IDP. Revista de Internet, Derecho y Política
E-ISSN: 1699-8154
dbindexing@uoc.edu
Universitat Oberta de Catalunya
España

Delerue, François
Civilian Direct Participation in Cyber Hostilities
IDP. Revista de Internet, Derecho y Política, núm. 19, octubre, 2014, pp. 3-17
Universitat Oberta de Catalunya
Barcelona, España

IDP
REVISTA D'INTERNET, DRET I POLÍTICA
REVISTA DE INTERNET, DERECHO Y POLÍTICA

www.uoc.edu/idp

ARTICLE

Civilian Direct Participation in Cyber Hostilities*

François Delerue
Ph.D. researcher in International Law
at the European University Institute (Florence)

Published: October, 2014

Abstract
This article studies the application of a well-known notion of international humanitarian law, civilian direct participation in hostilities, to cyber warfare.

According to the principle of distinction, civilians and combatants must be distinguished in times of armed conflict. The shift of hostilities from the real world into cyberspace affects neither the definition of combatants nor the negative definition of civilians. However, beyond the classical approach of the principle of distinction, the changing character of warfare also concerns cyber warfare. Indeed, the distinction between battlefields and civilian areas is increasingly less clear and a rising number of non-combatants directly participate in hostilities in various ways. Cyber means, and the development of cyber warfare, offer numerous new possibilities for non-combatants who want to take part in hostilities. It has never been this easy for civilians to get involved in hostilities and most civilians are ignorant of the consequences of their actions.

Recently, two groups of experts have released documents partly related to this topic with divergent conclusions: the first one is the *Interpretive Guidance on the Notion of Direct Participation in Hostilities under International Humanitarian Law* adopted by the ICRC in 2009. The second one is the *Tallinn Manual on the International Law Applicable to Cyber Warfare* written at the behest of NATO's Cooperative Cyber Defence Centre of Excellence. As these documents differ in their approach to reading of the topic, part of this article will analyze their divergences.

Keywords
cyber warfare, International Law, civilians, armed conflict, internet

Topic
cyber warfare

* This article constitutes the communication presented by the author at the International Conference on Internet Law and Politics 2014, and as such was included in the Proceedings of the Conference, which are available at <http://hdl.handle.net/10609/36801>.

Participación directa de la población civil en hostilidades cibernéticas

Resumen

Este artículo estudia la aplicación de un concepto muy conocido en el campo del derecho internacional humanitario, la participación directa de la población civil en las hostilidades, en la guerra cibernética.

De acuerdo con el principio de distinción, en periodos de conflicto armado hay que distinguir entre población civil y combatientes. La transición de las hostilidades del mundo real al ciberespacio no afecta a la definición de combatientes ni tampoco la definición negativa de población civil. Sin embargo, más allá del enfoque clásico del principio de distinción, el carácter cambiante de la guerra también es aplicable a la guerra cibernética. De hecho, la distinción entre campos de batalla y zonas civiles es cada vez menos clara y un número creciente de no combatientes participan directamente en las hostilidades de forma diversa. Los medios cibernéticos y el desarrollo de la guerra cibernética ofrecen numerosas nuevas posibilidades a los no combatientes que quieren participar en las hostilidades. Para la población civil nunca ha sido tan fácil participar, pero la mayoría de civiles desconocen las consecuencias de sus acciones.

No hace mucho, dos grupos de expertos han publicado sendos documentos, parcialmente relacionados con este tema, que llegan a conclusiones divergentes: por un lado, Interpretive Guidance on the Notion of Direct Participation in Hostilities under International Humanitarian Law, elaborado por el Comité Internacional de la Cruz Roja (ICRC) en 2009, y, por el otro, Tallinn Manual on the International Law Applicable to Cyber Warfare, redactado a petición del Centro de Excelencia de Cooperación en Ciberdefensa de la OTAN. Dado que estos documentos difieren en cuanto a la interpretación del tema, parte de este artículo analizará las divergencias.

Palabras clave

guerra cibernética, derecho internacional, población civil, conflicto armado, internet

Tema

guerra cibernética

1. Introduction

According to the principle of distinction, civilians and combatants shall be distinguished in times of armed conflict. However, the differentiation between battlefields and civilian areas is less and less clear, and an increasing number of civilians are taking direct part in hostilities in various ways. Cyber means, and the development of cyber warfare, offer numerous new possibilities for civilians who want to take part in hostilities.

First of all, cyber warfare clearly differs from conventional warfare simply because everybody can gain easy access to *cyber weapons*. Most people today have, indeed, access to a computer and Internet is full of websites, blogs, and forums, which describe how to design a cyber attack, and it is also easy to download ready-to-use computer codes for cyber attacks.

A good illustration of civilians taking direct part in hostilities can be found in the partisans characterized

by Carl Schmitt in his famous book titled *Theory of the Partisan*.1 Schmitt describes four traits that characterize a partisan: irregularity, intense political engagement, tactical versatility and speed, and a *telluric* character. Irregularity remains an important trait for a *cyber partisan*, as well as tactical versatility and speed, which is even truer and more relevant in the Internet age. Indeed, through the Internet a cyber partisan can act from wherever he wants, affecting a computer anywhere in the world. But, to me, it seems that the main challenge for Schmitt's theory in the information age concerns the intense political engagement. If a cyber partisan is fighting as a partisan in the real world, he might be aware of the risks; he can, indeed, be injured or even killed during the hostilities. Furthermore, it is not always easy and safe to acquire weapons for the classical partisan; as a consequence, his motives must be strong. But in the information age, it is very easy to be informed about what happens everywhere in the world and to find a way to act: access to *cyber weapons* is very easy. Also, cyber partisans are not physically engaged in the hostilities and due to the relative distance between them and the battlefield they can feel that they are immune from the consequences of their actions. This is perhaps one of the biggest shifts of civilian direct participation in cyber warfare. Indeed, it has never been so easy to get involved and most people tend to ignore the consequences of their actions.

Secondly, new technologies are omnipresent in modern battlefields. Consequently, States and their armies, as well as private actors, need the best technicians to use those technologies. It is the same in the case of cyber warfare, and this situation can lead to a large number of civilians involved in cyber hostilities. Sean Watts described this situation perfectly:

> Reports indicate that few information operations experts currently serve as active duty soldiers. Many private companies have employed the skills of those with expertise in the various weapons commonly used in CNA. For example, Panasonic hired a formally convicted computer hacker to monitor its cybersecurity. The government has also hired cybercriminals as "cyberwarriors" or for defensive purposes. Additionally, many of the individuals who conduct CNA attacks have been recruited from various disciplines with- in the military, including intelligence, operations, and communications.2

In order to review those different issues, this article addresses the question of civilian direct participation in cyber hostilities, firstly, by defining the notion; secondly, by focusing and comparing the divergent approaches of the *Interpretive Guidance on the Notion of Direct Participation in Hostilities under International Humanitarian Law* (hereafter the *Interpretive Guidance*), adopted by the ICRC in 2009, and the *Tallinn Manual on the International Law Applicable to Cyber Warfare* (hereafter the *Tallinn Manual*) written at the behest of NATO's Cooperative Cyber Defence Centre of Excellence. Moreover, the article addresses two issues more specific to cyber warfare: civilians acting from outside the geographical limits of the armed conflict or who are unaware of their participation.

2. The Notion of Direct Participation in Hostilities

Nowadays, "civilians saturate the modern battlefield, often engaging in activities that have traditionally been performed by members of the armed forces (combatants)".3 Under international humanitarian law (IHL) this participation is characterized by the notion of civilian direct participation in hostilities. This notion is not clearly defined in IHL (2.2.) and its foundation can be found in various instruments (2.1.)

2.1. Sources and legal value of the notion of direct participation in hostilities

The notion of civilian direct participation in hostilities is deduced from treaty law applicable to international and non-international armed conflicts. Article 51(3) of Protocol Additional I,4 which deals with international armed conflicts, and article 13(3) of Protocol Additional II,5 which deals with

1. C. Schmitt (2007).
2. S. Watts (2012, p. 160).
3. M. N. Schmitt (2010).
4. Protocol Additional to the Geneva Conventions of 12 August 1949, and relating to the Protection of Victims of International Armed Conflicts, 8 June 1977 (hereafter Protocol Additional I).
5. 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 (hereafter Protocol Additional II).

non-international armed conflicts, prescribe that civilians shall enjoy general protection against dangers arising from military operations "unless and for such time as they take a direct part in hostilities". The two additional protocols offer the best expression of this limitation of civilian immunity; however, the common article 3 to the Geneva Conventions limits the protection granted to civilians to "persons taking no active part in the hostilities" and that can be seen as an embryonic version of the concept of civilian direct participation in hostilities. The concept of civilian direct participation in hostilities also appears in other instruments, e.g. in the Rome Statute of the International Criminal Court.

This concept is based on the two additional protocols of the Geneva Conventions of 1949. Conversely to the Geneva Conventions, these additional protocols are not universally ratified and they are not entirely recognized as customary international law. For party States, these two articles are thus binding treaty law, but not for the others.

Several States are not party to these two additional protocols, including important military powers that are currently involved in war situations, notably the United States and Israel. However, the provisions of article 51(3) of Protocol Additional I, and of article 13(3) of Protocol Additional II are considered as part of customary international law.6 Even States that are not party to the additional protocols recognize this customary value. For example, the Supreme Court of Israel has recognized that "all of the parts of article 51(3) of The First Protocol express customary international law."7 In addition, it should be highlighted that "numerous military manuals state that civilians are not protected against attack when they take a direct part in hostilities",8 and recognize the notion of civilians taking direct part in hostilities and its consequences.9

In sum, as article 51(3) of Protocol Additional I and article 13(3) of Protocol Additional II reflect customary international law, the notion of civilian direct participation in hostilities has the same legal value and content for all States in the world.

2.2. Lack of definition

Although it is agreed that the notion of civilian direct participation in hostilities is customary international law, most authors also agree that there is a lack of definition of this notion.

Courts and tribunals can work without definition. As the International Criminal Tribunal for the former Yugoslavia (ICTY) pointed out, courts and tribunals try cases *a posteriori*, which allows for an assessment of situations on a case-by-case basis.10 Also, the work of the International Committee of the Red Cross (ICRC) on this notion started with a case-by-case approach.11 However, this lack of definition can become problematic during an armed conflict, as Nils Melzer highlighted, since "it leaves military commanders operating in situations of armed conflict without satisfactory guidance as to the legal standards governing the force used in response to civilian violence".12 Consequently, the practitioners of IHL need a definition or clarification of this notion. This situation explains why the ICRC decided to set up a reflection on this notion.

3. The Icrc's *Interpretive Guidance on the Notion of Direct Participation in Hostilities under International Humanitarian Law* Applied to Cyber Warfare

Civilians have become increasingly present and active in battlefields, in two roles in particular: firstly as mercenaries or members of private military companies, and secondly

6. L. Doswald-Beck and J. M. Henckaerts (2004, rule 6, pp. 19-24). See also ICTY, Trial Chamber II, *Prosecutor v. Pavle Strugar* (Judgment), IT-01-42-T, 31 January 2005, 101, § 220.
7. Israel, Supreme Court, Public Committee against Torture in Israel and Palestinian Society for the Protection of Human Rights and the Environment v Israel and others, Judgment, Case HCJ 769/02, 14 December 2006, reproduced in Oxford Report on International Law and Domestic Courts 597 (IL 2006) [hereafter Targeted Killing Case], § 30.
8. L. Doswald-Beck and J. M. Henckaerts (2004, rule 6, p. 20); M. N. Schmitt (2010b, p. 13).
9. M. N. Schmitt (ed.) (2013, rule 35).
10. ICTY, Trial Chamber, *the Prosecutor v Dusko Tadic a/k/a 'Dule'* (Opinion and Judgment), IT-94-1-T, 7 May 1997, § 616.
11. M. N. Schmitt (2010a, p.704).
12. N. Melzer (2010).

as civilians taking up arms against a perceived enemy. Additionally, it was in reaction to this increased civilian participation that the ICRC decided to set up a project on the notion of civilian direct participation in hostilities, resulting in the publication of the ICRC's *Interpretive Guidance* in 2009.13

The ICRC and the T.M.C. Asser Institute conducted this project from 2003 to 2008. Around 50 experts, participating in their private capacity and coming from different backgrounds (academic, military, governmental, and non-governmental), took part in five informal meetings. The project and the *Interpretative Guidance* did "not endeavour to change binding rules of customary or treaty IHL, but reflect[ed] the ICRC's institutional position as to how existing IHL should be interpreted in light of the circumstances prevailing in contemporary armed conflicts."14 The project was designed in order to address three questions: first, "[w]ho is considered a civilian for the purposes of the principle of distinction?"; second, "[w]hat conduct amounts to direct participation in hostilities?"; and third, "[w]hat modalities govern the loss of protection against direct attack?".15

The notion of civilian direct participation remains, however, highly controversial, and the final version of ICRC's *Interpretative Guidance* is far from non-contentious; it seems that it was impossible to reach the planned output to produce a consensus document.16 Some participants pushed for a more restrictive understanding of this notion, which would lead to a more protective status for civilians, and others "consistently advocated for a more permissive targeting regime than is proposed in the Interpretative Guidance".17 Also, as specified in the *Interpretative Guidance*, it is "an expression solely of the ICRC's views".18 As the *Interpretative Guidance* is highly controversial, the ICRC "took the unusual step of publishing the *Interpretative Guidance* without identifying the participants".19 Finally, it is important to note that this is a non-binding document for States20 even if it could influence States' practice.

Nils Melzer, the editor of the *Interpretative Guidance*, has replied to some of the criticisms that have been directed against it. 21 Notably, he listed some issues remaining controversial in his eyes:

> (1) the criteria for distinguishing civilians from members of organized armed groups; (2) the so-called "revolving door" of protection according to which civilians can repeatedly lose and regain protection against direct attack; and (3) the restraints imposed on the use of force against legitimate military targets. Finally, although the three defining elements of "direct participation in hostilities," the core piece of the Guidance, were far less controversial, their application to certain activities, such as voluntary human shielding and hostage taking, still gave rise to significant disagreement among the participating experts.22

As noted even by those who criticize the *Interpretative Guidance*, "[t]he work effectively identifies and frames the issues and offers a sophisticated departure point for further mature analysis".23 One of the critics, Michael N. Schmitt, for instance, highlighted that "the constitutive elements of

13. N. Melzer (2009a).
14. Ibid 9; *contra* see K. Watkin (2009). See also: W. H. Parks (2009, pp. 794-795); see also the reply formulated in N. Melzer (2009, pp. 893-894).
15. *Interpretive Guidance* (n. 13, p. 13).
16. N. M. Schmitt (2010b, n. 8, p. 6).
17. N. Melzer (2009b, n. 14, p. 834); in this way Michael N. Schmitt, among others, considers that 'the Interpretive Guidance repeatedly takes positions that cannot possibly be characterized as an appropriate balance of the military needs of states with humanitarian concerns. [...] Counter-intuitively, non-state actors, who enjoy no combatant privilege, benefit from greater protection than do their opponents in the regular armed forces. It is similarly disturbing that individuals who directly participate on a recurring basis enjoy greater protection than lawful combatants. [...] Unfortunately, the Interpretive Guidance, the product of tireless efforts on the part of the ICRC and the experts involved, sets forth a normative paradigm that states that actually go to war cannot countenance.' (M. N. Schmitt, 2010b, n. 8, p. 44); see also K. Watkin (2009, n 14, pp. 693-694).
18. Interpretive Guidance, (n. 13, p. 6).
19. N. M. Schmitt (2010b, n. 8, p. 6); see also W. H. Parks (2009, n. 14, p. 784).
20. *Interpretive Guidance* (n. 13, p. 6.
21. N. Melzer (2009b, n. 14, pp. 831-916).
22. *Ibid*. 834.
23. Schmitt (2009b, n. 8, p. 44).

direct participation, although not bereft of flaws, represent a useful step forward in understanding the notion."[24]

That being said, this article on direct participation in hostilities and cyber warfare is built upon the *Interpretative Guidance*, as it is the most thorough work on this topic to date. It is important to note that the editor of the *Interpretative Guidance* has published several articles on cyber warfare and, moreover, the *Interpretative Guidance* itself made some references to cyber warfare. These references have been analyzed in a critical perspective by Georg Kerschischnig,[25] for whom "the examples mentioned in the ICRC Guide are not convincing."[26] In addition, the dispositions on civilian direct participation of the *Tallinn Manual* are compared in this article to those of the *Interpretative Guidance*.

This article partly follows the outlines of the *Interpretative Guidance* but it will address only the points that are relevant for its topic: the ICRC's constitutive elements of the notion of direct participation in hostilities (3.1.), its temporal scope (3.2.), and the modalities governing the loss of protection (3.3. and 3.4.).

3.1. The constitutive elements of the notion of direct participation in hostilities

The fifth recommendation, on the *constitutive elements of direct participation in hostilities*, is the heart of the *Interpretative Guidance*.[27] Three cumulative criteria have been formulated: the threshold of harm (3.1.1.), the direct causation (3.1.2.), and the belligerent nexus (3.1.3.).[28] It must be noted that even though "various experts entertained specific concerns about particular facets of the constitutive elements, most viewed them as, in a very general sense, reflecting the group's broad understanding."[29]

3.1.1. Threshold of harm

The first cumulative criterion is called the *threshold of harm* and it requires that the "act must be likely to adversely affect the military operations or military capacity of a party to an armed conflict or, alternatively, to inflict death, injury, or destruction on persons or objects protected against direct attack".[30] On one hand, the effects of the act "must be likely to" produce the required consequence but must not necessarily have produced it. Thus, the *threshold of harm* can be reached without or before the materialization of the harm; the probability of the harm is sufficient.[31] In this way, "wherever a civilian had a subjective *intent* to cause harm that was objectively identifiable, there would also be an objective *likelihood* that he or she would cause such harm."[32] On the other hand, the *threshold of harm* can be reached alternatively by causing "harm of a specifically military nature or by inflicting death, injury, or destruction on persons or objects protected against direct attack."[33] It should therefore be highlighted that the threshold is higher when the target is not military.

The *Interpretative Guidance* states that "the interruption of electricity, water, or food supplies, [...] the manipulation of computer networks, [...] would not, in the absence of adverse military effects, cause the kind and degree of harm required to qualify as direct participation in hostilities."[34] This is particularly interesting and relevant in the context of cyber warfare. Indeed, cyber attacks can lead to death, injuries or destruction but usually this is an indirect consequence. Therefore, the civilian side of the *threshold of harm*, that is to say causing "death, injuries or destruction on persons or objects protected against direct attack", seems to be difficult, or even impossible, to fulfil by a cyber operation.

The *Interpretative Guidance* focuses on harm that occurs in the real world.[35] If death and injuries are impossible in

24. *Ibid*. 43; see also B. Boothby (2009, p. 768); it should be noted that most of articles and books studying the notion of direct participation in hostilities and cyber warfare refer to the *Interpretative Guidance*, see e.g. *The Tallinn Manual* (n 9) rule 35.
25. G Kerschischnig (2012, pp. 207-213).
26. *Ibid. p.* 209.
27. N. Melzer (2009b, n. 14, p. 856).
28. *Interpretive Guidance* (n. 13) p. 16, pp. 46-64; see also M. N. Schmitt (2010a, n. 3, passim).
29. M. N. Schmitt (2010b, n. 8, p. 27).
30. *Interpretive Guidance* (n. 13, p. 16, pp. 46-50).
31. *Ibid.* 47; M. N. Schmitt (2010a, n. 3, pp. 724-725).
32. M. N. Schmitt (2010a, n. 3, p. 725).
33. *Interpretive Guidance* (n. 13, p. 47).
34. *Ibid*. 50.
35. J. M. Prescott (2012, p. 253).

cyberspace, there can still be damage and destruction of data. Destruction or damage of data could, I think, be seen as reaching the requirement of the *threshold of harm* but we should be very careful on this point. If we include it, we should add a threshold of intensity to it. Indeed, the destruction of patient's data in a hospital database would, from my point of view, reach the *threshold of harm,* whereas the participation in cyber operations against the website of a private company would not.

The military side of the *threshold of harm*, that it must "adversely affect the military operations or military capacity", seems easier to reach by cyber operations. However, we should bear in mind that this military side can be too permissive against civilians participating in hostilities without being aware of what could be the consequences of their acts. Cyberspace offers an easy way to express protest against military operations, and perpetrators might not be aware of the legal repercussions of their actions. Also, it seems that most cyber operations might be likely to harm the military but with a very low consequence and the qualification of direct participation can be seen as disproportionate in this context.

3.1.2. Direct causation

The second cumulative criterion is called *direct causation* and it requires that "there must be a direct causal link between the act and the harm likely to result either from that act, or from a coordinated military operation of which that act constitutes an integral part".[36] The content of this criterion remains controversial.[37] I will not address the debate on this criterion here but solely analyze it in the perspective of cyber warfare.

Generally, to satisfy the criterion of *direct causation*, a specific act must directly cause or be expected to cause, the harm that satisfies the first criterion (*threshold of harm*) by itself or as an integral part of a collective operation.
Cyber operations do not particularly challenge this criterion. Indeed, even if the act can be perpetrated far from the battlefield that does not affect the causal link between the act and the caused or expected harm.[38]

However, some situations are specific to cyber warfare, for example a civilian who produces a cyber weapon. At first, it seems that this can be compared to the example of assembling and storing of an improvised explosive device (IED) given by the *Interpretative Guidance*.[39] Even if the assembling and storing of the IED "may be connected with the resulting harm through an uninterrupted causal chain of events, but, unlike the planting and detonation of that device, [it] do[es] not cause that harm directly."[40] But, in the case of cyber warfare, cyber weapons must be, in most cases, designed for a specific cyber operation. As a consequence, it seems that there is direct causal link between the production of the cyber weapon and the expected harm, and so the producer of it can be qualified as taking direct part in hostilities.

In cases of a collective cyber operation, even if a civilian's contribution does not satisfy the causal link on its own, the civilian can be considered as taking direct part in hostilities due to his involvement in the collective operation. The *Interpretative Guidance* illustrates this with the example of people involved in an attack carried out by unmanned aerial vehicles.[41]

3.1.3. Belligerent nexus

The third and last cumulative criterion, the *belligerent nexus* is "the less controversial of the three".[42] The *belligerent nexus* requires that the "act must be specifically designed to directly cause the required threshold of harm in support of a party to the conflict and to the detriment of another".[43] Also, this criterion is not specifically transformed or affected by cyber warfare and does not require further discussion on it.

However, it should be noted that the *belligerent nexus* requires that direct participation be distinguished from

36. *Interpretive Guidance* (n. 13, p. 16, p. 46, pp. 51-58).
37. See notably M. N. Schmitt (2010a, n. 3, pp. 725-735).
38. *Interpretive Guidance* (n. 13, p. 55).
39. *Ibid* 54; *contra* M. N. Schmitt (2010b, n. 8, p 31); see also M. N. Schmitt (2010a, n 3, pp. 731-732).
40. *Interpretive Guidance* (n. 13, p. 54).
41. *Ibid*.
42. N. M. Schmitt (2010a, n. 3, 735).
43. *Interpretive Guidance* (n. 13, p. 16, pp. 46, 58-64); see also N. M. Schmitt (2010a, n. 3, pp. 735-736).

individual self-defence[44] and also from opportunistic criminal activities. In cyberspace this criterion should be analyzed with great care; indeed, direct participation in hostilities and criminal activities can be closely linked and difficult to tell apart.[45]

3.2. Temporal scope of the direct participation in hostilities

The temporal scope can be divided into two questions: what is encompassed in direct participation in hostilities (3.2.1.) and what is its duration (3.2.2.).

3.2.1. Preparatory measures, deployment and return

According to section VI of the *Interpretative Guidance* titled *Beginning and end of direct participation in hostilities*, "[m]easures preparatory to the execution of a specific act of direct participation in hostilities, as well as the deployment to and the return from the location of its execution, constitute an integral part of that act."[46]

Preparatory measures must be distinguished whether they aim "*to carry out a specific hostile act* or aim *to establish the general capacity to carry out unspecified hostile acts*"; only the former constitutes an act of direct participation.[47] As noted before, in most cases it seems that the creation of cyber weapons will constitute an act of direct participation, as each cyber weapon needs to be adapted to its target. It is important to note that the temporal or geographical distance from the target of the civilian who directly participates does not affect the qualification of direct participation in hostilities.[48] It should be noted that a few authors criticize this approach and plead for an extension of the qualification of the direct participation in hostilities to acts aiming to increase the general capacity of a belligerent and not only to those linked to a specific act.[49]

On the question of the *deployment and return*, the *Interpretative Guidance* specifically mentions the question of cyber warfare:

> Where the execution of a hostile act does not require geographic displacement, as may be the case with computer network attacks or remote-controlled weapons systems, the duration of direct participation in hostilities will be restricted to the immediate execution of the act and preparatory measures forming an integral part of that act.[50]

It seems that the duration of cyber direct participation is restricted, therefore, to the execution of the act. The trip to the place from where a civilian will launch a cyber attack, and the return from it, cannot be qualified as deployment and return and are not part of the direct participation in hostilities, conversely to the launch of the cyber attack. Nevertheless, the *Tallinn Manual* takes the opposite position:

> Any act of direct participation in hostilities by a civilian renders that person targetable for such time as he or she is engaged in the qualifying act of direct participation. All of the Experts agreed that this would at least include actions immediately preceding or subsequent to the qualifying act. For instance, travelling to and from the location where a computer used to mount an operation is based would be encompassed in the notion.[51]

This distinction seems justified in cases of sporadic acts amounting to direct participation in hostilities. This leads us to the question of the duration of the participation.

3.2.2. Duration

The duration is one of the most controversial parts of the *Interpretative Guidance*.[52] According to the two additional protocols to the Geneva Conventions, a civilian loses his civilian protection "for such time" as he takes direct part

44. N. Melzer (2010, n. 12, § 10).
45. J. M. Prescott (2012, n. 35, p. 254).
46. *Interpretive Guidance* (n 13, p. 65).
47. *Ibid*. 66.
48. *Ibid.*
49. K. Watkin (2009, n . 14, pp. 660-662); B. Boothby (2009, n. 24, pp. 750-751).
50. *Interpretive Guidance* (n. 13, p. 68.)
51. *The Tallinn Manual* (n. 9, rule 35, § 7).
52. N. M. Schmitt (2010b, n. 8, p. 16).

in hostilities. Although this phrasing is highly controversial, it is considered customary international law.[53]

The *Interpretative Guidance* distinguishes between civilians who take part in hostilities sporadically and those who are members of an organized armed group:

> 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.[54]

The loss of protection for each specific act amounting to direct participation in hostilities, and its corollary, the regaining of the protection between each act, is generally called the *revolving door*. This revolving door is a controversial notion,[55] and in this way the *Interpretative Guidance* specifies that "[t]he *revolving door* of civilian protection is an integral part, not a malfunction, of IHL."[56] The experts taking part in the *Tallinn Manual* process were divided on this issue and no consensus was found.[57]

The revolving door is something important that should not be abolished. I believe that it is, indeed, the best way to address the problem of civilians taking direct part in hostilities without giving a disproportionate advantage to either side. However, this notion seems to be very difficult to apply to civilians taking part through cyber means for two reasons. Firstly, cyber attacks can be very quick to launch and so the direct participation of the civilian seems to be difficult to address given the short duration span. Secondly and especially, most cyber attacks are detected after their perpetration, when the civilian perpetrator has already regained his civilian protection.[58]

However, the purpose of the notion of direct participation in hostilities is not to punish the civilian who takes direct part, as a sanction for criminal behaviour, "but a consequence of military necessity in the conduct of hostilities."[59] At the end of his direct participation, a civilian regains his civilian protection and shall again enjoy general protection against dangers arising from military operations. But, this civilian does not enjoy immunity for his acts. Indeed, he remains "subject to criminal prosecution for violations of international or domestic law [he] may have committed during such participation".[60]

Based on this framework, it can be assumed that the difficulty highlighted just before, arising from the short duration of the cyber direct participation, perfectly aligns with the objective of the notion of direct participation that is to end the threat arising from the participation. The civilian remains subject to criminal law and shall be brought to justice for his acts. It is important to note that it is not the direct participation *per se* that is criminalized but the acts perpetrated by the civilian who has directly participated in hostilities.[61] Also, as explained later in this article, there are other ways to end the threat represented by a civilian taking direct part in hostilities through cyber means. It is the same for civilians who were members of an organized armed group and who have regained their civilian protection.

Contrarily, civilians who are "[m]embers of organized armed groups belonging to a non-state party to the conflict cease to be civilians for as long as they remain members by virtue of their continuous combat function."[62] As a consequence, civilians who are assuming continuous combat function as members of an organized armed group, even if only through cyber means, lose their protection as long as they stay with this armed group and do not end their continuous combat function, and this applies not only for each cyber attack they perpetrate.

53. *Ibid* 37-38.; Israel, Supreme court, *Targeted Killing Case*, (note 93), § 38; ICTY, Appeals Chamber, *Blaškić* (Judgment), IT-95-14-A, 29 July 2004, § 157.
54. *Interpretive Guidance* (n 13, p. 17, 70).
55. B. Boothby (2009, n. 24, pp. 753-759); K. Watkin (2009, n. 14, pp. 686-690); M. N. Schmitt (2010b, n. 8, pp. 37-38; M. N. Schmitt (2011, p. 102).
56 *Interpretive Guidance* (n. 13, p. 70).
57. *The Tallinn Manual* (n. 9, rule 35, § 10).
58. See e.g. J. M. Prescott (2012, n. 35, pp. 258-259); M. N. Schmitt (2010, n. 55, p. 102).
59. *Interpretive Guidance* (n. 13, p. 62).
60. D. Fleck (2007, p. 261, § 519).
61. *Interpretive Guidance* (n. 13, pp. 83-85).
62. *Ibid*. 71.

3.3. Presumption of non-participation in case of doubt

According to article 50(1) of Protocol Additional I, "[i]n case of doubt whether a person is a civilian, that person shall be considered to be a civilian". Section VIII of the *Interpretative Guidance* extends this presumption to the determination of whether a civilian is taking direct part in hostilities.[63] Accordingly, when a person is considered to be a civilian, that person can belong in three different categories: civilians who are not taking part in hostilities, "civilians directly participating in hostilities on a spontaneous, sporadic, or unorganized basis", or members of organized armed groups.[64] If there is a doubt over whether that person has directly participated in hostilities, or whether that person is a member of an organized armed group, that person shall be considered to be a civilian not taking direct part in hostilities.[65] The experts taking part in the *Tallinn Manual* were split and did not find a consensus on the existence of this presumption.[66] The position of the ICRC expressed in the *Interpretative Guidance* is, from my point of view, the most accurate one.

3.4. Restraints on the use of force in direct attacks

Section IX of the *Interpretative Guidance* titled *Restraints on the use of force in direct attacks*[67] seems to be one of the most controversial.[68] As we will see below, this section is also one of the most interesting for the application of the notion of direct participation to cyber warfare.

The idea of this section can be found in the famous statement of Jean Pictet, who wrote that "[i]f we can put a soldier out of action by capturing him, we should not wound him; if we can obtain the same result by wounding him, we must not kill him. If there are two means to achieve the same military advantage, we must choose the one which causes the lesser evil".[69] In this way, this section means that "[i]n addition to the restraints imposed by international humanitarian law on specific means and methods of warfare, and without prejudice to further restrictions that may arise under other applicable branches of international law, the kind and degree of force which is permissible against persons not entitled to protection against direct attack must not exceed what is actually necessary to accomplish a legitimate military purpose in the prevailing circumstances."[70]

Footnote 221 of the *Interpretative Guidance* delivers some interesting information for this article:

> During the expert meetings, it was generally recognized that the approach proposed by Pictet is unlikely to be operable in classic battlefield situations involving large-scale confrontations (report DPH 2006, pp. 75 f., 78) and that armed forces operating in situations of armed conflict, even if equipped with sophisticated weaponry and means of observation, may not always have the means or opportunity to capture rather than kill (report DPH 2006, p. 63).

I find this statement very interesting in light of cyber warfare. Indeed, in the confusion of a classic battlefield it can be difficult to adapt the force and the means of an attack to each civilian taking direct part in hostilities. However, the situation is different with cyber direct participation in two ways.

Firstly, those civilians are not on the battlefield, they are not directly and physically threatening soldiers by holding a weapon and targeting them. So there is no direct threat to the soldier that can lead to the necessity to shoot the civilian before he shoots the soldier.

63. *Ibid*. 74-76.
64. *Ibid*. 74.
65. *Ibid*. 75-76; for some scholars '[t]here is no presumption that civilians are not directly participating', see e.g. Boothby (2009, n. 24, p. 766); for some other scholars, in case of doubt, civilian should be presumed to be directly participating in hostilities, see e.g. M. N. Schmitt (2004, p. 509); see also M. N. Schmitt (2010a, n. 3, pp. 737-738); see also the reply from N. Melzer to those criticisms (2009b, n 14, p. 857).
66. *The Tallinn Manual* (n. 9, rule 35, § 12).
67. *Interpretive Guidance* (n. 13, pp. 77-82).
68. See e.g. W. H. Parks (2009, n. 14) *passim*; M. N. Schmitt (2010b, n. 8, pp. 39-43); *contra* see N. Melzer (2009b, n. 14, pp. 895-896 ('While Parks rightly points out that, during the expert discussions, several participating experts were extremely critical of Section IX, he fails to note that just as many experts strongly supported its inclusion in the Interpretive Guidance, and several others even argued that Section IX was not sufficiently restrictive, but should be complemented by human rights standards on the use of force.').
69. *Interpretive Guidance* (n. 13. p. 82, footnote 221).
70. *Ibid.* 77.

Secondly, following Jean Pictet we can say that it is better to capture the civilian than to wound him, and it is better to wound him than to kill him. In the case of civilians directly participating through cyber means, there is another possibility: that is to target their access to cyberspace and to launch cyber attacks. Indeed, without access to cyberspace (computer, network, or even electricity access) this civilian is no longer a danger for the military. I think that this possibility is to be seen as equivalent to the option of capturing them, and is to be understood as better than wounding or killing them. However, this solution can be criticized in that the civilian can find another computer, or if the network is disabled, another way to access it and perpetrate cyber attacks. But this issue can be addressed easily by capturing the civilian at the same time in order to remove the cyber threat they represent.

4. A challenge for the notion of direct participation in hostilities and cyber warfare: the application *ratione loci* of ihl and the spatial limits of armed conflicts

IHL is applicable solely in the context of an armed conflict.[71] As a consequence, IHL is only applicable in the geographical limits of the armed conflict. Mary E. O'Connell perfectly illustrates this in relation to the war in Afghanistan:

> In addition to exchange, intensity, and duration, armed conflicts have a spatial dimension. It is not the case that if there is an armed conflict in one state—for example, Afghanistan—that all the world is at war, or even that Afghanis and Americans are at war with each other all over the planet. Armed conflicts inevitably have a limited and identifiable territorial or spatial dimension because human beings who participate in armed conflict require territory in which to carry out intense, protracted, armed exchanges. International armed conflicts involving sovereign states inevitably implicate the territory controlled by those states,[72]

Civilian direct participation in hostilities is an IHL notion, and so, as a consequence, it is only relevant and applicable in the context of an international or non-international armed conflict. The ICRC's *Interpretative Guidance* mentions the problem of the evolution of means of warfare and the increasing ability to perpetrate attacks far away from the target.

> The requirement of direct causation refers to a degree of *causal* proximity, which should not be confused with the merely indicative elements of *temporal* or *geographic* proximity. For example, it has become quite common for parties to armed conflicts to conduct hostilities through delayed (i.e. temporally remote) weapons-systems, such as mines, booby traps and timer-controlled devices, as well as through remote-controlled (i.e. geographically remote) missiles, unmanned aircraft and computer network attacks.[73]

We can identify three main evolutions in the way that civilians can directly participate in hostilities. The first one, the most simple, is physical participation in the battlefield. Civilians take up weapons and fight the enemy physically.

The second one is the result of the use of vessels and drones by armies. Drones and vessels used for warfare are very expensive and difficult to access for civilians outside of the context of armies, and so those civilians are mostly taking direct part in hostilities from a localization encompassed in the geographical limits of the armed conflict. This situation was perfectly identified by Ryan Goodman and Derek Jinks:

> The concept of DPH has had to bear an especially heavy weight in contemporary armed conflicts. Technological developments have expanded the capacity of individuals to apply lethal force while remaining located thousands of miles away from their targets. States have increasingly relied on private contractors to maximize military power.[74]

The third and last evolution is cyber warfare and cyber means: civilians can take part in hostilities from everywhere in the world with great ease. This situation challenges the geographical limitation of the conflict as civilians can take part in hostilities from outside of the geographical limits of

71. See e.g. D. Fleck (2007, n 60, p. 45, § 201).
72. M. E. O'Connell (2009, p. 858).
73. *Interpretive Guidance* (n. 13, p. 55).
74. R. Goodman and D. Jinks (2009, p. 637).

the armed conflict. As a consequence, those civilians can be located in a place where IHL, and *a fortiori* the notion of direct participation in hostilities, are not applicable.

On the geographical limitations of cyber operations, the *Tallinn Manual* notes:

> As a rule, cyber operations may be conducted from, on, or with effects in the entire territory of the parties to the conflict, international waters or airspace, and, subject to certain limitations, outer space. Cyber operations are generally prohibited elsewhere. Of particular importance in this regard is the law of neutrality because cyber operation can transit neutral territory and may have unintended effect therein.[75]

However, cyberspace offers the possibility for civilians to design and launch cyber attacks against belligerents from the territory of a State not involved in the armed conflict. During an armed conflict, only the cyber operations perpetrated by civilians within the geographical limits of the armed conflict and of the application of IHL, can be regarded as direct participation in hostilities. However, civilians located outside of the geographical limits of the armed conflict can perpetrate cyber operations that reach the three cumulative criteria of DPH identified by the ICRC, namely the threshold of harm, a direct causation and the belligerent nexus. What about those civilians? Are they immune from prosecution for their acts?

Clearly, the answer is no. Civilians taking part in hostilities from outside the geographical limits of the armed conflict cannot be qualified as taking direct part under IHL but they can be prosecuted for their acts. Applying the notion of direct participation to all civilians taking part from outside of the geographical limits of the armed conflict would lead to extrajudicial killing of civilians that must not become the way of addressing this issue.[76]

Targeted States cannot act directly against those persons. These States must ask the State from which the cyber operation originated to prosecute the civilian perpetrator. This situation leads to a practical critique: this procedure can be long, and during this time the civilian can continue their cyber operation. The hosting States may also be uncooperative.

In the previous paragraphs, I have described briefly the legal framework of the spatial scope of the application of IHL and the notion of direct participation in hostilities. I will not go further in this reflection at this point, although it is safe to state that the issue deserves further and broader analysis on how cyber warfare is challenging the *ratione loci* application of international law.

However, I will describe here a few short reflections on this issue. Firstly, nowadays, States are trying in different ways to address the various situations challenging this spatial limitation of armed conflict. The best example seems to be the United States and the global *war on terror*, notably though the use of drones for targeted killing in States with which the United States is not engaged in an armed conflict. Several theories have been developed and used in this way, for example the *unwilling or unable* States or the notion of *transnational armed conflict*; those notions are far from being fully accepted and uncontroversial under international law and by the international community. It will be very interesting to compare those notions to cyber warfare. Nonetheless, cyber warfare differs significantly from existing battlefields. If we analyze the two situations from a practical, and not legal, standpoint, it must be highlighted that targeted killings usually take place in States close to the battlefield, but not parties to the armed conflict, and often those States are qualified as *unable or unwilling* to act against the threat. Conversely, the use of computer technology allows people to take part in hostilities from all around the world, even very far from the targeted States, and in States where it seems more difficult to violate their sovereignty without consequence.

For example, what if someone in France designed a cyber operation against Israel, in support of Hamas or Hezbollah in their conflicts with Israel? Even if the operation perpetrated by this person located in France reached all the criteria of the direct participation in hostilities, it seems impossible to qualify it as direct participation under IHL as the perpetrator is not located within the spatial limits of the armed conflict. The only solution for Israel would be to ask France to act against this person, and if France refused it would be difficult, or even impossible and dangerous, for Israel to use the *unable or unwilling* State theory against this person within French territory.

75. *The Tallinn Manual* (n. 9, rule 21, § 1).
76. R. Goodman and D. Jinks (2009, n. 74, p. 639).

This distinction between those who take part within the geographical limits of the armed conflict and those who act from outside is particularly relevant in cyber warfare. Here we can analyze the conduct of people on the Internet and notably those who claim to be part of the group Anonymous. Regularly, people from all around the world, acting on the behalf of Anonymous, are taking direct part in different armed conflict (e.g. the Israeli-Arab conflict or the conflict between the two Koreas). In light of the previous development, it seems that most of these actions cannot be qualified as direct participation in hostilities under IHL and can be solely addressed under law enforcement procedure.

5. The participation in hostilities of unaware civilians

Internet makes it possible to work remotely, in collaboration with people all around the world. Thus, people involved in a cyber operation can be unaware of the final or real purpose of the operation and of their work. In cyber criminality, it is common to recruit people through the Internet with tempting offers to make money easily, without meeting them face-to-face and without them knowing the real purpose of this job. These people are usually totally unaware that they have become involved in a complex and illegal action that can lead them to prosecution under criminal law.[77] Then this way of involving people can, it might be assumed, be transposed to cyber warfare.

Surfing on the Internet, it is easy to find on blogs or forums people asking for help in order to develop computer codes or even to be involved in cyber operations. But, it is less easy to discover what the real use and consequences of your participation will be, and one could thus easily be unwittingly involved in cyber warfare. Internet is a collaborative world and this situation – giving help on a forum without knowing the real purpose of what we do – is very common.

The ICRC's *Interpretative Guidance* distinguishes the belligerent nexus from the subjective intent of the civilian who takes direct part in hostilities. However, it specifies that:

> Only in exceptional situations could the mental state of civilians call into question the belligerent nexus of their conduct. This scenario could occur, most notably, when civilians are totally unaware of the role they are playing in the conduct of hostilities (e.g. a driver unaware that he is transporting a remote-controlled bomb) [...]. Civilians in such extreme circumstances cannot be regarded as performing an action (i.e. as doing something) in any meaningful sense and, therefore, remain protected against direct attack despite the belligerent nexus of the military operation in which they are being instrumentalized. 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.[78]

In the light of this, the possible situations described above cannot lead to the qualification of civilian direct participation in hostilities. Nonetheless, it seems very difficult to prove the civilian's awareness, or lack of awareness, of the final purpose of the cyber operation in which they are involved. As stated before in this article, if there is doubt as to whether a civilian is taking direct part in hostilities or not, this person shall be considered to be a civilian not taking direct part in hostilities.

Another issue arising from cyber warfare is the question of the use of the computer without the civilian owner knowing it, or against his or her will. Indeed, many cyber operations used one or more botnets, which is a collection of compromised computers named bots. Such compromised computers have usually been infected by a malware that allows someone to use and control them remotely without the knowledge of their owners or users. In this situation, the owners of the compromised computers cannot be seen as taking direct part in hostilities. However, this situation raises many questions on the difficult dissociation between the owner, the user and the computer and their legal qualifications.

In sum, the notion of direct participation is challenged by cyber warfare but remains applicable. The practice will need to find how to address and fix the specific issues concerning civilian direct participation in hostilities in relation to cyber warfare.

77. See notably the TV documentary D. Herbst, *In Den Fängen Der Internet-Mafia* (Arte 2013).
78. *Interpretive Guidance* (n 13) 60.

6. Conclusion

This article demonstrates that IHL applies and is sufficient in most cases of civilian direct participation in cyber hostilities. It proves, consequently, that the assertion according to which cyber warfare is not controlled by international law is wrong. In some specific cases, however, there is a need for a new interpretation or creation of IHL rules as demonstrated in this article.

By analyzing together the divergent approaches of the ICRC's *Interpretative Guidance* and the *Tallinn Manual*, this article highlights the diversity of possible approaches.

Bibliography

BOOTHBY, B. (2009). "And For Such Time As: The Time Dimension to Direct Participation in Hostilities". *New York University Journal of International Law and Politics*, n. 43, p. 741.

DINNISS, H. H. (2012). *Cyber Warfare and the Laws of War.* Cambridge Studies in International and Comparative Law. Cambridge University Press.

DOSWALD-BECK, L.; HENCKAERTS, J.M. (2004). *Customary International Humanitarian Law: Volume 1, Rules.*

FLECK, D. (2007). *The Handbook of International Humanitarian Law.* Oxford University Press.

GOODMAN, R.; JINKS, D. (2009). "ICRC Interpretive Guidance on the Notion of Direct Participation in Hostilities under International Humanitarian Law: An Introduction to the Forum".*New York University Journal of International Law and Politics*, no. 42, p. 637.

HERBST, D. (2013). *In Den Fängen Der Internet-Mafia* (Arte).

KERSCHISCHNIG, G. (2012). *Cyberthreats and International Law*. Eleven International Publishing 2012.

MELZER, N. (2009a). *Interpretive Guidance on the Notion of Direct Participation in Hostilities under International Humanitarian Law.* International Committee of the Red Cross.

MELZER, N. (2009b). "Keeping the Balance Between Military Necessity and Humanity: A Response to Four Critiques of the ICRC's Interpretive Guidance on the Notion of Direct Participation in Hostilities." *New York University Journal of International Law and Politics*, n. 42, p. 831.

MELZER, N. (2010). "Civilian Participation in Armed Conflict". *Max Planck Encyclopedia of Public International Law*.

O'CONNELL, M. E. (2009). "Combatants and the Combat Zone". *University of Richmond Law Review,* n. 43. p. 845.

PARKS, W. H. (2009). "Part IX of the ICRC Direct Participation in Hostilities Study: No Mandate, No Expertise, and Legally Incorrect". *New York University Journal of International Law and Politics*, no. 43, p. 769.

PRESCOTT, J. M. (2012). "Direct Participation in Cyber Hostilities: Terms of Reference for Like-Minded States?". *4th International Conference on Cyber Conflict (CYCON)*.

SCHMITT, C. (2007). *Theory of the Partisan: Intermediate Commentary on the Concept of the Political* (first published 1963, translated by G. L. Ulmen). Telos Press Publishing.

SCHMITT, M. N. (2004). "Direct Participation in Hostilities" and 21st Century Armed Conflict". *Crisis Management and Humanitarian Protection: Festschrift fur Dieter Fleck*.

SCHMITT, M. N. (2010a). "Deconstructing Direct Participation in Hostilities: The Constitutive Elements" *New York University Journal of International Law and Politics,* no. 42, p. 697.

SCHMITT, M. N. (2010b). "The Interpretive Guidance on the Notion of Direct Participation in Hostilities: A Critical Analysis". *Harvard National Security Journal*, no. 1, p. 5.

SCHMITT, M. N. (2011). "Cyber Operations and the Jus in Bello: Key Issues" *International Law Studies,* no. 87, p. 89.

SCHMITT, M. N. (2013) (ed.) *The Tallinn Manual on the International Law Applicable to Cyber Warfare* Cambridge University Press.

WATKIN, K. (2009). "Opportunity Lost: Organized Armed Groups and the ICRC Direct Participation in Hostilities Interpretive Guidance". New York University Journal of International Law and Politics, no. 42, p. 641.

WATTS, S. (2010). "Combatant Status and Computer Network Attacks". *Virginia Journal of International Law*, 50, p. 391.

Recommended citation

DELERUE, François (2014). "Civilian Direct Participation in Cyber Hostilities". *IDP. Revista de Internet, Derecho y Política*. No. 19, pp. 3-17. UOC. [Accessed: dd/mm/yy].
‹http://journals.uoc.edu/index.php/idp/article/view/n19-delerue/n19-delerue-en›
‹http://dx.doi.org/10.7238/idp.v0i19.2425›

About the author

François Delerue
francois.delerue@eui.eu
Ph.D. researcher in International Law
at the European University Institute (Florence)

François Delerue is a PhD candidate in international law at the European University Institute (EUI). He is studying international law and cyber warfare. He is a board member and the head of the international law section of the European Journal of Legal Studies (EJLS), member of the doctoral seminar of the Castex Chair of Cyberstrategy (Paris, France) and laureate of the French Institute of Higher National Defence Studies (IHEDN – Paris, France). He graduated with an LL.M. in Comparative European and International Law from the EUI in 2013 and a Research Master's in International Law and International Organizations from the Sorbonne Law School (Paris Panthéon-Sorbonne University) in 2011.

European University Institute
Badia Fiesolana
Via dei Roccettini 9,
I-50014 San Domenico di Fiesole (FI
Italy

UOC
Universitat Oberta de Catalunya