Cyberwarfare and International Humanitarian Law

By: Zen Chang

I. INTRODUCTION

The proliferation of cyber-attacks has shifted the paradigm of warfare. In May 2017, the world saw the first global cyber-attack where WannaCry ransomware affected thousands of civilian infrastructures (i.e. hospitals, transport services, energy services, etc.) in over a hundred nations. The WannaCry attack is the first instance where civilian lives were directly and intentionally endangered by a piece of malicious code. However, cyber-attacks transposing itself into the kinetic realm is not a new phenomenon. In July 2010, Iran’s nuclear facilities in Natanz was hit with the Stuxnet malware (“Operation Olympic Games”), which destroyed nuclear centrifuges and ultimately halted Iran’s nuclear ambitions. In the wake of WannaCry ransomware attack, calls have been made to codify a “Digital Geneva Convention.”

1 Master of International Law Candidate, Sydney Law School, University of Sydney; Bachelor of Laws (Hons), 2016, College of Law, Australian National University; Bachelor of International Relations, 2016, School of Politics and International Relations, Australian National University.
2 Attribution is unknown at time of writing, though inconclusive evidence points to North Korea.
cyberwars are not regulated by any international humanitarian law\(^7\) ("IHL") treaties, "their development and employment in armed conflict do not occur in a legal vacuum."\(^8\)

This paper seeks to explore the interaction between cyberwarfare and IHL. Whilst “the legal principles [of IHL] applies to all forms of warfare [including] those of the future,”\(^9\) how it is to apply remains contentious and subject to debate. This paper will critically analyse how the legal parameters of IHL, lex lata, apply in times of cyberwar. The scope of this paper is restricted to jus in bello in international armed conflicts ("IAC") (notwithstanding Section II.) Section II will argue how cyber-attacks are “armed conflict[s]” under Common Article 2 to the Geneva Conventions, to which IHL applies. Thereafter, how cyber-attacks are “attacks” under Additional Protocol I ("API") for relevant IHL restrictions to apply. Sections III, IV and V will explore how the principles of distinction, proportionally, and direct participation in hostilities, respectively, should apply in cyberwar. In arguing the above notion, this paper will attempt to interpret relevant provisions in the Geneva Conventions and API in the context of cyberwarfare. This paper seeks to explore, and perhaps show, the nuances in cyber-IHL which military commanders, and military legal advisors, ought to take note.

II. THE THRESHOLD OF “ARMED CONFLICT” AND “ATTACKS” IN CYBER OPERATIONS

A. CYBER “ARMED CONFLICT” IN IAC

The starting point is Article 2 common to all four Geneva Conventions which reads:

“[T]he present Convention shall apply to all cases of declared war or of any other armed conflict which may arise between two or more of the High Contracting Parties, even if the state of war is not recognized by one of them.”\(^10\)

\(^7\) Also known as Law of Armed Conflict.


\(^9\) Legality of the Threat or Use of Nuclear Weapons, Advisory Opinion, 1996 I.C.J. 226 (July 8).

For an IAC to exist, the performing act must constitute an “armed conflict” under Common Article 2 to the four Geneva Conventions. If the act does not amount to an “armed conflict which may arise between two or more of the High Contracting Parties”, there is no IAC, to which IHL does not apply. This sentiment was also echoed in Tadić, which notes that an “international armed conflict arises ‘whenever there is a resort to armed force between States.’” Prima facie, an “armed conflict” denotes some sort of kinetic force, however, in cyber operations, there is a lack of kinetic force. The main contention is whether kinetic force is a necessary condition precedent for an “armed conflict” to exist under IHL.

What amounts to an “armed conflict” or “resort to armed force” is not defined in any IHL treaties, one has to look at the jurisprudence to tease out the definition of “armed conflict,” and if it covers the scope of cyberwarfare. Most scholars posit the view that if a cyber-attack is attributable to a State, and the cyber-attack has the same effects as would kinetic force, it would reach the threshold of an “armed conflict.” This view is consistent with Pictet’s commentary which adopts a broad view of the term “armed conflict” as “any difference arising between two States and leading to the intervention of armed forces is an armed conflict within the meaning of Article 2.”

---

11 Id.; see also NEW TECHNOLOGIES AND THE LAW OF ARMED CONFLICT, 80-81 (Robert McLaughlin & Hitoshi Nasu eds., 2014).
13 Droeg, supra note 12, at 543.
sufficient to trigger an IAC. The Courts in Tadić also adopts Pictet’s broad view.\(^{16}\) The International Committee of the Red Cross (“ICRC”) notes:

“By using the words ‘from the outset’ the authors of the Convention wished to show that it became applicable as soon as the first acts of violence were committed, even if the armed struggle did not continue. [...].

*Mere frontier incidents* may make the Convention applicable, […].”\(^{17}\)

With the lack of a *de minimis* level of intervention, it can be cogently argued that cyberwarfare falls within the scope of an “armed conflict,” insofar the cyber act produces the same outcome as would kinetic force. The *Tallinn Manual 2.0* also adopts this expansive view.\(^{18}\)

However, what if the computer attack falls short of the threshold to produce the same outcome as would kinetic force (which might lead to death and injury), and produce mere disruptions to affect the object’s functioning? Can the expansive interpretation of an “armed conflict” be triggered to account for disruptions? It is difficult to enquire this *non liquet* due to the lack of state practice.\(^{19}\) One has to take several approaches to answer this hypothetical.

One approach is to consider if cyber acts that lead to mere disruptions of objects constitute as “armed conflict.”\(^{20}\) The object and purpose of IHL is to avoid legal *lacunas* in the protection of the civilian population (amongst other things) from the harmful effects of war.\(^{21}\) This is evident in the explicit absence of a “violence threshold” for the existence of an IAC to occur. By extension, this negative definition would favour an interpretation to account for cyber disruptions to trigger an “armed conflict.” Furthermore, including mere disruptions into the definition of an “armed conflict” would serve the purpose of the apparent denial of a *de minimis* level of intervention to

---

\(^{16}\) Tadić, para.70.


\(^{19}\) Droege, *supra* note 12, at 541-552.


trigger an “armed conflict.” States would also be keen to adopt this view as a pre-emptive measure to protect their infrastructures from a cyber-attack. Without an “armed conflict,” the protective mechanisms in IHL will not apply, leading States to be more vulnerable in managing their critical infrastructures. This is inconsistent with the object and purpose of IHL.

Another approach is to examine the “mens rea” requirement for an “armed conflict” to occur. Various literature has examined different major incidents between States that have not been treated as “armed conflict” despite meeting the necessary threshold requirements of an IAC (i.e. Dogger Bank Incident; USS Liberty Incident; USS Stark Incident). In contrast to minor incidents where States have asserted that “such situations and their consequences fell within the scope of the Geneva Conventions” (e.g. Iran Air Incident). The only difference between the two scenarios was based largely on the perceived intentions and threat assessments of the other party, an assessment which is often influenced by realpolitik. Melzer notes, “in the absence of a formal declaration of war, an IAC requires a minimal transgression, which expresses the belligerent intent of the acting state against another.” It can be deduced that what amounts to an “armed conflict” is really based on intent, rather than the factual circumstance.

If the cyber-attack amounts to a Common Article 2 “armed conflict,” IHL applies. Once IHL applies, it is imperative to evaluate whether a cyber-operation is an “attack” within the meaning of Article 49(1) in API – to which relevant IHL restrictions (i.e. distinction, proportionality, precaution, etc.) apply.

B. Are Cyber-Attacks “Attacks”?

---

22 Droege, supra note 12, at 3.
25 Id. at 121, (citing Int’l Rev. of the Red Cross, External Activities: September–October 1987, 27(261) IRRC 650 (1987).
26 NILES MELZER, TARGETED KILLING IN INTERNATIONAL LAW 250 (2008) [hereinafter Melzer, Targeted Killing].
“Attacks” are defined in Article 49(1) of API, which is customary law, as “acts of violence against the adversary, whether in offence or in defence.” It is an effect-based approach. From the travaux préparatoires, “violence” means physical violence. Also, the Oxford English Dictionary defines “violence” as “behaviour involving physical force.” The ICRC takes the view of the travaux préparatoires that the term “attack” means “combat action,” which denotes a physical act. Hence “attacks” excludes dissemination of propaganda, embargoes, or other non-physical means of warfare (e.g. psychological, economical, or political). However, the drafters of API (and the Geneva Conventions) could not have predicted the proliferation of cyber-technology and its harmful effects. Fortunately, Article 36 of API provides the need to apply the rules, lex lata, to new “development, acquisition or adoption of a new weapon, means or method of warfare,” which may extend to cyber-weapons.

Given that “attack” denotes physical force, cyber operations which result in a physical outcome (i.e. blowing up a nuclear plant with malware) would be an “attack” as it is an “act of violence” under Article 49(1) API. It is not the means of attack, but the consequence of the attack. For example, the use of biological, radiological, and chemical weapons would constitute an “attack” even though it lacks physical force. This “consequential harm” approach is also

---

28 Protocol Additional to the Geneva Conventions of 12 August 1949, and relating to the Protection of Victims of International Armed Conflicts (Protocol I) art. 49, June 8, 1977, 1125 U.N.T.S. 3 [hereinafter API].
29 Melzer, Targeted Killing, supra note 26, at 23-25.
31 Violence, OXFORD DICTIONARY OF ENGLISH (3d ed. 2010).
32 COMMENTARY ON THE ADDITIONAL PROTOCOLS OF 8 JUNE 1977 TO THE GENEVA CONVENTIONS OF 12 AUGUST 1949 ¶ 1880 (Yves Sandoz et al. eds., 1987) [hereinafter AP Commentary].
33 Michael Bothe Et Al., New Rules for Victims of Armed Conflicts: Commentary to the Two 1977 Protocols Additional to the Geneva Conventions of 1949 325 (2013); See also API, supra note 28, at art. 52(2).
34 API, supra note 28, at art. 36.
35 See Dinniss, supra, note 24, at 62-74; Dörmann, supra note 27, at 5-6.
36 Schmitt, Tallinn Manual 2.0, supra note 18, at 415; see also Prosecutor v. Tadić, Case No. IT-94-1, Decision on the Defense Motion for Interlocutory Appeal on Jurisdiction, ¶120-24 (Int’l Crim. Trib. For the Former Yugoslavia Oct. 2, 1995) (recognizing a general consensus in the
supported by numerous articles of API; Article 51(1) states that “civilian population and individual civilians shall enjoy general protection against dangers arising from military operations”\(^\text{37}\), Article 51(5)(b) states “loss of civilian life, injury to civilians, damage to civilian objects, […]”.\(^\text{38}\) These articles reflect how the drafters had the intention for the “consequential harm” approach to qualify as an “attack.”

What if, the cyber operation does not produce any violent consequence, but mere disruptions and interference to the object without causing physical damage, does this constitute an “attack”? There are two legal lenses to examine this hypothetical.

Schmitt adopts a narrow approach and posits that:

“[a] cyber operation, like any other operation, is an attack when resulting in death or injury of individuals, whether civilians or combatants, or damage to or destruction of objects, whether military objectives or civilian objects.”\(^\text{39}\)

To Schmitt, “damage” only refers to physical damage,\(^\text{40}\) and cyber-attacks that do not result in any form of physical damage (cf. interruptions and inconvenience) does not constitute as “attack” insofar it does not cause human suffering or loss of lives.\(^\text{41}\) With respect, Schmitt’s narrow approach is too under-inclusive, it would be incongruous to posit that anything which falls short of “physical damage” is not damage at all. If the machine, or infrastructure, has lost its function to operate because of a cyber-attack due to cyber interference and/or disruptions, it is “damaged” to the extent that the purpose of the infrastructure has been hindered (i.e. WannaCry ransomware). An object does not need to be physically damaged to render it unusable. Schmitt’s approach would not be feasible during a cyber-attack given that data and information can always be restored (due

\(^{\text{37}}\) API, supra note 28, at art. 51(a).
\(^{\text{38}}\) Id. at art. 51(5)(b).
\(^{\text{40}}\) Id at 95.
\(^{\text{41}}\) Id.
to cloud computing); hence, there can really be no physical destruction of the object which leads to permanent loss of functionality or destruction. Thus, cyber-attacks which lead to disruptions and interference without physical damage or destruction falls within the corpus of Article 49(1) of API, even if the disruption is temporary.

A broader interpretation is more appropriate; cyber-operations constitute “attacks” even if they do not lead to the destruction of objects. This view turns on the intent of the drafters of API, as shown in Article 52(2), which states that a military objective is one “whose total or partial destruction, capture or neutralization, […] offers a definite military advantage.” The term “neutralization” denotes that it would be “irrelevant whether an object is disabled through destruction or in any other way.” This shows that the drafters held the view that “attacks,” may not only lead to the destruction of objects, but may also lead to the loss of functioning without necessarily destroying it. Furthermore, by examining the travaux préparatoires of API, it was observed that the laying of mines constituted an “attack” “whenever a person is directly endangered by a mine laid.” Analogously, a single penetration of a piece of malicious code which does not necessarily meet the threshold of harm required for there to be “damage” or “destruction of object,” constitutes as an “attack” under API.

Once an “attack” has occurred, the relevant restrictions in IHL apply. This will be discussed in the next few sections.

III. DISTINCTION IN CYBERSPACE

The principle of distinction is set out in Article 48 of API, which reads:

“[T]he Parties to the conflict shall at all times distinguish between the civilian population and combatants and between civilian objects and military objectives and accordingly shall direct their operations only

---

42 Dörmann argues for a broad approach.
43 Dörmann, supra note 27, at 5.
44 API, supra note 28, at art. 52(2).
45 Dörmann, supra note 27, at 6.
46 AP Commentary, supra note 32, at ¶ 1881, see also Final Record, supra note 30, at 443-44 (discussing the impropriety of employing prisoners of war to remove laid mines).
against military objectives.”

Furthermore, the International Court of Justice in its Nuclear Weapons Advisory Opinion held that the principle of distinction is the “cardinal principle contained in the texts constituting the fabric of humanitarian law.” Article 48 of API is customary international law. Reading Article 48 literally, attacks may only be directed against military objectives; hence, vis-à-vis the cyber-sphere, attacks which are directed at civilian cyber infrastructures would amount to a breach of Article 48. Conversely, a “lawful” cyber-attack is one which only attacks military cyber infrastructures which would confer a “definite military advantage.” Given the distinction between civilian and military cyber infrastructure is not as distinct as traditional infrastructures of war, one has to tease out the legality of cyber distinction. This paper will examine the notion of “military objective” as noted in Article 48 of API, and examine how the principle of distinction applies in an interconnected cyber space.

Under IHL, civilian objects are all objects that are not military objectives. Military objectives are defined in Article 52(2) of API, which reads:

“[…] 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.”

Pursuant to Article 52(3), “object which is normally dedicated to civilian purposes, […] is being used to make an effective contribution to military action, it shall be presumed not to be so used.” Unfortunately, most cyber infrastructures are dual-use – they have both a civilian and military

---

47 *API, supra* note 28, at art. 48.
50 *API, supra* note 28, at art. 52(2).
51 *Id.* at art. 52(1).
52 *Id.* at art. 52(2).
53 *Id.* at art. 52(3).
function. Data centres, for example, which are primarily used for civilians to store information (i.e. on the cloud computing), will inevitably carry with it military data and information. Under IHL, this makes the said civilian data centre a military objective, a legitimate target of attack. Under Article 52(2) of API, the moment that civilian data centre (or any civilian cyber infrastructure) is used for a military action, it loses its civilian status and becomes a military objective in its entirety. It is noted that 98% of US owned military data are stored in civilian data centres around the world. Hence, by virtue of Article 52(2), a majority of the data centres around the world are legitimate military objectives. What adds to this complexity is the “purpose” criterion in Article 52(2). The ICRC notes that “the criterion of “purpose” is concerned with the intended future use of an object”; if the object was intended to be used militarily in the future, “they become military objectives.” As also noted by the Air and Missile Warfare Manual (“AMW Manual”), “[t]he purpose criterion recognizes that an attacker need not wait until [a civilian] object is actually used for military ends before being allowed to attack it as a military objective.” Technically, all civilian data centres (and civilian cyber infrastructures) around the world would satisfy this “purpose” criterion and become legitimate military objectives. This problem was examined in the Tallinn Manual 2.0 (though no common consensus has been reached), which notes that “it may be impossible to know over which part of the network military transmissions, as distinct from civilian ones, will pass. In such cases, the entire network… qualifies as a military objective.” This would essentially lead to a global cyber war.

A narrower interpretation of “military objectives” needs to be adopted which will alleviate

---

55 Henckaerts, *supra* note 49, at 29-32 & 175-182 (acknowledging that the Additional Protocol I states that military objectives are limited to those objects that contribute to military action, and that destruction of public property pursuant to imperative military necessity is allowable).
56 *Id.; AP Commentary, supra* note 32, at ¶ 2020-23.
58 *AP Commentary, supra* note 32, at ¶ 2022.
60 THE PROGRAM ON HUMANITARIAN POLICY AND CONFLICT RESEARCH AT HARVARD UNIV., *HPCR MANUAL ON INTERNATIONAL LAW APPLICABLE TO AIR AND MISSILE WARFARE* 117 (2013) [hereinafter AMW Commentary].
the complexities of distinction vis-à-vis cyber-sphere, and aim to distinguish between civilian and military cyber objects. Unfortunately, States and legal scholars have been expanding the definition of “military objectives”; this was most evident in Operation Enduring Freedom where the United States expanded the definition to include “war-sustaining” objects, and the Commentary on the AMW Manual which drastically expanded “military objectives” to include a “temporal” element (“temporary military objectives by nature”). A new and narrower approach has to be adopted for cyber-distinction. This paper will seek recourse to Articles 56 and 58 of API as new interpretations for “military objectives.”

The entire scope of Article 56 delegitimizes “military objectives,” that they “shall not be made the object of attack, […] if such attack may cause the release of dangerous forces and consequent severe losses among the civilian population.” Analogous to cyber-space, major cyber infrastructures which civilians heavily rely on should not be made the object of attack even if it is a “military objective.” Though, it might, or might not, lead to “severe losses among the civilian population,” the effects of a global-outage of dual-use civilian cyber infrastructures would be to that effect – that the destruction or neutralization of civilian cyber infrastructures would result in significant civilian impact which would outweigh any military benefits. To argue otherwise would be against the object and purpose of IHL, which aims to balance between military necessity and humanity. Hence, to include dual-use civilian cyber objects within the ambit of “military objective” would not be a feasible option as doing so will give preference to military necessity over humanity. That is not to say that every civilian cyber object should not be made the target of attack, doing so will also give preference to humanity over military necessity (and getting State consensus over this narrow interpretation will be tenuous). Article 56(2) gives guidance as to when the “special protection against attack provided by paragraph 1 shall cease.” The ICRC

---

62 See also Geiß & Lahmann, supra note 54, at 390.
64 AMW Commentary, supra note 60, at 109.
65 Geiß & Lahmann, supra note 54, at 381.
66 API, supra note 28, at art. 56.
67 Geiß & Lahmann, supra note 54, at 381, 391.
68 API, supra note 28, at art. 56(2).
Commentary of Paragraph 2 notes that “the decision to deprive them of protection can only be taken at a high military level.”69 Likewise, any operations directed against a dual-use cyber infrastructure must be instructed by the highest military command (notwithstanding the proportionality principle). Also, Article 56(6) outlines, albeit idealistic, instructions for all “High Contracting Parties and the Parties to the conflict […] to conclude further agreements among themselves to provide additional protection for objects containing dangerous forces.”70

Article 58 of API provides the need to segregate “civilian objects […] from the vicinity of military objectives.”71 The same can be said for cyber infrastructures. All States shall endeavour to segregate military cyber infrastructures from civilian cyber infrastructures. This approach (albeit idealistic) is the best way to distinguish between civilian and military cyber objects. Within Article 58(a) (which reflects customary law),72 is the “sense of duty”73 and imposition for States to keep civilian and military cyber objects separated “to the maximum extent feasible.”74 The question therein is, how feasible is it to conduct large-scale segregation of civilian and military cyber objects? According to Droege, “[w]hile it might theoretically be feasible to do this, it would be so […] costly.”75 States who have the available means to conduct large scale segregation should endeavour to do so. Failing which would breach the customary rule in Article 58(a). Furthermore, Article 58(c) lays out the obligation to “take the other necessary precautions to protect the civilian population, individual civilians and civilian objects under their control against the dangers resulting from military operations.”76 This can include guarding civilian property.77 It may be cogently argued that states, vis-à-vis cyber sphere, “to the maximum extent feasible,” will be obligated to ensure continuing cyber functionality despite a cyber attack insofar that cyber infrastructure is crucial for the civilian population.78

69 AP Commentary, supra note 32, at ¶ 2159.
70 API, supra note 28, at art. 56(6).
71 Id. at art. 58.
73 AP Commentary, supra note 32, at ¶ 2247.
74 API, supra note 28, at art. 58.
75 Droege, supra note 12, at 575.
76 API, supra note 28, at art. 58(c).
77 Henckaerts, supra note 49, at 70.
78 Geiß & Lahmann, supra note 54, at 395.
A narrower approach should be adopted when interpreting “military objectives” vis-à-vis cyber space. This approach would be the first step in which States are able to “direct their operations only against military objectives” in cyberspace.

IV. PROPORTIONALITY IN CYBER SPACE

The principle of proportionality is one of the most contentious areas in IHL due to unavoidable civilian deaths, or destruction to civilian objects, as collateral damage in times of an armed conflict. The application of the proportionality principle is mostly settled when it comes to traditional kinetic warfare; the same cannot be said for cyber operations. Given the dual-use nature of most cyber infrastructures, the principle of proportionality is paramount in protecting civilians and civilian objects in the cyber domain. The proportionality principle is found in Article 51(5)(b) of API, which also reflects customary international law applicable in both IACs and NIACs. Under Article 51(5)(b), an attack is prohibited if it “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.”

As mentioned before, civilians can never be made the object of attack. However, the principle of proportionality is an “exception” insofar that civilians, though not an object of attack, are collateral damages as a result of a lawful attack. For a lawful attack to occur, the commander must determine that the death, injury, and damage are not “excessive in relation to the concrete and direct military advantage anticipated.” There has been universal acceptance that the proportionality principle applies to cyber conflicts that constitute attacks, but how it is to be applied remains contentious. For cyber operations that utilize the proportionality principle, two
major elements of the principle deserve a more nuanced understanding and approach: the “damage to civilian objects” threshold in Article 51(5)(b); and the issue of indirect effects.

A. **THRESHOLD OF “DAMAGE TO CIVILIAN OBJECTS”**

The “damage” element has been discussed in Section I, however, that is in the context of “attacks” under Article 49(1) of API. “Damage” in Article 51(5)(b) requires a different appreciation and understanding, though some overlapping principles might apply.

While there is no doubt, that the advent of technological advances might eventually result in “incidental loss of civilian life [and] injury to civilians”\(^88\) *vis-à-vis* cyberwarfare, most reported cyber operations only result in “damage to civilian objects”\(^89\) (i.e. WannaCry, Stuxnet, Flame, NetTraveler, etc.). It is therefore imperative to interpret the “damage” element required for commanders to factor into their proportionality analysis when conducting cyber operations. Does the mere penetration of a cyber payload into dual-use civilian cyber infrastructure constitute “damage”? As discussed in Section I, analogizing from a kinetic attack, if what occurs from a cyber operation results in the same effect as would a kinetic attack, then “damage” has occurred. Commanders (and States) would feel comfortable with this interpretation as applying the proportionality principle in cyberwarfare would be no different to kinetic warfare.

However, many of the reported cyber-operations have not resulted in the same effects as would kinetic force. In fact, almost all of the reported cyber-operations involve only mere modifications of cyber codes into civilian infrastructures.\(^90\) For example, the *Stuxnet* cyber incident involves a single penetration of the payload via a USB thumb-drive into Iranian’s digital servers, which then went on to modify certain programmable logic controllers which resulted in the destruction of nuclear centrifuges.\(^91\) Would the modification of the programmable logic controllers

---

\(^88\) *API*, *supra* note 28, at art. 51(5)(b).

\(^89\) *Id.*


by the Stuxnet malware amount to “damage to civilian objects” (even if, hypothetically, it does not result in the destruction of nuclear centrifuges)? The kinetic analogy approach would limit almost all of the reported cyber operations from the proportionality analysis since “damage to civilian objects” in Article 51(5)(b) is a high threshold test.92

Conversely, some scholars have argued that a mere change or modification in digital codes in a cyber infrastructure would amount to “damage” under the proportionality principle.93 This approach would require military commanders to consider any, and every, effects on a cyber infrastructure in his/her proportionality analysis.94 With respect, this approach is wrong. It fails to understand the fundamentals of the principle of proportionality. The proportionality principle recognizes that collateral damage to civilians and/or civilian objects may occur,95 and setting such low a threshold for “damage” to account for penetrations and modifications of digital codes would be extremely excessive. This approach also adds more burden on military commanders having to take into account every aspect of a cyber operation, and whether the cyber operation would alter/modify a single piece of digital code. Furthermore, not all military commanders are versed in computer science to understand the intricacies of computer systems under their proportionality analysis.

As a corollary to Section 1, this paper takes the view that “damage” encompasses serious interruptions in functionality insofar that “damage” has occurred if the act impairs or hinders the intended function of the object. In Operation Olympic Games, the penetration of the Stuxnet malware which modified the programmable logic controller to the extent that its function to regulate nuclear centrifuges has been impaired,96 constitutes as “damage to civilian objects” in the proportionality calculus. Thus, military commanders of Operation Olympic Games should factor potential loss of functionally of civilian objects into their proportionality calculus before launching the Stuxnet attack. If the functionality approach were to be adopted in cyber conflicts, the traditional kinetic approach would be of limited value. The functionality approach should be the

92 Jensen, Cyber Attacks, supra note 82, at 204-207.
94 Jensen, Cyber Attacks, supra note 82, at 204-207.
95 Jensen, Cyber Attacks, supra note 82, at 208.
96 McMillan, supra note 91.
preferred approach when dealing with the proportionality principle as it is neither too broad (i.e. kinetic approach), nor too narrow (i.e. mere change or modification approach).\footnote{97 See also Jensen, Cyber Attacks, supra note 82, at 208.} Also, the functionality approach is consistent with the general principles of IHL as already discussed in Section I. By focusing on functionality, commanders are able to understand and apply the proportionality principle during a cyber conflict.

B. INDIRECT EFFECTS

The issue of whether indirect effects of an attack should be factored into the proportionality calculus is contentious under IHL given multiple positions put forth by States.\footnote{98 Eric Boylan, Applying the law of Proportionality to Cyber Conflict: Suggestions for Practitioners, 50 VAND. J. TRANSNAT’L L. 217, 234-35 (2017).} Indirect effects are “the delayed and/or displaced second-, third-, and higher-order consequences of action, created through intermediate events or mechanisms.”\footnote{99 JOINT CHIEFS OF STAFF, JOINT PUB. 3-60, JOINT TARGETING I-10 (13 Apr. 2017).} There has been differing State practice on this issue. The United States takes the view that “remote harms resulting from the attack do not need to be considered in a proportionality analysis,”\footnote{100 DEP’T OF DEF., LAW OF WAR MANUAL 242 (June 2015) [hereinafter US DOD LOAC Manual].} while the United Kingdom maintains that “regard must also be had to the foreseeable effects of the attack.”\footnote{101 MINISTRY OF DEFENCE, THE JOINT SERVICE MANUAL OF THE LAW OF ARMED CONFLICT, ¶ 5.33.4, (UK) [hereinafter UK MOD LOAC Manual].} “In any event, there is no dispute that indirect effects cannot be taken into account if they are too remote or cannot be reasonably foreseen.”\footnote{102 AMW Commentary, supra note 60, at 97.} This notion was also echoed by the ICTY Trial Chamber in \textit{Galič}.\footnote{103 Prosecutor v. Galič, Case No. IT-98-29-A, Judgment, ¶58 (Int’l Crim. Trib. For the Former Yugoslavia Nov. 30, 2006).}

As Greenwood notes, the Gulf War of the 1990s highlighted the fact that indirect effects cause more harm to civilians than the direct effects of the attack itself.\footnote{104 Christopher Greenwood, The Law of Weaponry at the Start of the New Millennium, in THE LAW OF ARMED CONFLICT: INTO THE NEXT MILLENNIUM 185, 202 (Michael N. Schmitt & Leslie C. Green eds., 1998).} In cyberwarfare, indirect effect includes damage that was beyond the scope of the intended attack, but results from that attack.\footnote{105 Jensen, Cyber Attacks, supra note 82, at 207.} Unfortunately, most cyber infrastructures are dual-use infrastructures, and the effects of
a cyber-attack might lead to detrimental indirect effects. This was most evident in *Operation Olympic Games* when the *Stuxnet* malware affected multiple civilian infrastructures (i.e. banks) outside of Iran. While there are differing views as to whether indirect effects should be factored into the proportionality calculus in kinetic warfare, with regards to cyberwarfare, this paper believes that it should.

This indirect effect factor is consistent with the words “*may be expected to cause*” in Article 51(5)(b) of API. Schmitt argues that indirect effects which are reasonably foreseeable, no matter the “tier” of effects, must be factored into the proportionality calculus as it is consistent with the wording of “*may be expected.*” Conversely, indirect effects which are not expected to be excessive are excluded from the proportionally calculus. The military commander must have reasonably expected what the indirect effects to be, given the information he/she had at the time of the operation. Surprisingly, despite the United States’ position on the “indirect effect” factor, considerations of “indirect effects” prevailed in a cyber-operation during the 2003 Iraq War, where the United States called off a cyber-operation to disable Saddam Hussein’s financial accounts as the attacks may potentially effect European banking systems and have negative repercussions on the financial markets in Europe.

**V. DIRECT PARTICIPATION IN CYBER HOSTILITIES**

Direct participation in hostilities (“DPH”) is highly problematic for IHL due to the complex nature of the topic, and absent universal acceptance by States and legal scholars as to what amounts to DPH. Determining DPH in traditional warfare is complex, “this is *a fortiori* the case when it...”

comes to cyberwarfare.”

As a general principle, DPH denotes that civilians are not to be made the target of attack, “unless and for such a time as they take a direct part in hostilities.” The rule on non-combat immunity applies insofar as civilians do not take direct participation in hostilities. This principle is reflected in Article 51(3) of API, which is a “valuable reaffirmation of an existing rule of customary international law.” When debating the legal parameters of Article 51 of API, no precise definition of DPH was universally adopted by States. The Commentary to the Additional Protocols notes broadly that, “the immunity afforded individual civilians is subject to an overriding condition, namely, on their abstaining from all hostile acts.” With the lack of a universally accepted definition of DPH, the ICRC began a six-year study into the concept of DPH. The *Interpretive Guidance on the Notion of Direct Participation in Hostilities 2009* ("Interpretive Guidance") serves as a good starting point. Despite not being legally binding, the *Interpretive Guidance* may be considered a useful subsidy source of international law. Furthermore, military manuals have acknowledged that the “recommendations and approaches [in the *Interpretive Guidance*] are helpful,” though some States “has not accepted significant parts of the ICRC’s interpretive guidance as accurately reflecting customary international law.” With the lack of universal State acceptance, one can only theorize *lex ferenda*, how the *Interpretive Guidance* should apply *vis-à-vis* cyberwarfare.

112 *API, supra* note 28, at art. 51(3); Protocol Additional to the Geneva Conventions of 12 August 1949, and relating to the Protection of Victims of Non-International Armed Conflicts (Protocol II) art. 13(3), June 8, 1977, 1125 U.N.T.S. 609.
115 *AP Commentary, supra* note 32, at ¶ 1942.
117 Statute of the International Court of Justice, art 38, ¶ 1(d).
119 *US DOD LOAC Manual, supra* note 100, at ¶ 5.9.1.2.
The Interpretive Guidance notes that DPH shall be defined as a specific act that meets three cumulative elements (Overall-Test):

“(1) the act must be likely to adversely affect the military operations or military capacity of a party to an armed conflict or, alternatively, to inflict death, injury, or destruction on persons or objects protected against direct attack (threshold of harm), and;”

“(2) there must be a direct causal link between the act and the harm likely to result either from that act, or from a coordinated military operation of which that act constitutes an integral part (direct causation), and;”

“(3) the act must be specifically designed to directly cause the required threshold of harm in support of a party to the conflict and to the detriment of another (belligerent nexus).”

With regards to “Temporal Scope of the Loss of Protection,” the Interpretive Guidance adopts the approach in the Commentary to the Additional Protocols that civilians who take a direct part in hostilities will forfeit their civilian immunity for the duration of each act of direct participation (revolving-door approach); this includes “measures preparatory to the execution of a specific act of DPH, as well as the deployment to and the return from the location of its execution, constitute an integral part of the act.” For the purposes of this paper, this Section will not discuss the temporal elements in cyber DPH.

As noted by the Israeli Court in Targeted Killings, “it is possible to take part in hostilities without using weapons at all.” Cyber-attacks were discussed in the Interpretive Guidance, which notes that “electronic interference with military computer network could […] suffice [as DPH].” It remains uncertain as to how the ICRC’s “Overall-Test” applies in cases of

120 Interpretive Guidance, supra note 116, at 46.
121 Id. at 70.
122 Id. at 74-77.
123 Id. at 69.
124 HCJ 769/02 The Public Committee against Torture in Israel v. The Government of Israel [2006] (Isr.).
125 Interpretive Guidance, supra note 116, at 48.
cyberwarfare, this Section will attempt to interpret ICRC’s “Overall-Test” within the context of cyberwarfare.

A. THRESHOLD OF HARM

1. “likely to adversely affect the military operations or military capacity of a party to an armed conflict”

Generally, the “military harm” required is explained broadly as including “essentially any consequence adversely affecting the military operations or military capacity of a party to the conflict.”126 Having regard to cyberwarfare, the Interpretive Guidance states that “electronic interference with military computer networks could also suffice”127 to reach the required threshold of harm. However, the Interpretive Guidance also notes that “the manipulation of computer networks […] may have serious impact on public security, health, and commerce, […] However, they would not, in the absence of adverse military effects, […] qualify as DPH.”128 This turns on whether the electronic interference and/or manipulation results in “adverse military effects,” without which would not trigger the first part of the “threshold of harm” element. To amount to “adverse military effects,” military operations must be affected.129 This is consistent with the requirements in Article 51(2) of API which defines “military objectives.”130 Objects which do not contribute militarily, or grant the adversary a military advantage, fail to qualify. Echoing the sentiments in the Interpretive Guidance, psychological, political or economic advantage or contributions, fail to suffice.131 Using this definition, if a cyber operation does not amount to “adverse military effect[s],” reading consistently with Article 52(2) API, harm has not occurred.

2. “to inflict death, injury, or destruction”

Turning to the alternative part of the “threshold of harm” element, reading literally, if the “harm” amounts to only mere disruptions, it would not meet the threshold required as it would not

---

126 Report Direct Participation in Hostilities 2005, pp. 22 f., 31 cited in Interpretive Guidance, 47.
127 Interpretive Guidance, supra note 116, at 48.
128 Interpretive Guidance, supra note 116, at 50.
129 Id. at 47.
131 Interpretive Guidance, supra note 116, at 46-49.
have inflicted “death, injury, or destruction.” This high threshold is consistent with the purpose of Article 51(3) API (albeit broader in scope than the commentary in Article 51(3)), which gives weight to civilian protection immunity. It is important to note that the threshold of “harm” required, in the context of DPH, is higher than the threshold of “consequential harm” (low-threshold which accounts for disruptions) in the context of “attacks” in Article 49(1) API (discussed in Section I). Though there is a distinction between the level of “harm” required to trigger the respective thresholds, both are aimed at protecting civilians from the harmful effects of war. As noted in the Interpretive Guidance, “the manipulation of computer networks […] will not qualify as DPH.”\(^\text{132}\) It is clear that mere cyber disruptions would not reach the required “threshold of harm.”\(^\text{133}\) If, for example, the cyber operation results in “death, injury, or destruction,” the “threshold of harm” element is triggered, and DPH has occurred, subject to the second and third elements of the overall-test.

B. **Direct Causation**

Unlike the “threshold of harm” criterion, “direct causation” element is subject to numerous debates and conflicting literature expressing differing opinions.\(^\text{134}\) The Interpretive Guidance states that “direct causation should be understood as meaning that the harm in question must be brought about in one causal step […].”\(^\text{135}\) The ICRC’s approach has serious limitations to the cyber domain as most cyber-attacks will be indirect in effect, which is outside the scope of “one causal step.” As noted in a National Research Council Report, “the desired effects of a cyber-attack are almost always indirect, which means that what are normally secondary effects are in fact of central importance.”\(^\text{136}\) The Interpretive Guidance indicates that indirect effects would not fall within the ambit of “one causal step,” and the harm required must be *objectively likely*.\(^\text{137}\) In a cyber-

\(^\text{132}\) *Id.* at 50.
\(^\text{134}\) Schmitt, *Deconstructing Direct Participation in Hostilities*, supra note 130, at 725-35.
\(^\text{135}\) Interpretive Guidance, supra note 116, at 53.
\(^\text{136}\) Turns, *supra* note 111, at 288 (citing William A. Owens et al., Technology, Policy, Law, and Ethics Regarding U.S. Acquisition and Use of Cyberattack Capabilities 127 (William A. Owens et al. eds., 2009)).
\(^\text{137}\) Interpretive Guidance, supra note 116, at 58.
operation, the harm intended (and unintended) are likely to occur over several causal steps. For example, in *Operation Olympic Games*, the harm was brought about over three broad steps; penetration, exploitation, and modification. The single act of penetrating the infected USB thumb drive does not, in and of itself, cause direct harm to the system. Furthermore, the *Stuxnet* malware was routed through several data-streams and compromised computer networks before delivering the inflected payload into the programmable logic controller. This is too far removed from “one causal step” to constitute DPH. In such a circumstance, it remains unclear if cyberwarfare could meet the requirements of direct causation for DPH to occur. This would mean civilians could engage in cyber DPH without impunity.

Schmitt argues for a “but for” standard in his deconstruction of the constitutive elements. However, Schmitt’s approach is too broad; it would extend participation to almost every civilian who has made a causal contribution. Hypothetically, it would be unreasonable to target the cab-driver who drove the passenger(s) carrying the infected thumb-drive to the Nuclear facility in Natanz, as “but for” his/her contribution, nuclear centrifuges would not be destroyed. That said, it remains uncertain as what amounts to “direct causation” in cyberwarfare. With the lack of universal State practice and consensus, “direct causation” will remain to be decided on a case-by-case basis. This “case-by-case” sentiment was also echoed in *Targeted Killings*, *Tadić*, and

---

138*Turns, supra* note 111, at 288.
140*Id.*
141Schmitt, *Deconstructing Direct Participation in Hostilities, supra* note 130, at 727-29.
143HCJ 769/02 The Public Committee against Torture in Israel v. The Government of Israel, ¶ 34-37 [2006] (Isr.).
144Prosecutor v. Tadić, Case No. IT-94-1-T, Opinion and Judgement, ¶ 616 (Int’l Crim. Trib. For the Former Yugoslavia May 7, 1997).
Furthermore, the United States, United Kingdom, and Australian military manuals (and many others) cite the need for a “case-by-case” approach.

C. Belligerent Nexus

The “belligerent nexus” element is the least contentious amongst the three. The Interpretive Guidance is silent on how the “belligerent nexus” element applies in times of cyberwarfare. However, absent guidance, it is relatively straightforward to tease out the legal parameters of “belligerent nexus.” The Interpretive Guidance defines belligerent nexus as an act that “must be specifically designed to directly cause the required threshold of harm in support of a party to the conflict and to the detriment of another.” DPH is only restricted to those individuals whose acts are integral to the hostilities. Not all acts which result in harm can be linked to a party to the conflict as some civilian acts are driven by personal gain. If indirect consequence were intended, which results in foreseeable harm, “belligerent nexus” element would be satisfied, albeit failing the “direct causation” element. If the indirect consequence was neither intended nor foreseeable, but satisfies the “threshold of harm” element, DPH has not occurred as “belligerent nexus” and “direct causation” were not satisfied. The Interpretive Guidance gave the example of a bank robbery, in which a belligerent nexus does not exist as the act committed by the robber was not in support of a party to the conflict even though harm has been caused.

Schmitt argues that the “belligerent nexus” criteria should be defined differently as act in support of or to the detriment of the party (cf. “in support of a party to the conflict and to the detriment of another”). Schmitt’s approach, not surprisingly, was adopted in the Tallinn Manual 2.0 where “belligerent nexus” is interpreted as an act directly related to the hostilities. The AMW Manual

146 The Commander’s Handbook, supra note 63, at ¶ 11.2
147 UK MOD LOAC Manual, supra note 101, at ¶ 5.3.4.
149 Schmitt, Deconstructing Direct Participation in Hostilities, supra note 130, at 735.
150 Interpretive Guidance, supra note 116, at 58.
152 Interpretive Guidance, supra note 116, at 60.
153 Schmitt, Deconstructing Direct Participation in Hostilities, supra note 130, at 736.
154 Schmitt, Tallinn Manual 2.0, supra note 18, at 430.
adopts a similar position. The approach taken in the Tallinn Manual 2.0 would be more appropriate in times of cyberwarfare as there is an immediate nexus between the resulting act and the cyber operation. This author thinks that the nexus required in the Interpretive Guidance might be too far removed in times of cyberwarfare as there is no immediacy factor, it turns on whether the resulting act is “in support of [or detriment of] a party to the conflict.” What is “in support” of and “to the detriment of” a party to the conflict lacks the immediacy factor required for belligerent nexus to be established. It adds a layer of the resulting act being attached to the parties involved, instead of to the resulting harm. Uncontentiously, belligerent nexus could become relevant at an earlier stage of the cyberwarfare, if the malware was written and designed specifically for a particular operation or act. It may seem inconsistent with the Interpretive Guidance, but in times of cyberwarfare, a belligerent nexus can be established before the commission of the hostile act, rather than during, or after its been committed.

However, as before, States have agreed to take a “case-by-case” analysis of DPH. It remains to be seen if the Interpretive Guidance and/or the Tallinn Manual 2.0 will come to fruition in an actual cyber-conflict. The nature of IHL is always changing; one can only apply the law lex lata and theorize how the law should be applied in a given circumstance.

VI. CONCLUSION

This paper has critically examined the interaction between cyberwarfare and IHL. Section II argued how cyber-attacks are “armed conflict[s]” under Article 2 common to the four Geneva Conventions, which brings in the protective mechanisms of IHL. Furthermore, how cyber-attacks are “attacks” under Article 49(1) of API, to which relevant IHL restrictions apply. Section III has examined the difficulties of cyber-distinction due to dual-use objects in cyber-sphere. This paper has argued how a narrower interpretation of “military objectives” needs to be adopted which will alleviate the complexities of distinction vis-à-vis cyber-sphere. Section IV examined how the proportionality principle applies in cyberwarfare, and has critically examined two areas of the proportionality principle. In doing so, this paper argues that the functionality approach should be the preferred method when dealing with the proportionality principle, and that indirect effects

155 AMW Manual, supra note 60, at 15-16.
156 Turns, supra note 111, at 289
157 Id.
should be factored into the proportionality calculus. Lastly, Section V of the paper argued how DPH applies in cyberwarfare. In doing so, this paper has examined the constitutive elements in the *Interpretive Guidance* to interpret non-combat immunity as stipulated in API. This paper can only theorize *lex ferenda*, how the *Interpretive Guidance* should apply *vis-à-vis* cyberwarfare. With the “unprecedented”¹⁵⁸ May 2017 global cyber-attack, States are forced to re-examine their respective cyber-strategy. It is only a matter of time where a global cyber-war occurs. States are urged to adopt multilateral cyber-IHL treaties for the wars of tomorrow.

---