
Book Section:
Tsagourias, N. orcid.org/0000-0002-1701-2572 and Biggio, G. (2022) The regulation of cyber weapons. In: Myjer, E.P.J. and Marauhn, T., (eds.) Research Handbook on International Arms Control Law. Research Handbooks in International Law series . Edward Elgar , pp. 440-455. ISBN 9781788111898

https://doi.org/10.4337/9781788111904.00042

The Regulation of Cyber Weapons

Nicholas Tsagourias* and Giacomo Biggio**

[forthcoming in Eric Myer and Thilo Marauhn, *Research Handbook on Arms Control Law,* Elgar, 2021]

I. Introduction

Cyber weapons can be fielded only if they comply with international humanitarian law (IHL). However, the particular features of cyber weapons such as their non-physical nature and the non-physical consequences they produce have inevitably raised questions about the ability of long-standing IHL principles and rules to regulate their use. This chapter will not consider the application of IHL to the use of cyber weapons in the course of an armed conflict and, more specifically, whether their use can comply with the rules governing the conduct of hostilities but it will consider the earlier question of how cyber weapons can be made IHL compliant before they are fielded that is, at the stage of study, procurement, acquisition or adoption. It will also consider the related but broader question of how cyber armaments or the development and acquisition of certain cyber weapons can be regulated. The first issue directs our attention to the weapons review mechanism introduced by Article 36 of Additional Protocol (I) to the Geneva Conventions (API)[1] whereas the latter to forms of regulation ranging from total or partial ban of cyber weapons to restrictions in the production or sale of certain of their components. All these mechanisms are part and parcel of the broader regulatory regime applicable to cyber weapons.

*Professor of International Law, University of Sheffield
** Ph.D student, University of Sheffield

[1] Protocol Additional to the Geneva Conventions of 12 August 1949 and relating to the Protection of Victims of International Armed Conflicts (Protocol I), 1125 UNTS 3 (hereinafter 'API'), Art. 36. See also ICRC, *A Guide to the Legal Review of Weapons, Means and Methods of Warfare* (ICRC: Geneva, 2006) (hereinafter 'ICRC, Guide').

II. Article 36 and legal reviews of cyber weapons

Article 36 API introduces a review mechanism to determine that new weapons are IHL compliant before they are actually fielded. According to this article:

"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 this Protocol or by any other rule of international law applicable to the High Contracting Party'.

Before we examine the application of Article 36 to cyber weapons, it is important to say a few words about its legal status. The obligation to review weapons is a treaty-based obligation binding States-parties to API, but it is the case that not all States-parties to API carry out weapons reviews. Whether it represents a customary law obligation is debated, not only because it was omitted from the ICRC study on customary humanitarian law[2] but also because state practice, as was noted above, is not 'general and widespread', even if certain important states that are not parties to API (such as the US and Israel) conduct weapons reviews. Against this, it can be said that the obligation to conduct weapons reviews derives from the treaty and customary law obligation to 'respect and ensure respect of international humanitarian law'[3] as well as from the customary humanitarian law principles that apply to the means and methods of warfare.[4] As the ICJ said in its 1996

---

[2]Henckaerts, J-M., Doswald-Beck, L, *Customary International Humanitarian Law*, Volume I. Rules (Cambridge University Press 2005) available at https://www.icrc.org/en/doc/assets/files/other/customary-international-humanitarian-law-i-icrc-eng.pdf.

[3] Geneva Conventions, Common Article 1; Also Sandoz,,Y., Swinarski,C., Zimmermann, B. (eds.), Commentary on the Additional Protocols of 8 June 1977 to the Geneva Conventions of 12 August 1949, ICRC, 1987, Article 1 Additional Protocol I, para 41 https://ihl-databases.icrc.org/applic/ihl/ihl.nsf/Comment.xsp?action=openDocument&documentId=7125D4CBD57A70DDC12563CD0042F793

[4] API, Art. 35. ICRC, *Customary International Humanitarian Law,* Rules 70 and 71. ICRC, *Guide,* p. 4

Advisory Opinion on the *Legality of the Threat or Use of Nuclear Weapons*, the established principles and rules of humanitarian law apply to "all forms of warfare and to all kinds of weapons, those of the past, those of the present and those of the future".[5] Regardless of whether it is a treaty-based and/or a customary law obligation, the scope of the relevant obligation is quite broad and applies to states that manufacture new weapons but also to states that acquire them and requires 'a standing mechanism that can be automatically activated at any time that a state is developing or acquiring a new weapon'.[6]

We will now examine in more detail the requirements of Article 36 API as applied to cyber weapons. First, reviews should be conducted at the stage of the 'study, development, acquisition or adoption' of new weapons. Ensuring compliance with IHL at the acquisition or adoption stage of weapons is self-evident since their employment is the next logical step but not so at the study and development stage because actual use depends on many variables. For this reason, it is submitted that such review should take place at the most expedient moment in the cycle when the state actually procures new cyber weapons for deployment but reviews can be conducted at different stages as well and with different levels of intensity.[7] Second, reviews are required for 'new' cyber weapons. This means that they are required for future cyber weapons or for cyber weapons acquired for the first time but also for existing cyber weapons with modified or updated components or functions.[8] This is particularly important in the case of cyber weapons because they are designed to exploit

---

[5] International Court of Justice, *Legality of the Threat or Use of Nuclear Weapons,* Advisory Opinion, 8 July 1996, [1996] I.C.J. Rep. 226, para. 86

[6]Kathleen Lawand, 'Reviewing the Legality of New Weapons, Means and Methods of Warfare', *International Review of the Red Cross*, vol. 88, no. 864 (December 2006), 925–30 927. *Commentary on the Additional Protocols of 8 June 1977 to the Geneva Conventions of 12 August 1949,* p. 428, para 1476

[7] UK, Ministry of Defence, Development, Concepts and Doctrine Centre, *UK Weapon Reviews* (2016),available at https://assets.publishing.service.gov.uk/government/uploads/system/uploads/attachment_data/file/507319/20160308-UK_weapon_reviews.pdf.

[8] ICRC, *Commentary on the Additional Protocols of 8 June 1977 to the Geneva Conventions of 12 August 1949* pp. 427 – 428, para. 1476 – 1478; ICRC, *Guide*, pp. 10, 23 – 24

particular vulnerabilities and, therefore, they may need to be updated or, generally, to be modified in order to maintain their function. The immediate question is whether they need to be reviewed after each and every modification. In our opinion, only if the modification or updating alters significantly the properties or functions of a cyber weapon Article 36 review is required, otherwise operational review by the commander before it is used will be sufficient.[9]

Third, and perhaps most critically, the obligation to review concerns 'weapons', 'means' and 'methods' which raises questions of definition because how they are defined will determine the object of review. 'Method' refers to the way the weapon is used in a structural rather than a tactical sense, whereas 'weapon' refers to an instrument or device that causes harm. According to the ICRC, weapons are 'means to commit acts of violence against human or material enemy forces'.[10] The focus thus is on the produced effects which must be violent regardless of 'the mechanisms through which they produce destruction or damage',[11] whether they are kinetic or cyber. 'Means' seems to refer to the broader category of capabilities, but this may expand the scope of review exponentially. For this reason, the terms 'weapons' and 'means' should be read together in order to delineate the object of the review. This is the approach adopted by the Tallinn Manual according to which cyber weapons are cyber means of warfare, to wit, any cyber device, materiel, instrument, mechanism, equipment, or software which are designed, used, or intended to be used to cause either (i) injury to, or death of, persons; or (ii) damage to, or destruction of objects, in

[9] See API, Art. 82.

[10] *Customary International Humanitarian Law,* Volume I, Rule 6, p. 23. See also HPCR, *Manual on International Law Applicable to Air and Missile Warfare* (Cambridge University Press 2013), p. 49, according to which the main characteristic of a weapon is to cause either injury or death of persons or damage or destruction of objects.

[11] Roscini, M., *Cyber Operations and The Use of Force in International Law* (Oxford University Press 2014) p. 50.

other words produce the consequences required for the qualification of a cyber attack. [12] The Tallinn Manual rightly broadens the definition of weapons but then aligns the definition of cyber weapons with that of conventional weapons by focusing on the physical consequences they produce. Any external physical effects will not however be the direct and immediate consequence of the use of cyber weapons and therefore the indirect physical consequences should be taken into account. These include, according to the Tallinn Manual, any reasonably foreseeable consequences.[13] That said, cyber weapons may not produce physical effects but remove the functionality of a system. The Tallinn Manual has adopted the view that interference with the functionality of a system counts as damage if, and only if, replacement of the components or reinstallation of the system is required, whereas the mere disruption, deletion or alteration of digital data which may affect the functionality of the system does not suffice.[14] This, in our opinion, is too limited and just replicates the physical damage requirement for conventional weapons.

The difference can be illustrated by comparing the Stuxnet virus with the Shamoon virus. The Shamoon virus targeted the Saudi-owned oil company Saudi Aramco, deleting data from more than 30,000 workstations. According to the Tallinn Manual's definition, Shamoon would fall outside the scope of the definition, as it did not result in the physical destruction of objects but in the mere deletion of digital data. Consequently, it will not be the object of review. Regarding Stuxnet, it targeted the uranium-enrichment facility in Natanz, Iran, and resulted in the destruction of some two thousand turbines by interfering, *inter alia*, with their rotatory speed. Stuxnet manipulated the system and took control of the system, in this

[12] Schmitt, M.N., *Tallinn Manual 2.0 on the International Law Applicable to Cyber Operations* (Cambridge University Press 2017), Rule 103, p. 452, in particular para 2.
[13] Tallinn Manual Rule 92, p. 416, para 5.
[14] Tallinn Manual Rule 92, pp. 417-8, paras 10-13

way it did not cause any internal destruction to the system itself however it caused external destruction. That destruction was the indirect (second order) consequence of manipulating the rotation system and in fact it was the normal, expected and intended purpose of Stuxnet. The Stuxnet malware can thus qualify as a weapon and therefore be the object of review.

In view of the above, we submit that cyber weapons are cyber means (devices, materiel, instruments, mechanisms, equipment, or software) that are designed, used or intended to be used to cause death or injury to humans; destroy, capture or neutralise objects; or incapacitate humans or objects.[15] This definition is in line with the ICRC's view that non-lethal weapons should also be reviewed[16] as well as with the definitions of cyber weapons adopted by certain states.[17] The aforementioned definition is also aligned with the type of methods for which cyber weapons are used. Cyber weapons are used to deny access to a system or to disrupt a system without causing damage as for example in a DDoS attack; they are used to penetrate a system and access data in order to incapacitate or disable the system or to destroy and degrade the data; they are used to manipulate the system or alter the data in order to cause external damage as in the case of Stuxnet or cause injury and death. It should also be recalled that cyber weapons such as malware have multiple properties and can perform different actions (propagation, access, exploitation, execution of payload) separately or in combination using the same underlying technology but each action

[15] This definition is also in line with the definition of cyber capabilities which is any device or software payload intended to disrupt, deny, degrade, negate, impair or destroy adversarial computer systems, data, activities or capabilities. Secretary of the Air Force, Air Force Instruction 51-402, Legal Reviews of Weapons and Cyber Capabilities, 27 July 2011, Attachment 1. https://nsarchive2.gwu.edu/NSAEBB/NSAEBB424/docs/Cyber-053.pdf

[16] ICRC, *Guide,* p. 9.

[17] According to the US Department of the Air Force, weapons are 'devices designed to kill, injure, disable or temporarily incapacitate people, or destroy, damage or temporarily incapacitate property or materiel'. Secretary of the Air Force, Air Force Instruction 51-402, Legal Reviews of Weapons and Cyber Capabilities, 27 July 2011, Attachment 1. https://nsarchive2.gwu.edu/NSAEBB/NSAEBB424/docs/Cyber-053.pdf

being integral to the ultimate function of the cyber weapon. Disaggregating or isolating the specific technological properties required for each action or the actions themselves is extremely challenging. For these reasons, the aforementioned definition captures adequately the nature of cyber weapons.

Having provided a definition of what is a cyber weapon, the aim of the review is to ascertain the inherent lawfulness of the cyber weapon and the lawfulness of its normal, expected and intended use. This is done against the international law rules (treaty or customary) that bind the reviewing state and against the rules contained in API. Regarding the former, it includes specific treaties prohibiting or restricting particular weapons such as disarmament, arms control or trade in arms treaties (an issue discussed in the third section of this chapter) as well as customary law rules on the means and methods of warfare. In this regard it should be noted that certain states take into consideration likely future developments of the law[18] but this may not be relevant as far as cyber weapons are concerned as the next section will demonstrate. Whether cyber weapons should be assessed against IHRL is a rather vexed question and not all states review weapons against IHRL.[19] Although the formulation in Article 36 API is quite broad, in our opinion human rights should be taken into account only to the extent that they relate to weapons and by taking into account the fact that IHL is *lex specialis*. If cyber weapons are to be used extensively for law enforcement purposes, the argument that they should be reviewed under IHRL is more convincing.

The latter set of obligations includes rules found in API which prohibit weapons that are (1) of a nature to cause superfluous injury or unnecessary suffering; (2) indiscriminate by

[18] UK, British Ministry of Defence, *The Joint Service Manual of the Law of Armed Conflict*, (Joint Service Publication 383, 2004 Edition), 6.20.1

[19] Casey-Maslen, S., Corney, N. and Dymond-Bass, A., 'The Review of Weapons Under International Humanitarian Law and Human Rights Law', in Casey-Maslen, S. (ed.), *Weapons Under International Human Rights Law* (Cambridge University Press: Cambridge, 2014), 411

nature because they cannot be aimed at a lawful target or because their effects cannot be limited as required by IHL; (3) intended, or may be expected, to cause widespread, long-term and severe damage to the natural environment as discussed in this section.

In what follows, we will explain in more detail how this set of obligations applies to cyber weapons. First, the review needs to ascertain whether a cyber weapon by its nature or normal, expected and intended use will cause superfluous injury or unnecessary suffering.[20] This is injury or suffering inflicted on those attacked (combatants, members of armed groups or civilians directly participating in hostilities) that serves no military purpose or is clearly excessive compared to the military purpose for which the cyber weapon is normally intended to be used. As the ICJ said, it is 'harm greater than that unavoidable to achieve legitimate military objectives'.[21] Such judgement is difficult to be made in advance; for this reason, the review should rely on certain objective factors based on scientific, medical and health-related evidence as well as on subjective factors related to the military advantage such weapons can bring about.

In the case of cyber weapons, this requirement becomes relevant only if their indirect effects are taken into account. That having been said, cyber weapons – compared to conventional weapons - may be more 'humane'. For example, if a military command and control system is disabled instead of physically destroying the building, this would cause less suffering. For these reasons, this requirement may have little relevance in the case of cyber weapons.

[20] AP I, Art 35 (2).
[21] *Legality of the Threat or Use of Nuclear Weapons* (Advisory Opinion), para 78.

Second, cyber weapons need to be assessed against the principle of distinction which is a 'cardinal' principle of IHL.[22] The question here is, first, whether the cyber weapon can be directed against an individual military objective, in other words whether it is able to distinguish combatants and military objectives from civilians and civilian objects[23] and, second, whether its reverberating effects can be controlled in order not to cause excessive harm to civilians compared to the military advantage rendered by using the cyber weapon. The principle of discrimination as far as weapons is concerned thus combines the principle of distinction and the principle of proportionality in its general and abstract dimension than in the specific dimension it acquires in the law of targeting. What matters then for purposes of review is to ascertain whether a cyber weapon is inherently indiscriminate in light of its properties, design, and normal, expected or intended use. The indiscriminate use of a cyber weapon in an attack will equally violate the principle of distinction as will do the disproportionate effects on civilians of a specific attack using a cyber weapon but these are issues that concern the law of targeting.

It follows that malware which by nature, design and normal, expected and intended use cannot be directed against military objectives or produce uncontrollable effects impacting on civilians or civilian objects are inherently indiscriminate. Such an assessment is critical in view of the interconnected and dual-use nature of cyberspace and the fact that some malware can replicate autonomously. Thus, and in order to comply with this requirement, it is important to introduce a command and control capability to monitor the path the cyber weapon can take as well as its effects in order to terminate any unintended engagement. It is also important to introduce a self-destruct or self-deactivation property to malware not

[22] *Legality of the Threat or Use of Nuclear Weapons* (Advisory Opinion), para 78. API, Articles 48 and 51 (4)(b) (c)
[23] *Legality of the Threat or Use of Nuclear Weapons* (Advisory Opinion), para 78; ICTY**,** *Prosecutor v. Kupreškic´ et al.,* (Judgement), Case No. IT-95-16-T, Trial Chamber (14 January 2000), para 524.

only in order to prevent unintended consequences but also to prevent the malware to be adapted and reused by other actors following its initial use. The Stuxnet[24] malware is a good example of a discriminatory cyber weapon because it was designed to detect and attack a specific SCADA software, otherwise it deactivated itself. Although it replicated itself, its effects on other computers or systems was limited and below the threshold of damage.

Third, it should be determined whether cyber weapons can cause widespread, long term and severe damage to the environment[25] although not all states review weapons according to their impact on the environment.[26] The aforementioned requirements are cumulative and seem to set the threshold quite high. They refer to the intensity of the damage, its persistence in time, and the size of the geographical area affected by the damage, but the specific interpretation of these requirements can be subject to debate. For example, is long-term measured in years, decades or months? How wide should be an area in order to satisfy the widespread threshold? Moreover, can environmental damage be geographically limited? Is severity assessed in relation to human life and natural resources or also in relation to economic assets? Should the harm to the environment itself be taken into account? This raises the question of how 'environment' is defined. The commentary to Article 55 API states that [t]he concepts of the natural environment should be understood in the widest sense to cover the biological environment in which a population is living. It does not consist merely of the objects indispensable to survival (…) but also includes forests and other vegetation (…), as well as fauna, flora and other biological or climatic elements.[27] This

[24] Falliere, N. at al., *Symantec Security Response*, *W.32.Stuxnet.* DOSSIER 2 (2011),
[25] AP I, Article 35 (3); , *Customary International Humanitarian Law*, Volume I. Rule 45.
[26] USDOD, Office of the General Counsel, *Law of War Manual* (USDOD: Washington, DC, June 2015), para. 6.2.2
[27]Commentary on the Additional Protocols of 8 June 1977 to the Geneva Conventions of 12 August 1949, Protocol I, p. 662, para 2126.

means that it includes biotic or abiotic natural resources, but it does not mention anything about their interaction or about environmental values.[28]

That having been said, it is not at all evident how the high threshold of environmental impact can be met by the use of cyber weapons and more critically whether it is the cyber weapon itself that can cause the damage. For example, if a cyber weapon manipulates the supervisory control system of a nuclear reactor leading to nuclear fall-out or manipulates the supervisory control system of an oil refinery that causes leaks which poison water reserves, they relate to targeting and not to whether the cyber weapon itself can cause widespread, long term and sever environmental damage.

In addition to these three set of obligations, it is debated whether cyber weapons should be assessed against the principle of humanityor else the Martens Clause.[29] In its 1996 Advisory Opinion on the *Legality of the Threat or Use of Nuclear Weapons*, the ICJ held that the clause "proved to be an effective means of addressing the rapid evolution of military technology".[30] In our opinion, this principle may play some role as an interpretative device but not as a gap filler for purposes of review since there is relevant law that specifically applies to all 'new' weapons.

Be that as it may, it is important to note that certain characteristics of cyber weapons pose challenges to the effectiveness of Article 36 reviews. First, questions may be asked as to whether cyber weapons can be meaningfully tested before deployment because they can take paths in their deployment cycle that cannot always be predefined or foreseen and

[28] See for example the ILC definition of the environment in Principle 2(b) of its study on *International liability for injurious consequences arising out of acts not prohibited by international law* (international liability in case of loss from transboundary harm arising out of hazardous activities) A/61/10, 101.
[29] *Convention (II) with Respect to the Laws and Customs of War on Land and its annex: Regulations concerning the Laws and Customs of War on Land* (adopted 29 July 1899, entered into force 4 September 1900) preamble; AP I, Art. 1(2).
[30] ICJ, *Legality of the Threat or Use of Nuclear Weapons* (Advisory Opinion), para 78.

which are outside human control. This is compounded by the fact that information about the targeted system or its linkages at the stage of review can be quite sketchy or just be unavailable. Often modelling and simulation exercises are used but they cannot capture the specific context of the actual deployment of cyber weapons.[31] Secondly, cyber weapons are usually developed in order to exploit specific vulnerabilities. This means that they are based on intelligence concerning the potential target's vulnerabilities and states will be reluctant to disclose such information because, once disclosed, effective counters can be developed, and the targeted state may enhance its defences. Third, the 'life expectancy' of cyber weapons can be short and constant adaptations may be needed in order to prolong its 'life' and effectiveness. This is true after a cyber weapon has been used if it is to be re-used but also before it is used if patches or other defences are made available. Moreover, if a system vulnerability which a particular cyber weapon has exploited is corrected, the specific cyber weapon cannot be used at all against systems that have installed the patch. The short 'life expectancy' of cyber weapons and the constant need for adaptation raises the question of whether Article 36 reviews can be at all effective; whether cyber weapons should be reviewed after each and every modification or adaptation or whether, instead, operational reviews are more appropriate. Fourth, cyber weapons such as malware can be adapted by other actors after being used which raises questions as to whether reviews should take into account the post-use 'life' of a cyber weapon. Fifth, the fact that cyber weapons are developed by the private sector who holds the relevant data and information but also owns networks and servers, raises questions about the scope of review, the level of independent scrutiny and the level of cooperation and sharing of information between the state and the

[31] For example, see NATO, *Ready for the Predictable, Prepared for the Unexpected - M&S for Collective Defence in Hybrid Environments and Hybrid Conflicts*, STO-MP-MSG-143 (2016).

private sector that is needed for purposes of review. Sixth, because of the technical properties of cyber weapons and their sophistication, reviews require multi-level expertise and reliable data which raises questions about the calibre and expertise of those conducting reviews, their training and the sharing of information.

In addition to the specific challenges posed by cyber weapons, a number of factors relating to reviews themselves can also reduce the effectiveness of this mechanism. These refer to the fact that not all states conduct weapons reviews; the rules against which reviews are performed may differ from one state to another; reviews are not transparent (often for legitimate reasons of security); their methodology is not clear; the reports are not published; and compliance with their recommendations varies.

In order to conclude this section, it can be said that, whilst weapons reviews can ensure IHL compliance of cyber weapons, a number of factors relating to reviews in general and to cyber weapons in particular make this mechanism a weak form of regulation.

III. Regulation of cyber weapons through cyber arms control treaties and confidence building measures.

In view of the difficulties surrounding Article 36 this section will consider whether cyber weapons can be regulated though other complementary mechanisms such as cyber arms control treaties and confidence building measures.

1. Cyber Arms Control Treaties

The post Word War-era has witnessed the conclusion of many conventions addressing different categories of weapons, from anti-personnel landmines to weapons of mass destruction (WMDs) such as bacteriological, chemical and nuclear weapons, and imposing

total or partial bans to the testing and production of such weapons. As States and non-state actors keep on developing their cyber arsenals, the question is raised as to how the cyber-arms race can be regulated. Before discussing whether existing regulatory regimes can be effectively adapted to the cyber context, it is worth noting certain particular features of cyber weapons which may impact on the regulatory potential of such regimes.

Firstly, as was noted in the preceding section, cyber weapons are capable of producing effects which are different from those resulting from the employment of traditional weapons or WMDs; therefore, reaching consensus over what, essentially, amounts to a cyber-weapon is of critical importance. Secondly, identifying what part of the code constitutes a cyber-weapon is important for purposes of regulation. A malware can be dissected into three different elements: namely, propagation, exploit and payload.[32] Propagation 'is the means of transporting malicious code from origin to target', employing tools as simple as an email attachment or a USB stick to jump a physical 'air-gap' between computers,[33] as it happened in the case of the Stuxnet worm.[34] An exploit is designed to take advantage of vulnerabilities in computer systems or surrounding networks, in order to allow the perpetrator to perform unintended operations, by typically targeting the operating system of the target network of one its key applications.[35] An exploit thus enables both 'the propagation method and payload's operation'.[36] The payload is the essential component of malware designed to execute its commands on the targeted system to

[32] See Herr, T., Rosenzweig, P., 'Cyber Weapons and Export Control: Incorporating Dual Use with the PrEP Model' 8 (2) *Journal of National Security Law and Policy* (2015) 301-319 (hereinafter 'Cyber Weapons and Export Control'); Herr., T., 'PrEP: A Framework for Malware & Cyber Weapons' 13 (1) *The Journal of Information Warfare* (2014)

[33] Herr, T., 'Cyber Weapons and Export Control', p.305.

[34] Zetter, K., 'An Unprecedented Look at Stuxnet, The First Digital Weapon', *Wired*, 11 March 2014.

[35] Mell, P., Grance, T., 'The NIST Definition of Cloud Computing' National Institute of Standards & Technology (2011)

[36] Herr., T., 'PrEP: A Framework for Malware & Cyber Weapons', p.6.

achieve some predefined goal, such as compromising password files or deleting data.[37] In the case of Stuxnet, for instance, the payload consisted in manipulating the rotatory speed of the turbines until they were destroyed.[38] In the preceding section we have grouped weapons and means together for purposes of review in order to include weapons in the 'widest sense' as required by the ICRC commentary[39] and we have also included non-physical effects in the definition of weapons but the proposed definition may not be appropriate for arms control regulation because such regimes fulfil different purposes. If the payload is for example the component by which a malware can be qualified as a cyber-weapon for arms control purposes, regulatory attempts should then be predominantly focused around limiting the proliferation of payloads of malware, rather than on propagation and exploits.

Secondly, if consensus is reached as to what constitutes a cyber-weapon for regulation purposes, it would then be possible to establish a cyber-arms control treaty, modelling it on conventional arms control treaties, such as the Treaty on the Non-Proliferation of Nuclear Weapons (NPT). However, it must be pointed out that such cyber-arms control treaty will pursue different objectives than the NPT. Signed at the heyday of the Cold-War, the NPT operated within a system where States were the only actors involved. Therefore, the NPT was aimed at limiting both horizontal proliferation (the development or acquisition of nuclear weapons by non-nuclear States), as well as vertical proliferation (the increase of the nuclear arsenal by States already in possession of nuclear weapons).[40] At the same time, the

---

[37] *Id*.

[38] Albright, D., Brennan, P., Warlond, C., 'Stuxnet Malware and Natanz: Update of ISIS December 22, 2010 Report 7', Institute for Science and International Security (2011).

[39] Commentary on the Additional Protocols of 8 June 1977 to the Geneva Conventions of 12 August 1949, Protocol I, p. 398, para. 1402.

[40] See generally, Venturini, G., 'Control and Verification of Multilateral Treaties on Disarmament and Non-Proliferation of Weapons of Mass Destruction' 17 (2) *University of California Davis Law Review* (2011) 345-383,

drafters of the NPT were not concerned about the fact that non-state actors could gain control of nuclear weapons; as such, the NPT does not contain any provision addressing the issue of sub-state proliferation. But things are quite different in cyberspace, where non-State actors play an increasingly important role. This is because cyber-weapons are relatively easier to acquire when compared to other forms of weapons. Consider, in this regard, the case of the Duqu malware, which was believed by Symantec to have been developed either by the same authors as Stuxnet, or by actors who had access to its source code.[41] It has also been acknowledged that part of the source code of Stuxnet has been made freely available on the Internet after its discovery in 2011, thereby reinforcing the likelihood that malicious users can gain possession of cyber-weapons. Furthermore, non-State actors already have the capabilities to develop malware, as demonstrated by the 2007 DDOS against Estonia, carried out by a patriotic hacker group, named Nashi ('Youth'),[42] or those launched by Anonymous against Israel in 2014.[43] Therefore, given the increased relevance of non-State actors in the cyber domain, it is safe to say that an arms control treaty for cyberspace should be addressed both to State and non-State actors and, more specifically, should focus on preventing them from developing or acquiring cyber-weapons. As we will see, similar attempts have been made in the context of CMBs, but with unsatisfactory results.

---

p. 358. See also Treaty on the Non-Proliferation of Nuclear Weapons, A/RES/2373 (XXII) (June 12,1968), 729 U.N.T.S. 161, Art, 5 and Art. 6.

[41] See *Symantec Security Response*: *W.32 Duqu: The Precursor to the Next Stuxnet*, Version 1.4 (November 23, 2011), Symantec, p.1. Available at https://www.symantec.com/content/dam/symantec/docs/security-center/white-papers/w32-duqu-11-en.pdf.

[42] Richards, J., 'Thousands of cyber-attacks each day on key utilities', *The Times*, 23 August 2008.

[43] Gilbert, D., 'Anonymous Continues Cyber-Attacks on Israeli Government Websites Knocking Mossad and IDF Online, *International Business Times*, 4 August 2014. Available at https://www.ibtimes.co.uk/anonymous-continues-cyber-attacks-israeli-government-websites-knocking-mossad-idf-offline-1459689.

The most problematic issue in devising rules for non-proliferation of cyber weapons however lies in the verification process. It includes different methods to determine whether signatories are in compliance with their obligations under the Treaty, and is essential in building and maintaining trust among States-parties.[44] In the context of WMDs control treaties, different verification methods have been introduced.

First, States may adopt National Technical Means (NTMs), which include a wide array of 'technological capabilities, collection systems, and other intelligence and analytical resources that can be used to gather information about the activities' of signatories.[45] In the case of nuclear weapons, NTMs comprise the use of satellite, radars, seismic and nuclear radiation detection among others.[46] Clearly, there are some technical limitations to the successful application of such measures in the cyber domain. For instance, it has been argued that, in the cyber context, the use of network scanners and intrusion detection systems can be considered as functionally analogous to the use of satellites, as those measures serve the same purpose of ascertaining whether a State, or a non-state actor, has developed or acquired cyberwarfare capabilities. As such, a State may monitor its own network and analyse the traffic of *data* occurring between two different States, in order to gather information aimed at determining whether malicious activities, including cyber-attacks, are occurring.[47] It must be pointed out, however, that this method loses some of its effectiveness when great amounts of *data* has to be analysed, because it becomes more difficult to detect a cyber-attack.

---

[44] See United Nations General Assembly Resolution A/RES/43/81, 1988.

[45] Hodgson, G., 'Cyber Attack Treaty Verification' 12 (2) *I/S: A Journal of Law and Policy for the Information Society* 92016) 231-2260, p. 242; Scribner, R., *The Verification Challenge: Problems and Promise of Strategic Nuclear Arms Control Verification* (Birkhauser 1985) p.77.

[46] Hodgson, G., 'Cyber Attack Treaty Verification' p.242.

[47] Shackleford, D., 'Optimized Network Monitoring for Real-World Threats', SANS Institute (July 1, 2011), available at http://www.sans.org./reading-room/whitepapers/optimized-network-monitoring-real-world-threats-35040.

Different questions are also raised in the case of NMTs involving State intrusion into another State's computer systems or networks, either through a cyber-operation or with the consent of the latter: in this case, there is no doubt that such a measure could be successful as a verification method, but it would give the performing State the ability to conduct malicious acts against the other State, or gain access to sensible information. Furthermore, it has been observed that 'needing to launch a cyber attack to verify that another state is not launching cyber attacks seems to defeat the purpose of having a cyber treaty.'[48] Lastly, the structural features of cyber technology make detection of cyber-attack inherently difficult, since malware can be hidden anywhere, from a USB drive to a computer which may not be located in the territory of the inspected State. For these reasons, NTMs seem to be ill-suited as a verification method in the cyber context.

Another verification method is that of On-Site Inspections (OSI), which consist of direct access by a State to another State's military sites, if the former suspects that the latter is in violation of the terms of the treaty. The cyber analogous to OSIs are cyber investigations employing 'network taps'[49] for the purposes of session reconstruction, log inspection and data analysis of suspected States. In the case of session reconstruction, packets of data are correlated with each other in order to determine what information was sent between two computers.[50] Traffic analysis can be used for the purposes of finding anomalies in traffic patterns,[51] showing the average packet quantity transmitted by the network, their size, or

[48] Hodgson, G., 'Cyber Attack Treaty Verification, p.246.
[49] A network tap is an external monitoring device that mirrors the traffic that passes between two network nodes. See *Network Tap Definition,* TechTarget, available at http://searchnetworking.techtarget.com/definitionNetwork-tap.
[50] *TCP Session Reconstruction*, RedSplice (2016), available at https://redsplice.com/tcp-session-reconstruction/.
[51] Hodgson, G., 'Cyber Attack Treaty Verification', p.250.

how many connections per hour take place within the inspected network.[52] The major problem with cyber inspections is that, similarly to what could happen to network monitoring, they give the performing State the possibility of gaining access to sensitive information resident in another State's network. Certainly, the inspected State may protect itself with containment measures, such as encryption: but this would likely make the investigation not as effective. Therefore, a cyber-arms control treaty would need to find the proper balance between allowing the inspecting State to investigate treaty compliance and allow the inspected States to protect their sensitive information.

A third method of verification that has been successfully implemented in conventional arms control treaty is that of *data* exchanges: according to the U.N. General Assembly, 'request for inspections or information in accordance with the provisions of an arms limitation and disarmament agreement should be considered as a normal component of the verification process. Such request should be used only for the purposes of the determination of compliance, care being taken to avoid abuses.'[53] Measures in the context of nuclear weapons disarmament and proliferation included 'continuous data exchanges on the technical details of missiles', full access to telemetric information from missile flight tests, exchange of information on treaty-limited items and notifications of future development or modification to such items.[54] In this regard, data exchanges can be implemented in the cyber context only to a limited extent, because sharing knowledge about the source code of a cyber-weapon would likely make it ineffective. To explain, a malware in order to be successful must not only be designed to properly deliver the payload against its designated

---

[52] Northcutt, S., *Traffic Analysis*, SANS Technology Institute, available at http://sans.edu./research/security-laboratory/article/traffic-analysis.

[53] United Nations General Assembly Resolution 43/81 (December 7, 1988).

[54] Woolf, A., *Monitoring and Verification in Arms Control* 2(2011) available at http://fas.org./sgp/crs/nuke.R41201.pdf.

target but also to be able to remain undetected from anti-virus researches since once its code is discovered and analysed, antivirus programs are updated to recognize the malware and delete it as soon as it infects a computer system.[55] Therefore, because malware depend heavily on keeping its code secret, a cyber treaty that would include verification methods that require parties to exchange details about the cyber weapons they use, means that the cyber weapons themselves would become ineffective. A possible solution, in this regard, would be to limit the amount of information exchanged by States to only the most destructive types of attack, for instance those that target industrial infrastructures or other critical infrastructures such as financial services, nuclear reactors or dams. It remains to be seen, however, to what extent this is a practical solution.

2. Confidence Building Measures.

The above section has shown why a putative cyber treaty would perhaps be ineffective at preventing the proliferation of cyber weapons. Considering this, this section will discuss some issues related to the application of Confidence Building Measures in the cyber domain. In this regard, Confidence Building Measures (CBMs) are measures designed to mitigate the fear of attack by States in a situation of potential conflict; CBMs operate, then, 'as a form of reassurance that seeks to demonstrate intent among rivals, therefore (ideally) conveying a desire to maintain the status quo and foster a sense of security between otherwise threatened States.[56] CBMs can take different forms, from unilateral actions to bilateral or multilateral non-binding agreement. During the Cold War, one of the most successful examples of CBMs had been the Helsinki Final Act by the Conference on Security and Co-

[55] Elisan, C., *Malware, Rootkits & Botnets: A Beginner's Guide* (McGraw-Hill 2013) 102.
[56] Borghard, E.D., Lonergan, S.W., 'Confidence Building Measures for the Cyber Domain' *Strategic Studies Quarterly* (Fall 2018) 10-49, p. 12; Alford, J., 'Confidence-Building Measures in Europe: The Military Aspects' *Adelphi Papers* 19, no.149 (1979) 4-13.

Operation in Europe (nowadays OCSE), in 1975.[57] The Helsinki Final Act aimed at contributing to 'reducing the dangers of armed conflict and of misunderstanding of military activities which could give rise to apprehension', providing measures such as voluntary reporting of military manoeuvres, exchange of observers for such military manoeuvres, hosting of military delegations and exchange of information related to defence budgets and employment of new weapons systems.

Are CBMs applicable to the cyber domain and, if so, are they effective? In order to answer this question, one must distinguish between information-related CBMs, notification CBMs and arms stability CBMs.

Information-based CBMs focus on sharing defence-related information between interested parties. Given the nature of cyberspace, such measures should involve States and non-state actors alike, considering the fact that the primary target of cyber-attacks is the private sector, which operates and owns the majority of cyber infrastructures and, most importantly, has a high level of knowledge about cyberwarfare tactics and capabilities. Borghand and Lonergan argue that information CBMs should prioritize three elements: firstly, the identification of threat actors and emerging methods and means for exploitation;[58] secondly, the dissemination of system vulnerability reports, in order to improve network defense;[59] and finally, doctrines and national policies should be shared not only at the State level, but also involve the private sector and other stakeholders.[60] In this regard, one attempt to develop information related CMBS has been the United Nations Governmental Group of Experts ('UN GGE') on Development in the Fields of Information and

[57] Conference on Security and Co-operation in Europe Final Act, August 1975, 84. Available at https://osce.org/helsinki-final-act?download=true
[58] Borghard, E.D., Lone Lonergan, S.W., 'Confidence Building Measures for the Cyber Domain', p.21
[59] *Id.*
[60] *Id.*, p.23.

Telecommunication in the Context of International Security, which was convened for the first time in 2004 and the focus of which became information sharing in the areas of cybersecurity, reducing risks to critical national infrastructures and finding a consensus towards a common language for regulating cyberspace. The GGE was moderately successful in the pursuit of its objectives, reaching agreement on issues such as the applicability of international law, on the concept of sovereignty in cyberspace and on attribution of conduct for the purposes of establishing state responsibility, despite falling short on finding a consensus on how international law should apply.[61]

As opposed to information related CBMs, measures involving notification, observation and stabilization might be more difficult to apply in cyberspace, given the unique characteristics of cyber warfare.

Notification measures are generally aimed at notifying other States about a military exercise in order to show transparency and provide reassurance. The reason why notification measures are not as effective in cyberspace is because States are, understandably, very reluctant to notify cyber-exercises to other States, because that would potentially reveal the State's cyber capabilities and vulnerabilities which can potentially be exploited.[62] Furthermore, as shown above with regard to verification measures, cyber-weapons are most effective when there are kept secret; therefore, showing offensive cyber capabilities to observing States would likely made them ineffective. It can be pointed out in this regard that notification measures may be implemented between allies; however, as much as it might be true, this would diminish the relevance of CBMs, the purpose of which is to foster mutual trust between opposing States.

[61] Sukumar, A.M., 'The UN GGE Failed. Is International Law in Cyberspace Doomed as Well?' *Lawfare*, 4 July 2017. Available at https://www.lawfareblog.com/un-gge-failed-international-law-cyberspace-doomed-well.
[62] Borghard, E.D., Lonergan, S.W., 'Confidence Building Measures for the Cyber Domain' p.23.

A third category of CBMs are those involving cyber-arms control; they share the same purpose of arms control agreements (limit the proliferation of a certain weapon category), but they are on an entirely voluntary basis. Similarly, to arms control agreements, cyber-arms control CBMs face similar challenges. The main challenge, then, is *how* to limit sub-state proliferation in a multi-stakeholder environment. The Wassenaar Arrangement is a good example in this regard. Originally signed in 1996, the Wassenaar Arrangement is a multilateral arrangement the objective of which is to limit the export of conventional weapons and sensitive dual-use goods. It establishes a six-month periodic exchange of information about transfer to non-Wassenaar States of selected categories of weapons, munitions and dual-use technologies. The Agreement has been amended in 2015 in an attempt to curb the sale of malware to repressive governments. As such, it has focused on limiting the export of 'intrusion software', defined as 'software specifically designed and modified to avoid detection by 'monitoring tools', or to defeat 'protective countermeasures' and capable of performing 'extraction of data or information' and/or 'modification of the standard path of execution process in order to allow the execution of externally provided instruction.[63]' However, the amendment did not directly and specifically target 'intrusion software', but rather focused on supporting components, a notion that includes 'any software, systems, equipment, component or technology used to generate, operate, deliver, or communicate with intrusion software.'[64]

[63] Herr, T., 'Malware Counter-Proliferation and the Wassenaar Arrangement' *Proceedings of the 8th International Conference on Cyber Conflict* (2016) p.9.
[64] *Id.*

The Wassenaar amendment on cyber was met with much criticism by cybersecurity firms and other private sector actors,[65] as it restricted a wide range of cybersecurity-related tools that focused on testing a systems' vulnerability, such as penetration testing technology,[66] or so called 'bug bounty' programs.[67] Recently, the 2018 Plenary Session of the Wassenaar arrangement updated its rules, relaxing export control requirements on certain hacking tools primarily used by cybersecurity researchers involved in vulnerability disclosure and incident response. The change has been welcomed by private actors as a step forward but key issues still remain unsolved, such as the overly broad definition of 'intrusion software.' [68] This highlights the need for future CBMs to address the issue of sub-state proliferation while, at the same time, involve the private sector without prejudicing their positive efforts in the field of cybersecurity.

IV. Conclusion

The chapter has considered the application of certain regulatory mechanisms to cyber weapons. It first considered the mechanism of weapons review introduced by Article 36 API. This is a mechanism that prevents the fielding of cyber weapons that violate IHL by ensuring that they are IHL compliant at the earlier stage of their study, development, adoption or acquisition. It is a general mechanism that applies to all new weapons including cyber

[65] Brandom, R., 'Google Says Controversial Exports Proposal Would Make the World 'Less Secure'', *The Verge*, 20 July 2015. Available at http://theverge.com/2015/7/20/9005351/google-wassenaar-arrangement-proposal-comments.

[66] The term 'penetration testing' describes an authorized simulated cyber-attack on a computer system, performed to evaluate the security of the system. See US Department of Interior, *Penetration Testing* (retrieved 28 Mar 2019). Available at https://www.doi.gov/ocio/customers/penetration-testing.

[67] A 'bug bounty' program is a deal offered by software development companies where individuals are rewarded for reporting bugs related to exploits and vulnerabilities.

[68] Cross, T., 'New Changes to Wassenaar Arrangement Export Controls Will Benefit Cybersecurity' *Forbes*, 16 January 2018. Available at https://www.forbes.com/sites/forbestechcouncil/2018/01/16/new-changes-to-wassenaar-arrangement-export-controls-will-benefit-cybersecurity/#65923a4f5ed6.

weapons but, as was noted, there are certain issues that impact on its effectiveness, with the particular features of cyber weapons exacerbating any problems that already exist. These refer to the lack of consensus on the legal definition of cyber 'weapons'; the multilevel expertise, training, data and resources required to conduct reviews; the need to adapt the reviewing methodology to take into account the features of cyber weapons and the difficulties in actually testing them; the need to recalibrate the relationship between the state and the private sector in view of the fact that cyber weapons are often dual-use, they are manufactured by the private sector who holds data and information and has the relevant expertise; the need for states to exchange information in view of the different levels of expertise and technological knowledge they possess and to share good practices in view of the different review processes they follow.[69] If the challenges posed by cyber weapons are addressed, if states' attitudes towards weapons review change and if there is reasonable cooperation between and among states and the private sector, weapons review can provide an effective mechanism for regulating cyber weapons.

The chapter then moved on to discuss the regulation of cyber weapons through non-proliferation mechanisms and CBMs. It identified three issues that prevent cyber-arms proliferation from happening. The first is terminological uncertainties related to what constitutes a cyber weapon and what parts should be regulated. This replicates the definitional problems mentioned in relation to Article 36 but raises the additional point of definition for armed control purposes. Secondly, the virtuality and secrecy that shrouds the development and use of cyber weapons make verification measures either unfeasible or unlikely to be implemented, given the legitimate concerns of States that revealing their

[69] UK, Ministry of Defence, Development, Concepts and Doctrine Centre, *UK Weapon Reviews* (2016) https://assets.publishing.service.gov.uk/government/uploads/system/uploads/attachment_data/file/507319/20160308-UK_weapon_reviews.pdf

cyber capabilities to an adversary can be advantageous to the latter and a similar point can be made with regard to notification-based CBMs. Finally, the cyber domain differs in that non-State actors are greatly involved and must play an active role in the establishment of CBMs and rules that attempt to limit the proliferation of cyber-weapons. Modelling such measures over past successful attempts in the context of conventional arms control will continue to prove ineffective and therefore a new approach may be needed.