2017 9th International Conference on Cyber Conflict
Defending the Core
H. Rõigas, R. Jakschis, L. Lindström, T. Minárik (Eds.)

The Misuse of Protected Indicators in Cyberspace: Defending a Core Aspect of International Humanitarian Law

Jeffrey Biller, Lt Col, USAF
Stockton Center for the Study of International Law
U.S. Naval War College
Newport, RI, USA
jeffrey.biller@usnwc.edu

Abstract: International humanitarian law (IHL) imposes a complex array of laws regarding the use of markings, signals, symbols and other indicators. Protections related to indicators are also directly implicated in the laws of perfidy and ruses. Although these laws are generally well accepted in principle, practitioners struggle to apply these rules in the newer, man-made domain of cyberspace. Despite recent steps forward in the application of IHL to cyber, questions surrounding enemy, neutral, and protected indicators remain largely unresolved. This paper seeks to answer these thorniest of issues related to military cyber operations during international armed conflicts.

The article is divided into two sections. The first addresses protected and specially recognized indicators, particularly those of the UN and the Geneva Conventions. The IHL rules regarding these symbols are defined and applied in the context of cyber operations. This section also discusses perfidy and proximate causation in the cyber context. The second turns to the improper use of national indicators in cyberspace, particularly the definition of military emblems, which draws on a separate body of law than protected or specially recognized emblems. Although the misuse of indicators may also implicate international criminal law, this article focuses exclusively on IHL applicability.

Keywords: *cyberspace, markings, indicators, emblems*

1. INTRODUCTION

A core principle of international humanitarian law (IHL) is the protection of civilians and civilian objects.[1] This protection includes aid organizations such as the International Committee of the Red Cross (ICRC) and observer organizations such as the United Nations (UN). These groups, in addition to neutral states not party to the conflict, are distinguished through the use of various indicators[2] governed by an extensive body of law dating back to ancient times.[3] IHL divides these indicators into two primary categories: first, the protected and recognized indicators of the Geneva Conventions (GC), the UN, the white flag of truce, and other internationally recognized protective emblems, signs or signals;[4] and second, indicators of nationality, such as the uniforms or military equipment markings of neutral or adversary nations.[5] Related to the use or misuse of indicators under IHL is the prohibition on perfidy, which outlaws killing or injuring by resort to acts inviting the confidence of an adversary leading to a reliance on protection under the rules of IHL with the intent to betray that confidence.[6]

The basic notion of extending the body of IHL regarding these indicators into cyberspace is uncontroversial.[7] However, a full agreement does not yet exist as to what constitutes recognized indicators in the cyber domain and how to realize the protections signaled by these indicators.[8] This article examines the treaty and customary rules related to the use of indicators and then applies those rules to many of the network characteristics currently used to identify entities in cyberspace. Although some characteristics meet the definitions of relevant indicators under IHL, this article identifies gaps where markings indicating a trusted party could be used to conduct offensive cyber operations.

To that end, the article is divided into two sections. The first addresses protected and specially recognized indicators, particularly those of the UN and the Geneva Conventions (GC). The IHL rules regarding these symbols is defined and applied in the context of cyber operations. This section also discusses perfidy and proximate causation in the cyber context. The second section turns to the improper use of national indicators in cyberspace, particularly the definition of military emblems, which draws on a separate body of law than protected or specially recognized emblems. Although the misuse of indicators may also involve international criminal law, this article focuses exclusively on IHL applicability.

1 Protocol Additional to the Geneva Conventions of 12 August 1949, and Relating to the Protection of Victims of International Armed Conflicts art. 48, June 8, 1977, 1125 UNTS 3 (hereinafter AP I).

2 The word "indicators" will be used throughout this paper to generally encompass the set of uniforms, emblems, flags, etc. that are used to indicate nationality, status, special protections, or particular categories.

3 Commentary on the Additional Protocols of 8 June 1977 to the Geneva Conventions of 12 August 1949, ¶ 1526 (Yves Sandoz, Christophe Swinarski & Bruno Zimmermann eds., 1987) (hereinafter AP I Commentary).

4 Convention (I) for the Amelioration of the Condition of the Wounded and Sick in the Armed Forces in the Field art. 44, 53, Aug. 12, 1949, 6 UST 3114, 75 UNTS 31 (hereinafter GC I); AP I, art. 38; Convention for the Protection of Cultural Property in the Event of Armed Conflict, May 14, 1954, 249 UNTS 240 see also 1 Customary International Humanitarian Law, r. 58-61 (Jean-Marie Henckaerts & Louise Doswald-Beck eds., 2005) (hereinafter CIHL Study).

5 AP I, art. 39; CIHL Study, r. 62-63.

6 CIHL Study, r. 65.

7 See *Tallinn Manual on the International Law Applicable to Cyber Warfare* r. 60-65 (Michael N. Schmitt ed., 2013) (hereinafter *Tallinn Manual*).

8 Id.

2. PROTECTED AND RECOGNIZED INDICATORS

A. The Law Regarding Improper Use of Protected and Recognized Indicators

The long-standing IHL rules against the improper use of protected and recognized indicators such as the emblems of the GCs and the UN are well-established.[9] This law developed as recognition of the need to protect certain classes of individuals, organizations, and locations on the battlefield from targeting by combatants. As such, the law focuses primarily on these emblems' use as concrete, visible representations.[10] Although it is unlikely that the use of protected indicators in a purely electronic environment was initially envisaged, the language within the relevant articles is broad enough to encompass its extension into the cyber domain.

The First Geneva Convention (GC I) defines the emblem of the Red Cross and delineates its permissible use.[11] Specifically, GC I states that the emblem, and the words "Red Cross" or "Geneva Cross,"[12] "may not be employed either in time of peace or in time of war, except to indicate or to protect the medical units and establishments, the personnel and material protected by the present Convention and other Conventions dealing with similar matters."[13] Similarly, Article 38 of Additional Protocol I (AP I) prohibits the "improper use of the distinctive emblem of the red cross, red crescent or red lion and sun or of other emblems, signs or signals provided for by the Conventions or by this Protocol" and also "to make use of the distinctive emblem of the United Nations, except as authorized by that Organization."

The 2016 Commentary to GC I (GC I 2016 commentary) notes that the GC emblems may serve both as a protective device indicating protection under the Convention and as an indicative sign demonstrating a connection to the organization of the International Red Cross and Red Crescent.[14] Although the indicative use does not imply that the bearer holds protections under the Convention, its improper use is still prohibited.[15] AP I does not address the indicative use, focusing on the protective use,[16] which provides "a visible sign of the protection conferred by international law on certain persons and objects."[17] Unlike misuse of the emblem as an indicative sign, a misuse of the protective function could implicate the prohibition on perfidy.[18]

GC I Article 53 further expands the law relating to the GC emblems, prohibiting their use "by individuals, societies, firms or companies either public or private, other than those entitled

9 See *CIHL Study*, r. 59-60; GC I, art. 44, 53. AP I, art. 38; Regulations Respecting the Laws and Customs of War on Land, annexed to Convention No. IV Respecting the Laws and Customs of War on Land art. 23(f), Oct. 18, 1907, 36 Stat. 2227, TS. No. 539 (hereinafter Hague Regulations).

10 Commentary to Geneva Convention I for the Amelioration of the Condition of the Wounded and Sick in the Armed Forces in the Field 325-330 (Jean Pictet ed., 1952) (hereinafter GC I 1952 Commentary).

11 GC I, art. 38, 44, 53.

12 Hereinafter, only the words "Red Cross" will be referenced, but both phrases are implicated.

13 GC I, art. 44.

14 International Committee for the Red Cross, Commentary to Geneva Convention I for the Amelioration of the Condition of the Wounded and Sick in the Armed Forces in the Field ¶ 2661 (2d ed. 2016), art. 44 (https://ihl-databases.icrc.org/applic/ihl/ihl.nsf/Comment.xsp?action=openDocument&documentId=EC42EEC3274A7323C1257F7A0056F100) (hereinafter GC I 2016 Commentary).

15 Id.

16 AP I Commentary, ¶ 1540.

17 International Committee for the Red Cross, Regulations on the Use of the Emblem of the Red Cross or the Red Crescent by the National Societies art. 1 (1991) (https://www.icrc.org/eng/resources/documents/article/other/57jmbg.htm).

18 AP I Commentary, ¶ 1532.

thereto under the present Convention, of the emblem or the designation "Red Cross" or "Geneva Cross," or any sign or designation constituting an imitation thereof, whatever the object of such use."[19] By including "imitations thereof," Article 53 broadens the prohibition and suggests that abbreviations or approximations of the words "Red Cross" that are meant to imitate an official representation would violate this prohibition.[20] Although many practitioners view Article 53 as primarily applying to peacetime misuse of the emblem in its protective sense, the GC I 2016 commentary states that Article 53 applies "both in situations of armed conflict and in times of peace."[21] Additionally, the GC I 2016 commentary notes that Article 53 "encompasses both misuse of the emblem in its protective sense and when used as an indicative sign."[22]

AP I also prohibits the unauthorized use of the distinctive emblem of the UN.[23] However, the treaty law governing the UN emblem is less expansive than that of the Red Cross emblems, primarily in the scope of its definition. Unlike the words "Red Cross" under GC I, IHL protects neither the words "United Nations" nor approximations thereof.[24] Therefore any application of law regarding the GC and UN emblems in the cyber context must start with an understanding that protections against GC emblems will have broader applicability.

Additional categories of protected emblems, signs, and signals established under international law include the Hague IV and AP I prohibition against the "improper use of a flag of truce"[25] and the AP I prohibition against deliberate misuse in an armed conflict of "other internationally recognized protective emblems, signs or signals."[26] Recognized protected indicators include those markings that indicate objects or locations such as installations containing dangerous forces, cultural property, POW or civilian internee camps, civil defense, and hospital, safety, or neutralized zones.[27]

Unlike the prohibition on perfidy (see below), there is an absolute character to these prohibitions, meaning that there is no requirement for a particular result following the prohibited misuse.[28] Examples of the improper use of protected emblems, signs, and signals include their use by other than intended personnel while engaging in attacks, to favor one's own military operations, or to impede enemy military operations.[29] These examples encompass almost the entire potential range of military operations, as any relevant action undertaken by a military is likely to either favor their own, or disfavor the enemy's, operations.

One section of law that would initially appear relevant to cyber operations is the rules contained in Annex I to AP I governing "distinctive signals." However, the signals referenced in Annex I are very specific types of radio communication and light signals.[30] Although some

19 GC I, art. 53.
20 GC I 2016 Commentary, ¶ 3065, art. 53 (https://ihl-databases.icrc.org/applic/ihl/ihl.nsf/Comment.xsp?action=openDocument&documentId=57F199148260B5AFC1257F7A00579E9B).
21 GC I 2016 Commentary, ¶ 3066, art. 53.
22 Id.
23 AP I, art. 38(2).
24 Id.
25 Hague Regulations, art. 23(f).
26 AP I, art. 38.
27 AP I, art. 59-60, 66; Office of the General Counsel, U.S. Department of Defense, *Law of War Manual* (hereinafter *DoD Law of War Manual*), para 5.24; CIHL Study, r. 61.
28 AP I Commentary, ¶ 1532.
29 *DoD Law of War Manual* § 7.15.4 (2016).
30 AP I, Annex I, art. 6-14.

conceivable use in cyber operations is plausible, there are no generally applicable rules to cyber operations that flow from them and therefore will not be discussed. With an understanding of the prohibition related to protected and recognized indicators in place, the following section analyzes its application in cyberspace.

B. Improper Use of Protected and Recognized Indicators in Cyberspace

Extension of the basic rule prohibiting "mak[ing] improper use of the protective emblems, signs, or signals that are set forth in the law of armed conflict" into the cyber domain is uncontroversial.[31] However, protected indicators signal the ability to trust, and trust plays a prominent role in network security systems, which depends on forming trust relationships between parties before allowing access and sharing information. Masquerading as a party known to be trusted by a target system is a frequently used method of defeating network security. Therefore, rules related to the cyber indicators of trusted parties, such as the ICRC and UN, require detailed understanding. This section examines particular cyber methods that involve violations of the trust relationship, including the variations on phishing, Internet Protocol (IP) spoofing, and domain name spoofing, as a way of contextualizing and exploring these rules.

First is the use of phishing, a type of social engineering, to manipulate authorized system users into providing information and thus allowing unauthorized system access.[32] This manipulation occurs in the purely cyber context through the use of e-mail, e-messaging, or online communications. The Tallinn International Group of Experts (IGE) addressed this situation, citing the example of an adversary sending an email with the "bare assertion that the sender is a delegate of the International Committee of the Red Cross."[33] The IGE found no misuse in this example, despite the use of the words "Red Cross." Although GC I Article 44 specifically protects these words from unauthorized use, the presumed argument is that the operator's use of the words "Red Cross" is not formal enough to be considered as an emblematic identifier. However, if the words were employed in a more formal manner, such as an email signature block, letterhead to an attachment, or another manner formally indicating an official Red Cross document, there is a much stronger argument that the use violates the GC I Article 44 prohibition on use of the words "except to indicate or to protect the medical units and establishments, the personnel and material protected by the present Convention and other Conventions dealing with similar matters."[34]

The difference between the simple statement as to ICRC affiliation and a signature block is the formality involved. Signature blocks are typically used in official correspondence and serve to indicate the sender's representative status of the named organization, whereas simple statements, while misleading, lack that level of formality. As a Red Cross signature block "is used to show that a person or object has a connection with one of the organizations of the Movement, without implying protection under the Geneva Conventions or any intent to invoke them," this would be considered an indicative use, as opposed to protective use, of the words "Red Cross." Nonetheless, it constitutes a violation of GC I Article 44 and the customary IHL related to the use of the distinctive GC emblems.[35]

31 *Tallinn Manual*, r. 62.
32 Michael Gregg, *Certified Ethical Hacker Guide* 513 (2014) (hereinafter *CEH Guide*).
33 *Tallinn Manual*, cmt. accompanying r. 62, ¶ 4.
34 GC I, art. 44.
35 CIHL Study, r. 59.

The second type of operation is a related type of phishing campaign, but with the aim of tricking the target operator into taking cyber-based self-defeating actions. This method uses spoofed emails, social media messaging, and websites to induce the target into either downloading malicious attachments or following web links to malicious websites.[36] Like other types of social engineering, these attacks rely on the target operator trusting the e-mail, website or attachment such that they will take the desired action. Protected emblems could easily be implanted into the e-mail, message or website to induce trust in the target. For example, the use of the UN emblem as a watermark in a downloadable document or the Red Cross emblem as a social media avatar could induce a target to follow a link to a site containing malware. As the actual protected emblem is clearly used in an unauthorized manner, this is a clear IHL violation. The Tallinn IGE came to the same conclusion on this question.[37]

A third method illustrating misuse of emblems is IP spoofing. Here, cyber operators attempt to gain unauthorized system access by creating a malicious message that appears to originate from a trusted machine, imitating its IP address.[38] For example, spoofing an IP address associated with the ICRC to defeat a firewall that relies on IP addresses for filtering.[39] The primary question is whether IP addresses should be viewed as a legal indicator of a protected organization. This appears logical, given the widespread use of IP addresses as a trust indicator by cyber operators. For example, a defensive operator may specifically program a firewall to permit connections from ICRC or UN IP addresses during an armed conflict. These connections may allow communications regarding the treatment of wounded or prisoners or war. If an adversary were to spoof these IP addresses, the network operator may be forced to block communications from these previously trusted sources.

Permitting a party to a conflict to represent a communication as coming from the ICRC or UN appears to run counter to the intent of IHL. Article 1 of AP I states it is a general principle that in cases not covered by AP I or other international agreements, customary law, "principles of humanity," and "dictates of public conscience" apply. Additionally, Article 31(1) of the Vienna Convention on the Law of Treaties (VCLT) states that a treaty should be interpreted partly "in the light of its object and purpose."[40] However, the VCLT also states that treaties should be interpreted "in accordance with the ordinary meaning to be given to the terms of the treaty in their context." Provisions governing use of the emblems suggest an element of general awareness or recognition of the emblem as such.[41] Thus, it is unlikely that a spoofed ICRC IP address could be considered an imitation of the emblem under the Article 53 standard given the lack of general awareness as to what the sequence of numbers in an IP address specifically indicates. The ordinary person is unlikely to mistake an IP address for an imitation of the words "Red Cross," even if a cyber operator may understand the connection. Despite their importance in identifying organizations in cyberspace, IP addresses do not meet any definition of the GC emblems. Given that the provision governing use of the UN emblem is less broad than that of the GC emblems, it is also doubtful UN-associated IP addresses would be considered as a protected indicator.

36 Ed Skoudis & Tom Liston, *Counter Hack Reloaded* 566 (2006).
37 *Tallinn Manual*, cmt. accompanying r. 63, ¶ 2.
38 Skoudis, at 470.
39 Id.
40 Vienna Convention on the Law of Treaties art. 31(1), May 23, 1969, 1155 UNTS 331.
41 The United Nations Flag Code and Regulations, ST/SGB/132, United Nations, January 1967; GC I 2016 Commentary, ¶ 2543, art. 38 (https://ihl-databases.icrc.org/applic/ihl/ihl.nsf/Comment.xsp?action=openDocument&documentId=8BF732335A87E6DCC1257F15004A27A5).

The fourth method for analysis involves the spoofing of e-mail addresses or domain names. By spoofing an email address, such as @ICRC.org, in the recipient's "From" field, the operator hopes to induce either the target system to allow the email through a firewall, or a target individual to trust the contents of the email. Once this trust is established, the operator may then use this connection to conduct the next phase of a cyber operation. Similarly, a Domain Name System hijacking operation may send an unwitting target who accesses the ICRC.org or UN.org websites to a spoofed website containing malicious links or false information.[42]

Here, the focus is on domain names which serve to provide users with a recognizable identity to resources found on the Internet. Although related to IP addresses, domain names differ in that they often contain an organization's name or abbreviation, as opposed to the numerical designator of an IP address. The narrower protection for the UN emblem, which does not include the name United Nations or approximations, eliminates its applicability from this analysis. The relevant question as to the Red Cross is whether a spoofed email address or domain name containing the words Red Cross, the acronym ICRC, or similar abbreviation, would constitute "an imitation thereof."[43] The Tallinn IGE "struggled with the issue," and laid out two potential approaches.[44]

The first approach argued that email address and domain names are not protected indicators because they do not constitute "electronic reproductions of the relevant graphic emblems."[45] This approach overlooks the prohibitions in Articles 44 and 53 on the unauthorized use of the words "Red Cross" or "an imitation thereof" when they function as an indicative or protective emblem. The 1952 Commentary to GC I expresses this concern well:

> It was obviously not enough merely to prohibit misuse of the red cross emblem. Protection had also to be extended to the words which form the official title of the great humanitarian institution known as the Red Cross. These words are as familiar to the public as the emblem, and must enjoy the same prestige.

The second approach found the key factor to be the "use of an indicator upon which others would reasonably rely in extending protection provided for under the law of armed conflict"[46] Thus, the imitation of the ICRC.org domain name or email address would be an unauthorized use because, as the IGE states, it "invites confidence as to the affiliation of the originator."[47] Although it does not reference Article 53, this view would be consistent with that article's inclusion of any "sign or designation constituting an imitation thereof." Given the ubiquitous use of the acronym ICRC, it would be hard to argue that it does not constitute an "imitation thereof." Therefore, the second approach of the IGE appears to be a more accurate reflection of IHL.

The various methods of phishing and spoofing are not the only types of cyber operations that implicate the rules against misuse of protected emblems. However, they highlight the most likely ways in which protected indicators may be used in a remote access cyber operation. They

42 Skoudis, at 220-221.
43 GC I, art. 53.
44 Tallinn Manual, cmt. accompanying r. 62, ¶ 6.
45 Id.
46 Id., cmt. accompanying r. 62, ¶ 7.
47 Id.

also serve to help identify which cyber indicators could constitute protected indicators and reveal gaps where adversaries could take advantage of the trusted nature of organizations such as the ICRC and UN to conduct offensive cyber operations.

C. Perfidy and Proximate Causation in Cyberspace

A discussion of the use of protected emblems naturally raises the issue of perfidy. Although the improper use of protected indicators is itself an IHL violation, it may also constitute an element of perfidy. Several questions surround the application of the rule against perfidy in the cyber context. However, this article focuses specifically on proximate causation between the perfidious act and the injury or killing of the adversary. Proximate causation is an act that directly produces an effect and without which the effect would not have occurred. Although there is no requirement for proximate causation written into the treaty law on perfidy, both the Tallinn Manual and commentators have extrapolated such a requirement from the definition of perfidy.[48]

Perfidy is well-defined in treaty and customary law. AP I Article 37(1) defines perfidy as:

> […] acts inviting the confidence of an adversary to lead him to believe that he is entitled to, or obliged to accord, protection under the rules of international law applicable in armed conflict, with intent to betray that confidence.

Perfidious acts that result in the killing or wounding of the enemy constitute a war crime.[49] Included in the ICRC CIL study's list of acts that constitute perfidy is the "simulation of protected status" through the use of the Red Cross, United Nations emblems, or other protected emblems.[50] Simulation of status can be accomplished in any manner intending to inspire confidence, and is therefore distinct from the "improper use" standard discussed previously. Despite this distinction, as with the improper use of protected emblems, the concept of violating trust makes perfidy an important concern in cyberspace.

The proximate causation requirement appears to come from the language of AP I Article 37(1), stating that it is "prohibited to kill, injure or capture an adversary by resort to perfidy." This language suggests a direct link between the perfidious act and the end result, and is reflected in the *Tallinn Manual* commentary.[51] The proximate cause example in the *Tallinn Manual* focuses on proximate causation after the kinetic effect produced by the perfidious cyber action.[52] However, the cyber operation should also be evaluated to determine whether the perfidious act was the proximate cause of the kinetic effect. If there is a distinct, additional phase to a cyber operation required between the perfidious act and the kinetic effect, then there is no proximate causation between the perfidious act and resulting death or injury.

Cyber operations are rarely single-step operations resulting in a malicious cyber effect. Rather, they are multi-stage efforts, with most stages resulting in no discernible effect on the target. Initial stages may be limited to developing an understanding of, and then gaining access to,

48 *Tallinn Manual*, cmt. accompanying r. 60, ¶ 5; Michael Bothe, Karl Josef Partsch & Waldemar A. Solf, New Rules for Victims of Armed Conflicts, Commentary on the Two 1977 Protocols Additional to the Geneva Conventions of 1949, at 244 (1982).
49 CIHL Study, r. 65.
50 Id.
51 *Tallinn Manual*, cmt. accompanying r. 60, ¶ 5.
52 Id.

a target network.[53] The perfidiously-gained access may be used for espionage purposes, for a sub-use of force non-cyber effect, or for no action whatsoever. Often a cyber operator has no other goal than to gain access to a system, with the use of that access to be determined at a later time. If the perfidious act, such as the use of a protected emblem, takes place during a phase of the cyber operation distinct from the action causing the non-cyber effect, then the perfidious act is unlikely to be the proximate cause of the death or injury. The amount of time between the perfidious act and the resulting death or injury is not determinative of proximate causation, though it may be an element in the analysis. Rather, the key question in determining proximate causation is whether the perfidious act directly produced the injury or death. If an intervening step is required to produce the injury or death, the misuse would not constitute prohibited perfidy. A similar category of indicators subject to IHL that could potentially implicate perfidy are enemy and neutral indicators, which will be discussed below in detail.

3. IMPROPER USE OF ENEMY AND NEUTRAL INDICATORS

A. The Law Governing Use of Enemy and Neutral Indicators

Indicators of national military status (uniforms, flags, and other military emblems) do not inherently indicate a protective status under international law. However, rules regarding the use these indicators are among the earliest found in warfare.[54] In current practice, Hague Regulation 23(f) and Article 39 of AP I are the primary treaty laws governing the use of these indicators. The Hague regulation prohibits the "improper use of a flag of truce, of the national flag or of the military insignia and uniform of the enemy." AP I Article 39(2) defines "improper use" as uses of adverse party indicators "while engaging in attacks or in order to shield, favour, protect or impede military operations." The commentary to AP I notes that this "includes the preparatory stage to the attack."[55]

In addition to treaty law, there is a variety of state opinion regarding enemy indicators. Many states consider the Article 39(2) rule to only apply in instances of attack, with varying reservations to the broader inclusion of those uses "in order to shield, favour, protect or impede military operations."[56] For example, Canada made such a reservation upon ratification of AP I[57], and the US believes that the prohibition extends to combat, but use outside combat would not be improper.[58] The ICRC CIL Study states that "it cannot be concluded, therefore, that the wearing of enemy uniforms outside combat would be improper." It should also be noted that the question as to whether use of enemy uniforms would constitute perfidy is unsettled, as no specific IHL protections attach to enemy uniforms.[59] However, the ICRC CIL Study notes that the wearing of such uniforms may invite the confidence of the enemy.[60]

Article 39(1) of AP I also covers the use of neutral party indicators, prohibiting the "use in an armed conflict of the flags or military emblems, insignia or uniforms of neutral or other States

53 CEH guide, at 42.
54 *See* AP I Commentary, ¶ 1526; and CIHL Study, r. 62.
55 AP I Commentary, ¶ 1586.
56 For discussion on state interpretations of the rule, *see* CIHL Study, r. 62.
57 Canada, Reservations and statements of understanding made upon ratification of Additional Protocol I.
58 *DoD Law of War Manual*, § 5.32.1.
59 Id. at § 5.32.1.1.
60 CIHL Study, r. 62.

not Parties to the conflict."[61] The AP I standard only requires connection to an armed conflict. According to the commentary, the prohibition includes the use for espionage purposes, "as this constitutes an intervention of a military nature."[62] The ICRC CIL Study found no contrary state practice or claims to the AP I Article 39(1) rule regarding neutral indicators.[63] The use of neutral military emblems may also be considered perfidious.[64]

Although these prohibitions are relatively well-defined, they do leave open the question of what constitutes an enemy or neutral indicator. Unlike the well-defined emblems of the Red Cross and the UN, the "military emblem" referenced in AP I is undefined in the text of the treaty, although the commentary does provide an in-depth discussion of the matter. There, emblems of nationality are described as "essentially customary in nature," meaning those uses in international society that "constitute a generally recognized language which is accorded the same respect as the spoken or written word in relations between individuals."[65] The commentary also notes that military emblems are "visible signs."[66] Though the description of military emblems is somewhat vague, it is clearly much broader than the definitions given the distinctive emblem of the Red Cross or of the UN. This broad definition is the key to determining the applicability of the law regarding enemy and neutral indicators in cyberspace.

B. Improper Use of Enemy Indicators in Cyberspace

The *Tallinn Manual* rule on enemy indicators holds that "[i]t is prohibited to make use of the flags, military emblems, insignia, or uniforms of the enemy while visible to the enemy during an attack, including a cyber attack."[67] This rule adds the AP I commentary requirement of visibility and drops the phrase "or in order to shield, favour, protect or impede military operations." In the cyber context, the meaning of the visibility requirement is key to determining the rule's applicability.

The Tallinn IGE postulated that "it is unlikely that improper use of the enemy uniforms and other indicators will occur during a remote-access cyber attack, as the cyber operators would not be in visual contact with the adversary."[68] However, under treaty provisions and the AP I commentary explanation of military emblems, there is no requirement that the attacker is physically visible to the adversary for its use to be improper, only that the military emblem is visible.[69] The minority position of the IGE pointed to the fact that such a provision "appears in neither Article 39(2) of Additional Protocol I, nor in the ICRC Customary IHL Study's discussion of that Article."[70]

The majority opinion is supported by historical precedent, and to some extent common sense, where concrete, visible signs attached to military uniforms and equipment allowed distinction between parties to the conflict.[71] However, IHL does not require that an attacker, much less

61 AP I, art. 39(1).
62 AP I Commentary, ¶ 1569.
63 CIHL Study, r. 63.
64 Id., r. 65.
65 AP I Commentary, ¶ 1562.
66 Id. at ¶ 1578.
67 *Tallinn Manual*, r. 64.
68 Id., cmt. accompanying r. 64, ¶ 3.
69 AP I, art. 39(2); AP I Commentary, ¶ 1562-1587.
70 *Tallinn Manual*, cmt. accompanying r. 64, ¶ 4.
71 Bothe et al., at 214.

the attacker's uniform or other military indicator, be visible to the intended target. Although cyber operations were not considered when the rule emerged, the development of other types of beyond visual range weaponry, such as artillery, certainly existed. Additionally, the AP I commentary notes that in modern warfare, military equipment is:

> […] supplied in large numbers throughout the world by a few manufacturers. They are all the same model, and very often it is only the emblems of nationality which unequivocally identify to which side they belong.[72]

The same rationale can be applied to the Internet and other networked information systems. National domains and systems are often uniquely identified at both the human and system interface levels.

If military emblems are to be extended to cyberspace, the next issue is what qualifies as a military emblem in cyberspace under IHL. The Tallinn IGE did raise the question of enemy marked computer hardware, finding the "rule has no application with regard to enemy marked computer hardware over which control has been remotely acquired and that is used for conducting attacks against the enemy." However, the Tallinn IGE did not extend the discussion to visible enemy network identifiers such as domain names.

Given the requirements for military emblems to be visible and identifiable in a "generally recognized language," cyber identifiers with no visual component, such as meta-data, and those that may be visible but are not generally recognized, such as an IP address, should not be considered as military emblems. However, commonly understood identifiers that appear at the human interface level, such as web and e-mail domain names, graphical symbols, and formal representations such as signature blocks in e-mails and electronic documents appear to qualify as military emblems. These identifiers provide official representation of a particular organization or individual. In this manner, they act just as a military emblem, designed to provide official notification of a combatant's status as a member of a particular party to the conflict. For example, there is a general trust that a domain name of www.af.mil is operated by, or at least under the control of, the United States Air Force. Similarly, an electronic letter bearing the official seal of the United States Navy and the signature block of the Chief of Naval Operations would indicate that the sender was a member of the United States Navy. Thus, if used in a cyber operation as part of an overall military attack campaign, they initially appear to constitute an improper use of an enemy military emblem. Prior to making this determination, however, the practitioner must analyze the issue of state practice concerning transmission of bogus signals, dispatches, and other communications considered as ruses of war.[74]

Ruses of war are defined in AP I as:

> […] acts which are intended to mislead an adversary or to induce him to act recklessly but which infringe no rule of international law applicable in an armed conflict and which are not perfidious because they do not invite the confidence of an adversary with respect to protection under that law.[75]

72 AP I Commentary, ¶ 1572.
73 *Tallinn Manual*, cmt. accompanying r. 64, ¶ 6.
74 Id., cmt. accompanying r. 64, ¶ 5.
75 AP I, art. 37(2).

For example, the UK *Manual* includes in its list of permissible ruses:

> […] transmitting bogus signal messages and sending bogus dispatches and newspapers with a view to their being intercepted by the enemy; making use of the enemy's signals, passwords, radio code signs, and words of command.[76]

These types of ruses find application in the cyber context, including most spoofing of enemy e-mail addresses, social messaging identifiers, and graphical symbols included in attachments, as these uses of military emblems are communicative in nature.[77]

Although most uses of enemy cyber identifiers would find function in permissible ruses, some cyber indicators that meet the definition of military emblems have the potential to fall outside this realm. Take, for example, the spoofing of a military web domain identified by a military domain address, such as "www.navy.mil." Domain names serve to identify networks, or portions thereof, as belonging to particular organizations. This is distinct from a communicative purpose such as relaying false information. Any example of an attack using a domain name to conduct an attack might be a weaponized honeypot using an enemy's spoofed domain name against their own forces, as that domain name represents a military emblem that falls outside the category of acceptable ruse. Here, it is not the communicative nature of the domain name that is used to conduct an attack, but rather the enemy's reliance on it as an indicator of their own forces.

Types and functions of cyber identifiers are likely to evolve at a rapid pace and any attempt to create an exclusive list of identifiers held to be military emblems would have limited utility. Fortunately, the definition described in the AP I commentary, describing emblems as "customary nature" allows the understanding of military emblems in cyber to evolve in CIHL as the technology changes. However, this understanding requires close coordination between legal practitioners and technical experts to determine what indicators should be considered as military emblems for IHL purposes. The definition of a military emblem also plays a key role in law regarding improper use of neutral indicators, to be discussed in the following section.

C. Improper Use of Neutral Indicators in Cyberspace

Similar to protected emblems, the prohibition on the improper use of neutral indicators is absolute "in an armed conflict."[78] Thus they cannot be used "for the promotion of the interests of a Party to the conflict in the conduct of that conflict."[79] There are no required elements of use in an attack, visibility, or any required result of the use.[80] Additionally, unlike false representations of enemy communications, those that are represented to come from neutral parties are not considered permissible ruses of war.[81] Therefore improper use of the whole potential range of previously discussed military emblems, such as web and e-mail domain names, graphical symbols, and formal representations such as signature blocks in e-mails and electronic documents are also prohibited under this rule.

76 United Kingdom Ministry of Defence, *The Manual of the Law of Armed Conflict* ¶ 5.17.2 (2004) (hereinafter UK Manual).
77 CIHL Study, r. 57; *Tallinn Manual*, cmt. accompanying r. 64, ¶ 5.
78 AP I, art. 39(1).
79 AP I Commentary, ¶ 1565.
80 Id.
81 CIHL Study, r. 57, 63.

This broad prohibition is reflected in the Tallinn Rule 65, stating that "[i]n cyber operations, it is prohibited to make use of flags, military emblems, insignia, or uniforms of neutral or other States not party to the conflict."[82] Curiously, the Tallinn IGE raise the issue of "employment of other reliable indicators of neutral status," by referencing the discussion of protected indicators in Rule 62 and the UN emblem discussed in Rule 63, rather than the more legally relevant discussion of military emblems in Rule 64.[83] Again, protected emblems and the UN emblem draw on separate law for definition and application than military emblems. Therefore, for example, discussion of the improper use of the ICRC.org domain name must necessarily be distinct from that of the navy.mil domain name, although they may sometimes come to the same conclusion.

Combining the broad definition of military emblems, the absolutist nature of the prohibition, and the lack of customary practice regarding use of neutral indicators in communications by parties to the conflict,[84] there is little room left for the possible legal use of neutral military emblems in cyberspace by parties to the conflict. Still excluded are indicators that do not meet the definition of military emblems, such as IP addresses. The only exception to this rule is matters not related to the conduct of hostilities, such as "matters of police or civil administration."[85]

4. CONCLUSION

The balancing of military necessity and humanity is not an equation that can be definitively solved. As nations, non-state actors, and the methods and means of warfare continue to evolve, the law seeking to provide that balance will need to evolve with it. However, there are some enduring provisions at the core of IHL that states must always defend. One of these is the protection of groups seeking to enable that IHL balance during an armed conflict. The protected indicators of the GC, the recognized emblem of the UN, and the military emblems that identify neutral parties are vital to protecting these categories. With warfare moving into the human-made domain of cyberspace, every effort must be made to identify what these indicators look like in the new domain.

This paper has argued that current cyber indicators such as domain names, e-mail addresses, electronic signature blocks, and graphically displayed emblems can constitute the traditionally understood indicators in warfare. However, there are also many gaps that could be exploited by parties to a conflict. For example, IP addresses cannot be understood as protected indicators because they are not generally understood by participants, but they can be instrumental in providing a trust relationship between information systems. This trust could be exploited by adversaries and therefore undermine the effectiveness of the organizations represented by the indicators. Given the broad acceptance of the importance of protected and recognized indicators, state should seek to close these gaps in the law and defend a core aspect of IHL.

82 *Tallinn Manual*, r. 65.
83 Id., cmt. accompanying r. 65, ¶ 4.
84 CIHL Study, r. 63.
85 AP I Commentary, ¶ 1565.

REFERENCES

Protocol Additional to the Geneva Conventions of 12 August 1949, and Relating to the Protection of Victims of International Armed Conflicts, June 8, 1977, 1125 UNTS.

Commentary on the Additional Protocols of 8 June 1977 to the Geneva Conventions of 12 August 1949 (Yves Sandoz, Christophe Swinarski & Bruno Zimmermann eds., 1987).

Convention (I) for the Amelioration of the Condition of the Wounded and Sick in the Armed Forces in the Field, Aug. 12, 1949, 6 UST 3114, 75 UNTS 31.

Convention for the Protection of Cultural Property in the Event of Armed Conflict, May 14, 1954, 249 UNTS. 240

Customary International Humanitarian Law (Jean-Marie Henckaerts & Louise Doswald-Beck eds., 2005).

Tallinn Manual on the International Law Applicable to Cyber Warfare (Michael N. Schmitt ed., 2013).

Regulations Respecting the Laws and Customs of War on Land, annexed to Convention No. IV Respecting the Laws and Customs of War on Land, Oct. 18, 1907, 36 Stat. 2227, T.S. No. 539.

Commentary to Geneva Convention I for the Amelioration of the Condition of the Wounded and Sick in the Armed Forces in the Field (Jean Pictet ed., 1952).

International Committee for the Red Cross, Commentary to Geneva Convention I for the Amelioration of the Condition of the Wounded and Sick in the Armed Forces in the Field (2d ed. 2016) [https://ihl-databases.icrc.org/applic/ihl/ihl.nsf/Comment.xsp?action=openDocument&documentId=EC42EEC3274A7323C1257F7A0056F100].

International Committee for the Red Cross, Regulations on the Use of the Emblem of the Red Cross or the Red Crescent by the National Societies (1991).

Office of the General Counsel, U.S. Department of Defense, *Law of War Manual* (2016).

Michael Gregg, *Certified Ethical Hacker Cert Guide* (2014).

Ed Skoudis & Tom Liston, *Counter Hack Reloaded* (2006).

Vienna Convention on the Law of Treaties, May 23, 1969, 1155 UNTS 331.

The United Nations Flag Code and Regulations, ST/SGB/132, United Nations, January 1967.

Michael Bothe, Karl Josef Partsch & Waldemar A. Solf, New Rules for Victims of Armed Conflicts, Commentary on the Two 1977 Protocols Additional to the Geneva Conventions of 1949 (1982).

Canada, Reservations and statements of understanding made upon ratification of Additional Protocol I.

United Kingdom Ministry of Defence, *Manual of the Law of Armed Conflict* (2004).