Direct Participation in Hostilities From Cyberspace

Collin Allan
DIRECT PARTICIPATION IN HOSTILITIES FROM CYBERSPACE

Collin Allan¹

ABSTRACT

As demonstrated by the cyber attacks against Georgia in 2008 and the cyber attacks against Aramco in 2012, civilians are increasing their participation in armed conflicts through cyber attacks. In 2009, the International Committee of the Red Cross (ICRC) published a document on how to determine when a civilian’s participation in armed conflict reaches the necessary level to render him or her targetable by one of the parties to the conflict. The Tallinn Manual was published this year to provide legal guidance in cyber situations. While professionals have written in this area previously, it is the first time that experts have compiled a manual of rules to indicate how international law applies to cyber situations. It includes a section on direct participation in hostilities through cyber means. This paper compares and contrasts the ICRC’s approach and the Tallinn Manual’s approach. The author reaches the conclusion that the Tallinn Manual’s approach has the general effect of lowering the standard for civilians’ actions in meeting the direct participation in hostilities bar. This makes a civilian participating in hostilities through cyber means targetable in more situations than a civilian participating in hostilities under the ICRC’s framework.

TABLE OF CONTENTS

I. INTRODUCTION ....................................................................................................................... 1
II. COMPARISON .......................................................................................................................... 1
   A. DPH Introduction ............................................................................................................... 1
   B. Threshold of harm ............................................................................................................ 1
   C. Direct Causation of Harm ............................................................................................... 1

¹ Collin Allan is a currently J.D. candidate at Brigham Young University. The author would like to thank Professor Eric Talbot Jensen for his support, guidance, and expertise.
I. INTRODUCTION

In the summer of 2012, a Saudi Arabian oil company was the victim of a devastating cyber attack. Nearly 30,000 computers were destroyed and information was removed from the computers. In the aftermath, Iran emerged as one of the likely culprits. It, of course, denied all accusations. While cyber attacks in general have provided international law with much to consider, one of the more revolutionary accusations to emerge from this attack, was one that placed the blame on the shoulders of ordinary civilians. This accusation is revolutionary because, if true, it presents one of the first of times that a cyber attack of this nature has been carried out by civilians. More and more, states are utilizing civilians to carry out attacks. In some cases, states temporarily utilize civilians to carry out attacks. States are also incorporating more civilians into their respective cyber commands because of the civilians’ expertise and access to software and equipment.

---

3 Id.
5 Id.
6 While cyber attacks have been carried out in other situations, this represents the first known time that a cyber attack was carried out against state actor by civilians, resulting in this magnitude of destruction. Previous cyber attacks may have disrupted the functioning of a state’s government or economic sector, but the destruction of a physical object as a result of the cyber attack has not recorded. See Paulo Shakarian, The 2008 Russian Cyber Campaign Against Georgia, 91 Military Review 63 (2011) (discussing the cyber attacks that occurred in connection with the Russia-Georgia conflict of 2008); Michael N. Schmitt, Cyber Operations and the Jus Ad Bellum Revisited, 56 Villanova Law Review 569, 570 and 594-595 (2011) (examining the cyber operations launched against Estonia in 2007).
8 Id.
Whenever civilians directly participate in hostilities, it creates myriad legal issues. These issues are exacerbated when the participating civilians are taking part in cyberspace.\textsuperscript{10} In 2009, the International Committee of the Red Cross (ICRC) issued a comprehensive “interpretive guidance” (Interpretive Guidance) on the legal issues surrounding civilians who directly participate in hostilities in an effort to address some of the legal questions arising from a civilian’s direct participation in hostilities.\textsuperscript{11} The Interpretive Guidance discusses three basic requirements (threshold of harm, direct causation, and belligerent nexus) a civilian’s actions are required to meet before that civilian will be deemed to be directly participating in hostilities, and therefore, targetable by the party that is a victim of that civilian’s actions.\textsuperscript{12} The Interpretive Guidance also discusses ancillary issues including the temporal scope of a citizen’s act, loss of protection, and provisions that govern a state’s response.\textsuperscript{13} While the Interpretive Guidance provides some insight into cyber activities,\textsuperscript{14} it focuses mostly on the general principle of direct participation in hostilities.

In 2013, a group of international legal scholars (Tallinn Experts) published the Tallinn Manual on the International Law Applicable to Cyber Warfare (Tallinn Manual) in order to address concerns similar to those contained in the Interpretive Guidance.\textsuperscript{15} In general, the Tallinn Manual was created to help states grapple with the difficult problem of applying international


\textsuperscript{12} \textit{Id.} at 46.

\textsuperscript{13} \textit{Id.} at 65-82.

\textsuperscript{14} \textit{Id.} at 48.

law, much of which was developed before the advent of the computer, to the increasingly active nexus between cyber activities and hostilities in general. The Tallinn Manual provides rules and commentary on how international law should be applied to issues, including the use of force, the definition of an attack, and neutrality. While legal scholars have considered the subject, the Tallinn Manual is the first major document of its kind that discusses the issue of civilians who directly participate in hostilities through cyber means.

This paper discusses the Tallinn Manual’s application of Direct Participation in Hostilities (DPH) principles in light of the Interpretive Guidance in an effort to determine how closely it conforms to the Interpretive Guidance. This paper concludes that, while the Tallinn Manual generally follows the ICRC’s framework, there are some areas where the Tallinn Manual differs in scope, including in its discussion of what satisfies the threshold of harm and belligerent nexus requirements and the temporal scope of direct participation in hostilities. This has the effect of establishing a lower standard for a civilian’s cyber actions than for a civilian’s kinetic direct participation in hostilities, making it easier for a civilian to satisfy the DPH requirements.

II. COMPARISON

A. DPH Introduction

Understanding whether or not a civilian’s actions rise to the level of direct participation in hostilities is an important analysis in which every combatant in a conflict, from the basic soldier to the commander, must engage. The purpose of the Law of Armed Conflict is to protect the lives of the people affected by any conflict. For example, when a wounded soldier has been captured by an opposing force, the first Geneva Convention will generally dictate that the soldier

---

16 Id.
17 Rules 10-12, 30, 91-95.
18 Rule 35.
should be treated humanely and outline the obligations a state has in protecting that soldier from further harm. The fourth Geneva Convention applies to civilians who are affected by armed conflict. In 1977, two Additional Protocols were introduced to provide further clarification regarding the obligations states have to protect the victims of both international and non-international armed conflicts. They address areas with which the Geneva Conventions previously had not dealt. While Common Article 3 serves as the written genesis of the DPH paradigm, the Protocols contain the current version, stating that civilians are protected from attack “unless and for such time as they take a direct part in hostilities.” This provision embodies the general presumption of protection for civilians in armed conflict, but it also serves as the basis for soldiers and commanders in determining when it is appropriate to target civilians.

Because this single provision is not totally clear, the Interpretive Guidance was drafted in an effort to clarify what actions will make a civilian targetable and for how long that civilian will be targetable. The Interpretive Guidance limits the scope of its discussion to actions that take

---

20 Convention (I) for the Amelioration of the Condition of the Wounded and Sick in Armed Forces in the Field. Geneva, 12 August 1949. The Geneva Conventions are generally considered customary international law. However, there is some discussion as to their the scope of their application. Some contend that the Geneva Conventions apply only to international armed conflict, conflict between states, and that only Common Article 3 governs non-international armed conflict, conflict usually characterized as internal conflict, although it may also include conflict between a state actor and non-state actors. See David Glazier, Playing by the Rules: Combating Al Qaeda Within the Law of War, 51 Wm. & Mary L. Rev. 957, (2009); Geoffrey Corn and Eric Talbot Jensen, Transnational Armed Conflict: A “Principled” Approach to the Regulation of Counter-Terror Combat Operations, Israel Law Review, Vol. 42, p. 46, 2009; Theodor Meron, The Geneva Conventions as Customary Law, 81 Am. J. Int'l L. 348 (1987).


22 Protocol Additional to the Geneva Conventions of 12 August 1949, and relating to the Protection of Victims of International Armed Conflicts, 8 June 1977 (AP I); Protocol Additional to the Geneva Conventions of 12 August 1949, and relating to the Protection of Victims of Non-International Armed Conflicts, 8 June 1977 (AP II).

23 Id.

24 Geneva Convention I-IV, Art. 3.

25 AP I, Art. 51.3; AP II, Art. 13.3.

26 Any time a civilian is targetable, any time anyone is targetable for that matter, the international legal principles of necessity and proportionality will govern any military response against that civilian. See Eric Talbot Jensen, Unexpected Consequences From Knock-on Effects: A Different Standard for Computer Network Operations?, 18 Am. U. Int'l L. Rev. 1145, (2003) (suggesting that the same principles of necessity and proportionality that apply to kinetic operations should and can apply during cyber operations.); Additional Protocol I, art 57.2(a)(iii), 62.1.
place during armed conflict stating that “the concept of direct participation in hostilities cannot refer to conduct occurring outside situations of armed conflict, such as during internal disturbances and tensions, including riots, isolated and sporadic acts of violence and other acts of a similar nature.”27 Furthermore, the Interpretive Guidance acknowledges that “even during armed conflict, not all conduct constitutes part of the hostilities.”28 That is, a civilian’s action must occur within an armed conflict and must constitute part of the hostilities. The Tallinn Manual agrees with these characterizations of the DPH regime.29

The Interpretive Guidance outlines three requirements that a civilian’s action must meet before it will qualify as direct participation in hostilities. It also discusses additional issues relevant to the determination of whether or not a civilian is targetable, including ascertaining the beginning and end points of the time for which a civilian is targetable.

First, a civilian’s act “must be likely to adversely affect the military operations or military capacity of a party to an armed conflict.”30 The first requirement may also be satisfied if the act is likely “to inflict death, injury, or destruction on persons or objects protected against direct attack.”31 This first requirement is characterized as the “threshold of harm.”

Second, for an act to qualify, “there must be a direct causal link between the act and the harm likely to result” from that act, independent of any other action taking place.32 This requirement may also be satisfied if there is a direct causal link between the act and the “harm likely to result . . . from a coordinated military operation of which that act constitutes an integral

27 Nils Melzer, at 41.
28 Id.
29 Tallinn Manual, Rule 35.1.3.
30 Nils Melzer, “Interpretive Guidance on the Notion of Direct Participation in Hostilities Under International Humanitarian Law,” May 2009. at 46. While the Interpretive Guidance and Tallinn Manual draw on relevant provisions from the Geneva Conventions and Additional Protocols, because the purpose of this article is to analyze only the differences and similarities between the two, this article will mostly cite to either the Interpretive Guidance or Tallinn Manual when establishing the basic law.
31 Id.
32 Id.
part.” This is referred to as the “direct causation” requirement.

Third, for an act to fall within the DPH scheme, it “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.” This is the “belligerent nexus” requirement.

Both the Interpretive Guidance and the Tallinn Manual discuss the temporal scope of targetability. That is, a civilian is only targetable while he or she is directly participating. This means that there is a start to direct participation and an end to direct participation, and thus a limit on when a civilian will be targetable.

The ICRC provides an explanation of each requirement, and the Tallinn Manual, in one of its sections, seeks to apply the DPH legal regime to cyber activities. The Tallinn Manual “generally agrees with the three cumulative criteria for qualification of an act as direct participation,” but there are important differences. The three requirements and other differences will be discussed below.

B. Threshold of harm

The Tallinn Manual’s discussion of the threshold of harm requirement marks the first, and arguably most important, area of divergence from the Interpretive Guidance. While the Interpretive Guidance employs an objective likelihood standard, the Tallinn Manual makes intent key to its definition. The Tallinn Manual’s reliance on intent has the effect of lowering the standard for civilians’ actions that may satisfy the threshold of harm requirement.

The first requirement of the DPH regime requires a civilian’s action to reach a certain threshold of harm. That is, there are many actions that a civilian may carry out during the course

---

33 Id.
34 Id.
35 Id.
36 Id.
37 Tallinn Manual, Rule 35.4.
of armed conflict, but if the action fails to meet the necessary threshold of harm, that action will not make a civilian targetable. The Interpretive Guidance’s requirement provides two different ways under which the threshold of harm requirement may be satisfied. The first way the “threshold can be reached [is] by causing harm of a specifically military nature.” The second way the threshold can be reached is “by inflicting death, injury, or destruction on persons or objects protected against direct attack.” This requirement does not actually require the realization of the harm; rather, it only requires the “objective likelihood that the act will result in such harm.” Therefore, actual harm does not have to occur to satisfy this requirement, rather “the relevant threshold determination must be based on ‘likely’ harm, that is to say, harm which may reasonably be expected to result from an act in the prevailing circumstances.”

The ICRC explains that as long as the harm is “military harm,” it will more than likely meet the necessary threshold of harm. The ICRC explains that military harm is “essentially any consequence adversely affecting the military operations or capacity of a party to the conflict.” The harm, so long as it is military harm, does not have to include death, injury, or destruction of military personnel or objects, nor is it required to include death, injury, or destruction on persons or objects protected against attack. The idea that military harm is sufficient is reflected in one of the Interpretive Guidance’s few references to cyber actions. This section states that a possible cyber action that “could” meet the required threshold of harm is “[e]lectronic interference with

38 Nils Melzer, at 47.
39 Id. The Interpretive Guidance cites the Hague Regulations as providing a description of what is protected from attack. The Hague Regulations are a body of international similar to the Geneva Conventions in that they provide regulations on the conduct of armed conflict and prohibit the “attack or bombardment, by whatever means, of towns, villages, dwellings, or buildings which are undefended.” Art 25 Hague Regulations IV. This is reflected in Common Article 3 in the Geneva Conventions and both Additional Protocols that requires the humane treatment of people taking no active part in hostilities. The IV Geneva Convention generally applies to the protection of civilians in Wartime.
40 Nils Melzer, at 47.
41 Id.
42 Id.
43 Id.
military computer networks,” regardless of whether it was “through computer network attacks (CNA) or computer network exploitation (CNE).”\textsuperscript{44} In this example, the military harm envisioned by the Interpretive Guidance was electronic interference.\textsuperscript{45} The Interpretive Guidance asserts that a cyber attack may rise to the threshold of harm if it achieves military harm; it does not explicitly extend this possibility to the second avenue, through which a civilian’s actions may qualify as directly participating in hostilities, that is by causing death, injury or destruction to persons or objects protected against attack.

The second way that a civilian’s action may reach the necessary threshold of harm is through the infliction of death, injury, or destruction on persons or objects protected against direct attack.\textsuperscript{46} This avenue allows for actions that do not inflict harm of a military nature, but rather focuses on harm caused to civilians and civilian objects.\textsuperscript{47} The Interpretive Guidance explains that International Humanitarian Law, another name for the Law of Armed Conflict, defines attacks as “‘acts of violence against the adversary, whether in offence or defence’.”\textsuperscript{48} The Interpretive Guidance further explains that the “phrase ‘against the adversary’ does not specify the target, but the belligerent nexus of an attack, so that even acts of violence directed specifically against civilians or civilian objects may amount to direct participation in hostilities.”\textsuperscript{49} As long as an action inflicts, or has the objective likelihood of inflicting, death, injury, or destruction on people or objects protected from attack, it will satisfy this the threshold

\textsuperscript{44} Id at 48. The Interpretive Guidance “tentatively” defines CNA as “operations to disrupt, deny, degrade, or destroy information resident in computers and computer networks, or the computer and networks themselves.” These attacks “may be conducted over long distances through radio waves or international communication networks.” The Interpretive Guidance admits that they may not cause “direct physical damage,” but that the “resulting system malfunction can be devastating.” A CNE is “the ability to gain access to information hosted with references and the ability to make use of the system itself.” (footnote 101)(quoting Background Doc. DPH 2003, pp.14-15.).

\textsuperscript{45} Nils Melzer at 48. (footnote 101)(quoting Background Doc. DPH 2003, pp.14-15.)

\textsuperscript{46} Id. at 46.

\textsuperscript{47} Id. at 49.

\textsuperscript{48} Id.

\textsuperscript{49} Id. The demands of the belligerent nexus requirement will be discussed more in-depth later in this paper.
of harm requirement.

The Interpretive Guidance uses “the manipulation of computer networks” as an example of a possible civilian action that would fail to satisfy either possible avenue for which a civilian’s action may meet the threshold requirement.\textsuperscript{50} If a cyber operation does not involve more than manipulation, the Interpretive Guidance suggests that it would not satisfy the threshold of harm requirement. This is because the manipulation of computer networks “neither cause[s] harm of a military nature nor inflict[s] death, injury, or destruction on protected persons or objects” and “cannot be equated with the use of means or methods of ‘warfare’ or, respectively, of ‘injuring the enemy’, as would be required for a qualification as hostilities.”\textsuperscript{51} The Interpretive Guidance envisioned the manipulation of non-military networks in this example, and, in doing so, it seems to be dismissing any cyber attack except for one against the military or one that causes death or physical damage or destruction. The Tallinn Manual, disagrees with the proposition that if a civilian does not a launch cyber attack against a military target, it would fail to satisfy the military harm prong.\textsuperscript{52} The Tallinn Manual also disagrees with the Interpretive Guidance’s requirement of the infliction of physical harm or material destruction.\textsuperscript{53}

The Tallinn Manual’s examination of the first DPH requirement generally echoes the Interpretive Guidance in stating that the action must have the “intention or effect of negatively affecting the adversary’s military operations or capabilities, or inflicting death, physical harm, or material destruction on persons or objects protected against direct attack.”\textsuperscript{54} However, it asserts that there “is no requirement for physical damage to objects or harm to individuals. In other

\textsuperscript{50} Id. at 50.
\textsuperscript{51} Id.
\textsuperscript{52} Tallinn Manual at Rule 35.4.
\textsuperscript{53} Id.
\textsuperscript{54} Id.
words, actions that do not qualify as a cyber attack\textsuperscript{55} will satisfy this criterion so long as they negatively affect the enemy military.\textsuperscript{56} The Tallinn Manual’s description of the first requirement differs in several ways from the Interpretive Guidance’s formation. The Tallinn Manual includes intention, leaves out the “likelihood” description, dismisses the requirement for physical harm or material destruction, and may extend the scope of the definition to actions that only aid one side. In its discussion of the three factors, the Tallinn Manual uses “a cyber operation [that] disrupts the enemy’s command and control network” as an example of a cyber activity that would satisfy this factor.\textsuperscript{57}

The Tallinn Manual states that a civilian’s action carried out with only the intent of negatively affecting the adversary’s military operations or capabilities or to inflict death, physical harm, or material destruction on persons or objects protected from attack is sufficient to meet the first prong of the DPH analysis.\textsuperscript{58} The Interpretive Guidance does not address intent in its threshold of harm analysis. Rather, it addresses intent in its belligerent nexus analysis and discounts it as unimportant to the DPH analysis except in extreme circumstances.\textsuperscript{59} This is an important difference between the Tallinn Manual and the Interpretive Guidance.

Including intent rather than likelihood in the Tallinn Manual’s characterization of this prong has the effect of lowering the bar for civilian actions that may satisfy the first requirement. When 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 that civilian being targetable by the victim of the attack. Here, as in the Interpretive Guidance, an attack does not

\textsuperscript{55} As will be discussed later, the Tallinn Manual defines cyber attack as a “cyber operation, whether offensive or defensive, that is reasonably expected to cause injury or death to persons or damage or destruction to objects.” See Tallinn Manual, Rule 30.

\textsuperscript{56} \textit{Id.} Rule 35.4.

\textsuperscript{57} \textit{Id.}

\textsuperscript{58} This idea is reflected in how the Tallinn Manual defines an attack. See Tallinn Manual, Rule 30.14, stating “A cyber operation need not actually result in the intended destructive effect to qualify as an attack.”

\textsuperscript{59} Melzer at 59.
actually have to occur, but whereas the Interpretive Guidance sets it bar at the objective likelihood of the attack, the Tallinn Manual’s bar has been set at the intent to carry out the attack. Furthermore, the Tallinn Manual’s wording suggests that if a civilian does not intend to negatively affect the adversary’s military operations or capabilities or inflict death, injury, or destruction on people or objects protected against direct attack, and yet the attack has that effect, the civilian can still be targeted by the victim of the attack. The Tallinn Manual states that the action must have “the intention or the effect” of negatively impacting the adversary’s military operations or capabilities or causing death, injury or destruction. This distinction is important because of the nature of cyber attacks and the methods through which they can be carried out. For example, cyber attacks can be launched from one computer located in one state, routed through servers located in a second state, having the effect of negatively affecting an adversary’s military operations and capabilities in a third state. The owner of the servers located in the second state may not have intended the attack, but their servers were key in negatively affecting the third state’s military operations and capabilities. The Tallinn Manual’s construction of this factor may make the civilians responsible for the servers in the second state targetable by the third state. Likewise, an individual whose computer has been slaved, through no fault of the owner, to carry out cyber attacks that have the effect of negatively affecting an adversary’s military operations or capabilities may be targetable by the victim state. The ability for victim states to target innocents may be resolved when the remaining two DPH requirements are applied, but, at least for the purposes of the first requirement, the second state and the ignorant civilians may be targetable.

The Tallinn Manual’s definition of cyber attack contains language similar to the

---

60 It will again be noted that any military response is tempered by the international law principles of necessity and proportionality. See note 24 for a more in-depth explanation.
Interpretive Guidance’s threshold of harm characterization. The Tallinn Manual describes a cyber attack as a “cyber operation, whether offensive or defensive, that is reasonably expected to cause injury or death to persons or damage or destruction to objects.”\textsuperscript{61} The Tallinn Manual’s cyber attack definition turns on whether or not a cyber operation is \textit{reasonably expected} to cause harm. This seems to be closer to the Interpretive Guidance’s objective likelihood standard than the Tallinn Manual’s intent or effect standard. If the Tallinn Manual stated that a civilian directly participating in hostilities was required to launch a cyber attack in order to meet the threshold of harm requirement, then the Tallinn Manual and the Interpretive Guidance would likely be more in line with each other. However, the Tallinn Manual itself explicitly states that a civilian’s actions are not required to rise to the level of a cyber attack to satisfy the threshold of harm requirement.\textsuperscript{62}

While neither version of the DPH regime explicitly requires an attack to meet the threshold of harm requirement, the Tallinn Manual’s definition of cyber attack may still be helpful in demonstrating what type of cyber action may satisfy the threshold of harm requirement. The Tallinn Manual focuses on the violent consequences of an attack, using the example of a cyber attack on an industrial control system that results in a fire. Because of this focus on the effects of an attack in determining what qualifies as an attack, it is helpful to discuss what the Tallinn Experts thought would satisfy the attack requirement as a standard of what types of civilian actions will definitely satisfy the threshold of harm requirement.

The Tallinn Manual’s approach to what types of damage and destruction qualify an action as an attack may widen the scope of what satisfies the threshold of harm requirement.\textsuperscript{63} Because the threshold of harm does not necessarily require an action to rise to the level of an

\textsuperscript{61} Tallinn Manual, Rule 30.
\textsuperscript{62} Id. at Rule 35.4.
\textsuperscript{63} Tallinn Manual, Rule 30.10
attack, the Tallinn Manual’s description of what qualifies as an attack, at the very least, demonstrates what type of civilian action will satisfy the threshold of harm. If a cyber activity qualifies as a cyber attack, it will satisfy the threshold of harm. The majority of the Tallinn Experts agreed that interference with the functionality of an object would rise to the level of damage or destruction necessary to constitute an attack when the replacement of physical components was necessitated as a result of a cyber operation.64 For example, a cyber operation that required the replacement of a hard drive would qualify as a cyber attack. The Tallinn Experts were split when it came to deciding if the reinstallation of an operating system as a result of a cyber operation constituted sufficient damage to qualify the action as an attack.65 So, it is unclear whether or not the reinstallation of an operating system would satisfy the threshold of harm because it does not definitely qualify as a cyber attack. However, because the Tallinn Manual does not require a civilian cyber activity to rise to the level of a cyber attack, necessitating the reinstallation of an operating system may satisfy the threshold of harm requirements if it intends or has the effect of adversely affecting military operations or capabilities or inflicting death, physical harm, or material destruction on persons or objects protected from attack. It is clear that a civilian cyber activity that necessitates the replacement of physical components, such as a hard drive, is going to satisfy the threshold of harm.

The Tallinn Manual’s characterization of the threshold of harm requirement, while similar to Interpretive Guidance’s, because of its reliance on intention, creates a lower standard for civilian actions that may satisfy this requirement. This means that depending on which standard a state uses, it will be easier for individuals who directly participate in hostilities through cyber means to satisfy the threshold of harm requirement than civilians directly

64 Id.
65 Id.
participating in hostilities through kinetic means. A lower standard results in civilian cyber participants losing their protected status in more circumstances, and thus becoming more easily targetable, than civilian kinetic participants.

\textit{C. Direct Causation of Harm}

There is little difference between the Interpretive Guidance’s characterization of the second requirement and the Tallinn Manual’s. This requirement focuses on the causal link between a civilian’s actions and the resulting harm. However, the Tallinn Manual’s invocation of “intended harm” in the fulfillment of this requirement results in a lower standard under the Tallinn Manual.

The second requirement in the DPH analysis is direct causation. The Interpretive Guidance requires there to be a “direct causal link between a specific act and the harm likely to result either from the act, or from a coordinated military operation of which that act constitutes an integral part.”\textsuperscript{66} The Interpretive Guidance makes a distinction between hostilities and “the general war effort” or “war-sustaining activities.”\textsuperscript{67} While the latter two may be “indispensable to harming the adversary,” they differ from hostilities in that hostilities are directly “designed to cause . . . the required harm.”\textsuperscript{68} Direct causation requires that “the harm in question . . . be brought about in one causal step.”\textsuperscript{69} That is, a civilian’s action that “merely builds up or maintains the capacity of a party to harm its adversary, or which otherwise only indirectly causes harm, is excluded from the concept of direct participation in hostilities.”\textsuperscript{70} Furthermore, the direct causation requirement does not demand that the action “be indispensable to the causation

\textsuperscript{66} Nils Melzer, at 51.
\textsuperscript{67} Id. at 51-52.
\textsuperscript{68} Id. at 52
\textsuperscript{69} Id. at 53.
\textsuperscript{70} Id.
of harm.” As an example, the Interpretive Guidance asserts that someone who assembles and stores an “improvised explosive device (IED) in a workshop,” or purchases or smuggles components for an IED would likely not qualify as someone directly participating in hostilities. This is because while those actions “may be connected with the resulting harm through an uninterrupted causal chain of events, . . . unlike the planting and detonation of that device, [they] do not cause that harm directly.”

The Interpretive Guidance also addresses direct causation in collective operations. This includes actions that cause harm “only in conjunction with other acts.” The Interpretive Guidance recognizes that actions, which on their own would not “directly cause the required threshold of harm,” satisfy the direct causation requirement where those actions constitute “an integral part of a concrete and coordinated tactical operation that directly causes such harm.” The Interpretive Guidance provides several examples of actions that may meet this characterization of the second requirement: “the identification and marking of targets, the analysis and transmission of tactical intelligence to attacking forces, and the instruction and assistance given to troops for the execution of a specific military operation.”

The Tallinn Manual is fairly silent on the second requirement of direct causation. There is no additional commentary that elucidates how a civilian’s actions may satisfy this requirement other than the Tallinn Manual’s assertion of the requirement. It requires that “a direct causal link between the act in question and the harm intended or inflicted” exist in order for a civilian’s

---

71 Id.
72 Id. at 54.
73 Id.
74 Id.
75 Id.
76 Id. at 54-55.
77 Id. at 55.
action to qualify him or her as a civilian directly participating in hostilities.\textsuperscript{78} The Tallinn Manual again uses intended harm instead of likely harm. While the focus of this requirement is the causal link between the action and the harm, including a causal link between a civilian’s action and the intended harm, again, can do nothing but lower the standard of what may satisfy the second requirement in a cyber attack. As long as there is a direct causal link between a civilian’s action and the intended harm, even if that harm is not realized, the second requirement is satisfied under the Tallinn Manual.

The Tallinn Manual would likely agree with Interpretive Guidance that the identification and marking of targets, the analysis and transmission of tactical intelligence to attacking forces, and the instruction and assistance given to troops for the execution of a specific military operation, even if conducted through cyber means, would all meet the direct causation of harm requirement. The Tallinn Manual would also agree that the planting and detonation of an IED, or a cyber equivalent, would satisfy the direct causation requirement. However, it is reasonably possible that the Tallinn Manual would conclude that the person who assembles and stores the IED would also satisfy the direct causation requirement. This is because the person who assembles and stores the IED intends the same harm that the person arming and detonating the IED does. Because the Tallinn Manual factors intended harm into its characterization, it has the effect of creating a lower standard for actions that may satisfy the direct causation requirement. This results in civilian cyber participants losing their protected status in more situations, and thus making it easier to target them, than civilian kinetic participants.

\textit{D. Belligerent Nexus}

The Tallinn Manual’s characterization of the belligerent nexus requirement differs from the Interpretive Guidance’s in its succinctness. This opens the door for a broader, and more

\textsuperscript{78} Tallinn Manual, Rule 35.4.
inclusive, application when civilians directly participate in hostilities.

The belligerent nexus requirement calls for an act to not only “be objectively likely to
inflict harm that meets the first two [requirements], 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.”\textsuperscript{79} The
Interpretive Guidance explains that “not every act” that meets the threshold of harm “necessarily
amounts to direct participation in hostilities.”\textsuperscript{80} Rather, “the concept of direct participation in
hostilities is restricted to specific acts that are so closely related to the hostilities conducted
between parties to an armed conflict that they constitute an integral part of those hostilities.”\textsuperscript{81}
Furthermore, according to the Interpretive Guidance, actions which are not “designed to harm a
party to an armed conflict,” or which are not “designed to do so in support of another party,
cannot amount to any form of ‘participation’ in hostilities taking place between these parties.”\textsuperscript{82}
That is not to say that engaging in one to the exclusion of the other will satisfy the belligerent
nexus. The Interpretive Guidance can only be saying that if it does not do one or the other, it
does not qualify as participation; to satisfy the belligerent nexus, it must do both.

The Interpretive Guidance’s belligerent nexus analysis rarely inquires into the actor’s
intent; rather, it focuses on the objective purpose of the act.\textsuperscript{83} A civilian’s mental state should be
considered only “in exceptional circumstances,” such as “when civilians are totally unaware of
the role they are playing in the conduct of hostilities” or when civilians “are completely deprived
of their physical freedom of action.”\textsuperscript{84} The Interpretive Guidance uses the examples of a “driver
unaware that he is transporting a remote-controlled bomb” or civilians who are “physically

\textsuperscript{79} Nils Melzer, at 58.
\textsuperscript{80} Id.
\textsuperscript{81} Id.
\textsuperscript{82} Id. at 59.
\textsuperscript{83} Id.
\textsuperscript{84} Id. at 60. The Interpretive Guidance refers to involuntary human shields physically coerced into providing cover in close combat.
coerced into providing cover in close combat” as illustrations of each of the aforementioned exceptional circumstances. The Interpretive Guidance posits that civilians in these “extreme circumstances cannot be regarded as performing an action . . . in any meaningful sense” and, because of this, “remain protected against direct attack despite the belligerent nexus of the military operation in which they are being instrumentalized.” The Interpretive Guidance discounts intent as an integral part of the DPH analysis and instead chooses to rely on the objective likelihood that the action will bring about the necessary harm to the advantage of one side of the conflict and the disadvantage of the other.

The Interpretive Guidance addresses the practical relevance of belligerent nexus, admitting that there could be “[m]any activities during armed conflict [that] lack a belligerent nexus even though they cause a considerable level of harm.” Among these types of activities, the Interpretive Guidance included the commission of crimes “for reasons unrelated to the conflict” and “the stealing of military equipment for private use” as examples of activities that could satisfy the threshold of harm requirement but not the belligerent nexus requirement because they “are not specifically designed to support a party to the conflict by harming another.” Even though they may cause the necessary level of harm, if the purpose of the action is not to support one party by harming another, it cannot fulfill the belligerent nexus.

The Tallinn Manual’s characterization of the belligerent nexus requirement is a broad statement: “Finally, the acts must be directly related to the hostilities.” One of the main differences between the Interpretive Guidance’s characterization of the belligerent nexus requirement and the Tallinn Manual’s, is that the Tallinn Manual is much broader in scope than

---

85 Id.
86 Id.
87 Id.
88 Id. at 61.
89 Tallinn Manual 35.4.
the Interpretive Guidance. The Interpretive Guidance states that the action must be in support of one party to the conflict and to the detriment of another. Under the Tallinn Manual, on the other hand, the action must be directly related to the hostilities, suggesting that as long as there is some direct connection between the act and the hostilities, the civilian’s action will be sufficient for the Tallinn Manual’s characterization of the belligerent nexus requirement.

Under the Tallinn Manual’s discussion of the threshold of harm requirement, some Tallinn Experts explicitly posited that acts, which do not directly harm an adversary may be sufficient to qualify as necessary harm. The argument is that any civilian’s cyber action that serves to “enhance one’s own military capacity [is] included in this criterion, as [it] necessarily weaken[s] an adversary’s relative position,” providing the example of “maintaining passive cyber defences of military cyber assets.” While the Tallinn Experts included this in their threshold of harm analysis, it is helpful to discuss it here because of the contradiction to the Interpretive Guidance’s belligerent nexus analysis. The Interpretive Guidance’s characterization of this requirement demanded an action that resulted in benefit to one side and, as a result, harm to the other. Some of the Tallinn Experts only required one, arguing that action would inherently result in support or detriment to the other party.

The Tallinn Manual’s characterization of the belligerent nexus requirement is clearly broader in scope than the Interpretive Guidance’s characterization of the belligerent nexus requirement that required an action to support one side while harming the other. The agreed upon rule, that requires only that an action be directly related to hostilities, seems to forgo this analysis

90 Id.
91 Id.
92 Because the Tallinn Experts only “generally” agreed with the Interpretive Guidance’s DPH paradigm, it is necessary for the sake of analysis to compare and contrast provisions even if this done outside of the authors’ intended placement. This is because of the similarity one provision under the Tallinn Manual’s characterization of a requirement may bear to a provision associated with a different requirement in the Interpretive Guidance.
altogether. As long as a civilian’s cyber action is directly related to the hostilities it does not seem to matter to the Tallinn Experts who receives support or detriment or that anyone receives support or detriment. This allows more civilian actions to fall within the scope of the third DPH requirement.

The Tallinn Manual commentary provides some clarification as to what exactly may not satisfy the belligerent nexus requirement.\textsuperscript{93} The Tallinn Manual states that “criminals who use cyber means to steal State funds belonging to a party to the conflict, but with a view to private gain, would not be direct participants in hostilities.”\textsuperscript{94} It seems that purely private motives, as long as the effects remain independent of the hostilities, will not satisfy the belligerent nexus under the Tallinn Manual characterization. This generally comports with the Interpretive Guidance. While an action may cause significant harm, if the crime was committed for a reason unrelated to the conflict it would not satisfy this requirement. Some members of the group of experts contended that if the operation was conducted to “finance particular military operations,” there would be a sufficient nexus to “constitute direct participation.”\textsuperscript{95} The Interpretive Guidance would also agree with this as long as the money was used to finance specific military operations in support of one party to the conflict and to the detriment of the other. If the money was not used to fund specific military operations, but rather, it was used to fund the general war effort, this would likely fail the Interpretive Guidance’s direct causation link. Likewise, if the money was used to fund military operations that supported both sides or harmed both sides, it would likely fail the Interpretive Guidance’s belligerent nexus requirement because it did not support one side to the conflict and harm another. However, it would likely satisfy the Tallinn Manual’s belligerent nexus requirement that a civilian’s actions be directly related to the hostilities.

\textsuperscript{93} Tallinn Manual, Rule 35.6.  
\textsuperscript{94} Id.  
\textsuperscript{95} Id.
E. Temporal Considerations

The Tallinn Manual starts by approaching temporal considerations in much the same way that the Interpretive Guidance does. However, due to delayed effects or repeated attacks, the Tallinn Manual may extend the time for which a civilian satisfies the DPH regime with the effect of making that civilian targetable for a longer period of time than the civilian would be under the Interpretive Guidance.

The Additional Protocols’ statement on civilians directly participating in hostilities, that they are protected from attack unless and for such time as they directly participate in hostilities, inherently raises questions of timing. There are actions inherent to direct participation in hostilities that occur both before and after the actual direct act, during which a civilian may be targetable, as long as his or her direct participation satisfies the DPH requirements. Both the Interpretive Guidance and the Tallinn Manual discuss the temporal scope of a civilian’s participation and potential problems that arise in applying the law to that civilian’s actions.

The Interpretive Guidance states that because “civilians lose protection against direct attack ‘for such time’ as they directly participate in hostilities, the beginning and end of specific acts amounting to direct participation must be determined with utmost care.”\footnote{Nils Melzer, at 65.} Furthermore, the Interpretive Guidance interprets this phrase to “clarif[y] that such suspension of protection lasts exactly as long as the corresponding civilian engagement in direct participation in hostilities.”\footnote{Id. at 70.} That is, a civilian is targetable only for such as time as he or she is directly participating in hostilities. As soon as that participation ceases, the civilian reassumes his or her protected status until he or she again directly participates in hostilities.

Determining whether or not preparatory actions qualify as directly participating in
hostilities is difficult.\textsuperscript{98} In discussing a civilian’s actions that lead up to the actual direct participation in hostilities, the Interpretive Guidance admits that whether “a preparatory measure amounts to direct participation in hostilities,” and, therefore, the ability to target a civilian, “depends on a multitude of situational factors that cannot be comprehensively described in abstract terms.”\textsuperscript{99} The distinction is made by examining whether or not the preparatory action is “aim[ed] to carry out a specific hostile act” or if the preparatory action is “aim[ed] to establish the general capacity to carry out unspecified hostile acts.”\textsuperscript{100} The Interpretive Guidance characterizes this criterion as whether or not the action plays “an integral part of a specific act” that amounts to “direct participation in hostilities.”\textsuperscript{101} The Interpretive Guidance does not consider it “necessary [or] sufficient for a qualification as direct participation that a preparatory measure occur immediately before” the civilian action.\textsuperscript{102} Likewise, the Interpretive Guidance does not consider it necessary that a “preparatory measure” occur “in close geographical proximity to the execution of a specific hostile act.”\textsuperscript{103} The Interpretive Guidance states that actions that are not “indispensable” to a specific hostile act do not disqualify those actions from satisfying the DPH regime.\textsuperscript{104} That is, preparatory measures can be separated from the actual direct participation in hostilities by time and space and still allow a civilian to be targetable. Furthermore, the Interpretive Guidance asserts that actions that are integral, but not indispensable, will still satisfy the DPH regime.

To illustrate these principles, the Interpretive Guidance, states that the “[equipping], instruction, and transport of personnel; gathering of intelligence; and preparation, transport, and

\textsuperscript{98} Id. at 65.
\textsuperscript{99} Id.
\textsuperscript{100} Id. at 66. (emphasis omitted).
\textsuperscript{101} Id. at 68.
\textsuperscript{102} Id. at 66.
\textsuperscript{103} Id.
\textsuperscript{104} Id.
positioning of weapons and equipment” would all satisfy the DPH regime if done with a “view to the execution of a specific hostile act.” On the other hand, the “purchase, production, smuggling and hiding of weapons; general recruitment and training of personnel; and financial, administrative or political support of armed actors” would not satisfy the requirements of the DPH regime. This means that if actions were not carried with the execution of a specific hostile act in mind, they would not satisfy the DPH requirements.

Regarding the closing of an operation, the Interpretive Guidance states that the ability to target an individual “ends once the individual in question has physically separated from the operation, for example by laying down, storing or hiding the weapons or other equipment used and resuming activities distinct from that operation.” As long as returning from the “execution of a hostile act remains an integral part of the preceding operation, it constitutes a military withdrawal,” and the civilian remains targetable. A civilian’s direct participation ceases when that civilian has removed himself or herself from the operation and taken up activities different from the operation.

The Tallinn Manual approaches the temporal aspect of DPH in much the same way as the Interpretive Guidance. It acknowledges that when a civilian’s action qualifies as directly participating in hostilities, that person is “targetable for such time as he or she is engaged in the qualifying act of direct participation.” The Tallinn Experts “agreed that this would at least include actions immediately preceding or subsequent to the qualifying act.” The Tallinn Manual forgoes any discussion of preceding or subsequent actions integral to a specific hostile

---

105 Id.
106 Id. at 66-67.
107 Id. at 67.
108 Id.
110 Id.
act. This may have the effect of lengthening the period of time for which a civilian is targetable. The Tallinn Manual states that some of the Tallinn Experts asserted that the time period of participation should extend for as long as there is a causal link.

The Tallinn Manual tentatively expands the scope of DPH application to actions that have a delayed effect or a civilian who repeatedly launches cyber attacks. A cyber operation’s delayed effects may be realized some time after a civilian’s actual direct participation ceases. A majority of the Tallinn Manual experts agreed that a civilian should be targetable only for the duration of that civilian’s participation, regardless of when the effects of that participation are realized.

Situations in which a civilian repeatedly launches cyber actions amounting to direct participation in hostilities have the potential to be quite frequent. The Tallinn Experts were split on this question. Some agreed with the Interpretive Guidance saying that each civilian action must be treated separately. Others stated that civilians should be targetable starting with the first qualified cyber operation and “continues throughout the period of intermittent activity.” The second approach seems to make more internal sense in light of the Tallinn Manual’s invocation of intention in its threshold of harm requirement. If all that is required is an intention of harm to one side’s military operations and capabilities, then a civilian may be targetable for quite some time prior to or subsequent to the actual direct participation in hostilities as long as intent exists. This is especially applicable in the case of a civilian who retains the intent to conduct cyber operations amounting to directly participating in hostilities

111 Tallinn Manual, Rule 35.8 and 35.10.
113 Id.
114 Tallinn Manual, Rule 35.10.
115 Id.
116 Id.
117 Id.
over a period of intermittent direct participation. As noted, this is a divergence from the Tallinn Manual’s conception the temporal scope of the DPH regime.

III. CONCLUSION

The Interpretive Guidance and the Tallinn Manual agree on several points. The Tallinn Manual adopts the basic DPH structure as outlined in the Interpretive Guidance with its three requirements. The Tallinn Manual also recognizes the same ancillary issues outlined by the Interpretive Guidance, including the temporal scope of a civilian’s direct participation.

However, the differences between the Interpretive Guidance and the Tallinn Manual have the effect of creating more circumstances where civilian cyber actions may satisfy the DPH requirements, making the civilian cyber participant more often targetable, than when the civilian is participating via kinetic means. The Tallinn Manual’s inclusion of intention in its threshold of harm and direct causation requirement expands the scope of the DPH requirements. The Tallinn Manual’s broad characterization of the belligerent nexus requirement also lowers the standard of what type of civilian action may satisfy it. The effect of these differences is that more civilians are targetable by the victim, or intended victim, of a cyber attack under the Tallinn Manual than are targetable under the Interpretive Guidance. With more civilians conducting cyber operations, this has great implications for their protected status in armed conflict.