UNIDIR RESOURCES

Cyberwarfare and International Law
2011

Cyberwarfare and International Law

Nils Melzer

IDEAS FOR PEACE AND SECURITY

The designations employed and the presentation of the material in this publication do not imply the expression of any opinion whatsoever on the part of the Secretariat of the United Nations concerning the legal status of any country, territory, city or area, or of its authorities, or concerning the delimitation of its frontiers of boundaries.

***

The views expressed in this publication are the sole responsibility of the author. They do not necessarily reflect the views or opinions of the United Nations, UNIDIR, its staff members of sponsors.

***

About the author

Dr. Nils Melzer, Research Director of the Centre for Business and Human Rights at the University of Zürich and former Legal Adviser of the International Committee of the Red Cross, is a participating expert in a process sponsored by the North Atlantic Treaty Organization's Cooperative Cyber Defence Centre of Excellence aiming to draft a Manual on International Law applicable to Cyber Conflict. This paper reflects the views of the author only and not necessarily those of the expert group or of any of the institutions he is or has been affiliated with.

I. Introduction

It is hardly an overstatement to say that the advent and global expansion of the Internet may prove to become the fastest and most powerful technological revolution in the history of mankind. In just 15 years, the number of individuals actively using the Internet has skyrocketed from an estimated 16 million in 1995 to more than 1.7 billion in late 2010.[1] Today, states, non-state communities, business, academia and individuals have become interconnected and interdependent to a point never imaginable before. At the same time, military reliance on computer systems and networks has increased exponentially, thus opening a "fifth" domain of war-fighting next to the traditionally recognized domains of land, sea, air and outer space.[2] This trend raises the question to what extent can existing international law be transposed to the cyber domain. Without any doubt, as a matter of principle, existing international law governs state activities wherever they are carried out, including in cyberspace. However, applying pre-existing legal rules, concepts and terminology to a new technology may entail certain difficulties in view of the specific characteristics of the technology in question.

It is the purpose of this paper to provide an overview: (a) of the potential restraints imposed on cyberwarfare by existing international law, (b) of the most important difficulties and controversies raised in the interpretation and application of international law to cyberwarfare, and (c) of the potential humanitarian impacts of cyberwarfare. In view of the constraints in terms of time and space, the envisaged overview cannot be exhaustive but will have to remain selective, focusing on providing a general understanding of the issues most relevant to contemporary state practice. Moreover, in view of the technical and legal complexity of the matter and the still rudimentary state of legal research, the ambition of this paper must remain limited to identifying issues and putting them into context, but cannot be to authoritatively resolve them. That said, this paper will focus on examining the following areas of international law:

- **Under the law governing the resort to force between states (*jus ad bellum*)**, it will have to be determined in what circumstances, if any, cyber operations can amount to (a) an internationally wrongful threat or use of "force", (b) an "armed attack" justifying the resort to necessary and proportionate force in self-defence, or (c) a "threat to international peace and security" or "breach of the peace" subject to UN Security Council intervention.
- **Under the law of neutrality**, the questions arise as to whether belligerents can lawfully use the telecommunications infrastructure of neutral states for the purpose of cyberattacks, and what the responsibilities of "neutral" states are with regard to non-state belligerents conducting attacks from within or through its territory or infrastructure.
- **Under the law of armed conflict (*jus in bello*),** here referred to as international humanitarian law (IHL), "cyberwarfare" must be distinguished from phenomena that are not necessarily governed by IHL, such as "cyber criminality" and

1 UK government, *"A Strong Britain in an Age of Uncertainty: The National Security Strategy"*, 2010, p. 29

2 For an express recognition of cyberspace as a separate "domain" of warfare see, for example, US Department of Defense, *The National Military Strategy for Cyberspace Operations*, 2006, p. 3; US Department of Defense, *The National Military Strategy*, 2004, p. 18.

"cyberterrorism". Where IHL does apply, it must be clarified to what extent its rules and principles, designed to govern traditional means and methods of warfare, can be transposed to cyberwarfare. In doing so, the focus will be on the rules and principles of IHL governing the conduct of hostilities rather than those governing the protection and treatment of persons in the hands of a party to an armed conflict, which is an area less relevant for cyberwarfare.

In examining these questions it should be kept in mind that, so far, there has not been a broad international dialogue on the interpretation and application of existing rules and principles of international law to cyberwarfare, and not even the technological implications and military potential of this domain have been fully explored. Although it may safely be assumed that cyber operations are not conducted in a legal vacuum, it is recommendable to adopt a cautious approach so as not to unnecessarily prejudge legal issues in this rapidly developing area. Attention should also be drawn to the ongoing efforts of a group of international experts working under the auspices of the North Atlantic Treaty Organization (NATO) Cooperative Cyber Defence Centre of Excellence to draft a "Manual on the International Law of Cyber Warfare" which, though not necessarily representative of a consolidated legal opinion of NATO or its members states, is likely to significantly contribute to the clarification of international law relating to cyberwarfare. While the author participates in this process in his capacity as an independent expert, the views expressed in this paper are those of the author alone and not necessarily those of the group of experts.

II. Specific Characteristics of Cyberwarfare

II.1. What is Cyberwarfare?

For the present purposes, the term "cyberwarfare" refers to warfare conducted in cyberspace through cyber means and methods. While "warfare" is commonly understood as referring to the conduct of military hostilities in situations of armed conflict,[3] "cyberspace" can described as a globally interconnected network of digital information and communications infrastructures, including the Internet, telecommunications networks, computer systems and the information resident therein.[4] Thus, for example, the infection of a belligerent adversary's computer network with a malicious virus would constitute an act of cyberwarfare, whereas the aerial bombardment of a military cybercommand would not.[5] The fact that cyberwarfare is conducted in cyberspace does not exclude that it may

3 On the notion of "armed conflict" see section V.1.

4 For other definitions of "cyberspace" see: The White House, *Cyberspace Policy Review*, 16 May 2011, p. 1, with reference to National Security Presidential Directive 54/Homeland Security Presidential Directive 23 (NSPD-54/HSPD23): "the interdependent network of information technology infrastructures, and includes the Internet, telecommunications networks, computer systems, and embedded processors and controllers in critical industries"; US Department of Defense, *The National Military Strategy for Cyberspace Operations*, 2006, p. 3: "A domain characterized by the use of electronics and the electromagnetic spectrum to store, modify, and exchange data via networked systems and associated physical infrastructures"; US Joint Chiefs of Staff, *Department of Defense Dictionary of Military and Associated Terms*, 2001, p. 41: "global domain within the information environment consisting of the interdependent network of information technology infrastructures, including the Internet, telecommunications networks, computer systems, and embedded processors and controllers"; and EastWest Institute, "The Russia–US Bilateral on Cybersecurity—Critical Terminology Foundations", 2011, p. 20: cyberspace is "an electronic medium through which information is created, transmitted, received, stored, processed, and deleted."

5 The same approach is taken by Marco Roscini, "World Wide Warfare—Jus ad bellum and the Use of Cyber

produce kinetic or other non-electronic effects outside the cyber domain and may even be specifically intended to do so by the attacker. For instance, targets of cyberwarfare may also include persons whose life, or objects the functionality of which, depends on computer systems, such as certain power stations, means of transport, or persons connected to various kinds of medical, military or professional life-support systems.

II.2. How is Cyberwarfare Unique?

When interpreting and applying existing international law to cyberwarfare, due consideration must be given to the specific characteristics of cyberspace. Most notably, cyberspace is the only domain which is entirely man-made. It is created, maintained, owned and operated collectively by public and private stakeholders across the globe and changes constantly in response to technological innovation. Cyberspace not being subject to geopolitical or natural boundaries, information and electronic payloads are deployed instantaneously between any point of origin and any destination connected through the electromagnetic spectrum. These travel in the form of multiple digitalized fragments through unpredictable routings before being reconstituted at their destination. While cyberspace is readily accessible to governments, non-state organizations, private enterprises and individuals alike, IP spoofing[6] and the use of botnets,[7] for example, make it easy to disguise the origin of an operation, thus rendering the reliable identification and attribution of cyber activities particularly difficult.[8]

II.3. Cyber Operations, Attacks, Exploitation and Defence

The term "cyber operation" or, synonymously, "computer network operation" (CNO) refers to the reduction of information to electronic format and the actual movement of that information between physical elements of cyber infrastructure.[9] Cyber operations can be categorized as "computer network attack", "computer network exploitation" and "computer network defence".[10] While computer network attacks (CNA) comprise all cyber operations aiming "to disrupt, deny, degrade, or destroy information resident in computers and computer networks, or the computers and networks themselves",[11] computer network exploitation (CNE) refers to "[e]nabling operations and intelligence collection to gather data from target or adversary automated information systems or networks".[12] Computer network defence (CND), in turn, refers to "[a]ctions taken to protect, monitor, analyse, detect, and respond to unauthorized activity within … information systems and computer networks" or, in short, the prevention of CNA and CNE through intelligence, counterintelligence, law enforcement and military

Force", in Armin Bogdany and Rüdiger Wolfrum (eds), *Max Planck Yearbook of United Nations Law*, vol. 14, 2010, p. 96.

6 "IP spoofing" refers to the creation of Internet Protocol (IP) packets with a forged source address with the purpose of concealing the identity of the sender or impersonating another computing system.

7 A "botnet" is an interconnected series of compromised computers used for malicious purposes. A computer becomes a "bot" when it runs a file that has bot software embedded in it.

8 On the characteristics and key features of cyberspace, see also US Department of Defense, *The National Military Strategy for Cyberspace Operations*, 2006, pp. 3–4.

9 At the time of writing, this was the preliminary definition of "cyber operations" accepted by the expert group working on the *Tallinn Manual*.

10 US Department of Defense, *The National Military Strategy for Cyberspace Operations*, 2006, GL-1.

11 Ibid.

12 Ibid.

capabilities.[13] This terminology, which is specific to operations conducted in cyberspace, must be carefully distinguished from existing technical terms of international law such as, for example, "force",[14] "armed attack"[15] and "attack".[16]

III. Cyber Operations and *Jus ad Bellum*

The *jus ad bellum* is that body of law which governs the resort by states to force in their international relations. Today, the most important source of *jus ad bellum* is the UN Charter. Certain aspects of that law, such as the precise modalities governing the use of force in a case of self-defence, for instance, are not regulated in the UN Charter and must be derived from customary law as reflected in state practice and *opinio juris* and identified in international jurisprudence. In this context, it will have to be examined in what circumstances, if any, cyber operations can amount to (1) an internationally wrongful threat or use of "force", (2) an "armed attack" justifying the resort to necessary and proportionate force in self-defence, or (3) a "threat to the peace", "breach of the peace" or "act of aggression" subject to UN Security Council intervention.

In practice, the first question is relevant because state-sponsored cyber operations qualifying as a use of "force" against another state would not only fall under the general prohibition of article 2(4) of the UN Charter, but would normally also trigger an international armed conflict. Cyber operations below the threshold of "force", even if otherwise prohibited under the customary principle of non-intervention, on the other hand, may represent lawful counter-measures in response to internationally wrongful acts not reaching the threshold of "armed attack" by another state. The second question is relevant because the occurrence of cyber operations amounting to an "armed attack" permits the attacked state to exercise its inherent right to self-defence through means otherwise prohibited by the Charter including, most notably, the resort to force. Lastly, the practical relevance of the determination that cyber operations amount to a "threat to the peace", "breach of the peace" or "act of aggression" is that it allows the UN Security Council to take forcible measures, including military force, in order to maintain or restore international peace and security irrespective of the qualification of the cyber operations in question as "force" or "armed attack" under articles 2(4) and 51 of the UN Charter.

III.1. Cyber Operations and the Prohibition of Interstate Force

III.1.1. Cyber Operations as "Force"

According to article 2(4) of the UN Charter, "[a]ll Members [of the United Nations] 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". The question thus arises to what extent cyber operations can qualify as "force" within the meaning of this prohibition. In the absence of a treaty definition, the concept of "force" must be interpreted in good faith in accordance

13 Ibid.

14 Art. 2(4).

15 UN Charter, art. 51.

16 Additional Protocol I to the Geneva Conventions, art. 49(1).

with the ordinary meaning to be given to the term in its context and in the light of the Charter's object and purpose.[17]

Although the ordinary meaning of "force" is clearly broad enough to include both armed and unarmed forms of coercion,[18] the overwhelming majority of commentators today consider the term "force" in article 2(4) of the UN Charter as practically synonymous to "armed" or "military" force.[19] This does not necessarily mean that the prohibition of interstate force is limited to the application of kinetic, chemical, biological or nuclear weaponry. According to the International Court of Justice, the prohibition applies "to any use of force, regardless of the weapons employed".[20] Indeed, it is relatively uncontroversial that cyber operations fall under the prohibition of article 2(4) of the UN Charter once their effects are comparable to those likely to result from kinetic, chemical, biological or nuclear weaponry.[21] This would certainly include the use of cyber operations as an offensive or defensive tool designed to cause death or injury to persons or the destruction of objects and infrastructure, irrespective of whether such destruction involves physical damage, functional harm, or a combination of both.[22] Conspicuous examples of a use of "force" within the meaning of article 2(4) of the UN Charter, therefore, would be cyber operations manipulating target computers systems so as to cause a meltdown in a nuclear power station, or opening the floodgates of a dam above a densely populated area, or disabling a busy airport's air traffic control during bad weather conditions, each with potentially horrendous consequences in terms of death, injury and destruction.

The real difficulty arises, however, with regard to the qualification as a use of "force" of cyber operations that do not, or not directly, cause death, injury or destruction. The *travaux préparatoires* of the UN Charter clearly show that the prohibition of "force" was not intended to extend to economic coercion and political pressures.[23] Also, article 41 of the UN Charter refers to "interruption of … communication" as a "measure not involving armed force", thus suggesting that certain denial of service attacks (DOS) would not fall under the prohibition of article 2(4). However, this does not warrant the conclusion that, absent violent effects, all cyber operations necessarily fall short of armed force.[24] While a

17 Art. 31(1), Vienna Convention on the Law of Treaties. The rules on interpretation codified in the Vienna Convention are generally considered to express customary international law. See Georg Ress, "The Interpretation of the Charter", in Bruno Simma (ed.), *The Charter of the United Nations: A Commentary*, vol. I, 2002, p. 18.

18 See Ian Brownlie, *International Law and the Use of Force by States*, 1963, p. 362; and Yoram Dinstein*, War, Aggression and Self-Defence,* 4th ed., 2005, pp. 80ff.

19 See Albrecht Randelzhofer, "Article 2(4)", in Bruno Simma (ed.), *The Charter of the United Nations: A Commentary*, vol. I, 2002, p. 117; Yoram Dinstein*, War, Aggression and Self-Defence,* 4th ed., 2005, p. 81; Ian Brownlie, *International Law and the Use of Force by States*, 1963, pp. 362, 431; N. Quoc Dinh et al., *Droit International Public*, 6th ed., 1999, pp. 893 and 906; Alfred Verdross and Bruno Simma, *Universelles Völkerrecht: Theorie und Praxis*, 1984, §§ 469, 476; and Marco Roscini, "World Wide Warfare—Jus ad bellum and the Use of Cyber Force", in Armin Bogdany and Rüdiger Wolfrum (eds), *Max Planck Yearbook of United Nations Law*, vol. 14, 2010, pp. 105ff.

20 See International Court of Justice, *Legality of the Threat or Use of Nuclear Weapons*, advisory opinion, 1996, § 39; and Ian Brownlie, *International Law and the Use of Force by States*, 1963, pp. 362, 431.

21 Michael Schmitt, "Computer Network Attack and the Use of Force in International Law: Thoughts on a Normative Framework", *Columbia Journal of Transnational Law*, vol. 37, 1999, p. 916.

22 At the time of writing (June 2011), this was also the lowest common denominator identified by the expert group drafting the *Tallinn Manual*.

23 A Brazilian proposal to extend the prohibition to "the threat or use of economic measures in any manner inconsistent with the purposes of the United Nations"' was rejected at the San Francisco Conference; *Documents of the United Nations Conference on International Organization*, vol. VI, 1945, pp. 559, 720–721.

24 See Michael Schmitt, "Computer Network Attack and the Use of Force in International Law: Thoughts on a

pioneering commentator has early on proposed a complex list of indicative criteria for the distinction between armed force and economic or political coercion in the cyber domain,[25] others have recently pointed out qualitative, quantitative and temporal weaknesses in these criteria which illustrate, rather than remove, the continued lack of clarity in this respect.[26]

Arguably, from a teleological perspective, the Charter can only achieve its overarching purposes of maintaining international peace and security (article 1), and "to save succeeding generations from the scourges of war" (preamble), if it prohibits the resort to any forcible measure likely to provoke military counter-force and, ultimately, the outbreak of international armed conflict.[27] As a matter of logic, the Charter cannot allow that the prohibition of interstate force be circumvented by the application of non-violent means and methods which, for all intents and purposes, are equivalent to a breach of the peace between the involved states.[28] Consider, for example the crippling effect of cyber operations disabling the electrical power grids of major cities, the incapacitation of systems controlling industrial production, or the infiltration of malware designed to "blind" an entire air defence system.

In this context, it should also be noted that article 2(4) of the UN Charter prohibits the resort to force between states regardless of magnitude or duration.[29] As the International Court of Justice (ICJ) clarified in its *Nicaragua Case*, even minor acts of interstate force fall under the general prohibition of article 2(4) of the UN Charter, regardless of whether they also qualify as acts of "aggression", or as "armed attacks" entitling the targeted state to resort to force in self-defence.[30] This interpretation is reinforced by the approach

Normative Framework", *Columbia Journal of Transnational Law*, vol. 37, 1999, p. 912.

25 Ibid., pp. 914–15, which proposes a non-exhaustive list of seven criteria (including severity, immediacy, directness, invasiveness, measurability, presumptive legitimacy and state responsibility) as indicative factors for the distinction of armed force from economic and political coercion.

26 See Marco Roscini, "World Wide Warfare—Jus ad bellum and the Use of Cyber Force", in Armin Bogdany and Rüdiger Wolfrum (eds), *Max Planck Yearbook of United Nations Law*, vol. 14, 2010, p.104.

27 This issue is not settled, however. Some authors go as far as considering the unwarranted detention of a ship and even the breaking into a diplomatic bag as examples of unarmed force prohibited by art. 2(4) of the UN Charter (see Yoram Dinstein*, War, Aggression and Self-Defence,* 4th ed., 2005, p. 174), whereas others show caution already when discussing the spreading of a flood or of fire across an international border (Ian Brownlie, *International Law and the Use of Force by States*, 1963, p. 362; Albrecht Randelzhofer, "Article 2(4)", in Bruno Simma (ed.), *The Charter of the United Nations: A Commentary*, vol. I, 2002, pp. 118ff.).

28 According to Frowein and Krisch, "Article 39 UN Charter", in Bruno Simma (ed.), *The Charter of the United Nations: A Commentary*, vol. I, 2002, p. 721, a "breach of the peace" is typically characterized by hostilities between armed units of two states, irrespective of duration.

29 International Court of Justice, *Corfu Channel (United Kingdom of Great Britain and Northern Ireland v. Albania) (Merits)*, separate opinion by Judge Alvarez, 1949, p. 47; International Law Commission, *Addendum—Eighth report on State responsibility by Mr. Roberto Ago, Special Rapporteur—the internationally wrongful act of the State, source of international responsibility (part 1)*, UN document A/CN.4/318/Add.5–7, 1980, §§ 58 and 86; Ian Brownlie, *International Law and the Use of Force by States*, 1963, pp. 214 and 432; Yoram Dinstein*, War, Aggression and Self-Defence,* 4th ed., 2005, pp. 175ff.; Albrecht Randelzhofer, "Article 2(4)", in Bruno Simma (ed.), *The Charter of the United Nations: A Commentary*, vol. I, 2002, p. 123; D.W. Bowett, *Self-Defense in International Law*, 1958*,* pp. 12ff*,* 273. For a sceptical perspective, see N. Quoc Dinh et al., *Droit International Public*, 6th ed., 1999, p. 898.

30 International Court of Justice, *Military and Paramilitary Activities in and against Nicaragua (Nicaragua v. United States of America)*, merits, 1986, §§ 191 and 195; International law Commission, *Report of the International Law Commission on the work of its Thirty-second session, 5 May–25 July 1980, Official Records of the General Assembly, Thirty-fifth session, Supplement No. 10*, UN document A/35/10, 1980, p. 44; Yoram Dinstein*, War, Aggression and Self-Defence,* 4th ed., 2005, p. 174ff; Ian Brownlie, *International Law and the Use of Force by States*, 1963, pp. 363ff, 366.

taken in IHL, according to which even minor instances of armed force occurring between states are sufficient to trigger an international armed conflict.[31] Indeed, it would hardly make sense for article 2(4) of the UN Charter, as the primary norm aiming to safeguard international peace and security, not to systematically prohibit all forms of interstate conduct sufficient to give rise to an international armed conflict within the meaning of article 2 common to the Geneva Conventions. In fact, the UN Charter even goes further and prohibits not only the actual use, but already the threat of force in interstate relations. While the UN Charter does not define what constitutes a wrongful "threat" of interstate force, the ICJ has held that:

> [t]he notions of "threat" and "use" of force under Article 2, paragraph 4, of the Charter stand together in the sense that if the use of force itself in a given case is illegal—for whatever reason—the threat to use such force will likewise be illegal. In short, if it is to be lawful, the declared readiness of a State to use force must be a use of force that is in conformity with the Charter.[32]

Overall, however, there still is no consensus as to the precise threshold at which cyber operations should amount to an internationally wrongful threat or use of force. In fact, there is not even an identifiable controversy with clear positions and conflicting criteria. The truth is that cyber operations, almost always falling within the grey zone between traditional military force and other forms of coercion, simply were not anticipated by the drafters of the UN Charter and, so far, neither state practice nor international jurisprudence provide clear criteria regarding the threshold at which cyber operations not causing death, injury or destruction must be regarded as prohibited under article 2(4) of the UN Charter.

It should also be noted that a cyber operation need not amount to "force" within the meaning of article 2(4) of the UN Charter to be internationally wrongful, nor would all cyber operations amounting to "force" necessarily be unlawful. First, the illegality of a cyber operation may result from the violation of any obligation under international law. For example, interstate computer network exploitation for the purposes of intelligence gathering, electronic dissemination of hostile propaganda, or denial of service attacks would each violate the sphere of sovereignty of the affected state and, thus, the customary principle of non-intervention, even if they do not qualify as a use of force within the meaning of article 2(4) of the UN Charter.[33] Similarly, non-destructive cyber operations intruding into computer-based archives, documents and correspondence of a foreign diplomatic mission, or interfering with the mission's free

31 See section V.1.

32 International Court of Justice, *Legality of the Threat or Use of Nuclear Weapons*, advisory opinion, 1996, § 47.

33 According to the "Declaration on the Inadmissibility of Intervention and Interference in the Internal Affairs of States" annexed to UN General Assembly resolution 36/103, 9 December 1981, the principle of non-intervention includes, inter alia, the following rights and duties: I(c) the right of states and peoples to have free access to information and to develop fully, without interference, their system of information and mass media and to use their information media in order to promote their political, social, economic and cultural interests and aspirations, based, inter alia, on the relevant articles of the Universal Declaration of Human Rights and the principles of the new international information order; II(j) the duty of a state to abstain from any defamatory campaign, vilification or hostile propaganda for the purpose of intervening or interfering in the internal affairs of other states; III(d) the right and duty of states to combat, within their constitutional prerogatives, the dissemination of false or distorted news which can be interpreted as interference in the internal affairs of other states or as being harmful to the promotion of peace, co-operation and friendly relations among states and nations.

communication, would violate international obligations of the receiving state under the law governing diplomatic relations.[34] Potentially relevant legal issues could also arise under international trade law, or under human rights law, for instance where denial of service attacks interfere with the freedom of expression of persons coming within the jurisdiction of the operating state.[35] The focus of the present analysis, however, are the restraints imposed by existing international law on cyberwarfare and not the international permissibility (or not) of cyber operations more generally.

III.1.2. The "Interstate" Dimension of Cyber Operations

Article 2(4) of the UN Charter is addressed to states only[36] and prohibits their resort to force exclusively in their mutual "international relations". This essentially means that the use or threat of force must be legally attributable to a state and directed against one or several other states. In international law, acts are attributable to a state when they are performed by persons or entities acting on behalf or with the authorization or endorsement of a state so as to engage its international legal responsibility for their behaviour. Such persons or entities are described as "state agents". Persons or entities who are not acting on behalf of a state or whose link to a given state is insufficient to engage its international legal responsibility, on the other hand, cannot be regarded as state agents and can be described as "non-state actors".

State agency, entailing the attributability of individual conduct to a state, must be determined based on the international law of state responsibility. This body of law has most recently and most comprehensively been restated by the International Law Commission in its *Draft Articles on Responsibility of States for Internationally Wrongful Acts* (2001).[37] Further aspects of the law of state responsibility are also regulated in special treaty provisions or have been clarified in international jurisprudence. A detailed discussion of the law of state agency would exceed the purpose and scope of this paper. Suffice it to note that, in practice, state agents likely to carry out cyber operations include not only government personnel, such as members of the armed forces or intelligence agencies (*de jure* state agents), but increasingly also other persons authorized to act on behalf of a state, such as private contractors (*de facto* state agents).

The use of force (including through cyber operations) by individual hackers and other non-state actors may be relevant under international humanitarian law and, in some cases, international criminal law, but is not prohibited by article 2(4) of the UN Charter.[38]

---

34 Vienna Convention on Diplomatic Relations, arts. 24, 27 and 45(a).

35 According to article 19 of the Universal Declaration of Human Rights, "[e]veryone has the right to freedom of opinion and expression; this right includes freedom to hold opinions without interference and to seek, receive and impart information and ideas through any media and regardless of frontiers". Similarly, article 19 of the International Covenant on Civil and Political Rights states: "2. Everyone shall have the right to freedom of expression; this right shall include freedom to seek, receive and impart information and ideas of all kinds, regardless of frontiers, either orally, in writing or in print, in the form of art, or through any other media of his choice. 3. The exercise of the rights provided for in paragraph 2 of this article carries with it special duties and responsibilities. It may therefore be subject to certain restrictions, but these shall only be such as are provided by law and are necessary: (a) For respect of the rights or reputations of others; (b) For the protection of national security or of public order (ordre public), or of public health or morals."

36 According to art. 4 of the UN Charter, only states can be "members" of the United Nations as referred to in the prohibition of art. 2(4) of the UN Charter.

37 See UN General Assembly resolution A/RES/56/83 of 12 December 2001 and its annex.

38 Albrecht Randelzhofer, "Article 2(4)", in Bruno Simma (ed.), *The Charter of the United Nations: A*

While states providing significant support to such non-state actors generally cannot be directly held responsible for cyber operations carried out by the latter, their assistance may in and of itself amount to "indirect" use of force in contravention of article 2(4) of the UN Charter and the principle of non-intervention. State agency, particularly *de facto* agency, should not to be confused with such "indirect" use of force. While the use of force by *de facto* state agents is directly attributed to the state on whose behalf they are acting, "indirect" use of force denotes a form of support by a state for non-state actors using force on their own behalf. In consequence, the supporting state is internationally responsible for the given assistance, but not for the force used by the assisted entities or persons.[39]

Finally, it cannot be excluded that the use of force by non-state actors may amount to a threat to international peace and security and require the Security Council to take or authorize measures of collective enforcement. Nevertheless, the prohibition of the actual resort to force by and among non-state actors is generally a matter of domestic criminal law and certainly is not the aim of Article 2(4) of the UN Charter.

III.2. Cyber Operations as "Armed Attacks"

III.2.1. The Difference between "Force" and "Armed Attack"

According to article 51 of the UN Charter, "[n]othing in the present Charter shall impair the inherent right of individual or collective self-defense if an armed attack occurs against a Member of the United Nations".[40] With respect to the use of force, this terminology suggests a gap between the prohibition of "force" under article 2(4) and the exception in case of "armed attack" of article 51 of the UN Charter. Indeed, the scope of article 2(4) of the UN Charter is wider than that of article 51 in that it prohibits not only armed, but also unarmed or indirect modes of force, and not only the actual use, but also the mere threat of force. In other words, not every threat or use of force prohibited by article 2(4) automatically also constitutes an armed attack justifying self-defensive action in derogation from the Charter regime.[41]

Commentary, vol. I, 2002, p. 121; Lassa Oppenheim, *International Law: A Treatise,* vol. II, 1921, § 57; Alfred Verdross and Bruno Simma, *Universelles Völkerrecht: Theorie und Praxis*, 1984, § 468.

39 In the *Nicaragua Case*, the ICJ ultimately concluded that the relations between the United States and the Contra rebels did not qualify as *de facto* agency but that the United States' conduct under review constituted "indirect use of force" (International Court of Justice, *Military and Paramilitary Activities in and against Nicaragua (Nicaragua v. United States of America)*, merits, 1986, §§ 115 ff, 205, 247). Conversely, in the *Tadić Case*, the International Criminal Tribunal for the former Yugoslavia concluded that the relations between the Federal Republic of Yugoslavia and the Bosnian Serb militia amounted to *de facto* agency, thus giving rise to an international armed conflict between Bosnia and Herzegovina on the one hand and the Federal Republic of Yugoslavia on the other (International Criminal Tribunal for the former Yugoslavia, Tadić judgement*,* 15 July 199, § 162); see also, inter alia, the territorial criteria discussed by the Tribunal in *§§* 138–140.

40 In French: "aggression armée". The ICJ confirmed this requirement for the right of both individual and collective self-defence also under customary international law. See International Court of Justice, *Military and Paramilitary Activities in and against Nicaragua (Nicaragua v. United States of America)*, merits, 1986, § 195.

41 See Albrecht Randelzhofer, "Article 51 UN Charter", in Bruno Simma (ed.), *The Charter of the United Nations: A Commentary*, vol. I, 2002, p. 790; Yoram Dinstein*, War, Aggression and Self-Defence,* 4th ed., 2005, pp. 167, 174; Ian Brownlie, *International Law and the Use of Force by States*, 1963, p. 278 ff. The existence of this "gap" was confirmed for customary international law in International Court of Justice, *Military and Paramilitary Activities in and against Nicaragua (Nicaragua v. United States of America)*, merits, 1986, § 191, where the Court considered that it was "necessary to distinguish the most grave forms

This discrepancy is not surprising, but is both necessary and intended. Though prohibited under article 2 (4) of the Charter, the threat or use of force below the threshold of "armed attack" is not of sufficient gravity to justify a response in derogation from the Charter regime of collective enforcement, prohibition of unilateral force and peaceful settlement of disputes.[42] The restriction of the derogatory clause of article 51 to cases of armed attack expresses the Charter's intent to prevent unnecessary escalation of interstate force and, in doing so, puts the common interest in the preservation of international peace and security before the interest of individual states in the absolute and immediate protection of their sovereign rights.[43] The lack of an express derogatory clause for situations where sates are confronted with the threat or use of force below the threshold of an armed attack does not preclude the injured state from exercising its right of self-defence through means not disallowed under the Charter, such as the interruption of network communication services, the introduction of domestic control and security measures, the mobilization and preparation of effective defence, or even counter-measures not involving the use of force.[44]

In one aspect, at least, the scope of article 51 may also exceed that of article 2(4), namely by derogating from the Charter restrictions *in all cases* where an armed attack occurs against a member state, irrespective of its attributability to another state. Arguably, therefore, an armed attack against a state carried out by non-state actors from within the territory of another state—although not, as such, prohibited under article 2(4)[45]—could potentially justify self-defensive action within that (territorial) state in derogation from Charter restrictions.[46] It should be pointed out, however, that the interpretation of the notion of armed attack to include acts carried out by non-state actors remains controversial and does not reflect universal consensus.

In any case, the practical relevance of the qualification of a state-sponsored cyber operation as an "armed attack" is that it would allow the injured state to take self-defensive action in derogation from the treaty restrictions otherwise imposed by the UN Charter including, most notably, the resort to military force both within and outside the cyber domain. In view of the difficulty of determining the precise threshold at which cyber operations should amount to a threat or use of "force" prohibited under article 2(4) of the UN Charter, similar problems are awaiting any attempt to transpose the notion of "armed attack" to the cyber realm.

---

of the use of force (those constituting an armed attack) from other less grave forms" (confirmed in ICJ, *Oil Platform Case*, § 51).

42 See Michael Schmitt, "Computer Network Attack and the Use of Force in International Law: Thoughts on a Normative Framework", *Columbia Journal of Transnational Law*, vol. 37, 1999, p. 929.

43 Therefore, contrary to Bowett's contention, this restriction to cases of armed attack is neither "unnecessary" nor "inconsistent with Art. 2 [4] which forbids not only force but the threat of force" (D.W. Bowett, *Self-Defense in International Law*, 1958, p. 191).

44 See ibid., pp. 23ff, which describes four possible constellations (1: forcible defence against force; 2: non-forcible defence against force; 3: forcible defence against non-forcible delict; 4. non-forcible defence against non-forcible delict) and doubts the lawfulness of the third due to lack of proportionality.

45 Article 2(4) of the UN Charter prohibits the threat or use of force by members of the United Nations only, not non-state actors. While article 2(4) of the UN Charter may prohibit territorial states from tolerating or supporting non-state actors resorting to force against a third state, it does not prohibit the non-state violence itself.

46 See, most notably, UN Security Council resolutions 1368 (12 September 2001) and 1373 (28 September 2001) reaffirming the inherent right to self-defence in the context of the terrorist attacks of 11 September 2001 against the United States.

III.2.2. Cyber Operations as "Armed" Attacks

First, from a textual perspective, the notion of "armed attack" necessarily implies the use of a weapon. In the advisory opinion on the legality of the threat or use of nuclear weapons, the ICJ clarified that article 51 of the Charter, just as articles 2(4) and 42, applies "to any use of force, regardless of the weapons employed".[47] While cyber attacks do not depend on the availability of traditional kinetic, biological, chemical or nuclear weaponry, they cannot be carried out without the requisite infrastructure making up cyberspace, thus raising the question of its qualification as a "weapon". In this respect, it has been convincingly noted that:

> it is neither the designation of a device, nor its normal use, which make it a weapon but the intent with which it is used and its effect. The use of any device, or number of devices, which results in a considerable loss of life and/or extensive destruction of property must therefore be deemed to fulfil the conditions of an "armed" attack.[48]

It would thus appear that cyber operations have the qualitative capacity to qualify as an "armed" attack within the meaning of article 51 of the UN Charter.

Beyond this conclusion, however, the criteria become murky.[49] In *Nicaragua*, the ICJ found it "necessary to distinguish the most grave forms of the use of force (those constituting an armed attack) from other less grave forms", such as a "mere frontier incident", based on the "scale and effects" of the force involved.[50] Unfortunately, however, the Court's subsequent failure to further explain and specify its reasoning provided for more confusion than insight and, today, does not prove particularly helpful in transposing the concept of "armed attack" to cyber operations.

---

47 International Court of Justice, *Legality of the Threat or Use of Nuclear Weapons*, advisory opinion, 1996, § 39.

48 Karl Zemanek, "Armed attack", in Rüdiger Wolfrum (ed.), *Max Planck Encyclopedia of Public International Law*, 2010, § 21.

49 As already a cursory review of legal doctrine illustrates, the controversy has remained vague and inconclusive. Some writers require that, in order for force to qualify as an "armed attack", it must be used on a relatively large scale: Albrecht Randelzhofer, "Article 51 UN Charter", in Bruno Simma (ed.), *The Charter of the United Nations: A Commentary*, vol. I, 2002, p. 796 ("relatively large scale"); and Christopher Greenwood, "War, terrorism and international law", *Current Legal Problems*, vol. 56, no. 1, 2003, pp. 516 ("certain level of gravity") and 518 ("of sufficient intensity""); while others doubt the validity of this criterion beyond the principle *de minimis non curat lex* ("the law is not concerned with trifles"), and see no reason to exclude small-scale armed attacks from the category of armed attack as long as the consequences reach a certain threshold, such as human casualties or serious destruction of property (see Yoram Dinstein*, War, Aggression and Self-Defence,* 4th ed., 2005, pp. 174–75, and, more specifically with regard to cyber operations, Yoram Dinstein, "Computer Network Attacks and Self-Defense", in Michael Schmitt and Brian O'Donnell (eds), *Computer Network Attack and International Law*, 2002, p. 105). Others limit themselves to admitting that, despite the terminology used by the ICJ, the dividing line between an armed attack and a frontier incident may often be unclear (see Robert Jennings and Arthur Watts (eds), *Oppenheim's International Law*, 9th ed., 1992, vol. 1, § 127 [p. 418, n. 6]), or simply tend to equate the notion of "armed attack" with the direct use of "armed force" against a state, irrespective of its scale and intensity (Michael Schmitt, "Computer Network Attack and the Use of Force in International Law: Thoughts on a Normative Framework", *Columbia Journal of Transnational Law*, vol. 37, 1999, p. 929).

50 International Court of Justice, *Military and Paramilitary Activities in and against Nicaragua (Nicaragua v. United States of America)*, merits, 1986, §§ 191 and 195.

III.2.3. Effects "Equivalent" to Death, Injury and Destruction

A good starting point may be to ask whether—as a minimum—every state-sponsored cyber operation intended to inflict death, injury or physical destruction within the sphere of sovereignty of another state would automatically also qualify not only as a "use of force" but also as an "armed attack" within the meaning of article 51 of the UN Charter. From a teleological perspective, this approach would certainly be convincing in that it would liberate states from the prohibition on the use of (counter-)force as soon as cyber operations directed against them are likely to result in destructive effects equivalent to those normally caused by the use of kinetic, chemical, biological or nuclear weapons. In view of the disruptive, rather than destructive, consequences of the vast majority of cyber attacks it remains unsatisfactory, however, to interpret the "scale and effects" criterion exclusively in terms of effects equivalent to physical destruction. The main problem is that, depending on what is considered to be "equivalent" to physical destruction, this approach will either end up being too restrictive (that is, including only cyber operations directly resulting in physical destruction but not, for example, the "mere" incapacitation of the entire national power grid, telecommunication network or air defence system) or too expansive (that is, including any large scale denial of service attack even against non-essential, purely civilian service providers such as, for example, online shopping services or telephone directories).

III.2.4. Cyber Attacks Incapacitating "Critical Infrastructures"

In order to come to an adequate interpretation of the "scale and effects" criterion in the absence of direct infliction of death, injury or destruction, reference could be made to so-called "critical infrastructures",[51] the protection of which has always been the key concern of states in their discussion of cybersecurity.[52] The advantage of the concept is that it is widely used by states and multilateral organizations in the discussion of cybersecurity and seems to address one of their key concerns in this respect. Moreover, although national interpretations of the term vary, the following examples illustrate that there is sufficient overlap and consistency to provide for a viable working definition:

- **UN General Assembly:** Critical infrastructures include "those used for, inter alia, the generation, transmission and distribution of energy, air and maritime transport, banking and financial services, e-commerce, water supply, food distribution and public health—and the critical information infrastructures that increasingly interconnect and affect their operations".[53]

---

51 See also Marco Roscini, "World Wide Warfare—Jus ad bellum and the Use of Cyber Force", in Armin Bogdany and Rüdiger Wolfrum (eds), *Max Planck Yearbook of United Nations Law*, vol. 14, 2010, p. 96.

52 See, e.g., UN General Assembly resolution 58/199 of 30 January 2004 ("Creation of a global culture of cybersecurity and the protection of critical information infrastructures"); US Presidential Decision Directive 63, "Critical Infrastructure Protection", 22 May 1998; The White House, *"The National Strategy for the Physical Protection of Critical Infrastructures and Key Assets*", 2003; and European Commission, *"Green Paper on a European Programme on Critical Infrastructure Protection*", document COM(2005) 576 final, 17 November 2005. See also, e.g., Eric Jensen, "Computer Attacks on Critical National Infrastructure: A Use of Force Invoking the Right of Self-Defence", *Stanford Journal of International Law*, vol. 38, 2002, pp. 207ff.; Sean Condron, "Getting It Right: Protecting American Critical Infrastructure in Cyberspace", *Harvard Journal of Law and Technology*, vol. 20, 2007, p. 403; Lesley Swanson, "The Era of Cyber Warfare: Applying International Humanitarian Law to the 2008 Russian-Georgian Cyber Conflict", *Loyola of Los Angeles International and Comparative Law Review*, vol. 32, 2010, p. 306.

53 UN General Assembly resolution 58/199 of 30 January 2004.

- **United States:** "Critical infrastructures are those physical and cyber-based systems essential to the minimum operations of the economy and government. They include, but are not limited to, telecommunications, energy, banking and finance, transportation, water systems and emergency services, both governmental and private".[54] "[T]he term 'critical infrastructure' means systems and assets, whether physical or virtual, so vital to the United States that the incapacity or destruction of such systems and assets would have a debilitating impact on security, national economic security, national public health or safety, or any combination of those matters".[55] "The critical infrastructure sectors consist of agriculture and food, water, public health, emergency services, government, the defense industrial base, information and telecommunications, energy, transportation, banking and finance, chemicals and hazardous materials, and postal and shipping".[56]

- **Shanghai Cooperation Organization:**[57] "'Critical structures'—public facilities, systems and institutions attacks on which may cause consequences directly affecting national security, including that of the individual, society and state".[58]

- **European Union:** "Critical infrastructure include those physical resources, services, and information technology facilities, networks and infrastructure assets which, if disrupted or destroyed, would have a serious impact on the health, safety, security or economic well-being of Citizens or the effective functioning of governments".[59] "Critical Information Infrastructure (CII): ICT systems that are critical infrastructures for themselves or that are essential for the operation of critical infrastructures (telecommunications, computers/software, Internet, satellites, etc.)".[60]

- **Australia**: "Critical infrastructure is defined as those physical facilities, supply chains, information technologies and communication networks which, if destroyed, degraded or rendered unavailable for an extended period, would adversely impact on the social or economic wellbeing of the nation or affect Australia's ability to ensure national security". Critical infrastructure sectors include: "banking and finance, communications, emergency services, energy, food chain, health (private), water services, mass gatherings, and transport (aviation, maritime and surface)".[61]

Even though the UN General Assembly rightly recognized "that each country will determine its own critical information infrastructures",[62] a determination likely to be

54 US Presidential Decision Directive 63, "Critical Infrastructure Protection", 22 May 1998, § I.

55 US Patriot Act of 26 October 2001 (42 U.S.C. 5195c(e)), § 1016(e).

56 The White House, *"The National Strategy for the Physical Protection of Critical Infrastructures and Key Assets"*, 2003, p. 35. See also The White House, *International Strategy for Cyberspace*, 2011, p. 3.

57 At the time of writing, the Shanghai Cooperation Organization included the following six member states: China, Russia, Kazakhstan, Kyrgyzstan, Tajikistan and Uzbekistan.

58 Annex I to the Agreement between the Governments of the Member States of the Shanghai Cooperation Organization on Cooperation in the Field of International Information Security of 16 June 2009.

59 European Commission, *Green Paper on a European Programme on Critical Infrastructure Protection*, document COM(2005) 576 final, 17 November 2005, annex 1, p. 20. An annexed "Indicative List of Critical Infrastructure Sectors" includes, with further specifications: energy, information and communication technologies, water, food, health, financial, public and legal order and safety, civil administration, transport, chemical and nuclear industry, and space and research (ibid., annex 2, p. 24).

60 Ibid., annex 1, p. 19.

61 Australian government, *Cyber Security Strategy*, 2009, p. 20.

62 UN General Assembly resolution 58/199 of 30 January 2004.

based on fluctuating considerations of national security,[63] it cannot be denied that the concept facilitates transposing the "scale and effects" criterion to the cyber domain. Accordingly, it could be argued that cyber attacks unlikely to result in death, injury or destruction could still amount to an "armed attack" if they aim to incapacitate "critical infrastructures" within the sphere of sovereignty of another state.

III.2.5. The Question of Intent

Although the ICJ's distinction between the "most grave" forms of the use of force and "mere border incidents" is intuitively convincing, a purely quantitative interpretation of the "scale and effects" criterion, even if extended to the incapacitation of critical infrastructures, remains unsatisfactory. On the one hand, it would certainly contradict the purposes of the Charter if every random "border incident" (or harmful cyber operation) could justify a military response in derogation from the prohibition of force and the duty of peaceful settlement of disputes. On the other hand, even extraterritorial commando operations resulting in the death of a single individual have been qualified as acts of aggression by the Security Council, which presumably would be of sufficient gravity to warrant self-defensive action.[64] What seems to be decisive for the distinction between "armed attacks" and less grave forms of the use of force, therefore, is not only the quantitative scale and effects of the operation, but also the degree to which it reflects the specific intent of the operating state's military or political leadership to violate another state's sphere of sovereignty.[65] This criterion would require that not only the cyber-attack in question, but also the aggressive intent inherent in the ordinary meaning of "attack", be legally attributable to the operating State. In doing so, it would not only reflect the maxim *de minimis non curat lex*, but would also avoid the accidental spreading of malware from being qualified as an armed attack based exclusively on the objective "scale and effects" of the accident.[66]

In the final analysis it should be recognized, however, that the threshold at which cyber operations may qualify as "armed attacks" triggering the derogatory clause in favour of self-defence has not been authoritatively clarified. Indeed, although the ICJ claimed already 25 years ago that "[t]here appears now to be general agreement on the nature of the acts which can be treated as constituting armed attacks",[67] this issue is far from resolved even today.

63 Marco Roscini, "World Wide Warfare—Jus ad bellum and the Use of Cyber Force", in Armin Bogdany and Rüdiger Wolfrum (eds), *Max Planck Yearbook of United Nations Law*, vol. 14, 2010, pp. 118–19.

64 On 16 April 1988, nine Israeli commandos killed PLO military strategist Khalil al-Wazir (Abu Jihad) in his home in Tunis. Tunisia brought the matter to the UN Security Council and, on 25 April 1988, the Council passed resolution 611 condemning the Israeli operation as an "aggression" in flagrant violation of the UN Charter, international law and norms of conduct.

65 Regarding the requirement of a specific intent to cause harm, see also International Court of Justice, *Oil Platforms (Islamic Republic of Iran v. United States of America)*, judgement, 6 November 2003, § 64.

66 See also Marco Roscini, "World Wide Warfare—Jus ad bellum and the Use of Cyber Force", in Armin Bogdany and Rüdiger Wolfrum (eds), *Max Planck Yearbook of United Nations Law*, vol. 14, 2010, p. 116.

67 International Court of Justice, *Military and Paramilitary Activities in and against Nicaragua (Nicaragua v. United States of America)*, merits, 1986, § 195. Instead of providing its own definition, however, the Court subsequently limits itself to argue on the basis of concrete examples.

III.3. Modalities Governing Self-Defence in the Cyber Domain

The basic function of the concept of self-defence in international law lies in protecting the legal order by balancing the rights of an attacking state against those of a defending state. In essence, it permits the defending state to take the measures necessary to repel an armed attack, even though this may require action otherwise prohibited under international law, most notably the use of interstate force. The justification for this permission is found in the initial wrongfulness of the offending state's conduct and the need to avert the harm likely to result from that wrongful conduct. Historically, a string of precedents such as the famous *Caroline* incident[68] have led to the formulation of certain modalities and principles which, although not expressly codified in the UN Charter, govern the exercise of the right of self-defence as a matter of customary international law.[69] These modalities comprise, most notably, the principles of necessity and proportionality.

While the *principle of necessity* defines the margins of lawful self-defence in terms of what is objectively necessary to avert or repel an armed attack, the *principle of proportionality* determines to what extent the harm to be prevented justifies the harm done by the defensive action. From a *qualitative* perspective, the principle of necessity requires that the self-defensive resort to an otherwise wrongful conduct, normally the use of force, be objectively necessary to avert or repel an armed attack (qualitative necessity). If this precondition is fulfilled, the principle additionally requires that the adopted measures be *temporally* and *quantitatively* necessary for the defensive action to achieve this legitimate purpose. From a *temporal* perspective, self-defensive action may not lawfully be carried out *before* it has actually become necessary to repel an armed attack, nor when it no longer is necessary for that purpose (temporal necessity). Indeed, the aim of self-defence is not to react to harm already done but to prevent the materialization of harm potentially resulting from a threat. It is therefore erroneous to claim that self-defensive action can be taken "after" an armed attack has occurred. Instead, it must be directed "against" an imminent or ongoing attack with the aim of preventing or repelling it. Note that the requirement of temporal necessity is sometimes also less convincingly discussed as a requirement of "immediacy" additional to the principles of necessity and proportionality. From a *quantitative* perspective, the principle requires that the kind and degree of force used in self-defence not exceed what is actually necessary to repel the armed attack in question (quantitative necessity).

The *principle of proportionality* (often confused with the quantitative aspect of the principle of necessity) additionally requires that the harm caused by the self-defensive action both to the attacking state and to uninvolved third states and individuals be justified by the gravity of the armed attack which the defensive action is designed to

68 On 29 December 1837, British troops made an incursion into US territory and destroyed the steamboat *Caroline*, which was being used by insurgents opposing British rule in Canada to ship recruits and equipment across the US–Canadian border. The subsequent diplomatic exchange between the two states led to the formulation of a number of conditions and modalities which had to be fulfilled to justify the British violation of US territorial sovereignty and which still form the basis for the customary principles governing self-defensive action today.

69 As stated by the ICJ, "[t]he entitlement to resort to self-defense under Article 51 is subject to certain constraints. Some of these constraints are inherent in the very concept of self-defense. Other requirements are specified in Article 51" (International Court of Justice, *Legality of the Threat or Use of Nuclear Weapons*, advisory opinion, 1996, § 40).

avert or repel. According to this modality, action taken in self-defence is legally justified only to the extent that the harm it is expected to cause remains in reasonable proportion to the harm it aims to prevent.

According to these modalities, self-defensive action in cyberspace is not permissible in response to harm which has already been caused by hostile cyber operations, but only with a view to preventing or repelling an imminent or ongoing attack, and only to the extent actually necessary for that purpose. Moreover, the harm likely to be inflicted on the attacker or uninvolved third states or individuals by self-defensive cyber action must always be justified by the seriousness of the harm to be prevented. The speed, unpredictability and clandestine nature of most cyber operations severely hamper the defending state's ability to react in time to detect and prevent or repel an imminent or ongoing attack, which may well be designed and timed to produce its harmful effects only months after the attacker's intrusion. In practice, cyberdefence must largely rely on automated systems, which render a human case-by-case assessment and verification both of the attacker's identity and the necessity and proportionality of self-defensive action extremely difficult. These specific characteristics of cyber operations, in conjunction with the fact that cyber attacks are increasingly conducted by non-state actors relying on series of small-scale operations, have provoked an extension to cyberspace of the continuing discussion on the permissibility of anticipatory self-defence.[70] Irrespective of how this question will eventually be resolved as a general matter, ensuring the compliance of self-defensive action in cyberspace with the fundamental requirements of necessity and proportionality will certainly continue to pose a significant challenge for some time to come. Lastly, it should also be recalled that while an armed attack within the meaning of article 51 of the UN Charter justifies self-defensive action in derogation from the UN Charter regime of peaceful dispute settlement and non-use of force, it does not relieve the defending state from its obligations under other applicable frameworks of international law such as, most notably, international humanitarian law.[71]

III.4. Cyber Operations and UN Security Council Enforcement

Within the United Nations system, the Member States "confer on the Security Council primary responsibility for the maintenance of international peace and security".[72] To the extent that cyber operations can adversely affect the international relations between states, there can be no doubt that the Council's responsibility also extends to maintaining international peace and security in cyberspace. When the Security Council determines the existence of a "breach of the peace", an "act of aggression" or, most commonly, a "threat to the peace", it can undertake or authorize such measures as may be necessary to maintain or restore international peace and security.[73]

70 See also Michael Schmitt, "Computer Network Attack and the Use of Force in International Law: Thoughts on a Normative Framework", *Columbia Journal of Transnational Law*, vol. 37, 1999, pp. 932–33; Eric Jensen, "Computer Attacks on Critical National Infrastructure: A Use of Force Invoking the Right of Self-Defence", *Stanford Journal of International Law*, vol. 38, 2002, pp. 221–24; and Marco Roscini, "World Wide Warfare—Jus ad bellum and the Use of Cyber Force", in Armin Bogdany and Rüdiger Wolfrum (eds), *Max Planck Yearbook of United Nations Law*, vol. 14, 2010, pp. 120–23.

71 See section V.

72 Art. 24, UN Charter.

73 Chp. VII, UN Charter.

Such measures may be limited to issuing recommendations,[74] or calling on the involved parties to comply with provisional measures,[75] but may also involve armed and unarmed enforcement.[76] The UN Charter lists as an example of unarmed enforcement the "complete or partial interruption of ... telegraphic, radio, and other means of communication",[77] thus providing an express basis for UN-sanctioned cyber blockades regardless of whether the relevant threat to the peace arises in cyberspace. Should the Council come to the conclusion that measures not involving armed force are or would be inadequate, "it may take such action by air, sea, or land forces as may be necessary to maintain or restore international peace and security".[78] While there can be no doubt that this provision also applies where the relevant threat to the peace arises in cyberspace, a textual reading of article 42 of the UN Charter seems to provide a basis for armed enforcement action only "by air, sea, or land forces". This may raise the question of whether forces operating in the separate domains of space and cyberspace are excluded. Clearly, however, the purpose of article 42 of the UN Charter was not to restrict the means of enforcement available to the Security Council but to extend them, where need be, to all armed services available to the leading military powers of the drafting period. From a teleological perspective, therefore, article 42 of the UN Charter cannot reasonably be interpreted as depriving the Security Council of the possibility to authorize the use of armed force in cyberspace.

Contrary to the notions of "force" and "armed attack", that of "threat to the peace" is a largely political concept,[79] which leaves the Security Council a broad measure of discretion. As a matter of law, the determination of a "threat to the peace" neither presupposes an internationally wrongful act,[80] nor a threat or use of "force" or the occurrence of an "armed attack" within the meaning of the UN Charter. In principle, therefore, the Security Council has the power to authorize enforcement action, including military force, against cyber threats far below the threshold required for self-defensive action or even for a qualification as interstate force. In determining whether a particular cyber operation constitutes a threat to the peace, the Security Council's discretion is not completely unlimited. At the very least, the Council is obliged to act in conformity both with the purposes and principles of the Charter[81] and, more generally, with the "principles of justice and international law".[82] Lastly, just as in the case of self-defensive action, it should be recalled that an authorization of armed enforcement by the UN Security Council merely provides a legal justification for the otherwise prohibited use of force against another state, but does not relieve enforcing states from their obligations under other applicable frameworks of international law such as, most notably, international humanitarian law.

74 Art. 39, UN Charter.

75 Art. 40, UN Charter.

76 Arts. 41 and 42, UN Charter.

77 Art. 41, UN Charter.

78 Art. 42, UN Charter.

79 International Criminal Tribunal for the former Yugoslavia, *Tadić Case* (IT-94-1), Decision on the Defence Motion for Interlocutory Appeal on Jurisdiction, 1995, § 29.

80 D.W. Bowett, *Self-Defense in International Law*, 1958, p. 186, n. 2.

81 See also International Criminal Tribunal for the former Yugoslavia, *Tadić Case* (IT-94-1), Decision on the Defence Motion for Interlocutory Appeal on Jurisdiction, 1995, § 29, and ibid.

82 Art. 1(1), UN Charter.

IV. Cyber Operations and the Law of Neutrality

As recognized by the ICJ, neutrality is a fundamental principle of international law that applies "whatever type of weapons might be used".[83] According to the principle, in a situation of international armed conflict, a neutral state is obliged to prevent its territory from being used by the belligerents,[84] a notion which can be interpreted to include cyberspace. The belligerents, in turn, must respect the inviolability of neutral territory and "are forbidden to move troops, or convoys of either munitions of war or supplies across the territory of a neutral Power".[85] The Convention further provides that neutral states are "not called upon to forbid or restrict the use on behalf of the belligerents of telegraph or telephone cables or of wireless telegraphy apparatus belonging to it or to companies or private individuals", as long as it applies the same policy towards all belligerents.[86] This provision is of particular relevance for the context of cyberwarfare because, although the target (or destination) of cyber operations can generally be determined with precision, their geographical routing cannot normally be controlled so as to completely avoid the use of neutral telecommunications infrastructure. The question is, therefore, whether the information and payloads transmitted by the belligerents through neutral cyber infrastructure constitute actual weapons systems (which would violate the law of neutrality) or mere communication data (which would be permissible). From a technical point of view the accurate answer is that, depending on the precise nature and design of the cyber operation in question, either option can be the case. For example, the large quantities of communication data used to flood selected servers in denial of service attacks could hardly, as such, be regarded as a "weapons system", whereas the contrary may have to be concluded in case of payloads being reconstituted at their destination with elements of local infrastructure and data to provide a destructive attack capability. The rationale of exempting neutral powers from controlling the use "on behalf of the belligerents of telegraph or telephone cables or of wireless telegraphy" was not, however, to distinguish between communications and weapons systems as much as it simply reflected the impossibility of the task of controlling the extraterritorially initiated use of publicly accessible transnational communications networks. Where the neutral state exercises territorial control, on the other hand, the Convention expressly prohibits that belligerents "(a) Erect on the territory of a neutral Power a wireless telegraphy station or other apparatus for the purpose of communicating with belligerent forces on land or sea; (b) Use any installation of this kind established by them before the war on the territory of a neutral Power for purely military purposes, and which has not been opened for the service of public messages".[87] Today, the exact same rationale underlying the Hague Convention would suggest that neutral states can be expected to prevent belligerent states from conducting cyber hostilities from within their territory, but not the routing of belligerent cyber operations through their publicly accessible communications infrastructure.

83 International Court of Justice, *Legality of the Threat or Use of Nuclear Weapons*, advisory opinion, 1996, § 89.

84 Art. 5, Rights and Duties of Neutral Powers and Persons in Case of War on Land (Hague V), 1907.

85 Ibid., art. 2.

86 Ibid., art. 8.

87 Ibid., art. 3.

Strictly speaking, the law of neutrality applies only in international armed conflict. Arguably, however, the pragmatic logic of its core principles has already found its way into the practice of non-international armed conflicts as well.[88] The practical consequences of non-state belligerents abusing "neutral" territory to conduct attacks against other states are not unlike those foreseen in the traditional law of neutrality and include, most notably, the loss of the neutral territory's inviolability. For example, as has been seen in connection with the attacks conducted by Al-Qaida against the United States from within Afghanistan, by Hezbollah against Israel from within Lebanon, and by the Fuerzas Armadas Revolucionarias de Colombia (FARC) against Colombia from within Ecuador, the attacked states have conducted extraterritorial military interventions directly against the respective groups because their "neutral" host states were either unable or unwilling to protect the attacked states' interests within their territory. While the permissibility of such extraterritorial incursions remains widely controversial in view of the UN Charter regime regulating the use of interstate force, the basic obligation of states to prevent hostile activities against other states from within their territory appears to be widely recognized, although normally expressed in terms of the principle of non-intervention rather than that of neutrality.[89]

V. Cyber Operations and *Jus in Bello*

IHL, sometimes also described as the "law of armed conflict" or *jus in bello*, applies exclusively in situations of armed conflict and regulates the conduct of hostilities between the belligerent parties, as well as the protection and treatment of those having fallen into the power of the enemy. Today, the most important sources of IHL are the four Geneva Conventions of 1949 (GC I–IV) and their first two Additional Protocols of 1977 (AP I and II), as well as the Regulations annexed to the Fourth Hague Convention of 1907 (H IV R) and a series of treaties prohibiting or restricting the use of certain weapons. Additionally, in the course of decades and centuries of warfare, a rich body of customary IHL has developed which proves helpful in cases not regulated by applicable treaty law.[90] In the following, "cyberwarfare" shall first be distinguished from phenomena not necessarily governed by IHL, such as "cyber criminality" and "cyberterrorism". Where IHL does apply, it shall be examined to what extent its most important rules and principles, designed to govern traditional means and methods of warfare, can be transposed to cyberwarfare. In doing so, the focus will be on the rules

88 See, for example, the Organization of American States Convention (1929) and Protocol (1957) on "Duties and Rights of States in the Event of Civil Strife". See also the International Committee of the Red Cross Official Statement of 8 March 2001 to the United Nations High Commissioner for Refugees Global Consultations on International Protection: "It is the ICRC's view that it [the Fifth Hague Convention] can also be applied by analogy in situations of non-international conflicts, in which combatants either from the government side or from armed opposition groups have fled into a neutral state".

89 See, for example, the "duty of a State to ensure that its territory is not used in any manner which would violate the sovereignty, political independence, territorial integrity and national unity or disrupt the political, economic and social stability of another State" and "to refrain from the promotion, encouragement or support, direct or indirect, of rebellious or secessionist activities within other States, under any pretext whatsoever, or any action which seeks to disrupt the unity or to undermine or subvert the political order of other States" reflected in the "Declaration on the Inadmissibility of Intervention and Interference in the Internal Affairs of States", annexed to UN General Assembly resolution 36/103 of 9 December 1981, §§ II(b) and II(f).

90 See, most notably, the International Committee of the Red Cross' extensive study on customary IHL, Jean-Marie Henckaerts and Louise Doswald-Beck, *Customary International Humanitarian Law*, vols. I and II, 2005.

and principles of IHL governing the conduct of hostilities rather than those governing the protection and treatment of persons in the hands of a party to an armed conflict, which is an area less relevant for cyberwarfare.

V.1. Cyber Operations as "Warfare"

V.1.1. Definition and Terminology

The question of whether cyber operations can amount to war, warfare, armed conflict or hostilities raises preliminary questions of definition and terminology. For the time being, the notions of "cyberwar", "cyberwarfare", "cyber hostilities" and "cyber conflict" have not been authoritatively defined for the purposes of international law. The only treaty definition that exists, by the regional Shanghai Cooperation Organization, concerns the wider concept of "information war", which is defined as:

> confrontation between two or more states in the information space aimed at damaging information systems, processes and resources, critical and other structures, undermining political, economic and social systems, mass psychologic[al] brainwashing to destabilize society and state, as well as to force the state to taking decisions in the interest of an opposing party.[91]

As usefully pointed out by a leading commentator, the term "information warfare" is often inaccurately used as a synonym for "information operations": while the latter can occur both in times of peace and of war, the former refers exclusively to information operations conducted in situations of armed conflict and excludes information operations occurring during peacetime.[92] Applied to the more specific context of cyber operations, this means that the use of the term "cyberwar", "cyberwarfare", "cyber hostilities" and "cyber conflict" should be restricted to armed conflicts within the meaning of IHL. Indeed, security threats emanating from cyberspace which do not reach the threshold of armed conflict can be described as "cyber crime", "cyber operations", "cyber policing" or, where appropriate, as "cyberterrorism" or "cyber piracy", but should not be referred to with terminology inviting doubt and uncertainty as to the applicability of the law of armed conflict.

V.1.2. Cyber Operations in Pre-Existing Conflicts

Today, it appears to be uncontested that IHL applies to cyber operations which are carried out in the context of a pre-existing international or non-international armed conflict.[93] It seems to be generally recognized that the fact that cyber operations did not exist at the time of the drafting and adoption of most contemporary instruments of IHL does not preclude their applicability to such operations. One of the most fundamental rules of IHL has always been that the right of belligerents to choose methods or means of warfare is

91 Annex I to the Agreement between the Governments of the Member States of the Shanghai Cooperation Organization on Cooperation in the Field of International Information Security of 16 June 2009.

92 Michael Schmitt, "Computer Network Attack and the Use of Force in International Law: Thoughts on a Normative Framework", *Columbia Journal of Transnational Law*, vol. 37, 1999, p. 891.

93 Already the participants in the 2004 Stockholm Expert Conference agreed that "International Humanitarian Law applies to computer network attacks (CNA) in an ongoing international armed conflict" (Proceedings, p. 181). At the time of writing, the same approach is taken (unanimously) in the draft *Tallinn Manual*.

not unlimited,[94] and article 36 of AP I expressly requires that: "In the study, development, acquisition or adoption of a new weapon, means or method of warfare, a High Contracting Party is under an obligation to determine whether its employment would, in some or all circumstances, be prohibited by [international law]." Thus, existing IHL clearly anticipates the application of its rules and principles to newly developed methods and means of warfare. It is not the precise nature of a means or method, but the context in which it is used, which subjects it to the rules and principles of IHL. Whether a cyber operation must be regarded as carried out in the context of an armed conflict does not necessarily depend on the territorial connection of the operation but, rather, on whether it is carried out for reasons related to an armed conflict or, in the words of the International Criminal Tribunal for the Former Yugoslavia (ICTY), whether it has a "nexus" with an ongoing armed conflict. This also means that cyber operations conducted for reasons unrelated to an armed conflict (lack of nexus) may qualify as cyber criminality, cyber policing etc., but are not governed by IHL, even if carried out by a belligerent party, or within a territory affected by an armed conflict.

V.1.3. Can Cyber Operations Trigger an Armed Conflict?

One of the most difficult questions is whether and, if so, in what circumstances cyber operations can give rise to an armed conflict without the parallel occurrence of conventional hostilities. In other words, can cyber operations, in and of themselves, trigger the applicability of IHL? This question must not be confused with the distinct questions of whether cyber operations can qualify as a "threat or use of force" or an "armed attack" within the meaning of the UN Charter.[95] According to a 2008 opinion paper of the International Committee of the Red Cross (ICRC), the currently prevailing legal opinion on the definition of "armed conflict" under IHL can be summarized as follows:

> 1. **International armed conflicts** exist whenever there is resort to armed force between two or more States.
>
> 2. **Non-international armed conflicts** are protracted armed confrontations occurring between governmental armed forces and the forces of one or more armed groups, or between such groups arising on the territory of a State [party to the Geneva Conventions]. The armed confrontation must reach a minimum level of intensity and the parties involved in the conflict must show a minimum of organisation.[96]

The ICRC's opinion paper also emphasizes that, while a situation can evolve from a non-international to an international armed conflict and vice versa, "[l]egally speaking, no other type of armed conflict exists".[97] Consequently, cyber operations can trigger the applicability of IHL to the extent that they can give rise to all required constitutive elements of an international or non-international armed conflict.

International Armed Conflict: As far as international armed conflicts are concerned, cyber operations must amount to the "resort to armed force between two or more States". The question of whether armed force occurs "between" states essentially turns on

94 Art. 22 H IV.R; art. 35(1) AP I.

95 See sections III.1. and 2.

96 ICRC, "How is the Term 'Armed Conflict' Defined in International Humanitarian Law?", 2008, p. 5.

97 Ibid., p. 1.

legal attributability as governed by the general international law of state responsibility. Accordingly, the applicability of IHL cannot be limited to acts committed by members of the state armed forces but must be extended to the conduct of any other person acting as a state agent, whether *de jure* or *de facto*, on behalf of a belligerent. While there is no reason to alter the application of the law of state responsibility in cyber space, the identification of the source or author of a cyber operation can pose particularly difficult evidentiary problems, which may have consequences for the presumptions or precautions to be applied in case of doubt.

The second question is whether cyber operations can be regarded as "armed force" (or, in non-international armed conflict, "armed confrontation") triggering the applicability of IHL even in the absence of the use of kinetic force. So far, there seems to be consensus that this is the case, at least wherever cyber operations cause the same effects as kinetic force, namely death, injury or destruction.[98] Obviously, however, not every use of force necessarily indicates the existence of an armed conflict and not all acts of war necessarily involve a use of force. Indeed, armed conflicts can even be triggered by formal declarations of war. Strictly speaking, therefore, the existence of an international armed conflict does not necessarily depend on the use of "force" between states but, at least in the absence of a formal declaration of war, on the occurrence of belligerent "hostilities" within the meaning of IHL. Accordingly, state-sponsored cyber operations would give rise to an international armed conflict if they are designed to harm another state not only by directly causing death, injury or destruction, but also by directly adversely affecting its military operations or military capacity.[99]

Non-International Armed Conflict: The constitutive elements for non-international armed conflicts differ from those for international armed conflict in that they involve at least one non-state belligerent showing a minimum degree of organization, and in that the armed confrontations (that is, hostilities) must show a minimum level of intensity ("protracted"). The first criterion requires organized collective action, which would certainly exclude cyber operations conducted by individual hackers from the notion of armed conflict. From a strictly theoretical perspective it cannot be excluded that even a small, but organized, group of hackers launching highly destructive cyber operations against, say, a state's military networks could trigger a non-international armed conflict. As long as such cyber operations emanate from within territory controlled by the attacked state, however, and as long as they are not accompanied by a threat or use of conventional military force which could prevent the state from exercising its territorial authority over the attackers, such operations would in practice most likely be regarded as a criminal threat to be addressed through law enforcement measures. A qualification of such operations as "hostilities" capable of triggering a non-international armed conflict becomes more likely when they occur repeatedly over a certain duration and emanate from territory where the attacked state cannot exercise its law enforcement authority, and where the local authority is unwilling or unable to intervene.

98 Michael Schmitt, "Cyber Operations and the Jus in Bello: Key Issues", *Naval War College International Law Studies*, 2011, p. 15; Knut Dörmann, "The Applicability of the Additional Protocols to Computer Network Attacks: An ICRC Viewpoint", in Karin Byström (ed.), *International Expert Conference on Computer Network Attacks and the Applicability of International Humanitarian Law, 17–19 November 2004, Stockholm, Sweden*, Swedish National Defence College, 2004, p. 142. At the time of writing, this approach is also taken in the draft *Tallinn Manual*.

99 On the notion of hostilities, see below section V.3.

For the time being, it is probably still too early to make definite statements as to the precise threshold at which cyber operations trigger a non-international armed conflict (a question unresolved even for non-international conflicts fought through traditional means and methods). As has rightly been stated already in the ICRC's contribution to the 2004 Stockholm Conference, "[w]hether CNA alone will ever be seen as amounting to an armed conflict will probably be determined in a definite manner only through future state practice".[100]

In any case, once the existence of an armed conflict has been determined, it will have to be determined to what extent traditional concepts and rules of IHL can be transposed to cyber operations conducted in the context of that conflict. In doing so, this paper will in the following focus on examining those concept and principles which are likely to be most relevant in practice, namely the concepts of "attack", "hostilities" and "direct participation" therein, as well as the rules and principles governing targeting and good faith in the conduct of hostilities.

V.2. Cyber Operations as "Attacks"

The term "attack" is an important technical term of IHL in that many of its fundamental rules on the conduct of hostilities are expressed in terms of attacks. For example, "the civilian population as such, as well as individual civilians, shall not be the object of attack";[101] "civilian objects shall not be the object of attack";[102] "indiscriminate attacks are forbidden";[103] and "attacks shall be limited strictly to military objectives".[104] The same applies, *inter alia*, to the rules regulating "precautions in attack" and "precautions against the effects of attack",[105] those protecting medical units,[106] persons *hors de combat*,[107] works and installations containing dangerous forces[108] against attack, as well as those obliging combatants to distinguish themselves from the civilian population during attack or military operations preparatory to an attack,[109] and those prohibiting the use of the flags or military emblems, insignia or uniforms of adverse parties during attack.[110]

As to the definition of the term, article 49(1) of AP I provides that "attacks means acts of violence against the adversary, whether in offence or defence." This definition has triggered significant discussion as to what extent cyber operations, in view of their non-kinetic nature, could be regarded as "acts of violence" and, therefore, as "attacks" within

100 Knut Dörmann, "The Applicability of the Additional Protocols to Computer Network Attacks: An ICRC Viewpoint", in Karin Byström (ed.), *International Expert Conference on Computer Network Attacks and the Applicability of International Humanitarian Law, 17–19 November 2004, Stockholm, Sweden*, Swedish National Defence College, 2004, p. 142. At the time of writing, this approach is also taken in the draft *Tallinn Manual*.

101 Art. 51(2), AP I.

102 Art. 52(1), AP I.

103 Art. 51(4), AP I.

104 Art. 52(2), AP I.

105 Arts. 57 and 58, AP I.

106 Art. 12(1), AP I

107 Art. 41(1), AP I

108 Art. 56, AP I

109 Art. 44(3), AP I

110 Art. 39(2), AP I

the meaning of IHL. Today, it seems to be generally recognized that "acts of violence" do not necessarily require the use of kinetic violence, but that it is sufficient if the resulting effects are equivalent to those normally associated with kinetic violence, namely the death or injury of persons or the physical destruction of objects (effects-based approach).[111] Strictly speaking, this approach does not "extend" the notion of attack beyond acts of violence, but simply recognizes that cyber operations triggering processes likely to directly cause death, injury or destruction are not only equivalent to, but constitute an integral part of, an "act of violence" within the meaning of article 49(1) of AP I.[112]

There is disagreement, however, as to whether the notion of attack also includes cyber operations aiming to merely capture or neutralize (that is, inhibit, hinder or hamper the proper exercise of its function)—rather than kill, injure or destroy—the target. The leading argument in favour of extending the effects-based interpretation of "attack" to cyber operations aiming to "neutralize" is that the treaty definition of military objectives in article 52(2) of AP I includes objects whose "capture and neutralization" would offer a definite military advantage and puts these two alternatives on the same level as total or partial destruction.[113] Those opposing this extension base themselves on a more literal interpretation of attacks as "acts of violence" and require that, if not the act itself, at least its consequences must be violent in order for it to be considered as an attack.[114] In support of their view they further point out that the principle of proportionality is formulated in terms of attacks causing "*loss* of civilian life, *injury* to civilians, *damage* to civilian objects, or a combination thereof"[115] but does not include capture or neutralization.[116]

While both arguments have their strong points, neither seems to provide an entirely satisfactory interpretation of the notion of attack in relation to cyber operations. On the one hand, it would hardly be convincing to exclude the non-destructive incapacitation of a state's air defence system or other critical military infrastructure from the notion of attack simply because it does not directly cause death, injury or destruction. On the other hand, it may well be exaggerated to extend the notion of attack to any denial of service attack against, for example, online shopping services, travel agents or telephone directories.[117]

---

111 Yoram Dinstein, "Computer Network Attacks and Self-Defense", in Michael Schmitt and Brian O'Donnell (eds), *Computer Network Attack and International Law*, 2002, p. 103; Michael Schmitt, "Wired Warfare: Computer Network Attack and Jus in Bello", *International Review of the Red Cross*, vol. 84, no. 846, 2002, p. 373; Michael Schmitt, "Cyber Operations and the Jus in Bello: Key Issues", *Naval War College International Law Studies*, 2011, pp. 6–7; Knut Dörmann, "The Applicability of the Additional Protocols to Computer Network Attacks: An ICRC Viewpoint", in Karin Byström (ed.), *International Expert Conference on Computer Network Attacks and the Applicability of International Humanitarian Law, 17–19 November 2004, Stockholm, Sweden*, Swedish National Defence College, 2004, text accompanying n. 15.

112 See also, in this respect, the discussion of direct participation in hostilities in relation to collective operations and preparatory measures in Nils Melzer, *Interpretive Guidance on the Notion of Direct Participation in Hostilities under IHL*, ICRC, 2009, pp. 54–55, 65–67.

113 See Knut Dörmann, "The Applicability of the Additional Protocols to Computer Network Attacks: An ICRC Viewpoint", in Karin Byström (ed.), *International Expert Conference on Computer Network Attacks and the Applicability of International Humanitarian Law, 17–19 November 2004, Stockholm, Sweden*, Swedish National Defence College, 2004, n. 16 and accompanying text.

114 Michael Schmitt, "Wired Warfare: Computer Network Attack and Jus in Bello", *International Review of the Red Cross*, vol. 84, no. 846, 2002, p. 377.

115 Arts. 51.5(b), and 57.2(a)(iii) and (b), AP I.

116 Michael Schmitt, "Wired Warfare: Computer Network Attack and Jus in Bello", *International Review of the Red Cross*, vol. 84, no. 846, 2002, p. 377. Michael Schmitt, "Cyber Operations and the Jus in Bello: Key Issues", *Naval War College International Law Studies*, 2011, pp. 5–8.

117 While Schmitt favours the restrictive interpretation, he recognizes the dilemma. See ibid., p. 7.

Although the term "attack" is a key notion of IHL, an analysis of the relevance of its rules on the conduct of hostilities for cyber operations cannot be limited to an examination of this notion. Recall, for example, that the basic treaty rule of distinction is not formulated in terms of "attacks" but in terms of "operations".[118] Similarly, treaty law protects the civilian population not only from direct attacks, but more generally from the "dangers arising from military operations"[119] and requires that, "[i]n the conduct of military operations, constant care shall be taken to spare the civilian population, civilians and civilian objects".[120] Also, at least for states party to AP I, the prohibition of perfidy applies not only for operations aiming to injure or kill, but also to those aiming to capture an adversary.[121] Most persuasive, however, is the fact that civilians lose their protection "for such time as they take a direct part in hostilities",[122] a notion that is generally considered to be wider than that of attack.[123] Therefore, although attacks certainly represent the predominant form of combat operation, it would be inaccurate to assume that cyber operations not amounting to an attack are not subject to IHL governing the conduct of hostilities. Accurately understood, the applicability of the restraints imposed by IHL on the conduct of hostilities to cyber operations depends not on whether the operations in question qualify as "attacks" (that is, the predominant form of conducting hostilities), but on whether they constitute part of the "hostilities" within the meaning of IHL.

V.3. Cyber Operations as "Hostilities" and "Direct Participation" therein

In the view of the ICRC, the concept of "hostilities" refers to the (collective) resort by the parties to the conflict to means and methods of injuring the enemy, and could be described as the sum total of all hostile acts carried out by individuals directly participating in hostilities.[124] In treaty IHL, the notion of "direct participation in hostilities" also describes the conduct which, if carried out by civilians, entails the suspension of their protection against direct attack.[125] Thus, for such time as civilian experts or individual hackers carry out cyber operations amounting to direct participation in hostilities, they are not only bound to comply with IHL governing the conduct of hostilities, but also become legitimate military targets just as if they were combatants. Moreover, civilians directly participating in hostilities do not have to be taken into account when taking precautions in attack, most notably with a view to avoiding or minimizing incidental harm (so-called "collateral damage").

According to the ICRC's official position, the notion of "direct participation in hostilities" goes beyond the notion of "attack" and includes not only the infliction of death, injury or destruction, but essentially any act likely to adversely affect the military operations or military capacity of a belligerent party (*threshold of harm*).[126] Additionally, in order

118 Art. 48, AP I.

119 Art. 51(1) and (3), AP I, and art. 13(1) and (3), AP II.

120 Art. 57(1), AP I.

121 Art. 37, AP I.

122 Art. 51(3), AP I, and art. 13(3), AP II.

123 Nils Melzer, *Interpretive Guidance on the Notion of Direct Participation in Hostilities under IHL*, ICRC, 2009, n. 97 with references and, more generally, pp. 47–50.

124 Ibid., pp. 43, 44.

125 Art. 51(3), AP I, and art. 13(3), AP II.

126 Nils Melzer, *Interpretive Guidance on the Notion of Direct Participation in Hostilities under IHL*, ICRC, 2009, pp. 47–48.

to constitute part of the hostilities within the meaning of IHL, the cyber operation in question must cause the required threshold of harm directly (*direct causation*), and it must also be designed to do so in support of a belligerent and to the detriment of another (*belligerent nexus*).[127] Whether the causal link between a specific operation and the resulting harm is "direct" or "indirect" depends, in essence, on whether it merely builds up the capacity of a belligerent party to harm the enemy (indirect) or whether it is an integral part of an operation using such capacity to actually inflict harm on the enemy (direct).[128] Accordingly, where cyber operations attributable to a belligerent party are designed to harm the adversary, either by directly causing death, injury or destruction, or by directly adversely affecting military operations or military capacity, such operations must be regarded as "hostilities" and, therefore, subject to all restrictions imposed by IHL on the choice and use of means and methods of warfare. If conducted by civilians, such operations also entail loss of protection against direct attacks.

In line with this interpretation, cyber operations aiming to disrupt or incapacitate an adversary's computer-controlled radar or weapons systems,[129] logistic supply or communication networks may not directly cause any physical damage, but would certainly qualify as part of the hostilities and, therefore, would have to comply with the rules and principles of IHL governing the conduct of hostilities.[130] The same would apply to cyber operations intruding into the adversary's computer network to delete targeting data, manipulate military orders, or change, encrypt, exploit, or render useless any other sensitive data with a direct (adverse) impact on the belligerent party's capacity to conduct hostilities.[131] Cyber operations causing neither death, injury or destruction, nor military harm, on the other hand, such as those conducted for the purposes of general intelligence gathering (no direct causation of harm), for purely criminal purposes or otherwise unrelated to the hostilities (no belligerent nexus), would fall short of the concept of "hostilities" and, thus, would not be governed by IHL on the conduct of hostilities and, if conducted by civilians, would not entail loss of protection against direct attacks.

The most difficult question that remains unresolved in this respect is whether "destruction" necessarily presupposes physical damage, particularly in the absence of military harm. In other words, while the non-destructive incapacitation of a military computer network would clearly amount to military harm and, thus, automatically also to "hostilities", the non-destructive incapacitation of a power station used exclusively for civilian purposes would cause neither military harm nor death, injury or destruction—unless the term "destruction" is interpreted as including harm other than physical damage. Again, the issue boils down to the dilemma between adopting either a too restrictive or a too permissive interpretation of the law. In the first case, even cyber operations causing the

127 Ibid., section V and commentary, pp. 46–64.

128 Ibid., pp. 52–53.

129 Ibid., pp. 47–50.

130 During the ICRC's clarification process on the notion of "direct participation in hostilities", the participating experts agreed that cyber operations directly causing military harm to the adversary in a situation of armed conflict amounted to direct participation in hostilities (see ICRC, *Third Expert Meeting on the Notion of Direct Participation in Hostilities*, summary report, 2005, p. 14).

131 According to the ICRC's Interpretive Guidance, examples of cyber operations qualifying as direct participation in hostilities would include electronic interference with military computer networks, whether through computer network attacks (CNA) or computer network exploitation (CNE), as well as wire-tapping the adversary's high command (Nils Melzer, *Interpretive Guidance on the Notion of Direct Participation in Hostilities under IHL*, ICRC, 2009,, p. 48).

incapacitation of major civilian electrical grids and communication networks could only qualify as part of the hostilities where they result in death, injury or physical destruction or military harm. In the second case, essentially any harm caused to the civilian population for reasons related to the conflict, including mere harassment or inconvenience, would have to be regarded as part of military hostilities, triggering not only the applicability of IHL on the conduct of hostilities, but also the loss of civilian protection for all those directly involved.

V.4. Targeting in Cyberspace

At the heart of IHL lies the principle of distinction, which requires belligerent parties to always distinguish between legitimate military targets and persons and objects protected against attack, and to direct their operations only against the former.[132] Derived from the principle of distinction, and indispensable for its faithful implementation, are the prohibition of indiscriminate attack and the requirements of precaution and proportionality.

V.4.1. Persons

As far as persons are concerned, legitimate military targets include, most notably, combatants, members of organized armed groups and civilians directly participating in hostilities, whereas civilians, medical and religious personnel and combatants *hors de combat* due to wounds, sickness, capture, surrender or any other reason must be spared and protected. While the implementation of the principle of distinction with regard to persons can pose significant practical difficulties, particularly with regard to the identification of decisive factors such as "direct participation in hostilities" and "membership" in irregularly constituted armed forces or groups, most of these problems are not cyber-specific and have been discussed in more detail elsewhere.[133] Aspects which, nevertheless, may require particular attention include the question of how targeting-relevant factors such as a group's "organization" or "membership" therein should be interpreted in cyberspace, where persons may act collectively without lasting affiliation or hierarchical command structure.[134] Also, how does the obligation of combatants "to distinguish themselves from the civilian population while they are engaged in an attack or in a military operation preparatory to an attack"[135] play out in cyberspace? Does it require hackers to wear uniforms even when far removed from the physical battlefield, or does it mean that their operations have to be recognizable as military operations to the adversary? How does this obligation relate to the distinction between (permitted) ruses of war and (prohibited) perfidy on the battlefield?[136] It is clear that these and other questions need urgent clarification if civilians exposed to

132 Art. 48, AP I; Jean-Marie Henckaerts and Louise Doswald-Beck, *Customary International Humanitarian Law*, vol. I, 2005, rule 1.

133 See, for example, Nils Melzer, *Interpretive Guidance on the Notion of Direct Participation in Hostilities under IHL*, ICRC, 2009. For critiques of the Interpretive Guidance and the ICRC's response, see also the series of articles contributing to the forum "The ICRC Interpretive Guidance on the Notion of Direct Participation in Hostilities Under International Humanitarian Law" in *Journal of International Law and Politics*, vol. 42, no. 3, 2010.

134 See also Michael Schmitt, "Cyber Operations and the Jus in Bello: Key Issues", *Naval War College International Law Studies*, 2011, p. 10.

135 Art. 44(3), AP I.

136 See section V.5.

cyberwarfare are to receive the protection they are entitled to under treaty and customary law. In the meantime, it may have to suffice to recall that, in case of doubt, any person must be presumed to be a civilian and, as such, as protected against direct attack.[137]

V.4.2. Objects

In so far as objects are concerned, "military objectives are limited to those objects which by their nature, location, purpose or use make an effective contribution to military action and whose total or partial destruction, capture or neutralization, in the circumstances ruling at the time, offers a definite military advantage".[138] Again, the challenge lies with the concrete implementation of this definition in cyberspace rather than with the adequacy of its basic elements.

First, it is characteristic for cyberspace that it relies heavily on civilian infrastructure spread and networked across the entire planet and that civilian and military cyber-infrastructure is tightly interconnected. Even more than in traditional warfare, therefore, military objectives in cyberspace are likely to be "dual use" objects. Also, the share or proportion of military versus non-military use of civilian cyber-infrastructure not only fluctuates, but also remains far below levels typical for traditional infrastructure such as power generating and distributing installations, industrial plants, or transport infrastructure such as ports, airports, roads and bridges. While this does not represent an absolute obstacle against attacking such objects, it requires a high level of precaution in identifying legitimate targets, as well as a comparatively sophisticated capability both of the attacker and the attacked for assessing, avoiding or controlling incidental harm likely to be inflicted on the civilian infrastructure and population.[139]

One of the most obvious problems in view of the prohibition of indiscriminate attacks is, of course, the question to what extent malware intended to damage military systems can be prevented from spreading to civilian infrastructure and causing havoc among the civilian population.[140] Even if the collateral effects can be controlled it may be asked to what extent it would be justified, for example, to incapacitate a domain name server directing global internet traffic, or to destroy a major intercontinental submarine cable, in order to prevent their use for hostile cyber operations if more than 90% of the data transmitted are of civilian nature and the consequences for global trade, traffic and communication would be debilitating.[141] It must also be recognized that the use of civilian cyber-infrastructure for military purposes may follow unpredictable patterns in terms of place (choice and location) and time (frequency and timing), thus making it extremely difficult to determine with sufficient precision and reliability the installations which

137 Art. 50(1), AP I. See also the broader discussion on the presumption of civilian protection in Nils Melzer, *Interpretive Guidance on the Notion of Direct Participation in Hostilities under IHL*, ICRC, 2009, section VIII.

138 Art. 52(2), AP I.

139 According to arts. 57 and 58, AP I.

140 According to art. 51(4), AP I, 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".

141 The definition of indiscriminate attack also includes those "which may be expected to cause incidental loss of civilian life, injury to civilians, damage to civilian objects, or a combination thereof, which would be excessive in relation to the concrete and direct military advantage anticipated"; art. 51(5)(b), AP I.

"make an effective contribution to military action" and the moment at which their "total or partial destruction, capture or neutralization … offers a definite military advantage".

Another key issue to be resolved is the question as to whether data, as such, constitutes an "object" within the meaning of IHL and, if so, what threshold of damage, modification, manipulation or interference would be required for the prohibition of attacks against civilian objects to be violated. The point is that virtually no cyber operation—not even espionage through computer network exploitation (CNE) or manipulations as simple as entering a password—can be carried out without at least temporarily deleting or changing data resident in the intruded systems. The correct answer is probably that, for the purposes of targeting, data should be regarded as an object which may not be directly targeted unless it fulfils all defining elements of a military objective.[142] The unavoidable (but incidental) deletion or modification of civilian data in the course of an operation pursuing a different aim, on the other hand, must be factored into the proportionality assessment, where the potentially temporary or minute nature of the inflicted harm can duly be taken into account.

It is therefore important to distinguish the actual aim of the operation from its incidental side effects. For example, the deletion or modification of civilian data in the course of an attack against military cyber-infrastructure would be equivalent to the kinetic causation of so-called "collateral damage". The manipulation or modification of access data to a civilian computer system in the course of an espionage or reconnaissance operation, on the other hand, could perhaps be compared to breaking the door or mailbox of a civilian house in the course of a search operation, but would not constitute an "attack" within the meaning of article 49 of AP I because neither the nature and effects nor the aim of the operation as such is equivalent to that of an "act of violence".[143] More difficult are examples such as the deletion or manipulation of data with the aim of disrupting civilian television broadcasts, which may be regarded as lawful by some,[144] whereas others would likely condemn it as a direct attack against a civilian object.[145]

The problem is that non-destructive measures such as blockades, border closures and economic sanctions, despite their potentially significant impact on the civilian population, are not as such prohibited under IHL, whereas direct attacks against civilian objects are outlawed regardless of the severity of the ensuing destruction. A possible clue towards an equitable solution may be the fact that, in certain circumstances, existing IHL nevertheless permits the intentional destruction of civilian property to the extent that it is rendered absolutely necessary for the purposes of military operations.[146] It is clear that

142 But see the rejection of this view in Michael Schmitt, "Cyber Operations and the Jus in Bello: Key Issues", *Naval War College International Law Studies*, 2011, p. 8.

143 Art. 49(1), AP I. As shown below, this does not exclude that such operations, as well as the destruction caused in their course, may still amount to an internationally wrongful act.

144 Michael Schmitt, "Cyber Operations and the Jus in Bello: Key Issues", *Naval War College International Law Studies*, 2011, p. 8; Michael Schmitt, "Wired Warfare: Computer Network Attack and Jus in Bello", *International Review of the Red Cross*, vol. 84, no. 846, 2002, p. 381.

145 Knut Dörmann, "The Applicability of the Additional Protocols to Computer Network Attacks: An ICRC Viewpoint", in Karin Byström (ed.), *International Expert Conference on Computer Network Attacks and the Applicability of International Humanitarian Law, 17–19 November 2004, Stockholm, Sweden*, Swedish National Defence College, 2004, n. 16, with reference, and accompanying text.

146 See art. 23(g), H IV R, which prohibits to "destroy or seize the enemy's property, unless such destruction or seizure be imperatively demanded by the necessities of war". See also art. 53, GC IV, which extends the same principle to the destruction of private property, albeit only for occupied territories.

any compromise norm which states may develop for cyberspace along these lines would imperatively have to include a proportionality assessment weighing the expected benefit of the operation against the harm inflicted by the deletion, modification or manipulation of civilian data.[147]

Other cyber-specific problems which need to be addressed include the question of how the computer-controlled systems of medical installations, transports and logistics such as hospitals, ambulances, ships and aircraft could be marked so as to ensure they are respected and appropriately protected from infection with military malware and other hostile cyber operations.[148] Similar problems also arise with regard to other specially protected objects (such as works and installations containing dangerous forces,[149] objects indispensable to the survival of the civilian population,[150] cultural objects and of places of worship[151] and the natural environment[152]) and areas (most notably non-defended localities[153] and de-militarized zones[154]).

V.5. Ruses and Perfidy (Good Faith) in Cyberspace

As has rightly been pointed out, the specific characteristics of cyberspace invite a plethora of opportunities and techniques to deceive the enemy with false information.[155] For example, belligerents can disguise the origin of their operations through the use of botnets or techniques like IP spoofing,[156] camouflage combat troops or vehicles as medical transports by using internationally recognized protective signals,[157] manipulate the enemy's reconnaissance data so as to wrongly make him believe that the opposing forces intend to surrender,[158] or even send seemingly innocent civilian email attachments to individual recipients at a military headquarters, causing them to inadvertently infect the computer system with malware.[159]

147 See also, for example, the commentary to art. 53, GC IV, which states: "whenever it is felt essential to resort to destruction, the occupying authorities must try to keep a sense of proportion in comparing the military advantages to be gained with the damage done"; Jean Pictet (ed.), *Commentary on the Geneva Conventions of 12 August 1949 relative to the Protection of Civilian Persons in Time of War*, ICRC, 1956 .

148 Art. 12(1), AP I.

149 Namely dams, dykes and nuclear electrical generating stations (art. 56, AP I).

150 For example, foodstuffs, agricultural areas for the production of foodstuffs, crops, livestock, drinking water installations and supplies and irrigation works (art. 54, AP I).

151 Art. 53, AP I.

152 Only in case of exposure to wide-spread, long-term and severe damage (art. 55, AP I).

153 Art. 59, AP I.

154 Art. 60, AP I.

155 See Knut Dörmann, "The Applicability of the Additional Protocols to Computer Network Attacks: An ICRC Viewpoint", in Karin Byström (ed.), *International Expert Conference on Computer Network Attacks and the Applicability of International Humanitarian Law, 17–19 November 2004, Stockholm, Sweden*, Swedish National Defence College, 2004, text accompanying nn. 46–56; Michael Schmitt, "Wired Warfare: Computer Network Attack and Jus in Bello", *International Review of the Red Cross*, vol. 84, no. 846, 2002, p. 395.

156 Marco Roscini, "World Wide Warfare—Jus ad bellum and the Use of Cyber Force", in Armin Bogdany and Rüdiger Wolfrum (eds.), *Max Planck Yearbook of United Nations Law*, vol. 14, 2010, p. 96.

157 Art. 11, annex 1, AP I (as amended on 30 November 1993). Michael Schmitt, "Wired Warfare: Computer Network Attack and Jus in Bello", *International Review of the Red Cross*, vol. 84, no. 846, 2002, p. 395; Knut Dörmann, "The Applicability of the Additional Protocols to Computer Network Attacks: An ICRC Viewpoint", in Karin Byström (ed.), *International Expert Conference on Computer Network Attacks and the Applicability of International Humanitarian Law, 17–19 November 2004, Stockholm, Sweden*, Swedish National Defence College, 2004, text accompanying n. 52.

158 Ibid., text accompanying n. 51.

159 Ibid., text following n. 56.

An important first distinction must be made between (permitted) ruses of war and (prohibited) perfidy. Ruses of war are defined as "acts which are intended to mislead an adversary or to induce him to act recklessly but which infringe no rule of (IHL) and which are not perfidious".[160] Prohibited perfidy, on the other hand, refers to the killing, injuring or capturing of an adversary by leading him to believe that he is entitled to, or is obliged to accord, IHL protection, and subsequently betraying that confidence. Treaty examples of perfidy include: (a) the feigning of an intent to negotiate under a flag of truce or of a surrender; (b) the feigning of an incapacitation by wounds or sickness; (c) the feigning of civilian, non-combatant status; and (d) the feigning of protected status by the use of signs, emblems or uniforms of the United Nations or of neutral or other states not parties to the conflict.[161] It should be recalled, however, that IHL prohibits the resort to perfidy only in connection with the killing, injuring or capturing of an adversary. Cyber operations limited to causing physical or functional damage to infrastructure and other forms of disruption or incapacitation, even if conducted by resort to perfidious deception, would not come under this prohibition.

More likely to be relevant for cyber operations, therefore, are the more broadly crafted prohibitions on misusing internationally recognized protective emblems (for example, Red Cross and Red Crescent Movement, flag of truce, cultural property),[162] the emblem of the United Nations[163] and the flags or military emblems, insignia or uniforms of neutral or other states not parties to the conflict.[164] Moreover, it is prohibited to use the flags or military emblems, insignia or uniforms of the adversary while engaging in attacks or in order to shield, favour, protect or impede military operations.[165] This would clearly outlaw any hostile cyber operation pretending to originate from a non-belligerent state, the ICRC or the United Nations, as well as attacks disguising themselves as operations conducted by "friendly forces" (that is, the attacked state or his co-belligerents).

V.6. The Status of Cyber Warriors

V.6.1. Combatants

Cyber operations are generally carried out by highly specialized personnel. To the extent that they are members of the armed forces of a belligerent state, their status, rights and obligations are no different from those of traditional combatants. According to treaty IHL, the armed forces of a belligerent state comprise all organized armed forces, groups and units which are under a command responsible to that state for the conduct of its subordinates.[166] This broad and functional concept of armed forces includes essentially

160 Art. 37(2), AP I.

161 Art. 37(1), AP I.

162 Art. 38(1), AP I.

163 Art. 38(2), AP I.

164 Art. 39(1), AP I.

165 Art. 39(2), AP I.

166 Art. 43, AP I, defines the armed forces as follows: "1. The armed forces of a Party to a conflict consist of all organized armed forces, groups and units which are under a command responsible to that Party for the conduct of its subordinates, even if that Party is represented by a government or an authority not recognized by an adverse Party. Such armed forces shall be subject to an internal disciplinary system which,' *inter alia*', shall enforce compliance with the rules of international law applicable in armed conflict. 2. Members of the armed forces of a Party to a conflict (other than medical personnel and chaplains …) are combatants, that is to say, they have the right to participate directly in hostilities."

all armed actors belonging to a belligerent state and showing a sufficient degree of military organization.[167]

V.6.2. Contractors and Civilian Employees

In recent decades, belligerent states have increasingly employed private contractors and civilian employees in a variety of functions traditionally performed by military personnel. Today, this also includes the support, preparation and conduct of cyber operations. As long as such personnel assume functions not amounting to direct participation in hostilities, they remain civilians and, if formally authorized to accompany the armed forces in an international armed conflict, are even entitled to prisoner-of-war status in case of capture.[168] However, where private contractors or civilian employees are expressly authorized by a state to directly participate in hostilities on its behalf, they become organized armed actors fighting on behalf of that state and, *de facto*, irregular members of its armed forces.[169] As such, they lose civilian status and, as long as they fulfil the so-called "four requirements", are entitled to combatant privilege and prisoner of war status.[170]

V.6.3. Levée en masse

The term "levée en masse" refers to the inhabitants of a non-occupied territory who, on the approach of the enemy, spontaneously take up arms to resist the invading forces without having had time to form themselves into regular armed units, provided they carry arms openly and respect the laws and customs of war.[171] Participants in a *levée en masse* are the only armed actors who are entitled not only to prisoner-of-war status, but also to the combatant privilege although, by definition, they operate spontaneously and lack sufficient organization and command to qualify as members of the armed forces. While this category of persons has become ever less relevant in traditional warfare, it may well come to be of practical importance in cyber warfare. Indeed, in cyber warfare, territory is neither invaded nor occupied, which may significantly prolong the period during which a *levee en masse* can operate. Also, cyber space provides an ideal environment for the instigation and non-hierarchical coordination of spontaneous, collective and unorganized cyberdefence action by great numbers of "hacktivists". The only question is, of course, how the requirement to "carry their arms openly" should be interpreted in cyber space. From a teleological perspective, a possible solution would be to consider this requirement as fulfilled when cyber operations are not conducted by feigning protected, non-combatant status within the meaning of the prohibition of perfidy.[172]

---

167 See also Nils Melzer, *Interpretive Guidance on the Notion of Direct Participation in Hostilities under IHL*, ICRC, 2009, p. 22.

168 Arts. 4(4) and (5), GC III.

169 Nils Melzer, *Interpretive Guidance on the Notion of Direct Participation in Hostilities under IHL*, ICRC, 2009, p. 39.

170 For members of irregular armed forces, entitlement to combatant privilege and prisoner-of-war status depends on meeting the following "four requirements": (a) that of being commanded by a person responsible for his subordinates; (b) that of having a fixed distinctive sign recognizable at a distance; (c) that of carrying arms openly; (d) that of conducting their operations in accordance with the laws and customs of war (art. 1(1), H IV R; art. 4 A.(2), GC III).

171 Art. 2, H IV R; art. 4(6), GC III.

172 Art. 37(1), AP I.

V.6.4. Members of Organized Armed Groups

In IHL governing non-international armed conflict, organized armed groups constitute the armed forces (that is, the armed wing) of a non-state belligerent and must not be confused with the belligerent party itself (for example, an insurgency as a whole, including its political or administrative wing) or with other supportive segments of the civilian population. Treaty IHL governing non-international armed conflict uses the terms civilian, armed forces and organized armed group without defining them. It is generally recognized, however, that members of state armed forces do not qualify as civilians,[173] and the wording and logic of article 3, GC I–IV, and AP II suggest that the same applies to members of organized armed groups.

Civilians may support a non-state party in various ways and may even directly participate in hostilities on a spontaneous, sporadic or unorganized basis. However, they cannot be regarded as members of an organized armed group unless it is their function to directly participate in hostilities on behalf of the non-state party. Such a combat function does not imply entitlement to combatant privilege, prisoner of war status, or any other form of immunity from domestic prosecution for lawful acts of war. Rather, it makes a strictly functional distinction between members of the organized fighting forces and the civilian population. For the present context this means that individuals conducting cyber operations on behalf of a non-state party lose their civilian status and become members of that party's "armed forces" only if their operations are conducted on a continuous basis and amount to direct participation in hostilities.[174]

V.6.5. Civilians

In IHL, the concept of civilian encompasses all persons who are neither members of the armed forces of a state or non-state party to an armed conflict, nor participants in a *levée en masse.* As civilians, they are entitled to protection against the dangers arising from military operations and, most notably, against attack. In cyberwarfare, this category is likely to include most non-state hackers not belonging to the military wing of an organized armed group. If and for such time as their operations amount to direct participation in hostilities, civilians lose their protection and may be directly attacked as if they were combatants. Contrary to combatants, however, they do not benefit from immunity from prosecution for lawful acts of war (so-called "combatant privilege") and, therefore, can be punished by their captor for any violation of national law. Civilians deprived of their liberty, including those having directly participated in hostilities, are entitled to humane treatment and fair trial guarantees as reflected in the various applicable instruments of IHL.[175]

---

173 See Jean-Marie Henckaerts and Louise Doswald-Beck, *Customary International Humanitarian Law*, vol. I, 2005, p. 19.

174 For the ICRC's position on this issue see Nils Melzer, *Interpretive Guidance on the Notion of Direct Participation in Hostilities under IHL*, ICRC, 2009, § II. See also *Prosecutor v. Martić*, judgement of 8 October 2008, ICTY, §§ 300–302.

175 In international armed conflict, civilians deprived of their liberty are protected by the GC IV, AP I and customary law, whereas in non-international armed conflict these protections are reflected in art. 3 common to the Geneva Conventions, AP II and customary law. Depending on the context, human rights law may additionally be relevant.

VI. Conclusions

As has been shown, as far as international law is concerned, the phenomenon of cyberwarfare does not exist in a legal vacuum, but is subject to well established rules and principles. That being said, transposing these pre-existing rules and principles to the new domain of cyberspace encounters certain difficulties and raises a number of important questions. Some of these questions can be resolved through classic treaty interpretation in conjunction with a good measure of common sense, whereas others require a unanimous policy decision by the international legislator, the international community of states. It has been attempted in this paper to identify the most important of these questions and to make suggestions as to possible avenues for their resolution. For the time being, cyberwarfare has not had dramatic humanitarian consequences, and it is to be hoped that this state of affairs will not change in the future. The potential for human tragedy, however, is already enormous, and it is likely to increase with our growing dependence on computer-controlled systems to sustain our daily lives. It is all the more important, therefore, that states be aware not only of their legal duty to examine whether new weapons and methods employed in cyberwarfare would be compatible with their obligations under existing IHL,[176] but also of their moral responsibility towards generations to come.

176 Art. 36, AP I.

UNIDIR RESOURCES

About UNIDIR

The United Nations Institute for Disarmament Research (UNIDIR)—an autonomous institute within the United Nations—conducts research on disarmament and security. UNIDIR is based in Geneva, Switzerland, the centre for bilateral and multilateral disarmament and non-proliferation negotiations, and home of the Conference on Disarmament. The Institute explores current issues pertaining to the variety of existing and future armaments, as well as global diplomacy and local tensions and conflicts. Working with researchers, diplomats, government officials, NGOs and other institutions since 1980, UNIDIR acts as a bridge between the research community and governments. UNIDIR's activities are funded by contributions from governments and donor foundations.