PAPER
DECEMBER 2020

UNIVERSITÀ DEGLI STUDI FIRENZE

USE OF FORCE AND ATTRIBUTION: THE APPLICABILITY OF INTERNATIONAL HUMANITARIAN LAW TO CYBER ATTACKS

GABRIELLA ACHILLE

Center for Cyber Security and International Relations Studies

CENTER FOR CYBER SECURITY AND INTERNATIONAL RELATIONS STUDIES (CCSIRS)

Centro Interdipartimentale di Studi Strategici,

Internazionali e Imprenditoriali (CCSSII)

Università degli Studi di Firenze

Via delle Pandette 2, 50127, Firenze

https://www.cssii.unifi.it/ls-6-cyber-security.html

Le dichiarazioni e le opinioni espresse nella presente relazione sono unicamente quelle dell'autore e non implicano l'approvazione da parte dell'Università di Firenze, del Centro Interdipartimentale di Studi Strategici, Internazionali e Imprenditoriali o del Center for Cyber Security and International Relations Studies.

Center for Cyber Security and International Relations Studies

USE OF FORCE AND ATTRIBUTION: THE APPLICABILITY OF INTERNATIONAL HUMANITARIAN LAW TO CYBER ATTACKS

Gabriella Achille

UNIVERSITÀ DEGLI STUDI FIRENZE

Paper

December 2020

ABOUT THE AUTHOR

Gabriella Achille is a Master student in International Relations at University of Florence. She previously graduated in Political Science, International Relations and European Studies at University of Bari. Her main field of interests concerns human rights, and she is writing her master thesis in International Human Rights Law with an insight on cyber security policies and strategy of China.

USE OF FORCE AND ATTRIBUTION: THE APPLICABILITY OF INTERNATIONAL HUMANITARIAN LAW TO CYBER ATTACKS

Introduction

Cyberspace is considered as the fifth domain, man-made, in which the actors are not limited to states, but also comprehend non-state actors, such as high-tech companies or organized groups (i.e. terrorists or cyber criminals). A legal framework still lacks, but this does not mean that they act in a legal *vacuum*, since many norms of international law are extended to cyber domain as well. Despite not being explicitly indicated in the United Nations (UN) Charter and other international (or regional) documents, rules can be interpreted in a broader way, in order to be applied to this new domain, not to leave it unregulated. These issues have been addressed by many experts of International Law or cyber security, such as Melzer (2011), Buchan (2012) and Shackelford (2009), among others, whose theses are endorsed in this paper.

The evolution of cyberspace is influencing the structure and the dynamics of international relations and it has opened the door to new challenges which jeopardize the international order. Indeed, new types of the use of force have emerged, some of them being considered full-fledged armed attacks, or even "acts of war" (Stone 2013), by some authors, as we will see later.

Other experts, such as Rid (2011), Valeriano and Maness (2015), claim that attacks occurring in cyberspace cannot be considered as acts of war, since they do not comply with the minimal requirements. In this paper, though, cyberwar is considered as a reality, since it can damage a modern state, even without drawing blood: an empirical example is the attack known as Stuxnet, which will be mentioned in the first section of this paper. Clearly, to evaluate them as armed attacks, cyber attacks need to reach a certain threshold of violence; once reached, they trigger the applicability of international law in general, but more precisely of international humanitarian law (IHL), which is also known as the law of armed conflicts. As in any other kind of conflict, actors shall distinguish

between “militants” and “civilians”, limiting the attacks to targeted persons and avoiding unfair injuries or even deaths.

The link between cyber security and human rights is a matter of great concern and this area of enquiry has been covered by several bodies of literature (Wagner, Kettemann and Vieth 2019; Della Morte 2018). The protection of human rights is a priority in all domains, even more in cyberspace, since it shall be used as a practical tool for ensuring an open and secure Internet, being this latter considered a human right in itself, even though not universally.

The debated issue this paper seeks to clarify is the applicability of international humanitarian law to cyber attacks. Some requirements and principles are analysed in order to understand to which extent this kind of attacks can be equated to traditional ones, and hence be subject to human rights rules in hostilities. Indeed, humanitarian consequences are not that dramatic yet, but the potential for a tragedy is already very high and it is likely to increase. For this reason, it is important to establish a framework in which rules and limits are set, in order to be prepared in the case of such an event.

Cyber attacks: the new type of the use of force

As anticipated, the phenomenon of cyberwar does not exist in a legal *vacuum*: it is subject to established international rules and principles, even though transposing them to cyber domain is a difficult task. A lot of important questions raise, and they can be tackled through treaty interpretation and common sense or with policy decisions by the international community (Melzer 2011, 36).

The evolution of norms in cyberspace is still in place. Shackelford (2009) has observed that cyber attacks can be indiscriminate, since they can affect the entire system by targeting even only a part of it, and for this reason they must be dealt with properly. The complexity of these attacks has increased significantly, and one of the main problems when dealing with them is the attribution of responsibility. It is possible to locate them geographically, but it is extremely difficult to determine the subject responsible, because of the high interconnection in the cyberspace.

The term “attack” has a considerable relevance especially in IHL, since it is used in most of the rules referring to the conduct in hostilities.

According to article 49.1 of the Additional Protocol (AP) I[1] to the Geneva Conventions, to define an operation as an attack, it shall entail violence against the adversary: attacks do not have to necessarily use kinetic violence, but it is sufficient that they cause equivalent effects, in particular death or injury of individuals or physical destruction of objects. However, no satisfactory interpretation of the notion of attacks is available when it comes to cyber operations (Melzer 2011, 26). Not all of them are cyber attacks, while it is true the other way around, whether the attack is offensive or defensive (Voitaşec 2015, 552-53). The Tallinn Manual – a non-binding document which addresses several important issues regarding the application of international law to cyberspace - defines cyber attacks as cyber operations that are expected to cause injury or death to people or destruction or damage to objects (Rule 30). Moreover, in the glossary of the Joint Publication by US Chiefs of Staff (2018) it is specified that cyber attacks create "noticeable denial effects (i.e., degradation, disruption or destruction) in cyberspace or manipulation that leads to denial effects in the physical domains. [...] Cyberspace attack actions are a form of fires [...] and are carefully synchronized with planned fires in the physical domains."

Cyber attacks, thus, can affect not only the virtual domain, but also the physical one and can be considered as military attacks. They are part of the general political and military situation. Indeed, cyber represents one of the tools which are part of states' capabilities and some strategic national documents deal with it[2].

What makes this domain a "soft ground" is still the absence of specific rules of the game. The applicability of IHL rules to cyber attacks depends on whether they are part of hostilities: the International Committee of the Red Cross (ICRC) defined armed attacks as the resort to means and methods that cause harm or injury to the adversary, and hence to the totality of acts perpetrated by participants directly active in the hostilities (ICRC 1977).To be considered as such, cyber operations must reach the required threshold of direct harm (*direct causation*) and must be carried out in support of one of the belligerent parties with the aim

[1] According to art. 49.1 "`Attacks' means acts of violence against the adversary, whether in offence or in defence."

[2] For instance, the *National strategic framework for cyberspace security* and the *National plan for cyberspace protection and ICT security*, both foreseen by the Prime Minister's "Decree Containing Strategic Guidelines for the National Cyber Protection and ICT Security" of January 24, 2013, aimed to build a comprehensive strategy to coordinate the efforts to tackle challenges of the cyberspace and to foster national interest in this domain (Presidenza del Consiglio dei Ministri 2013).

of destroying the other one(s) (*belligerent nexus*) (Melzer 2011, 28). It is possible to claim, then, that if cyber attacks are part of an ongoing conflict, they can trigger the applicability of IHL. If they occur as isolated cases, they need to fulfil specific requirements, such as the threshold of violence established for armed conflicts and they must cause more than a simple inconvenience or temporary shutdown of technological systems (Ambos 2015, 142).

When cyber attacks cause severe physical damages, they violate article 2.4 of the UN Charter, as any attack carried out with conventional weapons. The article in question cites as follow: "All Members shall refrain in their international relations from the threat or use of force against the territorial integrity or political independence of any state, or in any other manner inconsistent with the Purposes of the United Nations." Thus, it calls on Member States not to resort to the threat or use of force against another State and not to act in a manner that would result incoherent with the aims of the UN.

Some scholars and commentators are sceptical with regards to this comparison, claiming that the former fall out the international regulatory framework, since when it was established, cyberspace did not exist yet, and hence regulators were not referring to them when addressing the use of force. For instance, Benatar (2009, 377) talked about "scant legal impediments" to launch a "computer war", and Hoisington (2009, 440) stated that "[d]espite the potential lethality of cyber-warfare, the practice currently exists in a legal netherworld".

As opposed to this view, this research paper draws on the literature claiming that cyber attacks cause real concern, as much as traditional attacks, since they can provoke damages of the same magnitude, in particular Melzer (2011) and Voitaşec (2015). Nevertheless, not all cyber attacks represent a violation of article 2.4. In this vein, it can be useful to consider two empirical cases, Estonia 2007 and the Stuxnet attack, first uncovered in 2010, to understand the concept of lawful cyber attack.

The first one does not account to a violation of article 2.4 since it caused no physical damages, despite having provoked disruptive effects. Estonia suffered distributed denial-of-service attacks (DDoS) - which consist in a slowing down or crash of an Internet web page due to too many requests - and they did not result that dangerous (Buchan 2012, 218). On the other hand, a worm virus, called Stuxnet, was detected in

Iranian government's computers, specifically those controlling Natanz nuclear plant: the virus was created to change the operating mode of the centrifuges containing uranium (Buchan 2012, 219). According to Iranian declarations, Stuxnet was detected before it could effectively affect the process, hence few damages were caused; but determining whether this attack was unlawful is problematic, since an exact identification of its impact has not been possible. The findings of some experts were at odds with Iranian government's declarations: effects have been quite considerable and since Stuxnet caused the physical destruction of centrifuges at Natanz, there has been a violation of art. 2.4 (Buchan 2012, 221). Therefore, this can be considered as the first cyber attack with factual consequences.

Another issue to consider when dealing with cyber attacks is that they privilege indirect effects, and this make it difficult to distinguish whether the problem which affect a computer is due to an attack, an error or other kinds of malfunctioning (Van Puyvelde and Brantly 2019, 87). Furthermore, cyber attacks' effects are reversible, contrary to those resulting from traditional ones: this means that they can be annulled within a short time, without being able to adequately realize what is happening and possibly react (Van Puyvelde and Brantly 2019, 87).

The aim of the following sections is to clarify whether it is possible to apply IHL principles to cyber attacks and to what extent.

Three IHL core principles applied to cyber attacks

Principle of distinction

Article 48 of the AP I[3] – also known as "basic rule" - requires belligerents to distinguish military objectives from civilians and civil objects. The former can be targets of attacks because of their nature, purpose, location and because they make an effective contribution to military actions; hence their destruction or neutralization ensures a definite military advantage. This principle is part of International Customary Law and it was first introduced by the International Court of Justice in 1996 in an Advisory Opinion on the legality of the threat or use of nuclear weapons.

[3] Art. 48 states: "In order to ensure respect for and protection of the civilian population and civilian objects, the Parties to the conflict shall at all times distinguish between the civilian population and combatants and between civilian objects and military objectives and accordingly shall direct their operations only against military objectives."

Cyber weapons are different from traditional ones, since their use can result indiscriminate due to the difficulty in distinguishing potential targets from those who must not be made object of an attack (according to the rules of IHL). They result more like biological weapons, which can spread wide without choosing their victims. Therefore, a malware virus and a biological virus can absurdly have something in common, since they can both uncontrollably affect individuals (Dinstein 2012, 262).

Some computers are classified as military objectives by nature, for instance when they are designed as components of weapon systems or to produce them: in this case, they can be subject to military attacks. In other cases, computers generally created for the public can become military targets, when they are used by combatants or from a specific location (Dinstein 2012, 263).

Military use of a computer needs to be interpreted in a broad manner: it can comprehend different steps, from the planning of an attack to its actual realization. The software is the key to make an ordinary computer a military means, even though the hardware can remain contaminated after the removal of the former and acquire a military nature (Dinstein 2012, 263). A clear-cut distinction does not exist: civilian and military users are strictly entangled and many cyber capabilities can be turned in useful means to generate relevant military effects (Martino 2017, 53). In light of above, in cyberspace more than in traditional warfare, military targets are likely to have a dual use, thus it is not immediate to determine precisely which installation or system make an effective contribution to military purposes. Presumption shall be applied in case of doubts, but this can turn into a thorny issue.

It is controversial the question on considering data as “objects” in the meaning of the black letter of IHL. This debate exists because no cyber operation can be carried out without - at least temporarily - delete or change data (Melzer 2011, 31). In this case, it is necessary to consider them as military objective only if they present requirements to be considered as such: for this reason, it is important to clearly distinguish the actual aim of the attack and its possible incidental effects. Therefore, for instance, the modification of civilian data during a cyber attack against a military target can be equated to “collateral damages” caused by kinetic violence (Melzer 2011, 31).

It is possible to say that this principle has a minor relevance for cyber attacks, since military and civilian computer systems are highly

interconnected, and many cyber infrastructures can have a dual use (Ambos 2015, 131). The potentially nonlethal nature of cyber weapons intensifies the difficulty in applying the principle of distinction, which creates a legal "grey area", in warfare. This grey area is due to the fact that modern war makes it more difficult to distinguish between lawful and unlawful targets, and it "does not exclude targets whose destruction or neutralization do not directly advance the overall objective of the war, but do nonetheless degrade the enemy's ability and will to fight" (Kelsey 2008, 1439).

Principle of precaution

Article 57 of the AP I[4] provides for a series of active and passive precautionary measures to be taken in armed conflicts. It calls for constant care in the conduct of military operations and lists the adoption of a series of measures to spare civilian population and civil objects, but it is uncertain the extent to which they shall be applied. This is strictly linked to the principle of distinction, which could even be considered a sort of direct consequence, and it also recalls the principle of proportionality enshrined in article 51, discussed below.

As far as cyberspace is concerned, constant care consists in the verification of the targets, restraint of the effects to the greatest extent possible, and prior warning if civilian population might be affected. For instance, the attacker shall check the cyber network of the target in advance to limit the attack to military components and, if possible, to avoid collateral damages to civilian infrastructures (Ambos 2015, 131).

This principle can be applied to cyber attacks, but it encounters some difficulties in the effective feasibility and reasonability of the measures, since a high level of technical expertise is needed.

Principle of proportionality

The principle of proportionality is enshrined in article 51 of the AP I[5] and it calls for proportionate actions, meaning that they shall not excess the pursued aim. The principle of proportionality sets limits and prohibits

[4] As reported in art. 57.1: "In the conduct of military operations, constant care shall be taken to spare the civilian population, civilians and civilian objects."

[5] According to art. 51.4: "Indiscriminate attacks are prohibited. Indiscriminate attacks are: (a) those which are not directed at a specific military objective; (b) those which employ a method or means of combat which cannot be directed at a specific military objective; or (c) those which employ a method or means of combat the effects of which cannot be limited as required by this Protocol; and consequently, in each such case, are of a nature to strike military objectives and civilians or civilian objects without distinction […]"

attacks causing "superfluous injury or unnecessary suffering" and "widespread, long-term and severe damage to the environment" (Ambos 2015, 134). According to this, an indiscriminate attack will be considered as such when it causes excessive harm with respect to the military advantage, which must be precise and anticipated. The perpetrators must seek to determine the precise adverse effects of the attack and compare them with the military advantage, even though, in cyber attacks, anticipating the effects can result particularly difficult, because of the uncertainty which characterizes them.

The principle also establishes to what extent the harm to be prevented justifies damages caused with a self-defensive action. This is a remote possibility in cyberspace, since cyber attacks can be unpredictable and clandestine, and this make it more complicated to detect or repel them (Melzer 2011, 18).

The principle of proportionality provides for a better protection from harm and damages to civilian objectives and objects and it seems to find no great difficulties when applied to cyber attacks. However, as in traditional IHL, it is still debated whether a particularly great military advantage can be a proper justification for serious or extensive damage (Ambos 2015, 135-36).

The question of the attribution of responsibility

Finding the responsible of a cyber attack is one of the major issues in cyberspace and it has been treated by several authors (e.g. Healey 2011, Liu 2017). Shackelford (2009, 201), in particular, has observed that the interconnection which characterizes the Internet and the components of cyberspace makes it extremely difficult to identify true identities. In cyberspace, attacks coming from non-state actors are not easily attributable, because of many stratagems they use to hide their identity, and also because they could either act for their own interests or on behalf of the state. Indeed, this latter can resort to the so-called "plausible deniability" and seek to avoid accountability (Shackelford 2009, 233), ascribing the attacks to hacktivists or individual criminals. In fact, cyberspace allows actors to deceive their enemy through false information and deception: they can mislead the adversary with a series of tricks (such as IP spoofing) which then make it impossible to trace the actual responsible (Melzer 2011, 32).

In the very unlikely event of the identification of liable actors, another issue would arise: if cyber attacks are considered under IHL framework

and they reach a certain threshold of violence, the perpetrators can be prosecuted for war crimes. In the worst scenario, they could even be tried for genocide before the International Criminal Court (ICC), when attacks have large scale effects and they have specific intent of destruction (Shackelford 2009, 235). However, this issue is still highly confused and the applicability of IHL is questioned, since even when attackers are identified, it is necessary to clarify whether they are parties to a conflict or their acts can be attributed to a state.

The responsibility issue takes a pivotal place when a state suffers a cyber attack and invoke self-defence (as provided by article 51 of the UN Charter[6]), to justify a reaction aimed at safeguarding its security, if that attack reaches the level of an armed one (Shackelford 2009, 237)[7].

Furthermore, cyber attacks can be often political, but Rid (2012) claims that in the absence of attribution, they are automatically depoliticized (Van Puyvelde and Brantly 2019, 117) and this is one of the reasons why according to the author a *cyber* war is very unlikely

Some rules should be established, such as attributing responsibility to a given state if a cyber attack comes from its territory or if there are evidences about the linkage between the attack and the political relationship with the victim state, but these measures would meet discontent within the international community, because some states could take advantage of such rules to rage (perhaps wrongfully) against their adversaries.

Conclusions

In the light of the analysis provided in this paper, although there are enough elements to assert that IHL applies to cyber attacks, protection of the civilian population will depend on the interpretation of its rules.

So far, cyber attacks have not caused that much damage, but they are posing new challenges. Humanitarian risks do not seem immediate, but we are living in a very uncertain and tense situation, where it is extremely difficult to preserve the *status quo*. In the worst case scenario of a new world war, the most powerful states would employ their

[6] Art. 51 cites: "Nothing in the present Charter shall impair the inherent right of individual or collective self-defence if an armed attack occurs against a Member of the United Nations, until the Security Council has taken measures necessary to maintain international peace and security. […]"

[7] For instance, if Estonian national government had decided to act against Russia after the cyber attack suffered in 2007, this behaviour would have not been accepted by the International Community, first and foremost by Russia, since there were no evidences about its involvement.

advanced cyber weapons also in addition to other kinds of weapon (nuclear, biological, etc.), which together could cause irreversible damages. Cyber means and methods are constantly improving – the introduction of Artificial Intelligence (AI) is an example - and cyber attacks need to be monitored and regulated.

There is no real difference between cyber attacks and traditional ones when dealing with the application of IHL. It is important to attain to the principle of distinction, that remains of a high relevance, in order to spare unjust losses of lives. Although some IHL principles are difficult to apply to cyber attacks, because of their peculiarities and because cyberspace is still a partially unknown domain, this paper provided the conceptual framework to consider cyber attacks as armed attacks.

There is a need for an extensive interpretation of IHL norms: they were introduced when the fifth domain did not yet exist, but it is necessary to include it now, since it is becoming increasingly pervasive and in the next future it would acquire a considerable relevance for our lives. It is important to raise awareness of the necessity to evaluate the humanitarian impact of developing technologies, and to ensure that they are not employed under unlawful conditions.

References

Ambos, Kai. 2015. "International criminal responsibility in cyberspace." In *Research Handbook on International Law and cyberspace*, edited by Nicholas Tsagourias and Russel Buchan, 118-143. Cheltenham: Edward Elgar Pub.

Benatar, Marco. 2009, "The Use of Cyber Force: Need for Legal Justification?" *Gottingen Journal of International Law* 1(3): 375-396.

Buchan, Russel. 2012. "Cyber Attacks: Unlawful Uses of Force or Prohibited Interventions?" *Journal of Conflict & Security Law* 17(2): 211-227.

Buchan, Russel, and Nicholas Tsagourias, eds. 2015. *Research Handbook on International Law and cyberspace.* Cheltenham: Edward Elgar Pub.

Della Morte, Gabriele. 2018. *Big data e protezione internazionale dei diritti umani. Regole e conflitti*. Napoli: Editoriale Scientifica.

Dinstein, Yoram. 2012, "The Principle of Distinction and Cyber War in International Armed Conflicts." *Journal of Conflict & Security Law* 17(2): 261-277.

Healey, Jason. 2011. "The Spectrum of National Responsibility for Cyberattacks." *The Brown Journal of World Affairs, 18(1)*: 57-70.

Hoisington, Matthew. 2009. "Cyberwarfare and the Use of Force Giving Rise to the Right of Self-Defense." *Boston College. International & Comparative Law Review* 32(2): 439-454.

ICRC. 1977. *Commentary on the additional Protocols of 8 June 1977 to the Geneva Conventions of the 12 August 1949*.

Kelsey, Jeffrey T.G. 2008. "Hacking into international Humanitarian Law: The principles of distinction and neutrality in the age of cyber warfare." *Michigan Law Review* 106(7): 1427-1452.

Liu, Ian Yuying. 2017. "State responsibility and cyberattacks: Defining due diligence obligations." *Indonesian Journal of International & Comparative Law* 4(2): 191-260.

Martino, Luigi. 2017. "Cyberspace and International Relations: Diplomatic Initiatives to Avoid the Risk of Escalation in the Cyber Arena." *European Cybersecurity Journal* 3(3): 53-57.

Melzer, Nils. 2011. "Cyber warfare and International Law." *UNIDIR Resources Paper*. Accessed September 30, 2020. https://unidir.org/files/publications/pdfs/cyberwarfare-and-international-law-382.pdf.

Presidenza del Consiglio dei Ministri. 2013. *Quadro strategico nazionale per la sicurezza dello spazio cibernetico*.

Rid, Thomas. 2012. "Cyber War Will Not Take Place." *Journal of Strategic Studies* 35 (1): 5-32.

Shackelford, Scott J. 2009. "From Nuclear War to Net War: Analogizing Cyber Attacks in International Law." *Berkley Journal of International Law* 27(1): 193-251.

Stone, John. 2013. "Cyber War Will Take Place!" *Journal of Strategic Studies* 36(1): 101-108.

US Joint Chiefs of Staff. 2018. *Cyberspace Operations*. Joint Publication 3-12.

Valeriano, Brandon, and Ryan C. Maness. 2015. *Cyber War versus Cyber Realities: Cyber Conflict in the International System.* Oxford: Oxford University Press.

Van Puyvelde, Damien, and Aaron F. Brantly. 2019. *Cybersecurity: Politics, Governance and Conflict in cyberspace.* Cambridge: Polity Press.

Voitaşec, Dan-Iulian. 2015. "Applying International Humanitarian Law to cyberattacks." *Challenges of the knowledge society* 5 (1): 552-556.

Wagner, Ben, Matthias C. Kettemann, and Vieth Kilian, eds. 2019. *Research Handbook on Human Rights and Digital Technology: Global Politics, Law and International Relations.* Cheltenham: Edward Elgar Pub.

CENTER FOR CYBER SECURITY AND INTERNATIONAL RELATIONS STUDIES (CCSIRS)

Centro Interdipartimentale di Studi Strategici,

Internazionali e Imprenditoriali (CCSSII)

Università degli Studi di Firenze

Via delle Pandette 2, 50127, Firenze

https://www.cssii.unifi.it/ls-6-cyber-security.html

Center for Cyber Security and International Relations Studies