must have already failed or must be reasonably expected to fail.[49] Once the use of

---

[42] US Department of Defense, Office of the General Counsel, Law of War Manual (June 2015, updated December 2016), paras. 1.11.5.2, 16.3.3.1 (emphasis added).
[43] See the national positions adopted by France (2019), Norway (2021) and Singapore (2021), available at: https://cyberlaw.ccdcoe.org/wiki/Self-defence.
[44] Michael N. Schmitt, 'Cyber Operations in International Law: The Use of Force, Collective Security, Self-Defense and Armed Conflicts', in *Proceedings of a Workshop on Deterring Cyberattacks: Informing Strategies and Developing Options for U.S. Policy*, by National Research Council (Washington DC: National Academies Press, 2010).
[45] See the national positions of France (2019), Germany (2021), Italy (2021) and Norway (2021), available at: https://cyberlaw.ccdcoe.org/wiki/Self-defence.
[46] Michael N. Schmitt, 'Autonomous Cyber Capabilities and the International Law of Sovereignty and Intervention,' in *Autonomous Cyber Capabilities under International Law*, by Rain Liivoja and Ann Väljataga (Tallinn: NATO CCDCOE Publications, 2021), page 148.
[47] *Ibid.*
[48] Although Article 51 does not expressly mention the principles of necessity, proportionality and immediacy, the ICJ noted in the Nicaragua case that these conditions are 'well established in customary international law'. See Military and Paramilitary Activities in and against Nicaragua (Nicaragua v. United States of America) (Judgement), ICJ Reports 1986, para. 176; Legality of the Threat or Use of Nuclear Weapons (Advisory Opinion), ICJ Reports 1996 226 (International Court of Justice 1996), para. 41; Oil Platforms (Islamic Republic of Iran v. United States of America) (Judgement), ICJ Reports 2003 161 (International Court of Justice 2003), para. 198 ; Armed Activities on the Territory of the Democratic Republic of the Congo v. Uganda (Judgement), ICJ Reports 2005 168 (International Court of Justice 2005), para. 233.
[49] The 2012 US Presidential Policy Directive 20, for example, requires the United States to consider whether law enforcement or passive network defence techniques are sufficient to repel or defeat the hostile attack, before resorting to active cyber defence. See, in this respect, the US Presidential Policy Directive 20, 2012; and Marco

force in self-defence is considered necessary, the level of force employed in the response must be proportionate to the purpose of repelling or defeating the armed attack. This means that the scale, scope, duration and intensity of the response shall not exceed that required to end the situation that has triggered the right to self-defence.[50] That said, while the level of force used in self-defence must always be proportionate, it may not necessarily be quantitively equivalent to that used in the armed attack; more force may be necessary, or less force may be sufficient to respond.[51] Finally, since self-defence is intended to end the malicious cyber operation, it must be launched while the armed attack is still ongoing or immediately after. Whenever such temporal proximity is lacking, the victim state's forceful response would be unnecessary and disproportionate and would amount to an unlawful act of retaliation. Yet, there is no fixed period within which a state should respond in self-defence. Rather, the state under attack must be allowed a 'reasonable window of time' to respond, which may vary according to the context and the preparedness of the victim state.[52]

Whenever autonomous cyber capabilities are used in the act of self-defence, their response must comply with the principles of necessity, proportionality and immediacy.[53] These assessments, however, heavily rely on circumstances and on the judgement of the decision-makers, who should determine whether or not there are any alternatives to the use of force and carefully consider the intensity and timing of the response, also considering how the situation is evolving. Since autonomous cyber capabilities currently lack the necessary situational awareness and human judgement to make such decisions, they cannot be expected to comply with these principles. However, should an autonomous cyber capability violate one of these principles, the response in self-defence would amount to a violation of international law. Therefore, some degree of human control should be retained in the definition of whether or not a response in self-defence is necessary, proportional and immediate.

(iii) Attribution

Article 51 does not include any reference to the potential author of an armed attack against which a state may use force in self-defence. This may be simply because in 1945 only a nation-state may have used such a level of force against another. Since then, however, numerous non-state actors have acquired the capacity to perpetrate actions that may amount to armed attacks in their scale and effects. This has led to a longstanding debate as to whether states can exercise the right to self-defence against non-state actors, which is still ongoing.[54]

According to the traditional interpretation of the UN Charter and customary international law, the right of self-defence can only be exercised by a state against another state. This was the position of the ICJ in the Advisory Opinion on the Wall, where it stated that 'Article 51 of the Charter thus recognises the

---

Roscini, *Cyber Operations and the Use of Force in International Law* (Oxford: Oxford University Press, 2014), page 148.

[50] See Peter Margulies, 'A Moment in Time: Autonomous Cyber Capabilities, Proportionality and Precautions', in *Autonomous Cyber Capabilities under International Law*, by Rain Liivoja and Ann Väljataga (Tallinn: NATO CCDCOE Publications, 2021), 152–80, page 162. On the principle of proportionality under *jus ad bellum*, see also Enzo Cannizzaro, *Il Principio della Proporzionalità nell'Ordinamento Internazionale* (Milano: Giuffré Editore, 2000), pages 278-296.

[51] Schmitt, *Tallinn Manual 2.0 on the International Law Applicable to Cyber Operations,* Rule 72, Commentary, para. 5.

[52] *Ibid*, Rule 73, Commentary, para.4.

[53] Schmitt, 'Autonomous Cyber Capabilities and the International Law of Sovereignty and Intervention', page 150.

[54] Andrew Clapham, *War* (Oxford: Oxford University Press, 2021), pages 127-128. For an overview of the debate on self-defence against cyber attacks by non-state actors, see Roscini, *Cyber Operations and the Use of Force in International Law,* pages 80-88.

existence of an inherent right of self-defence in the case of armed attack by one state against another state'.[55] The only exception would be that of an armed attack carried out by a non-state actor acting by or on behalf of a state. According to the Articles on the Responsibility of States for Internationally Wrongful Acts, a state is responsible for the conduct of a private person or group whenever it is 'acting on the instruction of, or under the direction or control of, that state in carrying out the conduct'.[56] In this case, the sponsoring state would be considered responsible for the armed attack and the victim state would be entitled to use force against it to end the attack.[57]

An alternative approach that is gaining consensus in the international community and that has been endorsed also by the majority of experts working on the Tallinn Manual[58] advocates the recognition of the right to self-defence against non-state actors that are not operating on behalf of another state. According to this view, therefore, the right to self-defence could be equally exercised against states and non-state actors as long as their conduct amounts to an armed attack.[59] However, whether a state can use force on the territory of another state to terminate an attack launched by a non-state actor without violating the host state's sovereignty is still the subject of debate.[60] The majority of experts working on the Tallinn Manuals have opined that self-defence in these circumstances is permissible as long as the host state is 'unable (e.g., because it lacks the expertise or technology) or unwilling' to prevent or terminate an armed attack launched by a non-state actor from its territory.[61] A minority of experts, however, contends that using force in self-defence on the territory of a state to which the armed attack is not attributable is, in the absence of the state's consent or the authorisation of the United Nations Security Council, unlawful.[62]

Although the debate is not settled, a few states have explicitly accepted the exercise of the right of self-defence against non-state actors in their national position on cyberspace, when their conduct is not attributable to any state.[63] Should the traditional approach prevail, therefore, those states deploying autonomous cyber capabilities must first ascertain whether these technologies are capable of legally attributing an armed attack to the responsible state.[64] Nonetheless, while autonomous cyber capabilities can facilitate technical attribution (by, for example, identifying the source computer from its serial number, MAC address or IP address), the legal attribution of an armed attack to the responsible state

---

[55] Legal Consequences of the Construction of a Wall in the Occupied Palestinian Territory (Advisory Opinion), ICJ Reports 2004 136 (International Court of Justice 2004), para. 194. The same conclusion was reiterated by the ICJ in Armed Activities on the Territory of the Democratic Republic of the Conngo v. Uganda (Judgement), ICJ Reports 2005,
para. 146.
[56] International Law Commission, Draft Articles on the Responsibility of States for Internationally Wrongful Acts, 2001, Article 8.
[57] This was confirmed also by the ICJ in Military and Paramilitary Activities in and against Nicaragua (Nicaragua v. United States of America) (Judgement), ICJ Reports 1986, para. 195; and Armed Activities on the Territory of the Democratic Republic of the Conngo v. Uganda (Judgement), ICJ Reports 2005, paras. 146-147.
[58] Schmitt, *Tallinn Manual 2.0 on the International Law Applicable to Cyber Operations,* Rule 71, Commentary, para. 19.
[59] Dinstein, *War, Aggression and Self-Defence*, pages 245-249.
[60] Schmitt, 'Autonomous Cyber Capabilities and the International Law of Sovereignty and Intervention', page 149.
[61] Schmitt, *Tallinn Manual 2.0 on the International Law Applicable to Cyber Operations,* Rule 71, Commentary, para. 25. See also Roscini, *Cyber Operations and the Use of Force in International Law,* pages 81-82.
[62] Schmitt, *Tallinn Manual 2.0 on the International Law Applicable to Cyber Operations,* Rule 71, Commentary, para. 25.
[63] States that accept the exercise of the right to self-defence in cyberspace against non-state actors that are not acting on behalf of a state are Germany, Israel, Italy, the Netherlands, Poland and the US, available at: https://cyberlaw.ccdcoe.org/wiki/Self-defence.
[64] On the legal attribution of an internationally wrongful act to a state, see International Law Commission, 'Draft Articles on Responsibility of States for Internationally Wrongful Acts', Articles 4–11.

seems to be beyond autonomous cyber capabilities. The determination of whether a state is responsible for an armed attack is particularly complex and context-dependent.[65]

While it is true that neither the law on state responsibility nor customary international law offers clear standards or burdens of proof for attribution in cases of self-defence,[66] states are also increasingly aware of the need to support their claims with evidence.[67] They can rely on a wide range of evidence which may not always be found in cyberspace or processed by a computer.[68] Attribution can be also challenged by the fact that cyberspace offers numerous opportunities to hide or falsify the originator of a cyber operation.[69] For this reason, an autonomous cyber capability should be able to consider a wide range of factors including 'the reliability, quantum, directness, nature (e.g., technical data, human intelligence) and specificity of the relevant and available information'.[70] This seems to be impossible in the current state of technology since qualitative assessments and human judgement cannot be encoded into a computer program.

(iv) Anticipatory self-defence

Article 51 of the UN Charter only refers to those situations in which 'an armed attack occurs'. Nevertheless, states have often invoked the existence of a right to anticipatory self-defence under customary international law which would allow them to forcefully respond to threats before an armed attack occurs.[71] Whether and to what extent such a right exists under international law is one of the most controversial questions related to *jus ad bellum* and it remains unsettled. Neither states nor scholars have reached a consensus on this and the ICJ has so far avoided taking a clear position.[72]

Although the terminology on this topic is neither hegemonic nor consistent, it is possible to identify at least two forms of anticipatory self-defence: pre-emptive and preventive self-defence. The former concerns a threat that has not yet started but that is imminent; the latter concerns preventive actions

---

[65] Schmitt, *Tallinn Manual 2.0 on the International Law Applicable to Cyber Operations,* page 81. On legal and technical attribution in cyberspace, see also Delerue, *Cyber Operations and International Law,* pages 55 - 188.
[66] On the standard of proof for self-defence in international law, see James A. Green, 'Fluctuating Evidentiary Standards for Self-Defence in the International Court of Justice', *International and Comparative Law Quarterly* 58 (2009): 163–79.
[67] While some states claim they are under no obligation to disclose the evidence upon which attribution is made (France, the UK and the US), others have contended that self-defence is justified only when the origin of the attack and the identity of those responsible are sufficiently certain and that the burden of proof is a 'heavy one' (The Netherlands). The need to substantiate accusations with evidence was also highlighted in the 2015 Report of the United Nations Group of Governmental Experts on Developments in the Field of Information and Telecommunications in the Context of International Security, UN Doc. No. A/70/174, 22 July 2015, para. 28(f). On the standard or burden of proof in cyberspace, see Nicholas Tsagourias and Michael Farrell, 'Cyber Attribution: Technical and Legal Approaches and Challenges', *The European Journal of International Law* 31, no. 3 (2020): 941–67, pages 955-956.
[68] Delerue, *Cyber Operations and International Law,* pages 87-109; and Tsagourias and Farrell, 'Cyber Attribution: Technical and Legal Approaches and Challenges', pages 955-959.
[69] Delerue, *Cyber Operations and International Law,* pages 88-89. See also Schmitt, *Tallinn Manual 2.0 on the International Law Applicable to Cyber Operations,* Rule 15, Commentary, para. 15.
[70] Schmitt, *Tallinn Manual 2.0 on the International Law Applicable to Cyber Operations,* pages 81-82.
[71] On anticipatory self-defence, see Tom Ruys, *'Armed Attack' and Article 51 of the UN Charter: Evolutions in Customary Law and Practice* (Cambridge; New York: Cambridge University Press, 2010), pages 250-367.
[72] Although the ICJ has generally refused to address the issue of response to an 'imminent threat of armed attack' in its jurisprudence (Military and Paramilitary Activities in and against Nicaragua (Nicaragua v. United States of America) (Judgement), ICJ Reports 1986, para. 103; Armed Activities on the Territory of the Democratic Republic of the Congo v. Uganda (Judgement), ICJ Reports 2005, para. 222), it has consistently endorsed a conservative interpretation of the right to self-defence. Thus, the ICJ may be expected to be at least sceptical with respect to the recognition of a right to anticipatory self-defence.

against future threats which are neither imminent nor have fully materialised.[73] While preventive self-defence is generally considered unlawful,[74] pre-emptive self-defence has gained some attention in the public debate on self-defence, including at the United Nations level.[75] It relies on the standard of 'imminence' proposed by the US Secretary of State Daniel Webster in the famous Caroline incident, according to which self-defence is permitted when a state faces an imminent attack giving rise to a 'necessity of self-defence, instant, overwhelming, leaving no choice of means and no moment for deliberation'.[76]

The doctrine of pre-emptive self-defence has been incorporated also in the Tallinn Manuals. Rule 73 of the Tallinn Manual 2.0 relies on Webster's standard of imminence in stating that '[t]he right to use force in self-defence arises if a cyber armed attack occurs or is *imminent*'.[77] While there are several approaches to when a state can exercise its right to self-defence towards an imminent threat, the International Group of Experts adopted the 'last feasible window of opportunity' standard elaborated by Schmitt. Here, a state may act in anticipatory self-defence against an imminent armed attack whenever a 'failure to act at that moment would reasonably be expected to result in the state being unable to defend itself effectively when the attack actually starts'.[78] This criterion is quite flexible, as it does not refer to a fixed temporal frame but rather is context-dependent. However, some important limits cannot be breached. In particular, the experts stressed that states may exercise anticipatory self-defence only if they have reasonable grounds to believe that the attack is about to occur.[79] Until then, they may only respond with actions short of the use of force.[80] The same position has also been shared by some states in their national positions on cyberspace.[81]

Should a consensus over anticipatory self-defence emerge, autonomous cyber capabilities might facilitate a prompt response to an imminent armed attack thanks to their high speed. However, it might

---

[73] On the different categories of anticipatory self-defence, see Terry D. Gill, 'The Temporal Dimension of Self-Defence: Anticipation, Pre-Emption, Prevention and Immediacy,' in *International Law and Armed Conflict: Exploring the Faultlines*, by Michael N. Schmitt and Jelena Pejic, vol. Vol. 15, International Humanitarian Law Series, 2007, 113–55.

[74] Preventive self-defence was advocated by the US in the 'Bush Doctrine' which was adopted in the aftermath of the 9/11, but with the exception of Israel, most states and scholars reject this doctrine for being too broad and permissive. See O'Meara, 'Reconceptualising the Right of Self-Defence against 'Imminent' Armed Attacks', pages 283-284.

[75] In the 2004 Report 'A more secure world: our shared responsibility' (A/59/565, 02 December 2004), the UN Secretary General's High-level Panel on Threats, Challenges and Change referred to both pre-emptive and preventive self-defence, declaring the former lawful and the latter unlawful, unless authorised by the UN Security Council (paras. 188-192). This position was later endorsed also by the UN Secretary General Kofi Annan in his 2005 Report 'In larger freedom: towards development, security and human rights for all' (A/59/2005, 21 March 2005, para. 124), and by the Special Rapporteur on Extrajudicial, Summary or Arbitrary Executions in the 2020 Report on 'Use of armed drones for targeted killings' (A/HRC/44/38, 15 August 2020, para. 52).

[76] Daniel Webster, 'Letter to Henry Stephen Fox,' in *The Papers of Daniel Webster: Diplomatic Papers, Vol 1, 1841-1843*, by KE Shewmaker (Ed.), 1983, page 62. The Caroline doctrine was subsequently confirmed by the International Military Tribunal (Nuremberg Tribunal) in Prosecutor v. Goering et al (Judgement), 1 October 1946, para. 208.

[77] Schmitt, *Tallinn Manual 2.0 on the International Law Applicable to Cyber Operations,* Rule 73 (emphasis added).
Not all scholars, however, agree with this position: there is still a high scepticism with respect to anticipatory self-defence also in the context of cyberspace. See, e.g., Delerue, *Cyber Operations and International Law,* pages 465-477.

[78] Schmitt, *Tallinn Manual 2.0 on the International Law Applicable to Cyber Operations,* Rule 73, para. 4. The 'last feasible window of opportunity' standard was first formulated in Michael N. Schmitt, 'Responding to Transnational Terrorism Under the Jus Ad Bellum: A Normative Framework,' *Naval Law Review* 56, no. 1 (2008).

[79] Preventive strikes do not qualify as a lawful exercise of the right to anticipatory self-defence. Schmitt, *Tallinn Manual 2.0 on the International Law Applicable to Cyber Operations,* Rule 73, Commentary, para. 10.

[80] *Ibid*, para. 11.

[81] See the national positions on cyberspace of the following States: Australia (2020), Brazil (2021), France (2019), Germany (2021), Israel (2020), New Zealand (2020), Singapore (2021) and the United Kingdom (2021), available at: https://cyberlaw.ccdcoe.org/wiki/Self-defence.

be extremely difficult to exercise anticipatory self-defence in cyberspace. As underlined by Roscini, 'in the absence of visible indications, convincingly establishing the origin, nature and imminence of the cyber attack and the necessity and proportionality of the reaction may prove to be an impossible task'.[82] Autonomy is only going to further amplify this problem. At the current state of technological development, it seems unlikely that autonomous cyber capabilities will be able to conduct such assessments or to establish when the 'last feasible window of opportunity' to respond is, absent human intervention.

5.2 Countermeasures

To this day, no state or international organisation has qualified a cyber operation as reaching the threshold of an armed attack.[83] States also seem to prefer engaging in low-intensity cyber operations which are less expensive, easier to conduct and have less risk of a full-scale response (or any response at all) by the victim state.[84] As previously mentioned, unless they are part of a composite armed attack whose scale and effects reach the threshold under Article 51 of the UN Charter, low-intensity cyber operations do not ordinarily give rise to the right to self-defence.[85] Yet, states may still respond to cyber operations falling short of an armed attack by means of other remedies, including countermeasures.[86]

Countermeasures are unilateral 'measures that would otherwise be contrary to the international obligations of an injured state towards the responsible state if not taken in response to an internationally

---

[82] Roscini, *Cyber Operations and the Use of Force in International Law,* pages 79-80.
83 Although some experts and scholars argued Stuxnet reached the threshold under Article 51 of the UN Charter, since it caused some damage to the Natanz nuclear enrichment facility, no State or international organization has publicly qualified it as 'armed attack'. Given the unsettled nature of the armed attack threshold, states and international organisations usually prefer to refer to a more general colloquial notion of 'cyber attack'. Consider, by way of example, the cyber operation launched in 2022 against the Viasat's satellite KA-SAT network, or the cyber operations conducted in July 2022 against the Albanian national digital infrastructures, both defined as 'cyber attacks'. It is worth noting, however, that the notion of 'cyber attack' used in the latter two examples is colloquial and distinct from the specific meaning for 'cyber attack' under international law that is used, for example, in the Tallinn Manual 2.0. According to Rule 92, which reflects Article 49 (1) of the Additional Protocol I to the Geneva Conventions, a cyber attack is 'a cyber operation […] that is reasonably expected to cause injury or death to persons or damage or destruction to objects'. Thus, while a 'cyber attack' might be an 'armed attack' under Article 51 of the UN Charter or an 'attack' under *jus in bello*, these expressions are not necessarily coterminous. See Schmitt, *Tallinn Manual 2.0 on the International Law Applicable to Cyber Operations*, Rule 92; and Roscini, *Cyber Operations and the Use of Force in International Law*, page 17. On the satellite KA-SAT network case, see Council of the European Union, 'Russian Cyber Operations against Ukraine: Declaration by the High Representative on Behalf of the European Union', 10 May 2022, available at: https://www.consilium.europa.eu/en/press/press-releases/2022/05/10/russian-cyber-operations-against-ukraine-declaration-by-the-high-representative-on-behalf-of-the-european-union/; and the US Department of State, 'Attribution of Russia's Malicious Cyber Activity Against Ukraine', 10 May 2022, available at: https://www.state.gov/attribution-of-russias-malicious-cyber-activity-against-ukraine/. On the cyber operations against Albania, see United States, 'Statement by NSC Spokesperson Adrienne Watson on Iran's Cyberattack against Albania', 7 September 2022, available at: https://www.whitehouse.gov/briefing-room/statements-releases/2022/09/07/statement-by-nsc-spokesperson-adrienne-watson-on-irans-cyberattack-against-albania/; NATO, 'Statement by the North Atlantic Council Concerning the Malicious Cyber Activities against Albania', 8 September 2022, available at: https://www.nato.int/cps/en/natohq/official_texts_207156.htm.
[84] Roscini, *Cyber Operations and the Use of Force in International Law*, page 104.
[85] Low intensity cyber operations (e.g., acts of cyber intelligence gathering and cyber theft) do not normally reach the threshold of force required for an armed attack. Yet, when distinct (albeit related) low intensity cyber operations are considered together, they may reach the required threshold. For example, concurrent low intensity cyber operations that are intended to enable or facilitate a wider, concurrent conventional attack could be considered part of a composite armed attack. See *supra note* 37; Schmitt, *Tallinn Manual 2.0 on the International Law Applicable to Cyber Operations,* Rule 71, Commentary, paras. 8 and 11; and NATO, 'Allied Joint Doctrine for Cyberspace Operations (AJP-3.20)' (NATO Standardization Office, January 2020), page 20, para. 3.7.
[86] Military and Paramilitary Activities in and against Nicaragua (Nicaragua v. United States of America) (Judgement), ICJ Reports 1986, para. 249.

wrongful act to procure cessation and reparation'.[87] Several states and scholars have confirmed the applicability of the law of countermeasures to cyber operations[88] and there is no reason to believe it will not apply also to those carried out using autonomous cyber capabilities.[89] An example of countermeasures carried out by autonomous cyber capabilities would be an autonomous hack-back that would violate the sovereignty of the responsible state and would be intended to compel the responsible state to cease its prior internationally wrongful conduct.[90]

Given the exceptional character of countermeasures, they are considered justified under international law only when certain strict conditions are met. First, countermeasures can only be adopted by a state against another state in response to a previous international wrongful act that injured the state taking the countermeasure.[91] They must also have the specific intent to either put an end to an ongoing internationally wrongful act, secure reparations for one that has already occurred, or both. For this reason, countermeasures must be temporary in nature.[92] This is crucial, as the objective of countermeasures is the restoration of a condition of legality between the responsible and injured states and not the aggravation of the existing dispute.[93] As soon as the injured state succeeds in compelling the responsible state to cease its wrongful conduct or to grant reparation, the countermeasure must be terminated.[94] In the same way, countermeasures must be proportionate; that is, they must be 'commensurate with the injury suffered, taking into account the gravity of the internationally wrongful act and the rights in question'.[95] Proportionality limits the kind of responses that can be adopted by the victim state in response to an internationally wrongful act to those that are more tailored to fit the state's objective to enforce the respect of international law and/or to grant reparations.[96] Countermeasures should also be reversible, as far as possible, since inflicting considerable damage to the responsible

---

[87] International Law Commission, 'Draft Articles on Responsibility of States for Internationally Wrongful Acts', Commentary, Part III, Chapter II, para. 1. On the definition of countermeasures, see also Denis Alland, 'The Definition of Countermeasures', in *The Law of International Responsibility*, by James Crawford, Alain Pellet and Simon Olleson, Oxford Commentaries on International Law (Oxford: Oxford University Press, 2010): 1127-1136.
[88] See Schmitt, *Tallinn Manual 2.0 on the International Law Applicable to Cyber Operations,* Rule 20. See also the national positions on cyberspace respectively adopted by the following States: Australia (2020), Canada (2022), Estonia (2019, 2021), Finland (2020), France (2019), Germany (2021), Italy (2021), Japan (2021), the Netherlands (2019), New Zealand (2020), Norway (2021), Singapore (2021), Sweden (2022), Switzerland (2021), the United Kingdom (2018, 2021, 2022) and the United States (2016, 2020, 2021), available at: https://cyberlaw.ccdcoe.org/wiki/Countermeasures. Other States, however, were cautious in applying the law of countermeasures to cyberspace. See, by way of example, the Statement by the Chinese Delegation at the Thematic Debate of the First Committee of the 72th UNGA (16 October 2017), available at: http://un.china-mission.gov.cn/eng/chinaandun/disarmament_armscontrol/unga/201710/t20171030_8412332.htm.
[89] Schmitt, 'Autonomous Cyber Capabilities and the International Law of Sovereignty and Intervention', page 143.
[90] *Ibid.*
[91] International Law Commission, 'Draft Articles on Responsibility of States for Internationally Wrongful Acts', Article 49 (1). The fundamental prerequisite of a prior international wrongful act which injured the State taking countermeasures was also underlined by the ICJ in The Gabčcíkovo Nagymaros Project (Hungary vs. Slovakia) (Judgement), ICJ Reports 1997 7 (International Court of Justice 1997), para. 83. On this point, see also Alland, 'The Definition of Countermeasures', page 1135.
[92] International Law Commission, 'Draft Articles on Responsibility of States for Internationally Wrongful Acts', Article 49 (2). See also Schmitt, *Tallinn Manual 2.0 on the International Law Applicable to Cyber Operations,* Rule 21.
[93] International Law Commission, 'Draft Articles on Responsibility of States for Internationally Wrongful Acts', Article 49, Commentary, para. 7.
[94] *Ibid*, Article 49 (1). See also Maurice Kamto, 'The Time Factor in the Application of Countermeasures', in *The Law of International Responsibility*, by James Crawford, Alain Pellet and Simon Olleson, Oxford Commentaries on International Law (Oxford: Oxford University Press, 2010): 1169–76, page 1173.
[95] International Law Commission, 'Draft Articles on Responsibility of States for Internationally Wrongful Acts', Article 50. On this point, see also Roger O'Keefe, 'Proportionality', in *The Law of International Responsibility*, by James Crawford, Alain Pellet and Simon Olleson, Oxford Commentaries on International Law (Oxford: Oxford University Press, 2010), 1157–68; Michael Newton and Larry May, *Proportionality in International Law* (New York: Oxford University Press, 2014), pages 181-189.
[96] On the principle of proportionality under the law regulating countermeasures, see Cannizzaro, *Il Principio della Proporzionalità nell'Ordinamento Internazionale,* pages 359-427.

state would amount to a punitive retaliation.[97] Hence, whenever the injured state can select among a number of lawful and effective countermeasures, 'it should select one which permits the resumption of performance of the obligations suspended as a result of countermeasures'. [98] In addition, countermeasures cannot result in a violation of any peremptory norm of general international law, including the prohibition of threat or use of force under the UN Charter,[99] and the obligations relating to human rights and humanitarian law.[100] Finally, on a procedural level, the injured state must call on the responsible state to comply with its obligations under the law of state responsibility before taking countermeasures (summation), [101] and notify the responsible state of its intention to take countermeasures and offer negotiations,[102] unless the countermeasures are particularly urgent.[103]

If countermeasures are taken at the injured state's own risk,[104] when such decisions are delegated to an autonomous cyber capability, additional issues arise. First of all, when autonomous cyber capabilities are deployed, the assessment of the situation that is generally carried out by the injured state is entirely delegated to a computer. However, autonomous cyber capabilities currently lack the necessary situational awareness to understand whether an act or an omission amounts to an internationally wrongful act. This is highly problematic since a mistake in the assessment of the situation would

---

[97] *Ibid*, Article 49 (3). See also Kamto, 'The Time Factor in the Application of Countermeasures', pages 1174-1175.
As for the reversibility requirement of countermeasures, the International Group of Expert working to the Tallinn Manual was divided on whether States need to select the cyber countermeasure that is more likely to be reversed or simply one that is reversible: while the majority of experts was of the view that as long as both countermeasures are reversible then States may select either option, a few experts argued that States should adopt the countermeasure that is least likely to exacerbate the ongoing dispute. See Schmitt, *Tallinn Manual 2.0 on the International Law Applicable to Cyber Operations,* Rule 23, Commentary, para. 9.
[98] International Law Commission, 'Draft Articles on Responsibility of States for Internationally Wrongful Acts', Article 49, Commentary, para. 9.
[99] *Ibid*, Article 50 (1). The unlawfulness of forcible countermeasures was also confirmed by the ICJ in its jurisprudence and by the UN General Assembly in the Declaration on Principles of International Law concerning Friendly Relations and Cooperation among States in accordance with the Charter of the United Nations, Resolution 2625 (XXV), annex, first principle. Yet, few commentators have advocated for the possibility of using armed reprisals against uses of force short of armed attack (e.g., Judge Simma in his Separate Opinion to the Oil Platform case, para. 12). This position was also shared by a minority of the experts working to the Tallinn Manual, as underlined in Schmitt, *Tallinn Manual 2.0 on the International Law Applicable to Cyber Operations,* Rule 22, Commentary, paras. 10-13. On non-forcible countermeasures, see Alland, 'The Definition of Countermeasures', page 1130; and Charles Leben, 'Obligations Relating to the Use of Force and Arising from Peremptory Norms of International Law', in *The Law of International Responsibility*, by James Crawford, Alain Pellet and Simon Olleson, Oxford Commentaries on International Law (Oxford: Oxford University Press, 2010): 1197-1204. On the debate on forcible countermeasures in cyberspace, see also Delerue, *Cyber Operations and International Law,* pages 483 - 487.
[100] On this point, see Silvia Borelli and Simon Olleson, 'Obligations Relating to Human Rights and Humanitarian Law', in *The Law of International Responsibility*, by James Crawford, Alain Pellet and Simon Olleson, Oxford Commentaries on International Law (Oxford: Oxford University Press, 2010): 1177–1196.
[101] International Law Commission, 'Draft Articles on Responsibility of States for Internationally Wrongful Acts', Article 52 (1) (a).
[102] *Ibid*, Article 52 (1) (b). The extent to which this procedural condition should be a prerequisite for taking countermeasures was highly debated in the ILC. On the one hand, it was necessary to ensure appropriate safeguards against premature and unlawful countermeasures; on the other hand, it was imperative to not compromise the effectiveness of countermeasures. Eventually, it was decided that the injured State may be exempted from the prior notification requirement only when urgent countermeasures are necessary to protect its rights. See Yuji Iwasawa and Naoki Iwatsuki, 'Procedural Conditions', in *The Law of International Responsibility*, by James Crawford, Alain Pellet and Simon Olleson, Oxford Commentaries on International Law (Oxford: Oxford University Press, 2010), 1149–1155.
[103] International Law Commission, 'Draft Articles on Responsibility of States for Internationally Wrongful Acts', Article 52 (2). On the importance of urgent countermeasures in cyberspace, see Schmitt, *Tallinn Manual 2.0 on the International Law Applicable to Cyber Operations,* Rule 21, Commentary, paras. 11-12, and the national positions on cyberspace of Canada (2022), Israel (2020), Norway (2021), the United Kingdom (2021) and the United States (2020), available at: https://cyberlaw.ccdcoe.org/wiki/Countermeasures.
[104] Including, *inter alia*, the risk of escalation, which must be always taken into account when deciding whether and how, to engage in countermeasures. On this point, see Schmitt, *Tallinn Manual 2.0 on the International Law Applicable to Cyber Operations,* Rule 21, Commentary, para. 2.

jeopardise the lawfulness of the countermeasure.[105] The same holds for a misattribution of the internationally wrongful act to a third state that is not responsible.[106] As mentioned before, legal attribution is a complex process that cannot be carried out without human intervention. Moreover, in the context of countermeasures, it is often difficult to attribute malicious cyber operations to a particular state with unqualified certainty.[107] Should a state entirely delegate the attribution process to an autonomous cyber capability and should this mistakenly direct a countermeasure against a third party, it would result in a breach of international law.[108]

Secondly, current autonomous cyber capabilities lack the necessary situational awareness and human judgement to determine which kind of countermeasure to adopt, and whether to suspend or terminate it as soon as it achieves its objective. Neither can autonomous cyber capabilities yet foresee the effects that such countermeasures may have on the targeted system and thereby avoid those that might result in a violation of international law.[109] It is also unlikely that autonomous cyber capabilities will be able to carry out a proportionality assessment since it would rely on both quantitative (the injury suffered) and qualitative (the gravity of the injury) elements, which must be weighted one against the other.[110]

In the law of countermeasures, proportionality 'is best understood as a prohibition against excesses rather than a requirement for equivalence or mathematical equity'.[111] Computers cannot yet carry out such an assessment and the inability of autonomous cyber capabilities to undertake a proportionality assessment is particularly problematic in cyberspace due to its interconnected and interdependent nature.[112] If autonomous cyber capabilities cause excessive harm, the state taking the purported countermeasure will be considered responsible for a violation of international law.[113] The same holds if autonomous cyber capabilities launch a countermeasure that reverberates across transborder networks and violates a legal obligation owed to a third state.[114] Finally, as for the procedural limit, while it is true that in case of urgency, the injured state may avoid notifying the responsible state of its intention to adopt countermeasures, it should still call on the responsible state to terminate the internationally wrongful act and to offer reparations.[115] Thus, unless the injured state first calls for reparations,

---

[105] International Law Commission, 'Draft Articles on Responsibility of States for Internationally Wrongful Acts', Article 49, Commentary, para. 3.
[106] Schmitt, 'Autonomous Cyber Capabilities and the International Law of Sovereignty and Intervention', page 145.
[107] Schmitt, *Tallinn Manual 2.0 on the International Law Applicable to Cyber Operations,* Rule 20, Commentary, para. 15.
[108] International Law Commission, 'Draft Articles on Responsibility of States for Internationally Wrongful Acts', Article 49, Commentary, para. 4; Schmitt, *Tallinn Manual 2.0 on the International Law Applicable to Cyber Operations,* Rule 25, Commentary, paras. 3-5; and Johann-Christoph Woltag, *Cyber Warfare: Military Cross-Border Computer Network Operations under International Law*, International Law Series (Cambridge; Antwerp; Portland: Intersentia, 2014), page 161.
[109] Woltag, *Cyber Warfare: Military Cross-Border Computer Network Operations under International Law,* page 161.
[110] See O'Keefe, 'Proportionality', pages 1160-1166; and Michael Newton and Larry May, *Proportionality in International Law* (New York: Oxford University Press, 2014), pages 182-183.
[111] Newton and May, *Proportionality in International Law,* page 186.
[112] Schmitt, *Tallinn Manual 2.0 on the International Law Applicable to Cyber Operations,* Rule 23, Commentary, para. 6.
[113] Schmitt, 'Autonomous Cyber Capabilities and the International Law of Sovereignty and Intervention', page 144.
[114] Schmitt, *Tallinn Manual 2.0 on the International Law Applicable to Cyber Operations,* Rule 25.
[115] The so-called summation requirement is almost unanimously recognized as a mandatory procedural condition for taking countermeasures and is well established under customary international law. The ILC has explicitly referred to it both in the 1996 and 2001 Draft Articles on the Responsibility of States for Internationally Wrongful Acts and in the 2009 and 2011 Draft Articles on the Responsibility of International Organizations. It was also confirmed in the international jurisprudence in the Naulilaa case (Portugal vs. Germany), Vol. II Reports of International Arbitral Awards 1011 (Special Arbitral Tribunal 1928), page 1026; Air Service Agreement of 27 March 1946 between the United States of America and France, Vol. XVIII Reports of International Arbitral Awards 417 (Arbitral Tribunal 1978), paras. 85-87; and The Gabčcíkovo Nagymaros Project (Hungary vs. Slovakia) (Judgement), ICJ Reports 1997, para. 84. However, some scholars have expressed some doubts on the existence of a summation requirement under the law regulating countermeasures. According to Carlo Focarelli,

autonomous cyber capabilities might be not entitled to immediately launch a countermeasure in response to an incoming malicious cyber operation, even in an urgent situation.[116] Urgent countermeasures may also only be taken on an *ad hoc* basis in the light of the circumstances of the case.[117] Autonomous cyber capabilities should therefore be able to assess whether urgent countermeasures are necessary to preserve the rights of the victim state. However, autonomous cyber capabilities currently lack the necessary situational awareness and human judgement to make such a decision. Thus, while autonomous cyber capabilities may facilitate a rapid response in emergencies, the victim state should still be able to make a cogent argument that it was necessary to act without notice.[118]

5.3 Plea of necessity

A third basis upon which a state may respond against a malicious cyber operation not reaching the threshold of an armed attack by means of an autonomous cyber capability is that of necessity (*état de nécessité*). According to Article 25 of the Articles on State Responsibility for Internationally Wrongful Acts, in exceptional cases, a state can rely on necessity when that is the only way to safeguard an essential interest against a grave and imminent peril.[119] While only a few states have included the plea of necessity in their national positions on cyberspace,[120] the International Group of Experts working on the Tallinn Manual 2.0 agreed that, since the plea of necessity is 'customary in nature', it can be applied in the cyber context.[121]

The plea does not depend on a prior commission of an internationally wrongful act, but from a 'grave and imminent peril […to] an essential interest of the state'.[122] While there is no accepted definition of 'essential interest', the International Group of Experts has defined it as 'one that is of fundamental and great importance to the state concerned'.[123] The determination of whether an interest is essential to a state is contextual and may vary from state to state but the International Group of Experts has identified a few instances in which the plea of necessity is most certainly implicated; that is, 'when critical infrastructure is targeted in a manner that may have severe negative impact on a state's security, economy, public health, safety, or environment'.[124] Whenever these interests are threatened by a peril that is severe enough to interfere with a national interest in a fundamental way and that is 'objectively

---

for example, consuetudinary law does not require the summation requirement (or any other offer to settle the dispute) prior to the taking of countermeasures (Carlo Focarelli, *Le Contromisure Nel Diritto Internazionale* (Milano: Giuffré Editore, 1994).

[116] Woltag, *Cyber Warfare: Military Cross-Border Computer Network Operations under International Law,* page 162.

[117] International Law Commission, 'Draft Articles on Responsibility of States for Internationally Wrongful Acts', Commentary to Article 52, paragraph 6.

[118] Schmitt, 'Autonomous Cyber Capabilities and the International Law of Sovereignty and Intervention', pages 144-145.

[119] International Law Commission, 'Draft Articles on Responsibility of States for Internationally Wrongful Acts', Article 25 (1).

[120] See the national positions of France (2019), Germany (2021), Japan (2021), the Netherlands (2019), Norway (2021), Sweden (2022) and Switzerland (2021), available at: https://cyberlaw.ccdcoe.org/wiki/Plea_of_necessity. See also the Italian Position Paper on 'International Law and Cyberspace' (2021), page 4, available at: https://www.esteri.it/mae/resource/doc/2021/11/italian_position_paper_on_international_law_and_cyberspace.pdf .

[121] Schmitt, *Tallinn Manual 2.0 on the International Law Applicable to Cyber Operations,* Rule 26, Commentary, para. 1.

[122] The state of necessity may indeed be caused by a natural disaster or by other situations that do not implicate international legal norms. See Schmitt, *Tallinn Manual 2.0 on the International Law Applicable to Cyber Operations,* Rule 26, Commentary, para. 9.

[123] Schmitt, *Tallinn Manual 2.0 on the International Law Applicable to Cyber Operations,* Rule 26, Commentary, para. 2.

[124] *Ibid*, para. 5.

established and not merely apprehended as possible', the plea of necessity can be invoked.[125] Significantly, it can be equally exercised against states and non-state actors and the conduct of the non-state actor does not be to be attributed to a state.[126]

By invoking the plea of necessity, a state may temporarily act other than in conformity with its obligations under international law (unless they explicitly or implicitly provide otherwise), if this is the only way the state has to safeguard its essential interests.[127] However, if there are other lawful means available, the state must prefer them over necessity.[128] Given that states may rely on necessity only in very exceptional circumstances, Article 25 of the Articles on State Responsibility for Internationally Wrongful Acts poses two particular restrictions: first, the plea is excluded if the state has contributed to the situation;[129] and second, necessity is excluded when the temporary non-compliance with international obligations constitutes a serious interference with the essential interests of another state towards which the obligation exists, or of the international community as a whole.[130] Moreover, the state's power to unilaterally establish a new balance of interests is not unrestrained, but must be proportionate to the objective it pursues; that is, safeguarding its essential interest against a grave and imminent peril.[131] It is still unsettled, however, whether a state acting under the plea of necessity can adopt forcible measures. The International Law Commission took no position on the matter,[132] and the International Group of Experts could not reach a consensus. While some of them argued that the use of force in response to harmful cyber operations is only permissible pursuant to the law of self-defence under Article 51 of the UN Charter, others maintained that prohibiting forcible measures in necessity would mean limiting a state's response to cyber operations amounting to use of force but not crossing the threshold of armed attack.[133] The only state that took an explicit position on this issue was France, which expressly stated that actions taken pursuant to the plea of necessity must be peaceful.[134]

To some extent, autonomous cyber capabilities might seem better suited to operate in a plea of necessity circumstance than under the law regulating the use of force or countermeasures. While in the latter cases, autonomous cyber capabilities must be able to assess whether an internationally wrongful act or an armed attack has occurred, in the case of the plea of necessity there is no such requirement. Moreover, under the plea of necessity, an autonomous cyber capability is not required to attribute the malicious cyber operation to a specific state or even to the initiator of the operation, who may remain unknown.[135] Finally, by acting based on necessity, an autonomous cyber capability may even violate an obligation of international law, such as the sovereignty of a state, as long as doing so does not seriously

---

[125] International Law Commission, 'Draft Articles on Responsibility of States for Internationally Wrongful Acts', Article 25, Commentary, para. 15.
[126] Schmitt, *Tallinn Manual 2.0 on the International Law Applicable to Cyber Operations,* Rule 26, Commentary, para. 6.
[127] Unless international obligations explicitly or implicitly exclude reliance on necessity. International Law Commission, 'Draft Articles on Responsibility of States for Internationally Wrongful Acts', Article 25 (2) (a).
[128] *Ibid*, Article 25, Commentary, para. 15. See also The Gabčcíkovo Nagymaros Project (Hungary vs. Slovakia) (Judgement), ICJ Reports 1997, para. 55.
[129] International Law Commission, 'Draft Articles on Responsibility of States for Internationally Wrongful Acts', Article 25 (2) (b). See also The Gabčcíkovo Nagymaros Project (Hungary vs. Slovakia) (Judgement), ICJ Reports 1997, para. 57.
[130] International Law Commission, 'Draft Articles on Responsibility of States for Internationally Wrongful Acts', Article 25 (1) (b).
[131] Cannizzaro, *Il Principio della Proporzionalità nell'Ordinamento Internazionale,* pages 327-347.
[132] International Law Commission, 'Draft Articles on Responsibility of States for Internationally Wrongful Acts', Article 25, Commentary, para. 21.
[133] Schmitt, *Tallinn Manual 2.0 on the International Law Applicable to Cyber Operations,* Rule 26, Commentary para. 18.
[134] Ministry of Defense of France, International Law Applied to Operations in Cyberspace, 9 September 2019, para. 8.
[135] Schmitt, 'Autonomous Cyber Capabilities and the International Law of Sovereignty and Intervention', page 145.

impair the latter's essential interests.[136] At the same time, however, given that the grounds of necessity may be invoked only in very exceptional cases, it is difficult for states to delegate such decisions to autonomous cyber capabilities. As underlined by France in its national position on cyberspace, the decision to invoke necessity in response to a malicious cyber operation cannot be taken systemically, but on a case-by-case basis according to a discretionary political decision.[137] This is partly because the state determines which are its own essential interests and this determination relies heavily on contextual elements and cannot be prejudged.[138] Thus, it is unlikely that autonomous cyber capabilities will be able to understand whether an essential interest of the state is threatened by a grave and imminent peril since such assessment depends on qualitative elements and predictions.[139] In the same way, an autonomous cyber capability would probably not be able to understand whether its conduct is in violation of an international obligation that excludes necessity, or whether it impairs the essential interests of other states or the international community. Likewise, it would neither be able to comprehend whether the state has contributed to the situation for which it is invoking necessity. Should it respond in a way that would violate such limitations, the deploying state would be considered responsible.[140]

5.4 Retorsions

A final measure of self-help that can be invoked by a victim state is retorsion. An act of retorsion is an unfriendly, albeit lawful, measure that can be adopted by states to express their disapproval of the activities of another.[141] It may be adopted in response to any hostile act including an internationally wrongful act or any other unfriendly conduct. The only substantial condition is that they must not cross the threshold of legality. Apart from that, there are no restrictions concerning their purpose, duration or character.[142] A state need not legally attribute the injurious activity to another state before engaging in acts of retorsion.[143] Thus, they represent a flexible way for a state to respond to a wide range of hostile activities, regardless of whether they amount to a violation of international law.

To date, only a few states have included retorsions in their national positions on cyberspace.[144] Yet, states have often adopted acts of retorsion in response to malicious cyber operations.[145] Acts of

---

[136] *Ibid*, pages 145-146.
[137] Ministry of Defense of France, International Law Applied to Operations in Cyberspace, 9 September 2019, para. 8.
[138] International Law Commission, 'Draft Articles on Responsibility of States for Internationally Wrongful Acts', Article 25, Commentary, para. 15.
[139] By definition, in cases of necessity the peril has not yet occurred and must be assessed on the basis of the evidence reasonably available at the time. *Ibid*, para. 16.
[140] Schmitt, 'Autonomous Cyber Capabilities and the International Law of Sovereignty and Intervention', page 146.
[141] International Law Commission, 'Draft Articles on Responsibility of States for Internationally Wrongful Acts', Chapter II, Commentary, para. 3. See also Thomas Giegerich, 'Retorsion', in *Max Planck Encyclopedias of International Law*, September 2020, https://opil.ouplaw.com/display/10.1093/law:epil/9780199231690/law-9780199231690-e983.
[142] Jeff Kosseff, 'Retorsion as a Response to Ongoing Malign Cyber Operations', in *20/20 Vision: The Next Decade*, by T. Jančárková et al. (Tallinn: NATO CCD COE Publications, 2020): 9–23, page 15-17.
[143] Johnson Durward and Michael N. Schmitt, 'Responding to Proxy Cyber Operations Under International Law', *Cyber Defense Review* 6, no. 4 (Fall 2021): 15–30, page 22.
[144] See the national positions of the following States: Estonia (2021), Germany (2021), the Netherlands (2019), New Zealand (2020), Norway (2021), Poland (2022), Singapore (2021), Switzerland (2021), the United Kingdom (2022) and the United States (2016) (2021), available at: https://cyberlaw.ccdcoe.org/wiki/Retorsion.
[145] The US, for instance, responded to the hacking of Sony Picture Entertainment in 2005 and of the Democratic National Congress in 2016, which were attributed to North Korea and Russia respectively, with traditional measures of retorsion, including the removal of diplomats and the adoption of new sanctions. See François Delerue, 'Reinterpretation or Contestation of International Law in Cyberspace?', *Israel Law Review* 52, no. 3 (2019): 295–326; Delerue, *Cyber Operations and International Law,* pages 431-432; Alessandro Stiano, *Attacchi*

retorsion may take a variety of forms and be carried out both inside and outside cyberspace. Examples of traditional acts of retorsion include the prohibition or limitation of normal diplomatic relations, the imposition of trade embargoes or the withdrawal of voluntary aid programmes.[146] In cyberspace, states may undertake acts of retorsion by sending warnings to cyber operatives belonging to another state, observing the adversary's cyber activities on one's own network using tools such as honeypots or slowing down malicious cyber operations conducted by other states.[147]

Given their flexibility and lack of substantial constraints, it seems likely that autonomous cyber capabilities will be able to carry out acts of retorsion without requiring any form of human intervention. Since retorsions are lawful in nature, they can be taken in response to any form of malicious cyber operation, regardless of whether they amount to an internationally wrongful act or not. Therefore, autonomous cyber capabilities are not required to assess the nature of the malicious cyber operation. Neither are retorsions subject to the legal constraints concerning necessity, proportionality or duration. Thus, autonomous cyber capabilities do not need to be able to conduct such assessments.

In some circumstances, however, there is a risk that acts of retorsions by autonomous cyber capabilities could result in an unintentional breach of international law, such as a violation of another state's sovereignty or a breach of the principle of non-intervention. Consider, for example, an autonomous cyber capability that responds to a malicious cyber operation by hacking back the adversary's system to access information. To the extent this operation constitutes legal peacetime espionage, it might be considered an act of retorsion. However, if the act of espionage causes damage to data or computer systems, then it will be likely considered a violation of the state's sovereignty and thus not an act of retorsion.[148] Thus, autonomous cyber capabilities should only be able to undertake acts of retorsions that clearly do not risk an internationally wrongful act. For example, an autonomous cyber capability might be programmed to respond to a malicious cyber operation by autonomously granting access to the adversary's network without using it, leaving the state to decide how to respond to the hostile activity.[149]

Informatici e Responsabilità Internazionale dello Stato, Edizioni Scientifiche Italiane, Legal Culture and International Flows 9 (Naples, 2023), pages 230-233.

[146] International Law Commission, 'Draft Articles on Responsibility of States for Internationally Wrongful Acts', Chapter II, Commentary, para. 3. In 2022, Albania decided to interrupt its diplomatic relations with the Islamic Republic of Iran in the aftermath of a malicious cyber operation carried out against the digital infrastructure of the Albanian government. 'Video Message of Prime Minister Edi Rama, Regarding the Response of the Albanian Government to the Cyber Attack against the Digital Infrastructure of the Government of the Republic of Albania'.

[147] According to the Experts, an example of cyber retorsion would be that of a state that employs an access control list to prevent communications from another state because the former enjoys sovereignty over the cyber infrastructure on its territory. Schmitt, *Tallinn Manual 2.0 on the International Law Applicable to Cyber Operations,* Commentary to Rule 20, para. 4. On this point, see Kosseff, 'Retorsion as a Response to Ongoing Malign Cyber Operations'.

[148] While the prevailing view among scholars is that peacetime cyber espionage is lawful under international law, Buchan notes that states tend to consider cyber espionage as a violation of sovereignty. See Russell Buchan, 'The International Legal Regulation of State-Sponsored Cyber Espionage', in *International Cyber Norms: Legal, Policy & Industry Perspectives,* by Anna-Maria Osula and Henry Rõigas (Tallinn: NATO CCD COE Publications, 2016), 65–86.

[149] *Ibid*, pages 21-22.

6. Conclusions

States are increasingly investing in the research and development of so-called 'autonomous cyber capabilities', namely those that can perform some tasks without real-time human intervention. While such capabilities will certainly offer some operational advantages in terms of speed and the ability to deal with enormous quantities of data, they will also pose substantial challenges under international law provisions, especially when they are used for defensive purposes short of armed conflict.

At the current state of technological development, autonomous cyber capabilities lack the necessary situational awareness and human judgement to assess the nature of a malicious cyber operation; that is, whether it amounts to an armed attack, a use of force falling short of an armed attack, a violation of sovereignty or of the principle of non-intervention, or a lawful (albeit unfriendly) act. Thus, they remain unable to select the most appropriate form of self-help. The only measure of self-help that can always be adopted regardless of the nature of the malicious cyber operation is retorsion. All the others can be invoked only in response to very specific circumstances. Thus, retorsions aside, autonomous cyber capabilities will likely require some degree of human intervention to properly identify the nature of the malicious cyber operation and select the most appropriate form of self-help.

While autonomous cyber capabilities may certainly facilitate technical attribution, they are currently unable to legally attribute a malicious cyber operation to the responsible state as it would require a high degree of situational awareness and human judgement. They would be expected to have a relatively high degree of certainty that a particular state was behind the malicious cyber operation and to respond using countermeasures (or self-defence, according to a traditional interpretation of Article 51 of the UN Charter). However, the process of legal attribution is complex and based on a wide range of evidence which may not always be found in cyberspace or processed by a computer. Thus, current autonomous cyber capabilities cannot legally attribute a cyber operation to the responsible state without any form of human intervention. This said legal attribution is not always required under international law. In case of a plea of necessity or retorsion – and self-defence, should the Tallinn Manual's approach prevail – an autonomous cyber capability could respond to a malicious cyber operation without legally attributing it to the responsible state.

Autonomous cyber capabilities' current lack of situational awareness, reliability and predictability makes them unsuitable to calibrate their response in the light of the principles of necessity and proportionality and other requirements relevant to the selected measure of self-help. This is particularly problematic when autonomous cyber capabilities are expected to act on circumstances precluding wrongfulness. Since actions based on self-defence, countermeasures or plea of necessity involve what is otherwise a violation of international law, they are strictly regulated by context-based legal constraints. Moreover, proportionality has a different connotation under *jus ad bellum*, countermeasures and the plea of necessity. Hence, autonomous cyber capabilities would be expected to carry out different proportionality assessments, according to the measure of self-help adopted and the situation in which they are operating. The only exception would be that of retorsions: since acts of retorsion are lawful, albeit unfriendly, they are not subject to the same requirements as self-defence, countermeasures or plea of necessity. Finally, it should be noted that the principle of proportionality finds different applications also under international humanitarian law and international human rights law, and that autonomous cyber capabilities will be therefore expected to carry out different proportionality assessments also under these two legal regimes, depending on the applicable law and the circumstances of the case.[150]

[150] For a general overview of the principle of proportionality under international humanitarian law and international human rights law, see Cannizzaro, *Il Principio della Proporzionalità nell'Ordinamento Internazionale*.

In a future where autonomous cyber capabilities will be deployed for defensive purposes, it will be important for states to predetermine in what circumstances they can rely on full automation of cyber defence and when they will need some form of human intervention. Autonomous cyber capabilities raise important legal concerns over their deployment in circumstances involving self-defence, countermeasures and plea of necessity. By contrast, it seems that they may be used to carry out acts of retorsion without any form of human intervention. This is particularly relevant since states have thus far usually reacted to malicious cyber operations by retorsion, relying on political rather than legal attribution. This said, however, autonomous cyber capabilities may still result in unintentional unlawful conduct. Thus, states should be cautious in delegating acts of retorsion to autonomous cyber capabilities, especially when the response may result in a breach of international law.

7. References

Air Service Agreement of 27 March 1946 between the United States of America and France, Vol. XVIII Reports of International Arbitral Awards 417 (Arbitral Tribunal 1978).

Alland, Denis. 'The Definition of Countermeasures'. In *The Law of International Responsibility*, by James Crawford, Alain Pellet and Simon Olleson. Oxford Commentaries on International Law. Oxford: Oxford University Press, 2010.

Amoroso, Daniele. *Autonomous Weapons Systems and International Law: A Study on Human-Machine Interactions in Ethically and Legally Sensitive Domains*. Naples / Baden-Baden: Edizioni Scientifiche Italiane / Nomos Verlag, 2020.

Armed Activities on the Territory of the Democratic Republic of the Congo v. Uganda (Judgement), ICJ Reports 2005 168 (International Court of Justice 2005).

Bellasio, Jacopo and Erik Silfversten. 'The Impact of New and Emerging Technologies on the Cyber Threat Landscape and Their Implications for NATO'. In *Cyber Threats and NATO 2030: Horizon Scanning and Analysis*, by Amy Ertan, Kathryn Floyd, Piret Pernik and Tim Stevens, 88–107. Tallinn: NATO CCDCOE Publications, 2020.

Bendett, Samuel, Mathieu Boulègue, Richard Connolly, Margarita Konaev, Pavel Podvig and Katarzyna Zysk. 'Advanced Military Technology in Russia: Capabilities and Implications'. Research Paper. Russia and Eurasia Programme. London: Chatham House, September 2021. https://www.chathamhouse.org/sites/default/files/2021-09/2021-09-23-advanced-military-technology-in-russia-bendett-et-al.pdf.

Borelli, Silvia and Simon Olleson. 'Obligations Relating to Human Rights and Humanitarian Law'. In *The Law of International Responsibility*, by James Crawford, Alain Pellet and Simon Olleson, 1177–96. Oxford Commentaries on International Law. Oxford: Oxford University Press, 2010.

Buchan, Russell. 'The International Legal Regulation of State-Sponsored Cyber Espionage'. In *International Cyber Norms: Legal, Policy & Industry Perspectives,* by Anna-Maria Osula and Henry Rõigas, 65–86. Tallinn: NATO CCD COE Publications, 2016.

Cannizzaro, Enzo. *Il Principio Della Proporzionalità Nell'ordinamento Internazionale*. Milano: Giuffré Editore, 2000.

Clapham, Andrew. *War*. Oxford: Oxford University Press, 2021.

Council of the European Union. 'Russian Cyber Operations against Ukraine: Declaration by the High Representative on Behalf of the European Union', 10 May 2022. https://www.consilium.europa.eu/en/press/press-releases/2022/05/10/russian-cyber-operations-against-ukraine-declaration-by-the-high-representative-on-behalf-of-the-european-union/.

Cyber Security Intelligence. 'Japan's New AI-Based Cyber Defence System', 15 April 2020. https://www.cybersecurityintelligence.com/blog/japans-new-ai-based-cyber-defence-system--4907.html.

Delerue, François. *Cyber Operations and International Law*. 1st ed. Cambridge University Press, 2020.

Delerue, François. 'Reinterpretation or Contestation of International Law in Cyberspace?' *Israel Law Review* 52, no. 3 (2019): 295–326.

Dinstein, Yoram. *War, Aggression and Self-Defence*. Sixth Edition. Cambridge ; New York: Cambridge University Press, 2017.

Durward, Johnson and Michael N. Schmitt. ‘Responding to Proxy Cyber Operations Under International Law’. *Cyber Defense Review* 6, no. 4 (Fall 2021): 15–30.

European Defence Agency. ‘Cyber Defence R&T - CapTech Cyber’, 2020. https://eda.europa.eu/what-we-do/all-activities/activities-search/cyber-defence.

European Union. ‘Strategic Compass for Security and Defence’, 2022. https://www.eeas.europa.eu/sites/default/files/documents/strategic_compass_en3_web.pdf.

European Union Agency for Cybersecurity. ‘AI Cybersecurity Challenges: Threat Landscape for Artificial Intelligence.’, December 2020. https://data.europa.eu/doi/10.2824/238222.

Focarelli, Carlo. *Le Contromisure Nel Diritto Internazionale*. Milano: Giuffré Editore, 1994.

Giegerich, Thomas. ‘Retorsion’. In *Max Planck Encyclopedias of International Law*, September 2020. https://opil.ouplaw.com/display/10.1093/law:epil/9780199231690/law-9780199231690-e983.

Gill, Terry D. ‘The Temporal Dimension of Self-Defence: Anticipation, Pre-Emption, Prevention and Immediacy’. In *International Law and Armed Conflict: Exploring the Faultlines*, by Michael N. Schmitt and Jelena Pejic, 113–55. International Humanitarian Law Series, 2007.

Gray, Maggie and Amy Ertan. ‘Artificial Intelligence and Autonomy in the Military: An Overview of NATO Member States’ Strategies and Deployment’. Tallinn: NATO CCDCOE, 2021.

Green, James A. ‘Fluctuating Evidentiary Standards for Self-Defence in the International Court of Justice’. *International and Comparative Law Quarterly* 58 (2009): 163–79.

Guarino, Alessandro. ‘Autonomous Intelligent Agents in Cyber Offence’. In *5th International Conference on Cyber Conflict*, by K. Podins, J. Stinissen and M. Maybaym. Tallinn: NATO CCD COE Publications, 2013.

International Law Commission. ‘Draft Articles on Responsibility of states for Internationally Wrongful Acts’, 2001. https://legal.un.org/ilc/texts/instruments/english/commentaries/9_6_2001.pdf.

Italy. ‘Italian Position Paper on 'International Law and Cyberspace'', 2021. https://www.esteri.it/mae/resource/doc/2021/11/italian_position_paper_on_international_law_and_cyberspace.pdf.

Iwasawa, Yuji and Naoki Iwatsuki. ‘Procedural Conditions’. In *The Law of International Responsibility*, by James Crawford, Alain Pellet and Simon Olleson, 1149–55. Oxford Commentaries on International Law. Oxford: Oxford University Press, 2010.

Kamto, Maurice. ‘The Time Factor in the Application of Countermeasures’. In *The Law of International Responsibility*, by James Crawford, Alain Pellet and Simon Olleson, 1169–76. Oxford Commentaries on International Law. Oxford: Oxford University Press, 2010.

Kosseff, Jeff. ‘Retorsion as a Response to Ongoing Malign Cyber Operations’. In *20/20 Vision: The Next Decade*, by T. Jančárková, L. Lindström, M. Signoretti, I. Tolga and G. Visky, 9–23. Tallinn: NATO CCD COE Publications, 2020.

Kott, Alexander, Paul Théron, Martin Drašar, Edlira Dushku, Benoît LeBlanc, Paul Losiewicz, Alessandro Guarino, et al. ‘Autonomous Intelligent Cyber-Defense Agent (AICA) Reference Architecture, Release 2.0’. Devcom Army Research Laboratory, September 2019.

Lahmann, Henning. *Unilateral Remedies to Cyber Operations: Self-Defence, Countermeasures, Necessity and the Question of Attribution*. Cambridge ; New York: Cambridge University Press, 2020.

Leben, Charles. ‘Obligations Relating to the Use of Force and Arising from Peremptory Norms of International Law’. In *The Law of International Responsibility*, by James Crawford, Alain Pellet and Simon Olleson, 1197–1204. Oxford Commentaries on International Law. Oxford: Oxford University Press, 2010.

Legality of the Threat or Use of Nuclear Weapons (Advisory Opinion), ICJ Reports 1996 226 (International Court of Justice 1996).

Leyden, John. 'Japan Tasks Fujitsu with Creating Search-and-Destroy Cyber-Weapon'. *The Register*, 3 January 2012. https://www.theregister.com/2012/01/03/japan_cyber_weapon_research/.

Liivoja, Rain, Maarja Naagel and Ann Väljataga. 'Autonomous Cyber Capabilities under International Law'. Tallinn: NATO CCDCOE, 2019.

Margulies, Peter. 'A Moment in Time: Autonomous Cyber Capabilities, Proportionality and Precautions'. In *Autonomous Cyber Capabilities under International Law*, by Rain Liivoja and Ann Väljataga, 152–80. Tallinn: NATO CCDCOE Publications, 2021.

McFarland, Tim. 'The Concept of Autonomy'. In *Autonomous Cyber Capabilities under International Law*, by Rain Liivoja and Ann Väljataga, 12–35. Tallinn: NATO CCDCOE Publications, 2021.

Military and Paramilitary Activities in and against Nicaragua (Nicaragua v. United States of America) (Judgement), ICJ Reports 1986 14 (International Court of Justice 1986).

NATO. 'Allied Joint Doctrine for Cyberspace Operations (AJP-3.20)'. NATO Standardization Office, January 2020.

NATO. 'Statement by the North Atlantic Council Concerning the Malicious Cyber Activities against Albania', 8 September 2022. https://www.nato.int/cps/en/natohq/official_texts_207156.htm.

Naulilaa case (Portugal/Germany), Vol. II Reports of International Arbitral Awards 1011 (Special Arbitral Tribunal 1928).

Newton, Michael and Larry May. *Proportionality in International Law*. New York: Oxford University Press, 2014.

Noortmann, Math. *Enforcing International Law: From Self-Help to Self-Contained Regimes*. Aldershot: Ashgate, 2005.

Oil Platforms (Islamic Republic of Iran v. United States of America), ICJ Reports 2003 161 (International Court of Justice 2003).

O'Keefe, Roger. 'Proportionality'. In *The Law of International Responsibility*, by James Crawford, Alain Pellet and Simon Olleson, 1157–68. Oxford Commentaries on International Law. Oxford: Oxford University Press, 2010.

O'Meara, Chris. 'Reconceptualising the Right of Self-Defence against 'Imminent' Armed Attacks'. *Journal on the Use of Force and International Law* 9, no. 2 (3 July 2022): 278–323.

Perez, Louis. 'Is Stuxnet the next Skynet? Autonomous Cyber Capabilities as Lethal Autonomous Weapons Systems'. In *Artificial Intelligence and International Conflict in Cyberspace*, by Fabio Cristiano, Dennis Broeders, François Delerue, Frédérick Douzet and Aude Géry, 186–222. Routledge Studies in Conflict, Security and Technology. Oxon; New York: Routledge, 2023.

Roscini, Marco. *Cyber Operations and the Use of Force in International Law*. Oxford: Oxford University Press, 2014.

Ruys, Tom. *'Armed Attack' and Article 51 of the UN Charter: Evolutions in Customary Law and Practice*. Cambridge; New York: Cambridge University Press, 2010.

Sanchez, Salvador Llopis. 'Artificial Intelligence (AI) Enabled Cyber Defence'. *European Defence Matters*, no. 14 (2017): 18.

Scharre, Paul. *Army of None: Autonomous Weapons and the Future of War*. New York: W. W. Norton & Company, 2018.

Schmitt, Michael N. ‘Autonomous Cyber Capabilities and the International Law of Sovereignty and Intervention’. In *Autonomous Cyber Capabilities under International Law*, by Rain Liivoja and Ann Väljataga. Tallinn: NATO CCDCOE Publications, 2021.

Schmitt, Michael N. ‘Cyber Operations in International Law: The Use of Force, Collective Security, Self-Defense and Armed Conflicts’. In *Proceedings of a Workshop on Deterring Cyberattacks: Informing Strategies and Developing Options for U.S. Policy*, by National Research Council. Washington DC: National Academies Press, 2010.

Schmitt, Michael N. ‘Responding to Transnational Terrorism Under the Jus Ad Bellum: A Normative Framework’. *Naval Law Review* 56, no. 1 (2008).

Schmitt, Michael N, ed. *Tallinn Manual 2.0 on the International Law Applicable to Cyber Operations*. 2nd ed. Cambridge University Press, 2017.

Stiano, Alessandro. *Attacchi Informatici e Responsabilità Internazionale Dello Stato*. Edizioni Scientifiche Italiane. Legal Culture and International Flows 9. Naples, 2023.

Taddeo, Mariarosaria, Tom McCutcheon and Luciano Floridi. ‘Trusting Artificial Intelligence in Cybersecurity Is a Double-Edged Sword’. *Nature Machine Intelligence* 1, no. 12 (2019): 557–60.

Tammet, Tanel. ‘Autonomous Cyber Defence Capabilities’. In *Autonomous Cyber Capabilities under International Law*, by Ann Väljataga and Rain Liivoja, 36–50. Tallinn: NATO CCDCOE Publications, 2021.

The Gabčcíkovo Nagymaros Project (Hungary vs. Slovakia) (Judgement), ICJ Reports 1997 7 (International Court of Justice 1997).

Tsagourias, Nicholas and Michael Farrell. ‘Cyber Attribution: Technical and Legal Approaches and Challenges’. *The European Journal of International Law* 31, no. 3 (2020): 941–67.

United Kingdom. ‘National Cyber Strategy’, 2022. https://assets.publishing.service.gov.uk/government/uploads/system/uploads/attachment_data/file/1053023/national-cyber-strategy-amend.pdf

United Nations General Assembly. ‘Report of the Group of Governmental Experts on Developments in the Field of Information and Telecommunications in the Context of International Security’. United Nations General Assembly, 24 June 2013. https://documents-dds-ny.un.org/doc/UNDOC/GEN/N13/371/66/PDF/N1337166.pdf?OpenElement.

United States. ‘Statement by NSC Spokesperson Adrienne Watson on Iran’s Cyberattack against Albania’, 7 September 2022. https://www.whitehouse.gov/briefing-room/statements-releases/2022/09/07/statement-by-nsc-spokesperson-adrienne-watson-on-irans-cyberattack-against-albania/.

United States Cyberspace Solarium Commission. ‘Final Report’, March 2020. https://www.solarium.gov/report.

United States Department of Defence. ‘Military and Security Developments Involving the People’s Republic of China’. Annual Report to Congress, 2021.

United States Department of State. ‘Attribution of Russia’s Malicious Cyber Activity Against Ukraine’. https://www.state.gov/attribution-of-russias-malicious-cyber-activity-against-ukraine/, 10 May 2022. https://www.state.gov/attribution-of-russias-malicious-cyber-activity-against-ukraine/.

United States National Security Commission on Artificial Intelligence. ‘Final Report’, 2021. https://www.nscai.gov/2021-final-report/.

‘Video Message of Prime Minister Edi Rama, Regarding the Response of the Albanian Government to the Cyber Attack against the Digital Infrastructure of the Government of the Republic of Albania’, 7 September 2022. https://www.kryeministria.al/en/newsroom/videomesazh-i-kryeministrit-edi-rama-lidhur-me-

vendimin-e-qeverise-shqiptare-si-kunderpergjigje-ndaj-aktit-te-sulmit-te-rende-kibernetik-ndaj-infrastruktures-digjitale-te-qeverise-se-republikes-se-s/.

Webster, Daniel. ‘Letter to Henry Stephen Fox’. In *The Papers of Daniel Webster: Diplomatic Papers, Vol 1, 1841-1843*, by KE Shewmaker, 1983.

Woltag, Johann-Christoph. *Cyber Warfare: Military Cross-Border Computer Network Operations under International Law*. International Law Series. Cambridge; Antwerp; Portland: Intersentia, 2014.