2012 4th International Conference on Cyber Conflict
C. Czosseck, R. Ottis, K. Ziolkowski (Eds.)
2012 © NATO CCD COE Publications, Tallinn

Direct Participation in Cyber Hostilities: Terms of Reference for Like-Minded States?

Jody M. Prescott
West Point Center for the Rule of Law
West Point, New York, U.S.A.
jody.prescott@us.army.mil

Abstract: According to its recently published cyber strategy, the U.S. seeks to develop international consensus on how traditional law of armed conflict (LOAC) norms and understandings are modified and applied in cyberspace to help secure this global commons. Although the International Committee of the Red Cross's Interpretive Guidance on Direct Participation in Hostilities and the recent U.S. cyber strategy documents and policy statements are very different in many ways, examination of the relationships between their different aspects could be very useful in setting terms of reference framing the discussions which must occur to develop consensus on how LOAC rules and understandings regarding direct participation in hostilities could be adapted for use in cyberspace. This requires identification of their respective strengths and weaknesses, and potential areas of common ground between them. To be useful, this examination must include consideration of the significance of rules of engagement, formulations of hostile intent, and the proper inferences to be drawn from intelligence analyses as well as the legal standards by which direct participation in hostilities is determined.

Keywords: *direct participation, hostilities, cyber conflicts, law of armed conflict*

1. INTRODUCTION

The recently issued U.S. *International Strategy for Cyberspace* posits an end state in which cyberspace is "an open, interoperable, secure, and reliable information and communications infrastructure that supports international trade and commerce, strengthens international security, and fosters free expression and innovation."[1] To reach that goal, the U.S. foresees coordinated, international action as necessary to "build and sustain an environment in which norms of responsible behavior guide states' actions, sustain partnerships, and support the rule of law in cyberspace."[2] This end state would be fostered by norms resulting from the U.S.'s "work with like-minded states to establish an environment of expectations [...] that ground foreign

1 The White House, International Strategy for Cyberspace: Prosperity, Security, and Openness in a Networked World, 8 (May 2011), *available at* http://www.whitehouse.gov/sites/default/files/rss_viewer/international_strategy_for_cyberspace.pdf [hereinafter "International Strategy"].

2 *Id*.

and defense policies and guide international partnerships."[3] Working with "like-minded" states is important to the U.S. because it believes the current unsettled state of cyberspace has resulted in part from "governments seeking to exercise traditional national power through cyberspace" without "clearly agreed-upon norms for acceptable state behavior."[4] In addressing this situation, the U.S. believes that "[l]ong-standing international norms guiding state behavior – in times of peace and conflict – also apply in cyberspace," but that the "unique attributes of networked technology require additional work to clarify how these norms apply and what additional understandings might be necessary to supplement them."[5]

This paper suggests that a comparison of the INTERPRETIVE GUIDANCE[6] of the International Committee of the Red Cross (ICRC) on direct participation in hostilities could, in conjunction with the *International Strategy* and subsequent U.S. Department of Defense (DoD) cyber strategy documents and policy statements, help set terms of reference to frame the discussions concerning the application of the principle of direct participation in hostilities in cyberspace. This requires, however, a frank assessment of the conceptual weaknesses and strengths of each approach, where they differ, and where there may be common ground. Thus, this paper will first set out the main points of the INTERPRETIVE GUIDANCE, particularly noting its consideration of cyber conflict. Next, it will examine the shortcomings in the INTERPRETIVE GUIDANCE's approach to direct participation in modern armed conflicts. Against this backdrop, the apparent U.S. position will be examined to identify possible trends in the development of concepts related to direct participation in hostilities, and the ramifications of these trends were they to become operationalized. In conclusion, this paper will suggest that although the development of consensus among the "like-minded" on the topic of direct participation in hostilities will not likely be simple nor will it be smooth, its progress would be furthered by an understanding of how the relationships between the differences and the similarities in the ICRC and U.S. positions help set terms of reference for the discussions that must occur.

2. THE INTERPRETIVE GUIDANCE

The INTERPRETIVE GUIDANCE sets out three cumulative elements that must be met before an individual is deemed to have lost the presumption in favor of finding him to be a protected civilian in both international and non-international armed conflict: a threshold of harm, direct causation, and a belligerent nexus.

A. Threshold of Harm

As to the threshold of harm, the INTERPRETIVE GUIDANCE notes that if the reasonable result of an act would be "harm of a specifically *military nature*," this requirement would generally be met "regardless of the quantitative gravity" of the adverse effect.[7] As an example,

3 *Id*. at 9.
4 *Id*.
5 *Id*.
6 NILS MELZER, INT'L COMM. OF THE RED CROSS, INTERPRETIVE GUIDANCE ON THE NOTION OF DIRECT PARTICIPATION UNDER INTERNATIONAL HUMANITARIAN LAW 20 (2009), *available at* http://www.icrc.org/eng/assets/files/other/icrc-002-0990.pdf [hereinafter "INTERPRETIVE GUIDANCE"]. It was compiled on the basis of reports generated from meetings of international experts in the law of armed conflict (LOAC) held between 2003 and 2008. Id. at 8.
7 *Id*. at 47.

"electronic interference with military computer networks could […] suffice, whether through computer network attacks […] or computer network exploitation."[8] However, were the harm not military, the "specific act must be likely to cause at least death [or] injury, or destruction" of property.[9] Accordingly, although acts such as "the manipulation of computer networks [might] have a serious impact on public security, health and commerce," this impact itself would be insufficient to cross the threshold of harm.[10]

Some writers suggest that such a standard would be too restrictive, and that consistent with article 51.2 of Additional Protocol I[11] (prohibiting measures that terrorize civilian populations), injury should include "severe physical or mental suffering."[12] Further, the "loss of intangible assets (e.g., funds held electronically in a banking system) that are directly transformable into tangible assets (e.g., currency or purchasable objects) could be" within the definition of property.[13] The INTERPRETIVE GUIDANCE, however, focuses on harm that occurs in the geophysical world as a result of physical violence.[14]

B. Direct Causation

The INTERPRETIVE GUIDANCE notes that in keeping with the distinction set out in LOAC between direct participation in hostilities that would render an ordinarily protected civilian targetable and indirect participation (such as working in a munitions factory) which would not remove that protection, the difference between the two must "correspond […] to that between direct and indirect causation of harm."[15] Accordingly, "[i]n the present context, direct causation should be understood as meaning that the harm […] must be brought about in one causal step."[16] Examples of actions that would not meet this standard include capacity building through recruiting and training personnel.[17] The INTERPRETIVE GUIDANCE notes that not all of the experts agreed to this formulation, citing examples such as the building of improvised explosive devices (IEDs) and missiles by non-state actors as being more than "mere capacity building […] and becom[ing] measures preparatory to a concrete military operation."[18] As to the timeframe during which direct participation in hostilities exists, the INTERPRETIVE GUIDANCE states that actions in preparation for an "act of direct participation in hostilities, as well as deployment to and return from the location of its execution, constitute an integral part of that attack."[19] If, however,

> "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

8 *Id*. at 48.
9 *Id*.
10 *Id*. at 50.
11 Protocol Additional to the Geneva Conventions of Aug. 12, 1949, and relating to the Protection of Victims of International Armed Conflicts, June 8, 1977, 1125 U.N.T.S. 3 [hereinafter "AP I"].
12 Michael N. Schmitt, Heather A. Harrison & Thomas C. Wingfield, *Computers and War: The Legal Battlespace*, Background Paper prepared for Informal High Level Expert Meeting on Current Challenges to International Humanitarian Law, Cambridge, June 25-27, 5 (2004).
13 *Id*.
14 INTERPRETIVE GUIDANCE, supra note 6, at 20, 49-50.
15 *Id*. at 52.
16 *Id*. at 53.
17 *Id*. at 54.
18 *Id*. at 54 n.125.
19 *Id*. at 65. These acts must be of "a specific military nature and so closely linked to the subsequent execution of a specific hostile act that they already constitute an integral part of that attack." Id. at 65-66.

be restricted to the immediate execution of the act and preparatory measures forming an integral part of that attack."[20]

C. Belligerent Nexus

As to the third element, the purpose of the act being to directly cause an effect which crosses the required threshold of harm, the INTERPRETIVE GUIDANCE states that before an act could be considered direct participation, it must "be objectively likely to inflict harm that meets the first two criteria [and] specifically designed *to do so in support of a party to an armed conflict and to the detriment of another*."[21] The INTERPRETIVE GUIDANCE holds that such a group must belong to a party to the conflict; a status which "can be shown by conclusive behavior that makes it clear for which party the group is fighting."[22] As Professor Michael Schmitt has noted, this "would exclude those organized armed groups in an international armed conflict that might be directing cyber attacks against one of the parties for reasons other than support of the opposing party," such as unaffiliated patriotic hacker groups.[23]

The INTERPRETIVE GUIDANCE notes that not all uses of armed force in an armed conflict will necessarily be considered part of the on-going hostilities. For example, quelling civil unrest which is unrelated to the actual fighting in a combat zone would be excluded,[24] and armed forces engaged in such activities would find their use of force restricted to applications consistent with law enforcement standards and concepts of individual self-defense.[25] However, it also notes that in many armed conflicts, serious criminals may operate such that "it is difficult to distinguish hostilities from violent crime unrelated to, or merely facilitated by, the armed conflict."[26] In light of the increasing incidence of cybercrime, distinguishing between cyberspace actors who are directly participating in a conflict and those who are merely opportunistic criminals could prove even more challenging than in the geophysical world.

D. Continuous Combat Function

The INTERPRETIVE GUIDANCE also sets out the concept of "continuous combat function,"[27] by which individuals whose functions as part of organized non-state actor armed forces "involve […] the preparation, execution, or command of acts or operations amounting to direct participation in hostilities" may be targeted even if not actively participating in hostilities at the time they are engaged.[28] This is intended to distinguish them from "civilians who participate in hostilities on a merely spontaneous, sporadic, or unorganized basis, or who assume exclusively political, administrative or other non-combat functions."[29] This latter category of individuals

20 *Id*. at 68.
21 *Id*. at 58 (emphasis in original).
22 *Id*. at 35.
23 Michael N. Schmitt, *Cyber Operations and the Jus in Bello: Key Issues*, 87 INT'L LAW STUDIES, INTERNATIONAL LAW AND THE CHANGING CHARACTER OF WAR, 89, 100 (Raul A. Pedrozo & Daria P. Wollschlaeger eds. 2011) [hereinafter "*Cyber Operations*"].
24 INTERPRETIVE GUIDANCE, *supra* note 6, at 62-63.
25 *Id*. at 76.
26 *Id*. at 68.
27 *Id*. at 33.
28 *Id*. at 34.
29 *Id*.

could only be targeted for such time as they were taking a direct part in hostilities, as defined *supra*.[30]

The INTERPRETIVE GUIDANCE qualifies continuous combat function quite restrictively. First, for an individual to have membership in organized non-state actor armed forces, that person must assume a role that "corresponds to that collectively exercised by the group as a whole, namely, the conduct of hostilities on behalf of a non-state party to the conflict."[31] Second, the acts the individual commits in such a role must occur "in circumstances indicating that such conduct constitutes a continuous function rather than a spontaneous, sporadic, or temporary role assumed for the duration of a particular operation."[32]

The significance of the *group* purpose is fundamental to this concept, for as Professor Schmitt has noted, "the concept of armed forces makes no sense in the absence of a group purpose of violence."[33] Such a group could include "an on-line group [that has] a defined command structure and coordinate[s] its war-like activities" in cyberspace.[34] In Professor Schmitt's view, a group without a violent purpose "is but a collection of civilians", and its members only become targetable to the extent that their individual activities constitute direct participation in hostilities.[35] As a practical matter, however, given the fluid nature of identity in cyberspace, if intelligence showed that an individual member of such a group was directly participating in hostilities, and that similar groups ordinarily disguised their true purpose in part by vectoring war-like acts through a single member, it might be reasonably concluded that the requisite group purpose existed.

3. SHORTCOMINGS IN THE INTERPRETIVE GUIDANCE

A. The Standard of Decision

The first of the INTERPRETIVE GUIDANCE's four shortcomings lies in not following through to the logical conclusion that flows from its acknowledgment of the practical and situation-dependent standard to be used to determine whether an individual is a legitimate military target rather than a civilian. It notes that "all feasible precautions must be taken" to ensure that individuals who are targeted are in fact legitimate military targets, and not protected civilians. "[F]easible precautions" are "those which are practicable or practically possible taking into account all circumstances ruling at the time, including humanitarian and military considerations."[36] Accordingly, the INTERPRETIVE GUIDANCE notes that the standard of doubt to be applied in targeting decisions is not the same as that applied in criminal proceedings, and instead "must reflect the level of certainty that can reasonably be achieved

30 "Civilians lose protection against direct attack for the duration of each specific act amounting to direct participation in hostilities, whereas members of organized armed groups belonging to a non-state party to an armed conflict cease to be civilians[,] and lose protection against direct attack for as long as they assume their continuous combat function." *Id*. at 70.

31 *Id*. at 33.

32 *Id*.

33 *Cyber Operations, supra* note 23, at 99.

34 *Id*. at 98-99.

35 *Id*. at 99.

36 Final Report on the Meaning of Armed Conflict in International Law, Use of Force Committee, International Law Association, The Hague Conference, 75 (2010), *available at* http://www.ila-hq.org/en/publications/index.cfm.

in the circumstances."[37] The targeting decision must therefore consider factors such as "the intelligence available to the decision maker, the urgency of the situation, and the harm likely to result to the operating forces or to persons and objects protected against direct attack from an erroneous decision."[38]

These realities mean that the standard that is applied throughout the targeting process is in effect reasonable certainty under the circumstances.[39] Reasonable inferences will be developed as a result of continuing analysis of an incomplete and evolving intelligence picture, and the standard is therefore weighted towards providing significant latitude in the evaluation of the factors that establish direct participation in hostilities, and allowing action in response. Operationally, this reality tends to undermine the cumulative restrictions set out in the INTERPRETIVE GUIDANCE.

B. Dismissal of Hostile Intent

The second problem with the INTERPRETIVE GUIDANCE lies in its assessment of the concept of hostile intent as being too bound up with rules of engagement (ROE)[40] to be useful in determining the legal contours of direct participation in hostilities. Because the meeting of experts viewed hostile intent as a technical ROE term, and ROE as national political and command guidance on the use of armed force that did "not necessarily reflect the precise content of IHL", it was therefore "generally regarded as unhelpful, confusing or even dangerous to refer to hostile intent for the purpose of defining direct participation in hostilities."[41] However, the definition of hostile intent is completely relevant to a discussion of the definition of direct participation in cyber hostilities, because in many ways it sets the lowest threshold for activity that can be seen as justifying a lethal response from an opposing armed force in armed conflict involving unfriendly actors who do not necessarily identify themselves as being members of an organized armed force.

NATO ROE recognize that the different NATO member nations will have different interpretations of the right to engage in self-defense,[42] and to cross-level these inconsistencies ROE are provided for mission accomplishment that include the authority to respond to manifestations of hostile intent.[43] For example, NATO ROE Serial 421 provides that "[a]ttack against [designated] force(s) or [designated] target(s) demonstrating hostile intent (not constituting an imminent attack) against NATO/NATO-led forces is authorized."[44] The NATO ROE define hostile intent as having two elements: the "capability and preparedness of individuals, groups of personnel or units which pose a threat to inflict damage," and "evidence, including intelligence, which indicates an intention to attack or otherwise inflict damage."[45]

37 INTERPRETIVE GUIDANCE, *supra* note 6, at 76.
38 *Id*.
39 Joint Targeting Cycle and Collateral Damage Estimation Methodology (CDM), Briefing by DoD General Counsel, 26 (Nov. 10, 2009), *available at* http://www.nefafoundation.org/newsite/file/awlaki_DODUAVstrikes.pdf.
40 NATO defines ROE as "directives to military forces (including individuals) that define the circumstances, conditions, degree, and manner in which force, or actions which might be construed as provocative, may be applied." NORTH ATLANTIC TREATY ORGANIZATION, MILITARY COMMITTEE, MC 362/1, NATO RULES OF ENGAGEMENT, MC 362/1, 2 (June 30, 2003) [hereinafter "NATO ROE"].
41 INTERPRETIVE GUIDANCE, *supra* note 6, at 59 n.151.
42 NATO ROE, *supra* note 40, at 3-4.
43 *Id*. at ¶2, App. 1, Annex A.
44 *Id*. at A-19.
45 *Id*. at ¶3, App. 1, Annex A.

In illustrating this definition, the NATO ROE look in part to objective, physical indicators of ill intent, such as "manoeuvering into weapons launch positions," and non-tactical events such as the "increased movements of ammunition and the requisition of transport."[46] This definition also sets a threshold of harm to be used to help determine whether hostile intent is present, noting that "[i]solated acts of harassment, without intelligence or other information indicating an intention to attack or otherwise inflict damage, will not normally be considered hostile intent."[47]

The anonymity of cyber space, and the ability of unfriendly actors to "spoof" their true identities,[48] challenges the application of the principle of distinction to cyber actors. In those cases where the accurate identification of the cyber actor would be required before undertaking a certain response in the geophysical world, such as imposing economic sanctions or engaging the known digital infrastructure of a nation because its armed forces had apparently launched a cyber attack by proxy, attribution is of course a crucial issue. In the context of assessing whether an actor with an unknown identity is taking a direct part in hostilities as measured by an assessment of whether their intent is hostile, however, attribution to a particular state or non-state actor may not be necessary before engaging the threat.

C. Inaccurate View of the Intelligence Picture

The INTERPRETIVE GUIDANCE's third flaw is its inaccurate assumption of what targeting intelligence looks like, and its lack of discussion as to how reasonable inferences can be drawn from analyzing patterns of information that will work to fill in the gaps between actual data hard points. These inferences lend themselves to resolving doubt as to whether an individual is taking a direct part in hostilities without triggering the presumption of protected status, under the standard of reasonable certainty discussed *supra*. Although targeting intelligence may often be uneven in quality and depth, the INTERPRETIVE GUIDANCE appears to assume a very broad intelligence picture being available to militaries, one which is very detailed and capable of informing commanders and soldiers at various levels of the information they would need to make the informed decisions to comply with its recommendations. For example, in the determination of whether civilians meet the belligerent nexus element, it makes clear that it is not recommending assessing the subjective intent of the actor. However, it then provides the confusing example of civilians who might be unaware of the role they are playing in hostilities, by unknowingly transporting weapons for example. In this case, it states

> "[t]hey remain protected against direct attack despite the belligerent nexus of the military operation in which they are being instrumentalised. As a result, these civilians would have to be taken into account in the proportionality assessment during any military operation likely to inflict incidental harm on them."[49]

The chances of a targeting authority knowing that an individual transporting such a cargo was unaware of it are highly unlikely. The practical uselessness of this concept is demonstrated by the very fine distinction it attempts to draw between those who are executing a continuous combat function versus those whose war-like acts are "sporadic" or "spontaneous":

46 *Id*. at ¶4, App. 1, Annex A.
47 *Id*.
48 Jody Prescott, *War By Analogy: US Cyberspace Strategy And International Humanitarian Law*, 156 RUSI J. 32, 33-34 (Dec. 2010).
49 INTERPRETIVE GUIDANCE, *supra* note 6, at 60.

> "Where civilians engage in hostile acts on a persistently recurrent basis, it may be tempting to regard not only each hostile act as direct participation in hostilities, but even their continued intent to carry out unspecified hostile acts in the future. However, any extension of the concept of direct participation in hostilities beyond specific acts would blur the distinction made in IHL between temporary, activity-based loss of protection (due to direct participation in hostilities), and continuous, status or function-based loss of protection [...]."[50]

The INTERPRETIVE GUIDANCE provides no guidance that would help distinguish between reports of a series of war-like acts by an individual which are merely spontaneous as compared to reports on a person who commits the exact same sorts of acts but is exercising a continuous combat function. Instead, it posits that it is not operationally possible to "determine with a sufficient degree of reliability whether civilians not currently preparing or executing a hostile act have previously done so on a persistently recurrent basis and whether they have the continued intent to do it again."[51] Hypothetically, whether an individual has committed war-like acts in the past could be tracked by modern intelligence assets, if that information has been collected.[52] Communications intercepts or similar reports could indicate whether this person is participating in the planning of future war-like act. If "the principle of distinction must be applied based on information which is practically available and can reasonably be regarded as reliable in the prevailing circumstances,"[53] then the reasonable inferences that could be drawn from the information in this hypothetical would support an assessment of continuous combat function, rather than war-like spontaneity, on the part of the individual.

D. Too Restrictive Window of Direct Participation

The fourth shortcoming of the INTERPRETIVE GUIDANCE is its overly restrictive definition of the time frame within which those directly participating in cyber hostilities may be targeted. Restricting this attack window to just before, during, and immediately after a cyber event is at odds with the manner in which potential cyber attacks could occur. First, the nature of so-called "Zero Day"[54] defects in digital infrastructure means an unfriendly intrusion could evolve into a potentially catastrophic attack at near light-speed.[55] Second, at the moment it occurs, it is likely very challenging to quickly determine whether the intruder is an opposing state, a terrorist group, a cyber criminal, or a hacker.[56] Execution of a cyber attack might follow immediately after an intrusion, and the preparatory measures might either be invisible to the affected state or seem innocuous.[57] In Professor Schmitt's view, this means that "there may be no 'deployment' at all," since "only a computer, and not proximity to the target is required to

50 *Id*. at 45.
51 *Id*.
52 *See* Major General Michael T. Flynn, Captain Matt Pottinger & Paul D. Batchelor, *Fixing Intel: A Blueprint for Making Intelligence Relevant in Afghanistan, Voices from the Field*, CENTER FOR A NEW AMERICAN SECURITY, 7-8 (2010) (intelligence collection in Afghanistan focused on insurgent activity and identity).
53 INTERPRETIVE GUIDANCE, *supra* note 6, at 35.
54 William Jackson, *Malicious PDFs Exploit Zero-Day Vulnerability and Adobe Reader*, GOV'T COMPUTER NEWS, Feb. 20, 2009, *available at* http://gcn.com/articles/2009/02/20/pdf-zero-day-eploit.aspx.
55 *Cyber Operations, supra* note 23, at 102.
56 Committee on Offensive Information Warfare, *Technology, Policy, Law, and Ethics Regarding U.S. Acquisition and Use of Cyberattack Capabilities*, 18, 21, 90-91,William A. Owens, Kenneth W. Dam & Herbert S. Lin eds, NATIONAL RESEARCH COUNCIL (2009).
57 *Id*. at 90-93.

mount the operations."[58] Further, the cyber attack itself "might last only minutes, perhaps even seconds."[59] The Interpretive Guidance's restriction of direct participation in hostilities to the time of execution and just before or after would therefore "effectively extinguish the right to strike at direct participants."[60]

4. THE APPARENT U.S. PERSPECTIVE ON DIRECT PARTICIPATION

There is no single unclassified U.S. strategy document or policy statement that explicitly sets out how the U.S. understands and intends to apply the concept of direct participation in hostilities to cyber conflicts. Therefore, different unclassified strategy documents and policy statements must both be considered together and individually scrutinized to glean indications of how U.S. policy and thinking might be evolving in this regard. One fundamental theme runs through all the sources of the U.S. position, however: "cyberspace activities can have effects beyond networks; [and] such events may require responses in self-defense" and trigger "commitments [it has] with [its] military treaty partners [...]."[61]

A. The DoD Strategy for Operating in Cyberspace

Rather than focusing on the use of force, the unclassified version of the *DoD Strategy for Operating in Cyberspace* (DoD Strategy),[62] released two months after the publication of the International Strategy, instead describes complementary strategic initiatives which emphasize the need to create a well organized, trained and equipped cyber force structure; to develop partnerships with civilian governmental agencies, private industry, allies and other international partners; and the need to develop a national wellspring of talent and innovation to keep the U.S. military and industry competitive in the cyber arena. Although the DOD Strategy sets out the use of "active cyber defense" as an operating concept, it defines it in a fairly benign manner as the "synchronized, real-time capability to discover, detect, analyze and mitigate threats and vulnerabilities."[63]

To put the *DoD Strategy* into its proper perspective, however, it is useful to examine the statements made by U.S. officials regarding DoD's cyber strategy in general. First, the definition of "active cyber defense" in the *DoD Strategy* is not completely consistent with earlier statements made by U.S. officials that suggested that "active cyber defense" included operations within other nations' digital infrastructures.[64] Similarly, U.S. Deputy Secretary of Defense William J. Lynn remarked at the time the *DoD Strategy* was published that although he believed "destructive or disruptive cyber attacks that could have an impact *analogous* to physical hostilities" would

58 *Cyber Operations, supra* note 23, at 102.
59 *Id*.
60 *Id*.
61 *International Strategy, supra* note 1, at 11-14. NATO handles cyber incidents under the consultative procedures of Article IV of the NATO Treaty rather than as attacks under Article V. *NATO Agrees Common Approach to Cyber Defence*, EURACTIVE.COM, Apr. 4, 2008, *available at* http://www.euractiv.com/infosociety/nato-agrees-common-approach-cyber-defence/article-171377.
62 Department of Defense Strategy for Operating in Cyberspace, DOD, July 2010, *available at* http://www.defense.gov/news/d20110714cyber.pdf [hereinafter "DOD Strategy"].
63 *Id*. at 7.
64 Ellen Nakashima, *Pentagon considers preemptive strikes as part of cyber-defense strategy*, WASHINGTONPOST.COM, Aug. 28, 2010, available at http://www.washingtonpost.com/wp-dyn/content/article/2010/08/28/AR2010082803849_pf.html.

occur in the future, that "the vast majority of malicious cyber activity today d[id] not cross this threshold."[65] Deputy Secretary Lynn's use of the word "analogous" to describe the relationship between war-like acts in the geophysical world and significant ill-intended acts in cyberspace was likely deliberate, and it suggests that the classified version of the *DoD Strategy* does not reflect direct translation into it of LOAC rules and concepts applicable in the geophysical world. Prior to the *DoD Strategy*'s launch, statements by DoD officials had indicated instead that it would be based on a concept of "equivalence" between geophysical world hostilities and unfriendly acts in cyberspace to guide its use of force in the latter domain.[66] On the spectrum of similarity, "equivalence" would suggest a more literal adoption of LOAC concepts and applications than would "analogy".

B. The DoD Cyber Policy Report

In November 2011, DoD provided the U.S. Congress with a report on the status of DoD's efforts to operationalize LOAC concepts in cyberspace.[67] The *Cyber Report* recognized the importance of establishing the identity of unfriendly actors, because cyberspace's "unique characteristics [could] make the danger of escalation especially acute. For instance, the speed of action and dynamism inherent in cyberspace, challenges of anonymity, and widespread availability of malicious tools can compound communications and increase opportunities for misinterpretation."[68] It noted DoD's work "with international partners to bolster cyber forensics capabilities," and very intriguingly, its efforts to "assess the identity of [an] attacker via behavior-based algorithms."[69] Complementing these efforts, the *Cyber Report* noted DoD's intent "to expand and deploy applications that detect, track and report malicious activities across all DoD networks and information systems on a near real-time basis."[70]

The *Cyber Report* also described the scope of the challenge confronting intelligence specialists, noting that "[t]he often low cost of developing malicious code and the high number and variety of actors in cyberspace make the discovery and tracking of malicious cyber tools difficult."[71] Further, "most of the technology used in this context is inherently dual-use, and even software might be minimally repurposed for malicious action,"[72] which made it even more difficult to definitively recognize and effectively track unfriendly cyber actors. Despite these difficulties, it stated that as with military intelligence operations in general, cyber intelligence operations were "governed by long-standing and well-established considerations."[73] However, perhaps in an implicit nod to an aggressive theory of active cyber defense, the report noted "the possibility that those operations could be considered a hostile act."[74]

65 William J. Lynn, Remarks on the Department of Defense Cyber Strategy, speech made in Washington, D.C. (July 14, 2011), *available at* http://www.defense.gov/Speeches/Speech.aspx?SpeechID=1593 (emphasis added).

66 Siobhan Gorman & Julian E. Barnes, Cyber Combat: *Act of War*, WSJ.COM, May 31, 2011, *available at* http://online.wsj.com/article/SB10001424052702304563104576355623135782718.html.

67 DOD Cyber Policy Report Pursuant to Section 934 of the NDAA of FY 2011(Nov. 2011), *available at* http://www.defense.gov/home/features/2011/0411_cyberstrategy/docs/NDAA%20Section%20934%20Report_For%20webpage.pdf [hereinafter "Cyber Report"].

68 *Id*. at 5.

69 *Id*.

70 *Id*.

71 *Id*. at 8.

72 *Id*.

73 *Id*. at 7.

74 *Id*.

Regarding cyber ROE, the *Cyber Report* stated that response options available to the President "may include using cyber and/or kinetic capabilities,"[75] which means that any potential attacker of U.S. cyberspace interests must consider not just the possibility and risk of a U.S. cyber response, but also the possibility of individuals and units conducting the attack and their equipment being engaged in the geophysical world. The *Cyber Report* also stated that the U.S. cyber ROE reflect "the interconnectedness and the speed that defines cyberspace," and that therefore they "reflect: the implications of cyber threats; the operational demands of DoD's continuous, world-wide operations; and the need to minimize disruption from collateral effects on networked infrastructure."[76] Further, the *Cyber Report* noted that "[a]s in the physical world, a determination of what is a 'threat or use of force' in cyberspace must be made in the context in which the activity occurs, and it involves an analysis by the affected states of the effect and purpose of the actions in question."[77] Together, these statements emphasize the crucial importance of the internet to U.S. military operations, and suggest that the cyber ROE provide significant latitude to engage on the basis of hostile intent or hostile act.

The *Cyber Report* suggests that DoD is in fact operationalizing LOAC concepts in cyberspace in an "analogous" rather than an "equivalent" fashion. In general, it notes that "DoD will conduct offensive cyber operations in a manner consistent with the policy principles and legal regimes that the department follows for kinetic capabilities, including the law of armed conflict."[78] Importantly, this consistency is at high and abstract level, and consistency is itself a lesser state of compliance than conformance. The *Cyber Report*'s treatment of the issue of potential violations of third nations' sovereignty rights also suggests this. The *Cyber Report* states that in the case of a neutral third country finding itself involved in a cyber threat to the U.S., DoD would adhere to LOAC principles"[79], and that DoD's responses could "include taking actions short of the use of force as understood in international law."[80] However, a number of factors would need to be considered in each case, including the "[n]ature of the act, [the] role of the 3rd country, its ability and willingness to respond effectively, and potential issues of sovereignty."[81]

5. POTENTIAL RAMIFICATIONS OF THE U.S. CYBER STRATEGY

A. War by Analogy

If cyber conflict is seen as only analogous to war in the geophysical world, then the translation of geophysical LOAC rules and interpretations into cyber LOAC norms and understandings will likely reflect this perspective. If the assessment of the U.S. position *supra* is correct, then the U.S. application of this perspective regarding LOAC seems to be the inclusion of LOAC principles and rules as factors to be considered in whether to take action, along with very functional concerns of practical impact on U.S. interests. This approach presents two potential problems, the first of which is whether the U.S. would be able to persuade a coalition of the like-minded of sufficient international stature to not just agree to this approach, but to the

75 *Id*. at 4.
76 *Id*. at 6.
77 *Id*. at 9.
78 *Id*. at 5.
79 *Id*. at 8.
80 *Id*.
81 *Id*.

specific factors to be considered and any weighting of them in the decision making that would be required as well. Second, given that the U.S. reserves the right to respond to unfriendly cyber action by kinetic action in the geophysical world, and that actions in cyberspace could conceivably ripple into the geophysical world as well, conducting cyber war could become like a game of three-dimensional chess, with different rules on different levels.[82] This would require commanders and legal advisors to not just be familiar with the effects of technology in cyberspace and how the agreed-upon analogous norms applied; they would also need to be able to simultaneously track the effects and the traditional LOAC rules applicable to those effects in the geophysical world. The training, educational and experiential requirements that would need to be met by the individuals filling these positions, to say nothing of the doctrine and educational infrastructure that would need to be built to produce such soldiers, would require a significant investment by nations to create these capabilities.

B. Cyber Due Diligence

The *International Strategy* describes "cyber due diligence" as an emerging norm essential to cyberspace's proper use. This term is defined as states' obligations to protect their "information infrastructures and secure national systems from damage or misuse."[83] As noted *supra*, the *DoD Strategy* is based in part on the employment of "new defense operating concepts to protect DoD networks and systems," and this includes measures to better train DoD personnel and hold them accountable for the proper secure use of digital infrastructure and to prevent intrusions from occurring.[84]

Neutral states are required under international law to enforce their neutrality and prevent parties in armed conflicts from using their territories as bases from which the parties could launch attacks against one another. If a state does not protect its neutrality, whether through lack of will or capacity, it risks being seen by the party receiving attacks from its territory as a co-belligerent. The attacked party might then engage its attackers on the sovereign territory of the ostensibly neutral nation, and in that fashion the neutral nation finds that it has become a direct participant in the conflict.[85] As noted *supra*, the *Cyber Report* sets out a list of factors that would be considered in deciding whether to engage a cyber threat located in a third country, and whether the country is exercising cyber due diligence is arguably included within the factor of whether the country has the capability and willingness to deal with the threat effectively itself. Sovereignty as a consideration is expressed in terms of how the U.S. might handle potential sovereignty issues, which is a functional calculus quite different than the third country's sovereignty itself being a factor. The concept of cyber due diligence, therefore, may have the effect of expanding the concept of direct participation in hostilities through loosening the restrictions on infringing upon another nation's cyber sovereignty.

C. Hostile Intent and Hostile Acts

The U.S. Standing ROE allow its forces to respond with lethal force to acts they perceive to be hostile. "Hostile acts" are defined broadly as "attack[s] or other use[s] of force against

82 Prescott, *supra* note 48, at 35.
83 *International Strategy, supra* note 1, at 10.
84 *DoD Strategy, supra* note 62, at 7.
85 Tess Bridgeman, *The Law of Neutrality and the Conflict with Al Qaeda*, 85 N.Y.U. L. REV. 1186, 1200 n.75 (2010).

the [U.S.], U.S. Forces, or other designated persons or property."[86] The examples provided to illustrate the scope of acts considered hostile confirm this broad application, and "include[s] force used directly to preclude or impede the mission and/or duties of U.S. personnel or vital [U.S. government] property." "Hostile intent" is defined just as broadly,[87] and both U.S. definitions are less restrictive than their NATO ROE counterparts.[88]

Examination of the U.S. position suggests that U.S. cyber ROE provide significant latitude to engage perceived cyber threats. The *Cyber Report* appears to premise action in cyberspace largely upon perception of hostile intent, expressed or implied, and hostile acts.[89] Presumably, because of the speed with which cyber weapons could be deployed, relying only upon cyber due diligence presents too great a risk of intrusion by unfriendly actors into DoD networks. Determining whether an actor is demonstrating hostile intent may require cyber operators to conduct searches for certain malicious code in targeted software, regardless of where in the geophysical world those programs actually resided, as part of active cyber defense.[90] Thus, hostile intent might be deduced from a characteristic of malware's composition without it actually being employed. Interestingly, the U.S. appears to realize that such actions on its part could be perceived as hostile acts, which suggests that the U.S. could, were similar actions undertaken within its digital infrastructure, view them the same way.

D. Threshold of Harm

Although the INTERPRETIVE GUIDANCE appears to set a threshold of harm caused by action against military assets and capabilities lower than the U.S. position's, this may actually be an area of common ground between the two positions. The INTERPRETIVE GUIDANCE notes that if the reasonable result of an act would be "harm of a specifically *military nature*," the threshold of harm requirement would generally be met "regardless of the quantitative gravity" of the adverse effect.[91] The *Cyber Report*, however, states only that hostile acts must be significant to be actionable.[92]

Professor Nils Melzer notes that "it could be argued that cyber attacks unlikely to result in death, injury or destruction could still amount to an 'armed attack' if they aim to incapacitate 'critical infrastructures' within the sphere of sovereignty of another state."[93] In the absence of military harm, however, it is not clear that such actions would result in their perpetrators being targetable if the "attack" resulted in no observable destruction in the geophysical world.[94] The U.S., however, is apparently taking an assessment of effects approach to making such a determination across the board. Presumably, this means guidelines as to significance would

86 INSTRUCTION 3121.01B, STANDING RULES OF ENGAGEMENT/STANDING RULES FOR THE USE OF FORCE FOR U.S. FORCES, CHAIRMAN OF THE JOINT CHIEFS OF STAFF, ¶e, A-3, Enclosure A (Jun. 13, 2005).

87 *Id*. at ¶f, A-3, Enclosure A.

88 *See* NATO ROE, *supra* note 40, at ¶¶3-5, App.1, Annex 1.

89 *Cyber Report, supra* note 67, at 3-4, 6.

90 In response to a question whether the U.S. would be able to prevent a cyber attack before it registered in the U.S., General Alexander has testified before the U.S. Congress that he is seeking ROE "to protect and prevent" cyber attack. Shaun Waterman, *Cyberwarfare rules still being written*, WASHINGTONTIMES.COM, *available at* http://www.washingtontimes.com/news/2012/mar/20/cyberwarfare-rules-still-being-written/.

91 INTERPRETIVE GUIDANCE, *supra* note 5, at 47.

92 *Cyber Report, supra* note 67, at 4.

93 Nils Melzer, *Cyber Warfare and International Law*, UNIDIR Resources, 14-16 (2011), *available at* http://www.unidir.org/pdf/activites/pdf2-act649.pdf.

94 *Id*. at 28, 31.

be consulted in each case of hostile action, but given the speed at which activity moves in cyberspace, these assessments may be in large part driven by computers. This raises questions as to where accountable human commanders and their staffs would be included in the important processes that support decisions to strike direct participants in hostilities.

Traditionally, actions very harmful to the interests of nations that did not involve the actual use of armed force, such as economic sanctions or espionage, were not deemed to be attacks.[95] This understanding enjoys modern currency as well, as shown by the recent definition of the crime of aggression under the Rome Statute of the International Criminal Court. "Aggression" under the Rome Statute is "the use of armed force by a State against the sovereignty, territorial integrity or political independence of another State, or in any manner inconsistent with the Charter of the [UN]."[96] Accordingly, Professor Matthew Waxman notes that "[c]omputer based espionage, intelligence collection, or perhaps even preemptive cyber operations to disable hostile systems would not constitute prohibited force, because they do not produce direct or indirect destructive consequences analogous to a military attack,"[97] that is, damage in the geophysical world.

Cyber espionage under the U.S. approach could conceivably be so significant that it would be seen as analogous to a war-like act, and under the INTERPRETIVE GUIDANCE, the required casual link could possibly be established as well. Sophisticated cyber weapons are thought to be "[c]apable of providing remarkably adaptive payloads whose activation can be triggered in milliseconds or delayed for years."[98] Further, their "[p]ayloads may even be designed to receive instructions or mutate or change their mission either by remote message or upon satisfaction of certain embedded criteria."[99] Intrusions of this sort would appear to be "significant" under the *Cyber Report*, and there could be a causal link between the espionage and the damage sufficiently direct under the ICRC position. In the end, the conclusion as to whether someone was directly participating in hostilities through conducting this sort of potential sabotage, facilitated directly by espionage, might be the same under both the INTERPRETIVE GUIDANCE and the U.S. position.

E. Perceptions of Participation

As noted *supra*, DoD is undertaking efforts to improve its ability to accurately identify actors conducting cyber operations in part through the use of "behavior-based algorithms."[100] Presumably, these algorithms would be used to evaluate how certain software had behaved and then compare these findings against criteria that reflected the identified behavioral characteristics of different actors.[101] The *Cyber Report* does not explicitly state that these algorithms can only be used to evaluate programs that had intruded into DoD systems and had been isolated –

95 Further, depending upon the circumstances, some uses of armed force between states that resulted in damage or even loss of human life have not been deemed armed conflict. Final Report, *supra* note 36, at 14, 18-19, 26-27.

96 Rome Statute of the International Criminal Court, Rome, July 17, 1998, art. 8 bis.

97 Matthew C. Waxman, *Cyber Attacks as "Force" under UN Charter Article 2(4)*, 87 INT'L LAW STUDIES 43, 48 (2011).

98 Sean Watts, *Combatant Status and Computer Network Attack,* 50 VA. J. INT'L L. 391, 402 (2010).

99 *Id*.

100 *Cyber Report, supra* note 67, at 4.

101 *See* G. Narvydas, R. Maskeliunas, & R. Raudonis, *Goal Directed, State and Behavior Based Navigation Algorithm for Smart "Robosofa" Furniture*, 10 ELEC.AND ELEC. ENG'G J. 67, 69 (2011), *available at* http://www.ee.ktu.lt/journal/2011/10/15__ISSN_1392-1215_Goal%20Directed%20State%20and%20Behavior%20based%20Navigation%20Algorithm%20for%20Smart%20Robosofa%20Furniture.pdf (schematic of algorithm in which next steps in navigation process determined in part by assessment of robot's behavior in dealing with obstacles).

perhaps they could be deployed into other digital infrastructures to examine programs resident there to determine whether they posed a threat.[102] Although it recognizes that other nations could perceive such actions as hostile, it does not appear that the U.S. believes that so doing necessarily creates a state of hostilities, or that computer operators who are conducting such intrusions are taking a direct part in hostilities.

Professor Sean Watts points out, however, that "the argument that intelligence collection, or even intelligence analysis, constitutes taking a direct part in hostilities is far stronger when such information increases the destructive effects or lethality of an attack."[103] In terms of the conduct of an actual cyber attack, if cyber specialists provide real time updates and assessments, "their contributions to the computer network attack begin […] to look progressively more like direct participation in hostilities."[104] In terms of cyber reconnaissance, the same argument holds true. The armed forces of the state whose digital infrastructure has experienced an intrusion would be derelict in their duties if they did not view that penetration as potentially destructive until shown otherwise, and even if the intrusion's initial purpose was just to find malware, it could have a secondary purpose to find a Zero Day vulnerability that could be exploited destructively at some point in the future. Arguably, the better a state conducts its cyber due diligence, the less likely it is that a mere hacker or cyber criminal could find their way into that state's digital infrastructure. Any intrusion, therefore, would likely be assigned greater seriousness simply because it occurred. This risks unnecessary and potentially unmanageable escalation.

6. CONCLUSION

The INTERPRETIVE GUIDANCE and the U.S. position represent two very different approaches to addressing the issue of direct participation in hostilities in cyberspace. The INTERPRETIVE GUIDANCE is the result of a transparent, deliberate, consensus-driven and heavily academic process geared towards ensuring the appropriate protection of civilians, consistent with its proponent's special role in promoting the continuing and enhanced observation of LOAC.[105] The U.S. position, although cognizant of the need to achieve international consensus (at least among like-minded states), is the evolving product of a nation which is at this time possibly foremost in terms of its cyber capabilities, crafted under conditions of secrecy and heavy classification while likely requiring great internal consensus among operators and civilian and military leaders, and likely geared towards preserving core U.S. economic, political and military interests. Critical examination of the two very different approaches allows an assessment of the relationships between strengths and weaknesses of each; relationships that could help define a common platform of understanding upon which to continue the discussions which must take place to determine how to apply LOAC, and in particular the concept of direct participation in hostilities, to this crucial medium of human economic, political and social interaction.

What form should these discussions take? The U.S. understandings of how it believes it would apply LOAC to operations in cyberspace may have only recently been formalized,[106] suggesting

102 Professor Melzer would argue that "probably [ ] for the purposes of targeting, data should be regarded as an object which may not be directly targeted unless it fulfills all defining elements of a military objective." Melzer, *supra* note 93, at 31.
103 Watts, *supra* note 98, at 427.
104 *Id*. at 429.
105 INTERPRETIVE GUIDANCE, *supra* note 6, at 6.
106 Waterman, *supra* note 90 (U.S. expects standing cyber ROE to be implemented by June 2012).

that there is still an opportunity for reexamination and adjustment. However, given the speed at which both cyber technology and national legal frameworks for its use appear to be evolving,[107] the ordinary process of international conferences, workshops, and meetings of experts are unlikely to prove fruitful in narrowing the gap between classified national understandings and their implementers on the one hand and public scholarly interpretations and their proponents on the other. What is needed is a common experiential approach in which national cyber security personnel, including commanders, operators and lawyers, would work together with academics and representatives of international and non-governmental organizations in cyber situational training exercises. The purpose of these scenarios would not be to test whether particular cyber strategies and tactics would be successful; rather, they would place proponents of particular legal interpretations in the position of being forced to apply those interpretations to evolving simulations. The results of the different groups working through the simulations could then be analyzed and collectively compared by the participants, and this could lead to a better appreciation on everyone's part as to how legal inputs into cyber operations might actually play out. Otherwise, the divergence between classified understandings of LOAC's application in cyber conflict and their counterparts in the public domain will likely only widen, to the detriment of defending the democratic values inherent in the notion of a cyberspace commons.

107 Ellen Nakashima, *Pentagon is accelerating development of cyberweapons*, WASHINGTONPOST.COM, Mar. 19, 2012, *available at* http://www.washingtonpost.com/world/national-security/us-accelerating-cyberweapon-research/2012/03/13/gIQAMRGVLS_story.html.