

From Munitions to Malware: A Comparative Analysis of Civilian Targetability in Cyber Conflict

Colton Matheson *

ABSTRACT

Warfare is no longer limited to the bounds of physical space. Modern conflict now takes place not only on land, in air, or at sea, but also in cyberspace. While the international community has established widely accepted conventions governing armed conflict, including those designed to protect civilians, these conventions primarily apply to and envision scenarios embracing conventional warfare. The application of such conventions to conflicts in cyberspace raises novel questions, particularly questions regarding the targetability of civilians directly participating in cyber hostilities. This Article proceeds to examine and compare the proposed standards for direct civilian participation in cyber hostilities as promulgated by the ICRC's Interpretive Guidance and the Tallinn Manual 2.0. Specifically, this Article addresses the course of conduct that a civilian must undertake before he or she is considered a direct participant in a cyber hostility. Upon laying this groundwork, the Article then addresses the issue of duration—the time in which a civilian is considered a direct participant in hostilities and thus, targetable by his or her adversary. In applying the standards of both the ICRC's Interpretive Guidance and the Tallinn Manual 2.0 to hypothetical scenarios involving the direct participation of civilians in cyber hostilities, the Article aligns itself with a position advocated by the Tallinn Manual experts—that addressing the duration of direct civilian participation in cyber hostilities is superior because it properly considers the realities of modern conflict and effectively balances military necessity with humanitarian considerations.

* The author would like to thank Eric Talbot Jensen for his insightful comments and assistance in writing the Article.

I. INTRODUCTION

On July 20, 2008, cybersecurity experts in the United States observed intense Internet traffic directed at servers in former Soviet bloc Georgia.¹ The heavy traffic disrupted and rendered temporarily inoperable several Georgian servers in a cyberattack that is known as a distributed denial of service (or “DDoS”) attack.² The DDoS attack affected Georgian government websites, including the President’s,³ and critical civilian networks that included media and banking systems.⁴ Two weeks later, Russian military forces poured across the Georgian border into South Ossetia. In perfect harmony with Russia’s invasion, websites emerged that encouraged cyber operations against Georgia, identifying specific Georgian websites to attack along with instructions for conducting the attacks.⁵ Although the Russian Federation denied involvement in the cyberattacks that preceded its invasion, convincing evidence exists that indicates the Russian Business Network (RBN), a criminal organization based out of St. Petersburg, Russia, was largely responsible for the attacks.⁶ Further evidence suggests the Kremlin had a hand in the process as well.⁷ The wave of cyberattacks on Georgia in 2008 was the first time in history that a cyberattack coincided with a kinetic attack, and it set a precedent for what many experts agree will become a new era of warfare.⁸

The ubiquity of cyber warfare is taking hold across the world. In June of 2015, the Department of Defense published its long-awaited *Law of War Manual* defining “cyberspace” as an “operational domain in which the armed forces must be able to defend and operate, just like

¹ See John Markoff, *Before the Gunfire, Cyberattacks*, N.Y. TIMES, Aug. 12, 2008, at A1.

² *Id.*

³ David J. Smith, *Russian Cyber Strategy and the War Against Georgia* ATLANTIC COUNCIL (Jan. 17, 2014), <http://www.atlanticcouncil.org/blogs/nato-source/russian-cyber-policy-and-the-war-against-georgia>.

⁴ See generally David J. Smith, *Russian Cyber Capabilities, Policy and Practice*, JEWISH POL’Y CTR., Winter 2014, <https://www.jewishpolicycenter.org/2013/12/31/russian-cyber-capabilities>.

⁵ See Smith, *supra* note 3.

⁶ *Russian Invasion of Georgia*, *supra* note 4, at 5–7.

⁷ *Id.* at 2–10.

⁸ *Id.*; see Markoff, *supra* note 1.

the land, sea, air, and space domains.”⁹ Warfare has entered an era in which the typing of keys on a keyboard could be just as destructive as dropping bombs, albeit leaving much, if not all, of a nation’s physical infrastructure intact. But unlike traditional instruments of warfare that are almost exclusively operated by nation-states, cyber warfare can be, and has been, conducted by civilians far removed from the conflict.¹⁰ This shift in the paradigm of conventional warfare presents unsettling questions about the targeting of civilians in times of conflict. The laws of war are unambiguous when it comes to targeting a civilian who has taken up arms against a hostile national force—but those same laws, though still applicable, are muddled if the civilian is no longer aiming at soldiers through the scope of a rifle, but rather, sending malicious code through a global network of fiber optic cables.

Civilians engaged in cyber conflict raise a number of additional issues that hostile forces must consider before targeting those civilians. Are the civilians directly participating in hostilities? How long have the civilians engaged in a qualifying act of direct participation? How long before and after a cyberattack is a civilian considered to be directly participating in such hostilities? This Article explores these questions and related issues by comparing and contrasting two influential sources of legal thought on the law of war—the ICRC’s *Interpretive Guidance on the Notion of Direct Participation in Hostilities Under International Humanitarian Law*¹¹ and the *Tallinn Manual 2.0 on the International Law Applicable to Cyber Operations*.¹² While both sources address the

⁹ OFFICE OF GEN. COUNSEL, DEP’T OF DEF., DEPARTMENT OF DEFENSE LAW OF WAR MANUAL 1012 (2016) [hereinafter LAW OF WAR MANUAL].

¹⁰ See Shaun Waterman, *Civilian ‘Hacktivists’ Could be Lethal Targets in Cyberwar, NATO Study Says*, WASH. TIMES (Mar. 20, 2013) (emphasis added), <https://www.washingtontimes.com/news/2013/mar/20/civilian-hacktivists-could-be-lethal-targets-cyber> (describing how “[t]he DDoS attacks on [the Republic of] Georgia were led by loosely-knit groups of ultranationalist Russian hackers who coordinated the timing and targeting of their attacks and distributed software that could be downloaded by *volunteer users* so their computers could join the assault”).

¹¹ Int’l Comm. of the Red Cross [ICRC], *Interpretive Guidance on the Notion of Direct Participation in Hostilities Under International Humanitarian Law* (2009) (prepared by Nils Melzer), <https://www.icrc.org/en/doc/assets/files/other/icrc-002-0990.pdf> [hereinafter *Interpretive Guidance*].

¹² TALLINN MANUAL 2.0 ON THE INTERNATIONAL LAW APPLICABLE TO CYBER OPERATIONS (Michael N. Schmitt ed., 2d ed. 2017) [hereinafter TALLINN MANUAL].

issue of direct civilian participation in hostilities, the approaches they advocate are markedly different. This is due to divergent interpretations of an axiom of customary international law, which states that civilians should be protected from direct attack “unless and for such time as they take a direct part in hostilities.”¹³ The Article explores this divide in detail, and the consequences which stem from either interpretation.

The Article begins with a historical analysis of international law concerning the targeting of civilians, and then briefly exams the factors adopted by both the *Interpretive Guidance* and the *Tallinn Manual* for identifying civilian participants in hostilities. The Article dedicates a majority of its analysis on the duration of direct civilian participation in cyber hostilities and the temporal scope of the loss of protections that accompany participation, as proffered by the *Interpretive Guidance* and the *Tallinn Manual*. The Article concludes by contending that the latter interpretation, which limits the targeting of civilians to the duration of their direct participation, is superior to that of the former interpretation, because the *Tallinn Manual* properly considers the realities of modern cyber warfare while effectively balancing sovereign military necessity with humanitarian considerations.

II. TARGETING CIVILIANS IN CYBER CONFLICT

Throughout the brutality of World War II, one source estimates that forty-five million civilians lost their lives worldwide.¹⁴ Although civilians have never been immune from the ravages of war, the depths of suffering caused by the global conflict sent shockwaves through the collective human conscience. Responding to the need for civilian protections during times of war, nations from around the world adopted the Fourth Geneva Convention (or “Civilian Convention”) in 1949.¹⁵

¹³ Protocol Additional to the Geneva Conventions of 12 August 1949, and Relating to the Protection of Victims of International Armed Conflicts art. 1, § 3, June 8, 1977, 1125 U.N.T.S. 3 [hereinafter Protocol I].

¹⁴ *Research Starters: Worldwide Deaths in World War II*, NAT’L WWII MUSEUM, <https://www.nationalww2museum.org/students-teachers/student-resources/research-starters/research-starters-worldwide-deaths-world-war> (last visited Feb. 9, 2018).

¹⁵ Geneva Convention Relative to the Protection of Civilian Persons in Time of War, Aug. 12, 1949, 6 U.S.T. 3516, 75 U.N.T.S. 287; see *Civilians Protected Under*

The Civilian Convention formally provided legal protections for those “not belonging to armed forces or armed groups.”¹⁶

However, as warfare continued to evolve, so did the need for more robust civilian protections. In 1977, the international community again convened to supplement the 1949 Geneva Conventions with Additional Protocols I and II,¹⁷ prescribing clear protections of civilians involved in international armed conflicts (IAC) and non-international armed conflicts (NIAC) respectively.¹⁸

Specific as to the targeting of civilians, Articles 50 and 51 of Protocol I provide guidance concerning the scope of its protections. Article 50 defines “civilian,” clarifying to whom its protections apply,¹⁹

International Humanitarian Law, ICRC (Oct. 29, 2010), <https://www.icrc.org/eng/war-and-law/protected-persons/civilians/overview-civilians-protected.htm>.

¹⁶ *Id.*

¹⁷ See Protocol I, *supra* note 13, at 4 n.1 (entered into force Dec. 7, 1978); Protocol Additional to the Geneva Conventions of 12 August 1949, and Relating to the Protection of Victims of Non-International Armed Conflicts, June 8, 1977, 1125 U.N.T.S. 609, 610 n.1 [hereinafter Protocol II] (entered into force Dec. 7, 1978).

¹⁸ Compare Protocol I, *supra* note 13; with Protocol II, *supra* note 17.

¹⁹ See Protocol I, *supra* note 13, at 26 (citing the Geneva Convention Relative to the Treatment of Prisoners of War art. 4, Aug. 12, 1949, 6 U.S.T. 3316, 75 U.N.T.S. 135 [hereinafter Third Convention]), for the complete definition:

Article 50. DEFINITION OF CIVILIANS AND CIVILIAN POPULATION. 1. A civilian is any person who does not belong to one of the categories of persons referred to in Article 4 (A) (1), (2), (3) and (6) of the Third Convention and in Article 43 of this Protocol. In case of doubt whether a person is a civilian, that person shall be considered to be a civilian.

2. The civilian population comprises all persons who are civilians.

3. The presence within the civilian population of individuals who do not come within the definition of civilians does not deprive the population of its civilian character.

See also Third Convention art. 4, Aug. 12, 1949, 6 U.S.T. 3316, 75 U.N.T.S. 135, defining persons considered to be “prisoners of war” as:

- (1) Members of the armed forces of a Party to the conflict, as well as members of militias or volunteer corps forming part of such armed forces.
- (2) Members of other militias and members of other volunteer corps, including those of organized resistance movements, belonging to a Party to the conflict and operating in or outside their own territory, even if this territory is occupied, provided that such militias or volunteer corps, including such organized resistance movements, fulfill the following conditions:

while Article 51 outlines a variety of actions which civilians and the civilian population are to be protected against.²⁰ However, Section 3 to Article 51 conditions civilian protections by providing: “Civilians shall enjoy the protection afforded by this Section, *unless and for such time as they take direct part in hostilities*.”²¹ Nearly identical language is found in Section 3 to Article 13 of Protocol II.²² Thus, Article 51(3) and Article 13(3) clarify that the protections guaranteed to civilians in Protocols I and II do not extend to civilians “for such time” as they are “directly participating in hostilities.”

Although there has been some debate surrounding the scope of Articles 51(3) and 13(3) for some time, the advent of cyberspace and its battlefield applications have compounded the issue. Unfortunately, neither protocol defines the conduct that a civilian must engage to “directly participate in hostilities” nor do they elaborate on the duration of the temporal considerations that must be addressed in determining whether a civilian is protected by Protocol I or II. To better understand what constitutes a civilian’s “direct participation in hostilities,” as well as the duration during which a civilian may be said to be directly participating in hostilities, the Article proceeds by analyzing these two components in turn.

-
- (a) that of being commanded by a person responsible for his subordinates;
 - (b) that of having a fixed distinctive sign recognizable at a distance;
 - (c) that of carrying arms openly;
 - (d) that of conducting their operations in accordance with the laws and customs of war.
- (3) Members of regular armed forces who profess allegiance to a government or an authority not recognized by the Detaining Power.
- ...
- (6) Inhabitants of a non-occupied territory, who on the approach of the enemy spontaneously take up arms to resist the invading forces, without having had time to form themselves into regular armed units, provided they carry arms openly and respect the laws and customs of war.

²⁰ See Protocol I, *supra* note 13, at 26–27.

²¹ *Id.* art. 51(3) (emphasis added).

²² Protocol II, *supra* note 17, at 615 (emphasis added).

A. Direct Participation in Hostilities

On the modern battlefield, it can be difficult to delineate when a civilian's actions breach the realm of acceptable civilian conduct and enter into direct participation in hostilities. A civilian shooting at hostile forces or secreting an improvised explosive device ("IED") on the roadside is almost assuredly directly participating in hostilities, but what if instead that civilian engaged in a cyber operation against the hostile force? Does a cyber intrusion amount to direct participation in hostilities in the same manner as engaging in kinetic actions?

Both the *Tallinn Manual* and the *Interpretive Guidance* provide instruction concerning whether civilian conduct should be classified as directly participating in hostilities and consider substantially similar criteria.²³ Although every nation does not necessarily accept the criteria proffered by these two sources as the proper standard for analyzing civilian direct participation in hostilities, the following criteria are part of a widely accepted and applied study of civilian direct participation in hostilities.²⁴

²³ The qualifying language of the *Tallinn Manual* and the *Interpretive Guidance* vary slightly regarding the actions that satisfy the "threshold of harm" requirement for direct participation in hostilities by civilians. While slight, these subtle nuances have serious implications. See Collin Allan, *Direct Participation in Hostilities from Cyberspace*, 54 VA. J. INT'L L. 173 (2013), for an in-depth discussion on the issue of the duration of time in which a civilian directly participation in hostilities.

²⁴ See LAW OF WAR MANUAL, *supra* note 9, at 228–29 (2016) (internal citations omitted) for further discussion:

At a minimum, taking a direct part in hostilities includes actions that are, by their nature and purpose, intended to cause actual harm to the enemy. Taking a direct part in hostilities extends beyond merely engaging in combat and also includes certain acts that are an integral part of combat operations or that effectively and substantially contribute to an adversary's ability to conduct or sustain combat operations. However, taking a direct part in hostilities does not encompass the general support that members of the civilian population provide to their State's war effort, such as buying war bonds.

Whether an act by a civilian constitutes taking a direct part in hostilities is likely to depend highly on the context, such as the weapons systems or methods of warfare employed by the civilian's side of the conflict.

See also MINISTRY OF DEFENCE, THE JOINT SERVICE MANUAL OF THE LAW OF ARMED CONFLICT 53–54 (2004) (providing a perspective on the United Kingdom's legal approach to civilian direct participation in hostilities).

In determining whether a civilian is directly participating in hostilities, both the *Tallinn Manual 2.0* and the *Interpretive Guidance* consider three cumulative requirements: “(1) a threshold regarding the harm likely to result from the act, (2) a relationship of direct causation between the act and the expected harm, and (3) a belligerent nexus between the act and the hostilities conducted between the parties to an armed conflict.”²⁵ The foregoing requisite criteria are often simplified as (1) threshold of harm, (2) direct causation, and (3) belligerent nexus. All three criteria must be present before a civilian will be designated as directly participating in hostilities.

1. Threshold of Harm

According to the *Interpretive Guidance*, civilian actions weigh in favor of a finding of direct participation in hostilities when they are “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.”²⁶ Since modern cyber operations are almost exclusively relegated to the digital realm, cyberattacks rarely result in death, injury, or destruction to persons or objects, however, this is also not unprecedented.²⁷ More commonly, a cyber operation will negatively affect an adversary’s “military operations or military capacity.”²⁸

As previously mentioned, in order to meet the threshold of harm requirement under the *Interpretive Guidance*, a civilian 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.”²⁹ Conversely, the *Tallinn Manual* states that to satisfy the requirement, cyber operations need only “have the *intended or actual effect* of negatively affecting” enemy systems or causing harm to enemy objects

²⁵ *Interpretive Guidance*, *supra* note 11, at 46.

²⁶ *Id.* at 47.

²⁷ *Id.*; see Josh Fruhlinger, *What is Stuxnet, Who Created It and How Does It Work?*, CSO (Aug. 22, 2017, 2:29 AM), <https://www.csoonline.com/article/3218104/malware/what-is-stuxnet-who-created-it-and-how-does-it-work.html>.

²⁸ *Interpretive Guidance*, *supra* note 11, at 47.

²⁹ *Id.* (emphasis added).

or personnel.³⁰ While the wording is similar among the two sources, the intent-based threshold of harm criteria adopted by the *Tallinn Manual* alters the temporal considerations for determining if and when civilians may be considered directly participating in hostilities.

Collin Allan notes the significance of the *Tallinn Manual*'s inclusion of the intent-based criteria for determining whether a cyber operation meets the threshold of harm. Allan indicates that the effect of the intent requirement is to lower "the bar for civilian actions that may satisfy [the threshold of harm] requirement."³¹ Therefore, according to the *Tallinn Manual*'s characterization, "[w]hen a civilian intends to negatively affect the adversary's military operations or capabilities, even if the attack is not likely to be carried into effect, it will result in the civilian being targetable by the victim of the attack."³² In short, he argues the threshold of harm requirement under the *Tallinn Manual* "focuses on the civilian's intent to carry out the attack," whereas the *Interpretive Guidance* requirement "applies an objective likelihood of the attack standard."³³

The *Tallinn Manual* specifies a cyber operation disrupting an "enemy's command and control network" meets the threshold of harm requirement.³⁴ However, the intent requirement of the *Tallinn Manual* seems to further suggest that an unsuccessful operation *intended* to disrupt the enemy command and control network would also satisfy its threshold of harm requirement. Yet, to meet the same threshold under the *Interpretive Guidance*, a cyber operation intended to disrupt an enemy command and control network must also be objectively likely to achieve its intended purpose.³⁵ Thus, it is not enough for a civilian to engage in a cyber operation that is subjectively intended to cause harm to an enemy—the cyber operation must also be objectively likely to "adversely affect the military operations or military capacity of a party to an armed conflict" or "to inflict death, injury, or destruction on persons or objects protected against direct attack."³⁶

³⁰ TALLINN MANUAL, *supra* note 12, at 429 (emphasis added).

³¹ See Allan, *supra* note 23 at 182.

³² *Id.*

³³ *Id.*

³⁴ TALLINN MANUAL, *supra* note 12, at 429.

³⁵ See *id.*

³⁶ *Interpretive Guidance*, *supra* note 11, at 47.

Under these two regimes, the inclusion of intent-based criteria creates a structure in which it is easier for a civilian hacker to meet the threshold of harm requirement under the *Tallinn Manual* than under the *Interpretive Guidance*. Nevertheless, a civilian that meets the threshold will still not be considered as directly participating in hostilities unless the two other elements of the analyses are met.³⁷

2. Direct Causation

The impact that civilians have on modern conflict is enormous. Whether acting as medical personnel in military hospitals, maintaining critical infrastructure, or distributing the food and fuel required for a military to achieve its objectives, civilians aid militaries throughout the conduct of hostilities. These civilian aids perform vital functions that, without which, a military or armed group would be unable to carry out its hostile intent effectively. But generally speaking, civilians engaged in the aforementioned conduct or similar roles are not directly causing harm to the enemy.

For a civilian actions to be considered as directly causing harm, “there must be a direct causal link between a specific 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.”³⁸ Direct causation occurs when the actions of the civilian bring about the “harm in question . . . in one causal step.”³⁹ Thus, civilian actions that “merely build[] up or maintain[] the capacity of a party to harm its adversary, or which otherwise only indirectly cause[] harm, [are] excluded from the concept of direct participation in hostilities.”⁴⁰

In the context of cyber operations, a civilian that writes malware and subsequently disseminates it online may not be considered to be directly participating in hostilities since the malware is not specifically intended for any particular party. Even if a military or armed group may

³⁷ It should also be noted some *Tallinn Manual* experts consider civilian cyber conduct that “enhances one’s own military capacity” to satisfy the threshold of harm because it diminishes the adversary’s relative power; the *Interpretive Guidance* is silent on this issue.

³⁸ *Interpretive Guidance*, *supra* note 11, at 51.

³⁹ *Id.* at 53.

⁴⁰ *Id.*

use the civilian's malware to conduct hostilities against another armed group or nation, the civilian's conduct is not the direct cause of that harm.⁴¹ However, if a civilian's actions "constitute[] an integral part"⁴² of a military operation, then the civilian will be considered to have directly caused the harm inflicted by the operation. For example, if a civilian had been explicitly recruited to write the malware to exploit vulnerabilities in enemy systems, her actions will be viewed as an integral part of directly causing the harm because her actions make possible a specific attack.⁴³

A civilian's causal proximity to harm should not be confused with their temporal or geographic proximity.⁴⁴ The modern battlefield increasingly removes participants from the geographic and temporal proximity of the actual conflict, particularly so in cyber operations. Although a civilian may be geographically or temporally separated from the actual harm, the considerations of direct causation within the larger context of direct participation in hostilities are exclusively focused on the causal proximity of a civilian's actions to the harm. Thus, where a civilian deploys a cyber weapon against an enemy on another continent "[t]he causal relationship between the employment of [the cyber weapon] and the ensuing harm remains direct regardless of . . . geographical proximity."⁴⁵ The same holds true for a civilian that plants a "logic bomb" within an enemy system designed to unleash malicious code at a certain future date or at the execution of a certain system function.⁴⁶ The fact that the civilian may have planted the logic bomb weeks, months, or even years before its deployment does not necessarily diminish the civilian's direct causal proximity to the harm.

Therefore, where a civilian's actions are directly linked to the harm likely to result from those actions or constitute an integral part of a military operation designed to cause harm reaching the required threshold discussed in the previous section, that civilian is considered to have directly caused the harm regardless of geographical or temporal considerations.

⁴¹ TALLINN MANUAL, *supra* note 12, at 430.

⁴² *Interpretive Guidance*, *supra* note 11, at 51.

⁴³ *See* TALLINN MANUAL, *supra* note 12, at 430.

⁴⁴ *See Interpretive Guidance*, *supra* note 11, at 55.

⁴⁵ *Id.*

⁴⁶ *See* TALLINN MANUAL, *supra* note 12, at 431.

3. Belligerent Nexus

In addition to satisfying the requisite threshold of harm and direct causal link, a civilian act must also meet the belligerent nexus requirement. In order for a civilian act to amount to direct participation in hostilities under the ICRC's *Interpretive Guidance*, the act "must not only be *objectively likely* to inflict harm that meets the first two criteria, but it must also be *specifically designed to do so in support of a party to an armed conflict and to the detriment of another*."⁴⁷ In what facially appears to be an effort to redefine the belligerent nexus requirement, the *Tallinn Manual* simply states that civilian acts amounting to a belligerent nexus need only "be directly related to the hostilities."⁴⁸ However, in prescribing this rationale, the *Tallinn Manual* explicitly cites to the *Interpretive Guidance* requirements.⁴⁹ Thus, the ICRC's *Interpretive Guidance* and the *Tallinn Manual* actually agree on the component elements that underly the belligerent nexus requirement, despite the *Tallinn Manual*'s revised description.

But the belligerent nexus requirement should not be confused with subjective or hostile intent. While subjective or hostile intent concern the state of mind of the actor, the belligerent nexus focuses solely on the objective purpose of the act.⁵⁰ Therefore, in evaluating the belligerent nexus of a civilian action, military commanders must look to the purpose and design of the civilian's act, and not the "mindset of every participating individual."⁵¹

Determining the belligerent nexus of a civilian action can be difficult. Commanders must look to "information reasonably available" which is "deduced from objectively verifiable factors" before making targeting decisions concerning civilians.⁵² Ultimately, "the decisive question should be whether the conduct of a civilian, in conjunction with the circumstances prevailing at the relevant time and place, can reasonably be perceived as an act designed to support one party to the conflict by directly causing the required threshold of harm to another

⁴⁷ *Interpretive Guidance*, *supra* note 11, at 58.

⁴⁸ TALLINN MANUAL, *supra* note 12, at 430.

⁴⁹ *Id.* (citing *Interpretive Guidance*, *supra* note 11, at 58).

⁵⁰ See *Interpretive Guidance*, *supra* note 11, at 59–60.

⁵¹ *Id.* at 59.

⁵² *Id.* at 63.

party.”⁵³ Hence, considering the example of a civilian writing malware above, where the civilian is arbitrarily posting the malware online with no directions to attack specific targets, there would be no belligerent nexus since the act is not designed to support any one party. However, where the civilian specifically authors the malware in order to exploit a deficiency in the enemy’s system, there would be a belligerent nexus because the civilian’s actions are designed to support one party of a conflict to the detriment of the other.

4. Summary

Whether a civilian’s actions in a cyber operation satisfy the requisite threshold of harm depends upon which interpretation the State applies. As mentioned previously, where a civilian intends her actions to affect negatively enemy systems, they meet the threshold of harm under the *Tallinn Manual*. But intent alone is insufficient under the ICRC’s *Interpretive Guidance*. Under the latter, a civilian’s conduct must be objectively likely to negatively affect an enemy systems. Nevertheless, where a civilian’s actions satisfy the threshold under either interpretation, directly cause harm, and are designed to support one party to a conflict over another, that civilian is directly participating in hostilities and therefore, targetable. Thus, in considering whether a civilian cyber intrusion may amount to direct participation in hostilities in the same manner as engaging in kinetic actions, the answer is clearly “yes.” And for such time as a civilian is directly involved in those cyber hostilities, the civilian will remain targetable by the enemy. But exactly when a civilian directly participating in hostilities becomes targetable, how long the target remains, and when traditional civilian protections return, is a matter of dispute.

B. Duration of Participation and the Temporal Scope of the Loss of Civilian Protections Relating to Targetability

Although the *Tallinn Manual* and the *Interpretive Guidance* apply the same three criteria (albeit with slight variations) in evaluating whether a civilian is directly participating in hostilities, there is no consensus regarding the temporal limitations during which a civilian is

⁵³ *Id.*

considered to be directly participating in hostilities. The ICRC's *Interpretive Guidance* states that civilians directly participating in hostilities "lose their protection for the duration of each specific act amounting to direct participation in hostilities."⁵⁴ While some of the *Tallinn Manual* experts agree with the *Interpretive Guidance*, others argue that a civilian's direct participation in hostilities begins "with the [civilian's participation in the] first such cyber operation and continues throughout the period of intermittent activity."⁵⁵

In support of its approach on this issue, the ICRC's *Interpretive Guidance* points to customary international law.⁵⁶ However, in a study on customary international law conducted by the ICRC itself, the ICRC underscored that, under customary international law, civilians shall not be targeted "unless and for such time as they are directly participating in hostilities."⁵⁷ This language itself is undisputed. What *is* disputed, however, is the interpretation of "for such time as."

The primary split among the approached taken by the ICRC's *Interpretive Guidance* and the *Tallinn Manual* is the duration of time during which such a civilian is targetable.

1. The ICRC's Approach to the Temporal Scope of the Loss of Civilian Protections

The ICRC's *Interpretive Guidance* takes the position that the phrase "unless and for such time" means that any "suspension of protection lasts exactly as long as the corresponding civilian engagement in direct participation in hostilities. This necessarily entails that civilians lose and regain protection against direct attack in parallel with the intervals of their engagement in direct participation in hostilities (so-called 'revolving door' of civilian protection)."⁵⁸ The ICRC defends this interpretation by pointing to customary international

⁵⁴ *Id.* at 73.

⁵⁵ TALLINN MANUAL, *supra* note 12, at 432.

⁵⁶ See *Interpretive Guidance*, *supra* note 11, at 73.

⁵⁷ Jean-Marie Henckaerts, *Study on Customary International Humanitarian Law*, 87 INT'L REV. RED CROSS, no. 857, Mar. 2005, at 175, 198.

⁵⁸ *Interpretive Guidance*, *supra* note 11, at 70.

law,⁵⁹ but as noted above, the language actually accepted as customary international law is, in isolation, unclear, and up for interpretation.

Yet, in an almost antithetical position to that articulated above, the ICRC also concedes that the period of “engagement” in a qualifying act of direct participation in hostilities may extend beyond the period during which the civilian performs the actual hostile act to include activities that form an “integral part” of the operation.⁶⁰ This includes actions taken before and after the actual act of direct participation, such as deployment to, and returning from, the location of the hostile act.⁶¹ The ICRC’s *Interpretive Guidance* maintains this extension of civilian targetability when addressing the targetability of civilians involved in cyber operations:

Where the execution of a hostile act does not require geographic displacement, as may be the case with computer network attacks . . . *the duration of direct participation in hostilities will be restricted to the immediate execution of the act and preparatory measures forming an integral part of that act.*⁶²

Although facially expansive, this language is astonishingly limiting, since the immediate execution of a cyberattack may take only seconds and preparatory measures may be largely conducted before the actual hostilities commence.

The ICRC adopted the “revolving door” approach based on the premise that “it prevents attacks on civilians who do not, at the time, represent a military threat.”⁶³ From the ICRC perspective, the conduct of civilians varies widely and therefore, cannot be readily anticipated.⁶⁴ This perspective differs substantially from that of a number of the *Tallinn Manual* experts, who view civilians that repeatedly directly participate in hostilities as targetable “throughout the period of intermittent activity.”⁶⁵ But for the ICRC, “[e]ven the fact that a civilian

⁵⁹ See *id.*

⁶⁰ *Id.* at 67.

⁶¹ *Id.*

⁶² *Id.* at 68 (emphasis added).

⁶³ *Id.* at 70.

⁶⁴ See *id.* at 70–71.

⁶⁵ TALLINN MANUAL, *supra* note 12, at 432.

has repeatedly taken a direct part in hostilities, either voluntary or under pressure, does not allow a reliable prediction as to future conduct.”⁶⁶

The ICRC understood that this “revolving door” approach may make it difficult for opposing armed forces from an operational standpoint, but maintained its position, claiming that “it remains necessary to protect the civilian population from erroneous or arbitrary attack.”⁶⁷ According to the ICRC, the only civilians who lose their civilian protections against direct attack on a continuous basis are those civilians who assume a “continuous combat function” as a member of an “organized armed group belonging to a party to the conflict.”⁶⁸ Under such circumstances, the participating civilian’s traditional protections are suspended “for the duration of their membership, that is to say, for as long as they assume their continuous combat function.”⁶⁹ However, Cordula Droege, the head of the ICRC’s Operational Law Unit, has written that “[f]or a group to qualify as an organised armed group that can be a party to a conflict within the meaning of [International Humanitarian Law], it needs to have a level of organisation that allows it to carry out sustained acts of warfare and comply with IHL.”⁷⁰ Although some state-sponsored cyber operations groups may fall within this definition, it is a virtual certainty that individual hackers, small “hacktivist” groups, and even larger groups of hackers who may be just as capable of causing harm or interference to a nation’s vital systems, will be excluded from the narrow definition of an organized armed group. Therefore, these less organized groups will be protected from continuous targeting, and instead, will benefit from the “revolving door” of civilian protections. By limiting the continuous suspension of civilian protections against direct attack to only those civilians involved in organized armed groups and promoting the concept of a “revolving door” of protections for civilians directly participating in hostilities, the ICRC’s *Interpretive Guidance* fails to

⁶⁶ *Interpretive Guidance*, *supra* note 11, at 71.

⁶⁷ *Id.*

⁶⁸ *Id.* at 73

⁶⁹ *Id.*

⁷⁰ Cordula Droege, *Get Off My Cloud: Cyber Warfare, International Humanitarian Law, and the Protection of Civilians*, 94 INT’L REV. RED CROSS, no. 886, Summer 2012, at 533, 550.

address the nuances of the modern battlefield that are more thoroughly addressed in the *Tallinn Manual*.

2. The *Tallinn Manual*'s Approach to the Temporal Scope of the Loss of Civilian Protections

Like the ICRC's *Interpretive Guidance*, the *Tallinn Manual* group of experts accepts that direct participation in hostilities includes actions "immediately preceding or subsequent to the qualifying act."⁷¹ Although all of the *Tallinn Manual* experts accept that such acts include travel to and from the geographic location of the direct participation in hostilities, some experts also advocate the interpretation proposed by international law scholar Yoram Dinstein.⁷² According to Dinstein, "in demarcating the relevant time span in the course of which a civilian is directly taking part in hostilities, it is necessary to go as far as is reasonably required both 'upstream' and 'downstream' from the actual engagement."⁷³ The *Tallinn Manual* experts slightly altered Dinstein's interpretation, stating that "the period [during which a civilian directly participating in hostilities lost civilian protections] extended as far 'upstream' and 'downstream' as a causal link existed."⁷⁴ It should be noted that, although the Article employs the phrase "*Tallinn Manual* interpretation" for ease in distinguishing the various interpretations, only "[s]ome of the [e]xperts" agreed with the modified Dinstein interpretation, which could indicate that only a small minority of the *Tallinn Manual* experts actually subscribed to the idea.⁷⁵

The Dinstein interpretation is particularly significant in the context of cyber operations, where the actual duration of a civilian's engagement in a qualifying act may be fleeting. The execution of a cyberattack may be as easy as clicking a mouse. Under the ICRC's *Interpretive Guidance*, the civilian involved would likely only be

⁷¹ TALLINN MANUAL, *supra* note 12, at 431

⁷² *Id.*

⁷³ YORAM DINSTEIN, *THE CONDUCT OF HOSTILITIES UNDER THE LAW OF INTERNATIONAL ARMED CONFLICT* 177 (3d ed. 2016).

⁷⁴ TALLINN MANUAL, *supra* note 12, at 431 (emphasis added).

⁷⁵ *Id.*

targetable for a few moments.⁷⁶ However, applying the *Tallinn Manual* interpretation extends the period in which a civilian may lawfully be targeted. More specifically, under the *Tallinn Manual*, the loss of such protections may “begin once an individual began probing the target system for vulnerabilities, extend throughout the duration of activities against that system, and include the period during which damage is assessed to determine whether ‘re-attack’ is required.”⁷⁷ Applying the *Tallinn Manual* interpretation provides additional time for military commanders to investigate cyber operations and to take action against any civilian taking a direct part in cyber hostilities than the more narrow interpretation of the ICRC’s *Interpretive Guidance*.

III. THE TALLINN MANUAL APPROACH PROPERLY CONSIDERS THE REALITIES OF MODERN CONFLICT AND EFFECTIVELY BALANCES MILITARY NECESSITY WITH HUMANITARIAN CONSIDERATIONS

The narrow interpretation of the ICRC’s *Interpretive Guidance* concerning the duration of the loss of civilian protections against direct attack is insufficient and ineffective in the context of cyber operations. Whereas, the *Tallinn Manual* specifically contemplates cyber warfare and effectively tailors its interpretation to logically address the realities of modern conflict while simultaneously balancing military necessity with humanitarian considerations.

A. *The Tallinn Manual Interpretation Properly Considers the Realities of Modern Conflict*

Civilians are becoming progressively more common on the modern battlefield. Armies no longer frequently meet on distant fields, but rather in the streets of occupied towns and cities where civilian and combatant intermingle. This increasingly familiar dynamic has led to alterations in the manner in which wars are fought, particularly against

⁷⁶ It is unclear what the ICRC’s *Interpretive Guidance* considers as an “integral part of the act” and thus, the application of the ICRC’s interpretation may lead to civilian targetability of a more significant duration. See *ICRC Interpretive Guidance* at 58.

⁷⁷ TALLINN MANUAL, *supra* note 12, at 431.

non-state actors such as organized armed groups,⁷⁸ unprivileged belligerents,⁷⁹ as well as civilians directly participating in hostilities. Therefore, the manner in which nation-states address these threats is becoming more relevant to operational success. The *Tallinn Manual* interpretation regarding the targeting of civilians directly participating in hostilities addresses these modern threats pragmatically. The Article proceeds by examining the merits of the *Tallinn Manual* and the *Interpretive Guidance* interpretations, and comparing the implications that each presents under two scenarios likely to arise within the context of a cyber operation—a civilian engaged in repeated attacks and a civilian conducting a delayed effects attack.

1. The Repeat Offender

Under the *Tallinn Manual* interpretation, direct participation, and thus targetability, of a civilian engaged in repeated cyberattacks “begins with the first such cyber operation and continues throughout the period of intermittent activity.”⁸⁰ Thus, if a civilian hacker conducts multiple cyberattacks against an opposing force over the course of a month, the hacker would be targetable for the entire month.⁸¹

⁷⁸ Organized armed groups, according to Protocol II “must be ‘under responsible command, [and] exercise such control over a part of [a State’s] territory as to enable them to carry out sustained and concerted military operations and to implement this Protocol.’” See Peter Margulies, *Networks in Non-International Armed Conflicts: Crossing Borders and Defining “Organized Armed Group”*, 89 INT’L L. STUD. 54, 56 (2013) (alterations in original) (quoting Protocol II, *supra* note 18, art. 1(1)).

⁷⁹ See LAW OF WAR MANUAL, *supra* note 9, at 103, which in part:

Although seldom explicitly recognized as a class in law of war treaties, the category of unprivileged belligerent may be understood as an implicit consequence of creating the classes of lawful combatants and peaceful civilians. The concept of unprivileged belligerency, *i.e.*, the set of legal liabilities associated with unprivileged belligerents, may be understood in opposition to the rights, duties, and liabilities of lawful combatants and peaceful civilians. Unprivileged belligerents include lawful combatants who have forfeited the privileges of combatant status by engaging in spying or sabotage, and private persons who have forfeited one or more of the protections of civilian status by engaging in hostilities.

⁸⁰ TALLINN MANUAL, *supra* note 12, at 432.

⁸¹ See *id.*

Conversely, the *Interpretive Guidance* insists that, “as long as such participation occurs on a merely spontaneous, unorganized or sporadic basis,” the hacker is entitled to civilian protections against direct attack.⁸² Even where a civilian hacker engages in hostile acts on a “persistently recurrent basis,” the hacker retains that protection.⁸³ Considering the repeat hacker example above, under the *Interpretive Guidance* interpretation, even though the hacker engaged in multiple attacks during the month, the hacker would only be targetable while conducting each separate attack and would be protected from direct targeting in the interim.⁸⁴

According to the *Interpretive Guidance*, the only circumstance under which a civilian may be continuously targeted occurs when a civilian joins an organized armed group (OAG) that belongs to a party to the conflict.⁸⁵ Membership in an OAG “begins the moment when a civilian starts *de facto* to assume a continuous combat function for the group” and endures until the cessation of such function.⁸⁶ Essentially, the “revolving door” of civilian protections operates upon status.⁸⁷

However, as Ken Watkin observed, the *Interpretive Guidance* fails to provide “clear guidance on the number of times a civilian can walk back through the ‘revolving door.’”⁸⁸ Instead, it suggests that “a civilian can go through the revolving door on a ‘persistently recurring basis,’” repeatedly suspending and then reacquiring civilian protections against direct attack. However, the *Interpretive Guidance* also claims that a civilian qualifies for membership in an OAG and thus continual targeting by an opposition force, when the “civilian ‘starts *de facto* to assume a continuous combat function for the group.’”⁸⁹ Despite the significant implications of each, the *Interpretive Guidance* makes no meaningful distinction between what it means to engage in hostilities

⁸² *Interpretive Guidance*, *supra* note 11, at 71.

⁸³ *Id.* at 44.

⁸⁴ See TALLINN MANUAL, *supra* note 12, at 432.

⁸⁵ See *Interpretive Guidance*, *supra* note 11, at 72.

⁸⁶ *Id.*

⁸⁷ See *id.*

⁸⁸ Kenneth Watkin, *Opportunity Lost: Organized Armed Groups and the ICRC “Direct Participation in Hostilities” Interpretive Guidance*, 42 N.Y.U. J. INT’L L. & POL. 641, 661 (2010).

⁸⁹ *Id.*

on a “persistently recurring basis” or to undertake a “continuous combat function.” Under the *Interpretive Guidance*, members of an OAG may be targeted continuously, while those civilians directly participating in hostilities are granted a “revolving door” of civilian protections.⁹⁰ In short, the *Interpretive Guidance* proposes significantly greater protections to civilians directly participating in hostilities than to members of an OAG without properly defining when a civilian’s engagement in combat functions on a persistently recurring basis crosses the line into a continuous combat function and *de facto* membership in an OAG. Thus, the *Interpretive Guidance* bestows gratuitous advantages to civilians persistently participating in hostilities and facilitates a legal regime in which such a civilian may be a “farmer by day, fighter by night,”⁹¹ effectively creating “a form of immunity for insurgents.”⁹²

For example, during the Russian invasion of Georgia in August of 2008, a massive cyber operation was simultaneously ongoing against Georgian cyber infrastructure.⁹³ Several websites appeared, such as “stopgeorgia.ru,” listing websites to attack with instructions on how to complete the cyberattacks and including an “after-action report page.”⁹⁴ The purpose of these websites was ostensibly to encourage civilian participation in the Georgian cyberattacks. Assuming the pro-Russian websites encouraging the cyberattacks were successful in influencing Russian citizens to join in, it is undeniable that under the provisions of the *Interpretive Guidance*, these citizens were directly participating in hostilities. Their attacks were likely to cause harm to Georgian military operations, their actions were directly linked to the harm, and it was in support of the Russian invasion of Georgia. However, under the *Interpretive Guidance*’s “revolving door” approach, “the suspension of protection lasts exactly as long as the corresponding civilian engagement in the direct participation in hostilities.”⁹⁵ Therefore, each of these Russian civilians could go to work during the day as a civilian protected from direct attack and then return home in the evening to

⁹⁰ See *Interpretive Guidance*, *supra* note 11, at 44–45.

⁹¹ See *id.* at 72.

⁹² Watkin, *supra* note 88, at 689.

⁹³ See Smith, *supra* note 3.

⁹⁴ *Id.*

⁹⁵ *Interpretive Guidance*, *supra* note 11, at 70.

conduct cyberattacks against Georgian websites where they would then become subject to attack, and in effect, creating farmers by day and fighters by night.

More pragmatically, the *Tallinn Manual* approach permits the continuous targeting of civilians who persistently participate in such hostilities.⁹⁶ In so doing, the *Tallinn Manual* approach avoids the farmer by day, fighter by night dilemma, and adequately considers the reality of modern cyber conflict.⁹⁷

Revisiting the issue of Russian civilian participation in the Georgian cyberattacks, every Russian civilian cyber operation with the “intended or actual effect of negatively affecting” Georgia’s “military operations or capabilities” would result in the direct participation in hostilities of those civilians involved.⁹⁸ Many *Tallinn Manual* experts took the position that “direct participation begins with the first such cyber operation and continues throughout the period of intermittent activity.”⁹⁹ Thus, Russian civilians engaging in repeated cyberattacks against Georgia would be subject to direct attack until they provided a “clear indication that [they] would no longer engage in such attacks.”¹⁰⁰

Additionally, cyber weapons may be developed and staged long before a conflict begins. The actual implementation of cyber weapons may only take a matter of seconds and can be accomplished from any number of devices or locations. Therefore, a civilian engaging in a cyberattack may only be “directly participating” in an attack for a brief period of time, in a non-conflict area, and using a device that few would suspect to be an implementation of cyber warfare.¹⁰¹ This makes it virtually impossible for a nation to defend itself against civilian cyber attackers under the *Interpretive Guidance*. If the civilian continues the attacks, becoming a repeat offender, each subsequent implementation of the cyber weapon may likewise take only seconds, providing little

⁹⁶ See TALLINN MANUAL, *supra* note 12, at 432.

⁹⁷ It must be noted, however, that fundamental principles of targeting civilians, such as proportionality and distinction, must be considered in each situation.

⁹⁸ *Id.* at 429.

⁹⁹ *Id.* at 432.

¹⁰⁰ *Id.*

¹⁰¹ This statement considers the reality that cyber operations can be conducted from almost any location on the planet using mundane devices, such as a laptop, desktop computer, or even just a cellular phone.

time to the targeted nation to target the hacker responsible. The civilian hacker may plague an opposition force with intermittent cyberattacks that take only seconds to unleash and remain effectively shielded from targeting by the opposition force.

However, the *Tallinn Manual* would conceivably allow a nation to exercise force, up to and including deadly force, after the hacker is located, in order to pacify the hostile civilian who has demonstrated, through repeated action, a tendency and intent to engage in hostile cyber operations.¹⁰²

It is vital to note that a single act of direct participation in hostilities does not qualify a civilian hacker for continuous targeting. However, if it becomes clear an attack launched by a civilian directly participating in hostilities is the first in a long line of impending attacks, continuous targeting may be acceptable under the *Tallinn Manual*. As noted earlier, the *Tallinn Manual* focuses on the “civilian’s intent to carry out [an] attack” when determining the threshold of harm.¹⁰³ Thus, where a civilian hacker engages in a cyber operation and reliable intelligence indicates the hacker intends to conduct further malicious cyber operations, that civilian may be considered directly participating in hostilities because both the initial attack and the reliable intelligence indicate a continuing hostile intent. This assertion is not based solely upon the fact that the hacker engaged in an initial cyberattack against the victim nation. The *Interpretive Guidance* is clear that “[e]ven the fact that a civilian has repeatedly taken a direct part in hostilities, either voluntarily or under pressure, does not allow a reliable prediction as to future conduct”¹⁰⁴ but the *Tallinn Manual* makes no such claim to the contrary. Rather, the above assertion relies both on the intent of the civilian hacker to conduct an attack manifest in both the initial attack *and* the reliable intelligence suggesting the hacker intends to conduct future attacks.¹⁰⁵

¹⁰² Additional considerations, such as the geographical location of the civilian hacker and the viability of lesser means to pacify the hacker, may influence whether a majority of the *Tallinn Manual* experts would agree that deadly force targeting the civilian is appropriate under such a scenario.

¹⁰³ See Allan, *supra* note 23, at 182.

¹⁰⁴ *Interpretive Guidance*, *supra* note 11, at 71.

¹⁰⁵ Questions regarding the reliability of intelligence will still need to be evaluated on a case-by-case basis.

Generally, the initial treatment of a civilian's participation in cyber hostilities must be viewed independently, meaning an opposing force may not continuously target a civilian hacker after the hacker's initial cyberattack if that hacker has clearly indicated she will no longer engage in such attacks.¹⁰⁶ Barring any additional cyberattacks by that civilian hacker, the hacker will regain their civilian protection against direct attack. It is only after evidence arises of *continued and repeated* involvement in cyber operations against a party to a conflict that a civilian may be targetable "throughout the period of intermittent activity."¹⁰⁷ Adherence to this concept is fundamental because a civilian hacker engaging in only an initial cyberattack with no intent or manifestation to continue such attacks, who then clearly indicates a withdrawal from hostilities, has his or her civilian protections against direct attack restored. The deliberate targeting of a civilian is a violation of Protocols I & II,¹⁰⁸ both of which have largely been accepted as customary international law.¹⁰⁹ Thus, the force responding to the cyberattack must bear the burden of continuously revalidating the target's continuous participation in hostile cyber activities.

Modern conflict also exposes combat forces to areas in which civilians may be hostile or unfriendly to opposition forces.¹¹⁰ Because cyberattacks may be mounted from virtually any location, a civilian may position himself in a neighborhood unfriendly to opposition forces to launch one cyberattack after another. Per the ICRC's interpretation, opposition forces are largely limited to law enforcement responses because unless the civilian is actively engaged or preparing for a cyberattack, the hacker retains his civilian protections against direct attack, preventing a targeted use of force against the hacker upon locating him.¹¹¹ Therefore, unless the opposition force wishes to wait until the civilian hacker begins preparing for, or executing an attack which could result in harm to the systems attacked by the hacker, the

¹⁰⁶ See TALLINN MANUAL, *supra* note 12, at 432.

¹⁰⁷ *Id.*

¹⁰⁸ *Interpretive Guidance*, *supra* note 11, at 37, 90.

¹⁰⁹ Henckaerts, *supra* note 57, at 187–88.

¹¹⁰ See Dexter Filkins, *One by One, Iraqi Cities Become No-Go Zones*, N.Y. TIMES (Sep. 5, 2004), <http://www.nytimes.com/2004/09/05/weekinreview/one-by-one-iraqi-cities-become-nogo-zones.html>.

¹¹¹ *Interpretive Guidance*, *supra* note 11, at 71.

opposition force will be required to enter the hostile neighborhood and expose themselves and surrounding civilians to violence and needless danger. Yet, under the *Tallinn Manual* interpretation, once the civilian hacker is located, a well-placed, drone-fired missile may resolve the issue without exposing opposition forces and numerous innocent civilians to a law enforcement action in a hostile area that may devolve into a skirmish.¹¹² Perhaps more practically, the *Tallinn Manual* interpretation would also allow opposition forces to enter the hostile area and conduct its operation at a more operationally favorable time such as the middle of the night when fewer hostile forces may be present and innocent civilians will largely be off the streets.

2. Delayed Effects

Hackers may also deploy “delayed effects” cyber weapons, or cyber weapons that are uploaded into an adversary’s systems and lay dormant until activated by command or other directive.¹¹³ For example, a hacker may plant a computer virus in a nation’s radar system that he could activate prior to an airstrike or at the onset of an invasion of that nation. Such a virus may also be uploaded into other defense or early warning systems, like radar guided missile systems.¹¹⁴ A virus may be programmed to execute when the missile system is given the command to fire, thus resulting in a failure of the system or potential kinetic result.¹¹⁵ In both of these situations, the civilian hacker may have only been directly participating in hostilities for seconds when executing the cyberattack. Still, that hacker may have planted the virus months or even years before the attack was actually executed.

Under the *Interpretive Guidance*, “where the execution of a hostile act does not require geographic displacement, as may be the case with computer network attacks or remote-controlled weapons systems, the duration of direct participation in hostilities will be

¹¹² While the *Interpretive Guidance* may also permit the implementation of a drone strike in this situation, as previously mentioned, the use of a drone option would be limited to the preparation or execution of the actual attack, not upon locating the hacker.

¹¹³ See TALLINN MANUAL, *supra* note 12, at 431.

¹¹⁴ *Id.*

¹¹⁵ *Id.*

restricted to the immediate execution of the act and preparatory measures forming an integral part of that act.”¹¹⁶ The creation, development, and placement of software may be an integral part of the cyberattack,¹¹⁷ but as the example above illustrates, these elements may occur well before the execution of the attack and therefore the opening of hostilities. Thus, the *Interpretive Guidance* results in the hostile civilian only being targetable during the creation and placement of the virus, which may occur during peacetime, as well as during the actual deployment of the virus which may take only moments.

The *Interpretive Guidance* effectively ignores the civilian’s awareness of the attack and his belligerent intent. The civilian hacker knows that the malicious code is in place and is simply waiting for the optimal time to engage. Allowing this conduct is the equivalent of permitting a civilian to be targeted while placing an IED on a roadside frequented by opposition forces, but then restricting the targeting of that civilian when he resumes his civilian activities. The IED may remain undetected for days, weeks, or longer, awaiting detonation by the civilian at the opportune moment. Yet, because the civilian has returned to work or resumed normal responsibilities, the civilian is not targetable until he detonates the IED, even though he is aware of the impending attack and maintains the capacity to stop it. The assertion strains reason.

Conversely, a majority of the *Tallinn Manual* group of experts adopted the position that the duration of a civilian’s direct participation in hostilities begins with their involvement in planning the operation and ends when the civilian ceases any active role in the operation.¹¹⁸ Essentially, the civilian’s direct participation in hostilities would extend from the commencement of planning how to upload the malware onto the enemy system through the activation of the malware at the command of that civilian.¹¹⁹ This is a result, primarily, of the *Tallinn Manual*’s focus on the intent of a hostile civilian to conduct an attack. By stating that an act having the “intended or actual effect” of

¹¹⁶ *Interpretive Guidance*, *supra* note 11, at 68.

¹¹⁷ See *id.* at 66 (emphasis in original) (stating “it could be said that preparatory measures aiming to carry out a specific hostile act qualify as direct participation in hostilities”).

¹¹⁸ See TALLINN MANUAL, *supra* note 12, at 431.

¹¹⁹ *Id.*

adversely affecting the enemy's "military operations or capabilities,"¹²⁰ the majority of the *Tallinn Manual* experts recognize the threat of a civilian with belligerent intent and respond by allowing a civilian with his finger on the virtual trigger of a cyber weapon to be targeted before he has a chance to activate it.

However, the scenario can become more complex in situations in which a civilian implants malware into a system and then walks away from the situation without any intention of ever monitoring it or actively activating the malware. For example, a civilian technician in a munitions plant uploads malware into an anti-aircraft missile set for foreign distribution to prevent the proper function of the missile when it detects a certain aircraft's Identification, Friend, or Foe (IFF) code. The civilian does not intend to monitor the malware after uploading it because the malware will activate itself and, in any case, the malware may never be activated. May that civilian technician be considered directly participating in hostilities? There is no clear answer, but any answer would likely depend upon the circumstances.

According to Collin Allan, "[w]hen a civilian intends to negatively affect [an] adversary's military operations or capabilities, even if the attack is not likely to be carried into effect, it will result in that civilian being targetable by the victim of the attack."¹²¹ Given the missile example above, it is unlikely that any nation would sell anti-aircraft armaments to an adversary, but political climates shift and it is possible for a nation's own arms to be turned against it. Provided the intent-based focus of the *Tallinn Manual* in determining the threshold of harm required when evaluating civilian direct participation in hostilities, the civilian technician uploading the malware to the missile in our example would meet that threshold of harm because the civilian intends to negatively affect a nation's military capacity.¹²² Similarly, the civilian's actions would be the direct cause of any harm, thus meeting the direct cause requirement.¹²³ However, because the civilian's actions are not related to any hostilities, the civilian's actions cannot be said to have a belligerent nexus.¹²⁴ Therefore, a civilian

¹²⁰ *Id.* at 429.

¹²¹ See Allan, *supra* note 23, at 182.

¹²² See TALLINN MANUAL, *supra* note 12, at 429.

¹²³ See *id.*

¹²⁴ See *id.* at 430.

undertaking such conduct cannot be said to be directly participating in hostilities.

However, should hostilities occur between the two nations, and the victim nation is unable to utilize all or even a portion of their anti-aircraft missiles due to the conduct of the civilian technician, then the civilian's conduct is directly related to hostilities and is having the intended negative affect on the adversary's military capabilities. Thus, the civilian may be considered directly participating in hostilities. However, the civilian would not be targetable by the victim nation until those hostilities occurred. The duration of the civilian's targetability is indeterminable because each attempt by the victim nation to use an infected missile may reasonably be considered an activation of the civilian's malware, thus demonstrating arguable participation in the hostilities and therefore, targetability.

Nevertheless, a minority of the *Tallinn Manual* experts view emplacement and activation of the malware as separate actions.¹²⁵ Under that characterization, because the civilian did not take additional steps to activate the malware, there may be no subsequent participation in hostilities and therefore, no targetability.¹²⁶ Ultimately, there is no clear answer regarding this situation.

In another scenario in which a nation is aware of malicious code in its systems, it may be argued that the victim nation's knowledge of a hostile civilian's involvement in the placement of the code should allow the nation to remove it without resorting to targeting the civilian. Yet, this argument neglects consideration of the potential consequences of such an action. If the civilian hacker becomes aware that the virus he planted weeks earlier has been discovered and attempts are made to remove the virus, negate its effects, or arrest the hacker, the hacker might activate the virus prematurely. This consequence is most clearly illustrated using the IED example above. If a civilian who previously placed an IED learns that attempts are being made to remove or disarm his IED, or to arrest him, then the civilian may detonate the IED, threatening both combatant and civilian lives. However, directly targeting the civilian hacker responsible for the placement of a virus and who retains the capacity to activate it reduces the possibility of premature execution.

¹²⁵ See *id.*

¹²⁶ See *id.* at 431–32.

Although the protection of civilians is of paramount concern in treaty and customary law, effort must be undertaken to ensure that civilians repeatedly engaging in hostilities or conducting delayed effects attacks do not receive undeserved advantages. Thus, the approaches advocated by a portion of the *Tallinn Manual* group of experts should be adopted because they properly consider the realities of modern conflict by allowing national forces flexibility in targeting civilian hostiles while simultaneously ensuring real protections for uninvolved civilians.

*B. The Tallinn Manual Interpretation Properly Balances
Military Necessity and Humanitarian Considerations*

Critical to the modern law of armed conflict are two core principles: military necessity and humanitarian considerations.¹²⁷ The two principles endeavor to pragmatically regulate the manner in which nation-states conduct hostilities. For example, the principle of military necessity “permits a belligerent to apply only that degree and kind of regulated force, not otherwise prohibited by the laws of war, required for the partial or complete submission of the enemy with the least possible expenditure of time, life, and physical resources.”¹²⁸ In short, the principle of military necessity acknowledges that warring nations seek military advantage over one another and, in doing so, might make use of the application of deadly force. Nevertheless, this principle is counter-weighted with the principle of humanity, which prohibits the “employment of any kind or degree of force not necessary for the purpose of war, i.e., for the partial or complete submission of the enemy with the least possible expenditure of time, life, and physical resources.”¹²⁹ Consequently, while force may be applied in the pursuit to subdue an enemy, such force is limited to the minimum necessary to achieve the military objective.

¹²⁷ See Michael N. Schmitt, *Military Necessity and Humanity in International Humanitarian Law*, 50 VA. J. INT’L. L. 795, 798 (2010).

¹²⁸ ROBERT W. TUCKER, *THE LAW OF WAR AND NEUTRALITY AT SEA* 364 (1957).

¹²⁹ *Id.*

Modern international laws of armed conflict embrace a balance of these two principles.¹³⁰ As such, the duration during which a civilian is considered directly participating in hostilities should similarly reflect a balance of these fundamental principles of customary international law of armed conflict. Although the *Tallinn Manual* and the ICRC's *Interpretive Guidance* consider both military necessity and humanity, the *Tallinn Manual* more appropriately balances the two principles on the issue of the targeting of civilians for such time as they directly participate in hostilities.

Military necessity permits the application of force required for the submission of the enemy.¹³¹ The *Tallinn Manual* interpretation embodies military necessity by permitting the targeting of civilians repeatedly involved in hostilities for a duration beginning with the civilian's first cyberattack and continuing through the period of intermittent activity.¹³² The *Tallinn Manual* further holds that civilians involved in delayed effects attacks are directly participating in hostilities from the moment they begin planning the attack until the civilian terminates any further activity in support of the cyber operation.¹³³ These interpretations reflect a consideration of military necessity in that each allows for the targeting of a civilian directly participating in cyber hostilities for a duration of time sufficient to grant the targeting nation an ample period in which to consider the effects of the civilian's cyberattack and respond accordingly.

The interpretations further reflect an understanding of military necessity in that they allow a targeting nation to consider the intent of civilians involved in hostilities against it. Consider the *Tallinn Manual* interpretation regarding the targeting of civilians repeatedly involved in hostilities for a duration beginning with a civilian's first cyberattack and continuing through the period of intermittent activity. By allowing nations to target civilians directly participating in hostilities during the period of intermittent activity between hostile acts, the *Tallinn Manual* interpretation recognizes that civilians with the present, apparent capability and intent to repeatedly engage in hostilities are likely to

¹³⁰ See Schmitt, *supra* note 127 (stating that "IHL represents a carefully thought out balance between the principles of military necessity and humanity. Every one of its rules constitutes a dialectical compromise between these two opposing forces").

¹³¹ See TUCKER, *supra* note 128.

¹³² TALLINN MANUAL, *supra* note 12, at 432.

¹³³ *Id.* at 431.

continue. Thus, the *Tallinn Manual* acknowledges that where the intent and ability of the civilian to continue engaging in cyber operations against a state is clear, it is conceivable that the civilian will not be dissuaded from additional attacks except through the application of force proportional to achieve that military purpose.

By adopting this interpretation, the *Tallinn Manual* appears to apply an anticipatory self-defense approach to repeat civilian offenders. Where a civilian attacker has repeatedly demonstrated aggression and provided no indication that such attacks will cease, the targeting nation may target the civilian on the belief that another attack is imminent.¹³⁴ In making this observation, a *jus ad bellum* analysis is applied (anticipatory self-defense) to the *jus in bello* principle (targeting of civilian participants in hostilities). Although separate legal paradigms, the *Tallinn Manual* position appears to explore the intersectionality amid the two, though it does not expressly articulate such a purpose.

In the case of a delayed effects attack, by permitting nations to target civilian actors from planning to execution, the *Tallinn Manual* interpretation acknowledges that the civilian who planted the malware in the targeting nation state's system not only intended to do harm, but *still retains the capacity to do so*. Thus, in a delayed effects situation, a targeting nation knows that a civilian hacker has already demonstrated intent to harm, and now may, at will, cause that harm. It may, therefore, be a military necessity to quickly remove the threat of the delayed effects attack by removing the civilian hacker from the equation with physical force. By interpreting the targetability of a civilian directly participating in hostilities to "continue throughout the period of intermittent activity," the *Tallinn Manual* interpretation effectively provides military forces with the time necessary to respond to the threat posed by a civilian directly participating in cyber operations and in a manner that will remove or reduce the threat of future attacks.¹³⁵

¹³⁴ See Leo Van den hole, *Anticipatory Self-Defence Under International Law*, 19 AM. U. INT'L. L. REV. 69, 104 (2003) (emphasis added) (noting "[i]f, however, the threat or the attack in question consisted of a number of successive acts, and there is sufficient reason to expect a continuation of acts from the same source, the international community should view the requirement of the immediacy of the self-defensive action in the light of those acts as a whole").

¹³⁵ TALLINN MANUAL, *supra* note 12, at 432.

Yet, simultaneously the *Tallinn Manual* appropriately applies the principle of humanity throughout its interpretation. In the context of a delayed effects attack, a civilian is only targetable for such time as the civilian participates in the attack.¹³⁶ The *Tallinn Manual* is explicit that the key to determining targetability is ascertaining where the civilian's participation begins and ends.¹³⁷ Thus, a civilian who uploads a virus to the targeting nation's system and later delivers the activation mechanism to another is no longer targetable because the civilian's direct participation in that particular cyberattack is over.¹³⁸ In this manner, the *Tallinn Manual* prohibits the employment of any kind or degree of force not necessary for the purpose of war because it acknowledges that the civilian no longer possesses the capacity to do harm, thus it is not "necessary for the purpose of war" for the civilian to remain targetable.

Similarly, the *Tallinn Manual* captures the principle of humanity by restoring civilian protections to a civilian who repeatedly participates in hostilities when he or she clearly indicates that they are no longer participating in hostilities.¹³⁹ By so doing, the *Tallinn Manual* recognizes the civilian's intention to refrain from further combat actions and return to civilian life where they no longer pose a threat to the targeting nation.

Conversely, the *Interpretive Guidance* employs the "revolving door" approach to civilian protections.¹⁴⁰ In the ICRC perspective, this interpretation "prevents attacks on civilians who do not, at the time, represent a military threat."¹⁴¹ However, this interpretation weighs too heavily in favor of humanity at the expense of military necessity, particularly in the context of cyber warfare.

Consider a delayed effects attack by a civilian hacker. Under the *Interpretive Guidance*, once the malicious code is uploaded to a system, the hacker may not be directly targeted because the hacker no longer "represent[s] a military threat."¹⁴² However, this understanding is flawed in that it mischaracterizes a lack of harm as an absence of

¹³⁶ See *id.* at 431.

¹³⁷ See *id.*

¹³⁸ See *id.*

¹³⁹ See *id.* at 432.

¹⁴⁰ See *Interpretive Guidance*, *supra* note 11, at 70.

¹⁴¹ *Id.*

¹⁴² *Id.*

military threat. A combatant in possession of a loaded firearm presents a military threat because he maintains the immediate capacity to cause harm even though no harm has occurred, nor will harm occur until the trigger is pulled. Likewise, a civilian hacker that uploads malicious code to a system, leaves it there, returns to normal activities, but monitors the code while retaining the ability to activate it has not done any noticeable harm. However, that hacker poses a military threat because he demonstrated a willingness to cause harm by initially uploading the malicious code to the target system and retaining the capacity to cause harm at a moment's notice. The fact that the hacker knowingly and intentionally maintains that capacity indicates he has not fully returned to civilian life nor ceased directly participating in hostilities.

Many in the international community agree that Article 51 of the UN Charter does not prohibit anticipatory self-defense, but rather, where an attack is imminent, a nation may take action in advance of the attack.¹⁴³ As mentioned above, the concept of Article 51 anticipatory self-defense is a *jus ad bellum* doctrine and cannot, as a matter of law, be applied to the *jus in bello* principle of the targeting of civilians directly participating in hostilities. However, the same logic justifying anticipatory self-defense may be applied in the targeting of the civilian hacker. Even in situations in which the hacker has not, as of yet, caused any harm, he knowingly and intentionally maintains the ability to do so at will, thus posing a serious military threat. As such, a nation aware of the hacker's activities should not need to wait until the attack occurs to eliminate the threat.

The *Interpretive Guidance* further defends its "revolving door" interpretation claiming that civilian behavior changes constantly based on circumstances and therefore even if a civilian has repeatedly participated in hostilities, that prior conduct is not a reliable predictor of future conduct.¹⁴⁴ However, the restoration of civilian protections for a civilian hacker that has repeatedly participated directly in hostilities allows that hacker to be a farmer by day and a fighter by night, because each time the civilian ceases hostilities, they again regain civilian protections against direct attack. William Boothby notes that, under the ICRC's perspective, the only group to lose protected status on a

¹⁴³ See Van den hole, *supra* note 134, at 82.

¹⁴⁴ *Interpretive Guidance*, *supra* note 11, at 71.

continual basis are members of an organized armed group engaged in a continuous combat function, while the loss of protection for a civilian directly participating in hostilities is limited to the time he is engaged in direct participation.¹⁴⁵ Consequentially, the ICRC's interpretation "narrows excessively the class of those who lose protection on a continuous basis and the result is a distortion of the balance [between military necessity and humanity] inherent in international law."¹⁴⁶ The ICRC itself states in the *Interpretive Guidance* that the loss of civilian protections for only the duration of specific hostile acts was designed only for civilians and not organized armed groups because it would provide "significant operational advantage over members of State armed forces"¹⁴⁷ and encourage OAGs to "operate as farmers by day and fighters by night."¹⁴⁸ But civilians can abuse the restoration of civilian protections just as easily as an OAG. The position of the ICRC that civilians should be allowed a "revolving door" of civilian protections is inconsistent with its own logic.¹⁴⁹ Thus, the ICRC's *Interpretive Guidance* creates an imbalance between military necessity and humanity by providing civilians engaged in repeated participation in hostilities with a significant operational advantage over State armed forces while restricting the armed forces of a State from applying the degree and kind of force required for the partial or complete submission of the enemy.

¹⁴⁵ Bill Boothby, "And for Such Time As": *The Time Dimension to Direct Participation in Hostilities*, 42 N.Y.U. J. INT'L L. & POL. 741, 753 (2010).

¹⁴⁶ *Id.* at 753, 767.

¹⁴⁷ *Interpretive Guidance*, *supra* note 11, at 71–72.

¹⁴⁸ *Id.*

¹⁴⁹ See DINSTEIN, *supra* note 73 (explaining that the positions taken by the *Interpretive Guidance* "are inconsistent since the only practical way to foil the assumption of the double role of a soldier by night and peaceful citizen by day is to treat the person purporting to be both as an unlawful combatant at all times. Differently put, he must lose protection from attack even during the intermediate periods punctuating military operations. Civilian protection from attack cannot be regained 'until he unambiguously opts out through extended nonparticipation or an affirmative act of withdrawal.'") (internal quotation and citation omitted).

C. *Shortfalls of the Tallinn Manual*

Although the *Tallinn Manual*'s interpretation regarding the duration of the loss of civilian protections against direct attack for civilians directly participating in hostilities is appropriately tailored to the realities of modern warfare and more effectively balances military necessity with humanity, it does have some drawbacks. Specifically, the *Tallinn Manual* approach fails to identify with specificity how a civilian "clearly indicates" they will no longer engage in cyberattacks in order to regain civilian protections after a prolonged period of repeated participation in hostilities.¹⁵⁰

Although some experts argue that civilian protection from attack may only be regained through an unambiguous act of extended non-participation or an affirmative act of withdrawal, these definitions are not significantly more descriptive than that of the *Tallinn Manual*.¹⁵¹

Ironically, the *Interpretive Guidance* offers insight in its explanation of the process through which a civilian may regain their protections against direct targeting by leaving an OAG. With the exception of a few notable areas, the *Interpretive Guidance* is instructive, providing that "disengagement . . . need not be openly declared; it can also be expressed through conclusive behaviour [sic], such as a lasting physical distancing from the group and reintegration into civilian life or the permanent resumption of an exclusively non-combat function. In practice, assumption of, or disengagement from, a continuous combat function depends on criteria that may vary with the political, cultural, and military context."¹⁵² Despite providing greater clarity into actions amounting to a clear indication of a civilian's intent to cease directly participating in hostilities, the *Interpretive Guidance* in the context of cyber warfare makes it difficult for the targeting nation to know when the civilian has undertaken such actions.

For example, a civilian who had repeatedly launched mortar shells at enemy soldiers could clearly demonstrate their cessation of participation in such activities by never taking out the mortar again and distancing himself from enemy military positions. However, unlike the

¹⁵⁰ See TALLINN MANUAL, *supra* note 12, at 432 (explaining that "in the absence of a clear indication that the hacktivist was no longer engaging in [cyber] attacks, he or she [will remain] targetable").

¹⁵¹ See DINSTEIN, *supra* note 73.

¹⁵² *Interpretive Guidance*, *supra* note 11, at 72.

mortar example, in which the civilian is easily able to store and avoid the implementations of his direct participation in hostilities, a civilian engaged in cyberattacks may legitimately continue to use a computer or a cell phone, which was once his weapon of choice. Consequently, it is difficult for a targeting nation to know when a civilian has sincerely ceased to participate in cyber hostilities because the civilian's conduct during his return to normalcy may appear no different than his conduct during his period of direct participation in cyber hostilities.

Although the *Interpretive Guidance* does provide some clarity as to conduct clearly indicating a civilian's intent to refrain from further participation in hostilities, it should be noted that the ability of a civilian under the *Interpretive Guidance* to regain civilian protections through the permanent resumption of exclusively non-combat functions is intrinsically flawed.¹⁵³ Indeed, it may be said that after an attack every repeat offender permanently resumes non-combat functions—until he does not. More precisely, it is impossible to say when a previously repeat offender has permanently resumed an exclusively non-combat function because permanency only endures until the civilian again undertakes a combat function, even if years pass between each engagement in hostilities. Therefore, using the permanent resumption of non-combat functions as a mechanism through which a civilian may regain their civilian protections is inherently meaningless and again allows for farmer by day, fighter by night conduct.

The most instructive passage within the *Interpretive Guidance* is the suggestion that criterion for determining disengagement from a continuous combat function varies with political, cultural, and military contexts.¹⁵⁴ This suggestion properly alerts operational commanders, who will likely make the disengagement determination, to perspectives that should be considered in reaching such a determination. And although this Article advocates that operational commanders undertake the disengagement determination by considering relevant political, cultural, and military perspectives, the determination must still be unequivocally “based on concrete, objectively verifiable facts and on standards of good faith and reasonableness in the prevailing circumstances.”¹⁵⁵ Concisely, operational commanders should cogently

¹⁵³ See *id.*

¹⁵⁴ See *id.*

¹⁵⁵ Watkin, *supra* note 88, at 693.

consider the facts of a situation in the context of applicable political, cultural, and military perspectives when making the disengagement determination.

While no single source adequately defines or clarifies what actions clearly indicate a civilian's intent to withdraw from persistent participation in hostilities, the recognition of the *Interpretive Guidance* that each situation is unique, coupled with an emphasis on the facts, good faith, and reasonable consideration under the circumstances,¹⁵⁶ most appropriately balances the totality of the circumstances.¹⁵⁷

IV. CONCLUSION

The introduction of cyber warfare into the modern battlefield brought with it various unique characteristics. One of which is the ease with which civilians may participate in cyber hostilities. Every civilian with Internet access and computer knowledge may now engage in conflict against a State armed force. This growing reality of modern conflict highlights the need for direction in addressing the participation of civilians in cyber hostilities.

The *Interpretive Guidance* and the *Tallinn Manual* both offer compelling interpretations of the temporal duration during which a civilian is considered directly participating in hostilities and thus loses their civilian protections against direct attack. Yet, particularly in the context of cyber warfare, the *Tallinn Manual* interpretation recognizes the reality of civilian participants in modern war. State militaries are provided the legal breadth to appropriately eliminate civilian threats from the beginning of the civilian's participation in hostilities until a clear termination of their active role in the engagement, as well as target those repeat offenders until such offender clearly indicates a withdrawal from further hostile activity.

The *Tallinn Manual* interpretation is derived from balancing military necessity with humanitarian considerations, and recognizing that while innocent civilians require protection, the international law of armed conflict should not be so broad as to allow civilians to oscillate between status of a civilian and a combatant with impunity—thus perverting the protections that are intended for truly innocent civilians

¹⁵⁶ See *id.*

¹⁵⁷ See *Interpretive Guidance*, *supra* note 11, at 67.

refraining from participation in hostilities. In short, the *Tallinn Manual* prevents the ability of civilian participants in cyber conflict from being farmers by day and fighters by night.

As the world becomes increasingly connected through artificial intelligence, virtual reality, and the Internet of things, cyber warfare will continue to play an ever more critical role in global conflicts—and the *Tallinn Manual* creates a solid foundation for addressing the status of civilian participation in future conflicts of that nature.