The Internet *in Bello*: Cyber War Law, Ethics & Policy

Seminar held 18 November 2011, Berkeley Law

Kate Jastram and Anne Quintin[1]

V. Attack and Distinction

The second panel was chaired by Kate Jastram. The speakers, in order of presentation, were Anne Quintin, Lt. Col. Peter Hayden, and Sean Watts.

Jastram began by noting that the shift to a discussion of *jus in bello* principles was not intended to close off the debate from the first panel as to whether this body of law applies. Indeed, examining how well the existing body of laws works in cyber operations is an important part of that inquiry. She recalled that attack is defined as an act of violence against an adversary, while distinction is one of the most fundamental principles of international humanitarian law. The panel would explore the applicability of these norms to the cyber domain.

A. Comments by Anne Quintin

Public Affairs Officer, International Committee of the Red Cross

Quintin observed that cyber warfare is a recent area of research for the International Committee of the Red Cross. With the military potential of cyber only starting to be explored, it is important to assert the applicability of IHL to cyber operations in the context of armed conflict. We do not yet know what the humanitarian consequences might be. However, this should not be a decisive obstacle, as we can already imagine the potential large-scale effects on civilian populations, if for instance airports, transportation systems, nuclear power stations, and dams were to be attacked. The consequences are difficult to assess now, but we cannot afford to wait until it is too late to prevent worst-case scenarios.

There is, of course, no mention in IHL of cyber operations, or computer attacks, or related terms. These terms do not have an internationally agreed legal meaning and are used in different contexts, not always limited to armed conflicts, and with different meanings. Does this indicate that IHL is ill-adapted to regulate such operations? The answer is no, for the following reasons.

First, new technologies have always been developed, and IHL is sufficiently broad to encompass them. Some specific conventions have been drafted for specific weapons, such as chemical and biological weapons, and anti-personnel mines. There may be a need for specific conventions on cyber weapons, but before they are developed, IHL does provide some answers.

[1] Kate Jastram is a Lecturer in Residence and Senior Fellow, Miller Institute for Global Challenges and the Law, University of California, Berkeley, School of Law. Anne Quintin is a Public Affairs Officer at the International Committee of the Red Cross in Washington, D.C.

The prospect of new types of weapons, not thought of and not conceivable at the time the Conventions and the Additional Protocols were drafted, is clearly envisaged in Article 36 of Additional Protocol I to the Geneva Conventions.[2] This clearly demonstrates that cyber operations, when conducted during an armed conflict, fall under the scope of IHL.

On a more practical note, if cyber means and methods produce the same effects as kinetic operations, they are – and should be – governed by the same rules. For example, a cyber manipulation of the air traffic control system resulting in the crash of civilian aircraft would be governed by IHL. The legality of such an attack would be assessed through the traditional principles applicable to kinetic operations -- distinction, proportionality, and precautions. There is no legal vacuum in cyberspace.

It is, however, important to stress that IHL comes into play only when cyber operations are committed in the context of an armed conflict. Which brings us to a second consideration: when do we have an armed conflict? What if the first or only hostile act is a cyber attack? Can this be qualified as constituting an armed conflict within the meaning of IHL? The answer today can only be theoretical. There is not enough State practice for a customary international law rule to be determined. But we do have a few elements of the answer.

Consider the definition of armed attack in Additional Protocol I, Article 49.1.[3] The term "attacks" means acts of violence against the adversary, whether in offense or in defense. "Acts of violence" have been interpreted as meaning physical force. Based on that view, which ICRC shares, viruses or worms which cause physical damage to persons or objects that go beyond the computer program or data attacked, could be qualified as an attack. To go further, it is not necessary to reach destruction or damage to have an armed attack. Additional Protocol I, Article 52.2 on military objectives also refers to neutralization.[4] Thus, it has been argued that neutralization is sufficient for an attack and so falls within IHL. Disabling a power grid, for example, without destroying it, would qualify.

Quintin then turned to the challenges posed by cyber operations to traditional notions of IHL, namely, *distinction, proportionality*, and *precaution*. *Distinction* requires parties to distinguish at all times between civilians and combatants, civilian objects and military objectives. It is sometimes claimed that cyber operations can be directed against a broader range of targets than kinetic attacks, including objects usually considered to be civilian objects.

[2] API: 36. "In the study, development, acquisition or adoption of a new weapon, means or method of warfare, a High Contracting Party is under an obligation to determine whether its employment would, in some or all circumstances, be prohibited by this Protocol or by any other rule of international law applicable to the High Contracting Party."

[3] API:49.1. "'Attacks' means acts of violence against the adversary, whether in offence or in defence."

[4] AP1:52.2. "Attacks shall be limited strictly to military objectives. In so far as objects are concerned, military objectives are limited to those objects which by their nature, location, purpose or use make an effective contribution to military action and whose total or partial destruction, capture or neutralization, in the circumstances ruling at the time, offers a definite military advantage."

In ICRC's view, this claim is totally unfounded under IHL. The definition of a military objective is not dependent on the methods and means employed. The principle of distinction is entirely applicable during cyber attacks. It follows that attacks against civilian objects cannot be lawful, even if not leading to destruction.

To make distinction easier for parties, it has been suggested that one might verify the status of a data stream by distinctive markers, something similar to the marking of hospitals in kinetic warfare. So, for example, a computer facility could have a cyber marker to show that it is operated by a hospital. This is an idea that merits further discussion.

The principle of distinction also includes the prohibition against indiscriminate attacks. This is potentially the most serious problem in the context of cyber operations. Cyber space is characterized by interconnectivity. According to a recent Department of Defense report, DOD employees operate 15,000 computer networks with 7 million computers at hundreds of locations around the world. Nearly all military cyber infrastructure relies on civilian networks. As a consequence, the release of a virus against the military network could seep out into the civilian system of the targeted state, and could cross borders, and so would be indiscriminate. However, viruses and malware have reached a certain level of sophistication, allowing a high level of control over what is being targeted.

Therefore the question may rather be one of *proportionality*, as it is not possible to anticipate all the reverberating consequences. Using a concrete example of hospitals, "respect" means not only not attacking them but also not interfering with their work. In that regard, an attack against the military network that would also affect the network of the military medical services, for instance preventing them from continuing to give treatment to the wounded, would be disproportionate. But it should also be said that cyber attacks may also be the best way to minimize collateral damage. It may now be possible to switch something off instead of destroying it.

In considering *precaution*, parties are required to take all necessary precautions to reduce the effects of attacks. The obligation includes removing civilian objects from the vicinity of military objectives. In cyber space, the obligation becomes one of ensuring that military computer systems are sufficiently separate from the civilian network. But, as previously mentioned, information technology may also serve to limit incidental damage. It might be less damaging to disrupt than to destroy. Precaution arguably imposes on States the obligation to choose less harmful means to achieve military aims, and cyber operations may sometimes respond to that principle.

She concluded by confirming that ICRC will continue to follow cyber developments closely and to assess their potential humanitarian impact with a view to ensuring that IHL is respected.

B. Comments by Lt. Col. Peter Hayden

Deputy Legal Counsel (Operations), Office of the Chairman, Joint Chiefs of Staff

Hayden began with the disclaimer that he was presenting his own thoughts and was not speaking for the Department of Defense. He drew attention to a recent report in the *Washington Post* that DOD had advised Congress that it has the capability to conduct offensive operations in cyber space to defend the American nation, its allies, and its interests. If directed by the President, DOD will conduct offensive cyber operations in a manner consistent with policy, principles, and legal regimes the Department allows for kinetic operations, including the law of armed conflict.[5]

He noted that cyber is dangerous and exposes significant vulnerabilities, not least that we rely upon it so heavily, while it also offers tremendous opportunities in support of distinction and the concept of a military objective. With respect to the latter, he offered a comparison of the elements of the definition of a military objective contained in Additional Protocol I, Article 52.2 in the kinetic realm and in the cyber realm.

Nature In kinetic operations, it would be a tank. In the cyber realm, it is malware, key logging espionage malware, or something that could cause system failure. By nature, that is a legitimate military objective.

Location It can be a chokepoint, or high ground, or a bridge. In cyber space, it could be network nodes or key junctures.

Purpose and use It would be a pharmaceutical factory if it is manufacturing chemical weapons. In cyber, it could be many things. Michael Schmitt has noted that many things have a dual use in cyber.[6]

Cyber offers the opportunity to narrow the definition of a military objective with a greater degree of precision. "Military objective" is not a lawyer's creation, but a doctrinal definition for warfighters. Military strategic thinkers have adopted it straight out of Additional Protocol I; for instance, Milan Vego, a preeminent strategic military thinker, included it in his work on *Joint Operational Warfare*.[7] The aim is to figure out what exactly is necessary to weaken the enemy forces to the greatest extent possible.

[5] Ellen Nakashima, "Pentagon: Cyber offense part of U.S. strategy," *Washington Post*, Nov. 15, 2011, available at http://www.washingtonpost.com/national/national-security/pentagon-cyber-offense-part-of-us-strategy/2011/11/15/gIQArEAlPN_story.html?wprss=rss_politics (last visited 23 March 2012).

[6] Michael Schmitt, "Wired warfare: computer network attack and *jus in bello*," *International Review of the Red Cross*, Vol. 84, N°846 (June 2002) 365-98, available at http://www.icrc.org/eng/resources/documents/misc/5c5d5c.htm (last visited 23 March 2012).

[7] Milan Vego, *Joint Operational Warfare: Theory and Practice* (2007).

Hayden then presented the Stuxnet attack as a good illustration of what is possible under cyber and how to analyze concepts of military objective and attack. He stressed that all the information he had on Stuxnet was open source.

He recalled that in September 2010, a worm infected some 90,000 systems in several countries. After a while, it became clear that the worm was targeting Iran. More precisely, its aim was to get at a certain kind of programmable logic controller that ran centrifuges for a nuclear production facility in Iran. It targeted only those controllers which drove centrifuges manufactured by Siemens and which operated in certain parameters. Although Stuxnet infected 90,000 systems, most of them were not affected by it; Stuxnet simply replicated and moved on. Of nine thousand nuclear enrichment centrifuges, approximately one thousand failed, causing major damage and thus delaying Iran's nuclear production.

Going back to the concept of a military objective, Hayden used the hypothetical example of a hot war with "Antarctica", hence justifying the application of a LOAC framework. Antarctica's military objective is to degrade or take away Iran's nuclear capability. The subordinate military objective is to take down centrifuges. Subordinate to that is corruption of the programmable logic controllers.

Cyber can do only very limited things on its own: (1) alter a program's structure, or (2) give the program bad information. The first-level objective is to do something to the program or data, but the purpose is to cause physical destruction, breaking the controllers. So yes, this is probably an attack within the meaning of the law of war.

Under LOAC, the next question would be whether the attack violated other provisions. More specifically, did it cause incidental damage to civilians? There is a possibility that Iran's capacity to produce sufficient energy for its civilian population would have been compromised, in which case the attack could be considered to cause damage to the Iranian civilian population.

At this stage of the example, the hypothetical can be changed slightly: assume that among the affected centrifuges, some are used for military production and some are for civilian use. It is therefore a dual-use target. But Antarctica may have the ability to tailor its attack to avoid producing incidental civilian damage. To do that, Antarctica must have a great deal of knowledge about the enemy's systems. Cyber can allow those who use it to refine an attack and exclude harm to civilian objects.

Another question that arises is how did Stuxnet get to Iran? The Iranian systems were spread out across the country, sheltered and hidden away. Moreover, they were air-gapped, not connected to the Internet. As a consequence, the attack had to be carried via thumb drives, which was a very clever way to defeat the strategy. Did the Stuxnet author cause incidental damage or loss to civilian lives? Antarctica would say no, there was no loss of civilian life or damage to civilian objects. The virus replicated through civilian systems without causing damage to them. At the end of the operation, it shut itself off, and eliminated itself. It was designed so as not to cause damage to any other system. It was a clever cyber weapon, almost as if it were written by lawyers.

Stuxnet is only one kind of model. There are all kinds of cyber attacks, due to the creativity of technical people. This is an area where offense probably has the advantage. What about the reverse?

Revisiting the issues from a defensive point of view, what could be done by Iran to stop Stuxnet, or Duku, the cyber missile that carried it, which is now also getting press? The carrier itself (the nodes through which it travels) is a military objective. Iran would want to disarm the enemy. It could destroy computers and nodes with a kinetic attack. Such an action would be destroying a *location*, and would need to follow the principle of proportionality. It could cause disproportionate damage given the number of systems the worm was spread across.

What if you could attack just the program itself, using bad data, or corrupting the program? Bytes would be fighting bytes, ones and zeros versus ones and zeros. What is interesting in that scenario is whether it amounts to an attack in the first place. As a piece of code that offers the enemy an advantage, there is no doubt that Stuxnet is a legitimate military objective. Michael Schmitt says military objectives can be defined only after there is an attack. Hayden would reverse that and say a military objective occurs once a warfighter sees a need to get rid of something that would offer him an advantage. Attack is one tool to neutralize. There are other tools – for example, psychological operations, and propaganda. If, for example, Berkeley was in a war with San Francisco, the Bay Bridge would be a military objective. Or you could get airport baggage handlers to go on strike, in order to shut down the airport. There are ways of neutralizing objectives that are not attacks under international law.

So if we can go after data, if we can attack programs or data such that there is no injury or loss of life to civilian objects, cyber allows us to be "hyper distinctive" in the context of war. We might be able to activate it only when pursuing military purposes.

He then raised the question of going into neutral countries, as attack vectors. That is a concern, as the law is undefined. With respect to the doctrine of neutrality, Hague Convention V of course did not mention cyber, so the law is not settled as to what happens when you go through neutral countries.

Cyber is a remarkable enabling tool. Hayden expressed a concern with Anne Quintin's earlier statement regarding the obligation to take precautions in attack. He suggested thinking of cyber as the Ferrari in the garage. Just because you have one, does not mean you always take it out of the garage.

Elaborating on this point, he noted that cyber attacks go after "zero day" capabilities. Once the target country finds out that there has been an attack, then everyone knows. Thus, cyber is the Ferrari in the garage that gets taken out only on special occasions. Cyber attacks are a perishable tool. The *Washington Post* reported that there was apparently consideration of using cyber attacks against Libya.[8] If so, one consideration would have been whether it was worth it. With cyber, losing your Ferrari is one factor in making that decision. Simply because you can reduce civilian damage the furthest by using a cyber capability does not mean that you have to use a cyber capability. You can husband the capability for a more opportune time. The reverse side of that point is that if you shepherd it too long, it will become obsolete, but that is not a legal concern.

C. Comments by Sean Watts

Associate Professor, Creighton University School of Law

Watts began by noting that cyber warfare offers a tremendous opportunity for lawyers to go through both longstanding legal concepts and terms of art, but also to move beyond doctrinal questions and think normatively about the law of war in ways that have been neglected for some time, for example, rethinking the fundamental balance between military objectives and humanity.

Watts has worked in the legal cyber realm on questions of status, and is interested in the way existing rules transfer to the cyber world. He particularly appreciated this panel for the opportunity to discuss the applicability, timeliness, and relevance of rules to targeting. His remarks would focus on a single dimension of attack, the notion of perfidy.

In its simplest terms, perfidy is a betrayal of legal good faith, consisting of three elements:

1) a feigned protected status;
2) an invitation to the target to recognize that protected legal status; and
3) a betrayal of that status, in order to take a military advantage.

These three elements, as Watts sometimes explains to his students, may be presented in terms of the elements of contracts: offer, acceptance and breach.

There is however a further legal understanding, or appreciation. Acts that might meet these three elements are not prohibited as perfidy unless they rise to a certain level. Traditionally, this result has to be killing or wounding of the enemy. Mere exploitation or damage of objects does not rise to this level. Additional Protocol I altered the traditional definition and added "capture",[9] creating two implications. First, a distinction should be made between perfidy and prohibited perfidy. Second, perfidy is a concept for combat, in the sense that it anticipates, like the legal notion of attack, a resort to violence.

As additional legal background, this is an area where codified law has rejected the general rule of staying away from examples. The prohibitions contained in Article 37 of Additional Protocol I clearly cite feigning wounds, feigning surrender or truce, feigning civilian

[8] Ellen Nakashima, "U.S. cyberweapons had been considered to disrupt Gaddafi's air defenses," *Washington Post*, Oct. 17, 2011, available at http://www.washingtonpost.com/world/national-security/us-cyber-weapons-had-been-considered-to-disrupt-gaddafis-air-defenses/2011/10/17/gIQAETpssL_story.html (last visited 23 March 2012).

[9] API: 37.1. "It is prohibited to kill, injure or capture an adversary by resort to perfidy."

status, and feigning neutral status. These are the protected classes of three out of the four Geneva Conventions. Another eye-catching characteristic of the Article lies in its partly-negative definition: routine military deception, ruses, and camouflage do not count as perfidy.

To finish up this legal snapshot, Watts explained the reason and purpose behind the prohibition of perfidy. The explication is complicated, and has ancient roots. In the earliest treaties, treachery was the term used, with ”perfidy” emerging only in the mid-20th century. Treachery related to notions of military honor grounded in chivalry. Combatants had a right not to die by unfair or dishonorable means. By the mid-20th century, certainly by the 1970s, the object and purpose of the prohibition of perfidy is not to focus on combatants, but to focus on protected classes (the person whose status is being feigned). The logic goes like this: if soldiers are routinely exposed to feigned status, they will regard suspiciously and honor less frequently those protected statuses. Such an unfortunate consequence has been experienced in Iraq and Afghanistan. Battlefields on which civilian status is routinely feigned become very dangerous places for civilians. Finally, Watts added one last rationale behind the prohibition: belligerents must have some level of trust in each other in order for war to end. Repeated betrayal of trust jeopardizes chances for peace.

Perfidy in the cyber realm

Perfidy is an attractive way to explain what is troubling about cyber warfare. Cyber is sneaky, backhanded, and deceptive; like perfidy, it takes advantage of its target by appearing benign. On an intuitive level, cyber warfare and perfidy seem a close match. Watts found a confirmation of this intuition recently while attending events in Australia and in China. There were many questions regarding perfidy posed by very well-informed, well-placed, and – in a cyber sense – relevant actors.

However, not all cyber attacks amount to perfidy and thus violations of the law of armed conflict. It is necessary to move away from the intuitive reaction, and look at the question more rigorously. Three observations may give us pause about that intuitive relationship.

1. Perfidy has always been expressed in human terms. Recalling the distinction between perfidy and prohibited perfidy, the prohibited type of perfidy features interaction between humans. One could envision a cyber attack that could satisfy the three elements, but the nonhuman cases are only a small fraction of these attacks.

2. Applied to cyber, perfidy loses much of its crucial interactive component. If something is automated, or preprogrammed, or based on algorithms, we lose that affirmative misrepresentation. It is much more implied, and difficult to analogize. There are cyber contexts that do involve human interactions, but they are reduced in the cyber realm.

Professor Caron was correct in his introduction – analogies seem to fail us. It is hard sometimes in cyber to make really useful analogies. A few questions then arise.

When we lack this interactive component, do we really have offer and acceptance? To borrow contract thinking further, is there really a meeting of the minds that we envision in perfidy? If there is, is there a conversation that makes use of the legal terms relevant to perfidy, for example, a wounded person, or a potential prisoner of war?

3. Of course, perfidy exists. In the context of the *jus ad bellum*, these are the attacks that get all the press: the whole electrical grid, the whole air traffic control system. But Watts' understanding is that that is not the future of cyber attack. The future of cyber attacks are low intensity attacks, where you do not even realize that you have been attacked. It is below the defenders' threshold of reaction. All the defender knows is that he is a little more inefficient: things are not working well, but he cannot figure out why.

Catastrophic attacks in the future will be a rarity. Cyber space has a problem with perfidy, because prohibited perfidy is limited to attacks resulting in human death or injury, even if you can get over the legal definitional obstacle presented by attack, which Watts thinks is significant and is meaningful. Quintin shared with us a perspective articulated previously by Knut Doermann, that neutralization is captured in the prohibition on interactions with civilian objects.[10] Watts is more sympathetic to the view that there needs to be an attack before Additional Protocol I, Article. 52.2 is relevant. Be that as it may, we are dealing with a very rare set of events that can rise to the level required for prohibited perfidy. Watts concluded by noting that this portion of his remarks was a doctrinal sketch.

He then moved on to raise a few normative points: If this is the law, has the law of war not come up short? Maybe somehow it has missed something, if that is all we have. Perhaps the intuitive reactions are onto something. It is increasingly apparent that some of the law of armed conflict does not accommodate the significance of cyber. It is true when we apply the term "attacked." If a State is just exploiting or conducting espionage, it can get around IHL limitations. This is especially so considering that cyber stands to be opening shot of most future large-scale conflicts. Are those the first swirls of a downward spiral to unregulated warfare? If perfidy's purpose is to have some level of trust, we are setting the scene poorly.

In closing, Watts noted that he had presented a formalistic and skeptical application of perfidy to cyber. In spirit, perfidy does capture a great deal of what is troubling to us. But if the technical case for perfidy in cyber war is weak, how long will we tolerate the consequences of that gap? The positivist in him is not ready to concede that the law of armed conflict covers cyber war under perfidy. However, you can make a strong case under *lex ferenda* that this is where we need to develop the law.

[10] Knut Doermann, "Applicability of the Additional Protocols to Computer Network Attacks", *International Committee of the Red Cross* (2004), available at http://www.icrc.org/eng/resources/documents/misc/68lg92.htm (last visited 23 March 2012).

D. Discussion on Attack and Distinction

It was clarified that Doermann's article[11] was not really looking at neutralization as the attack that could trigger armed conflict. But when you are in a situation of armed conflict, and you are using neutralization as a way to bring about a military advantage and can assert that this is targeting a military objective, it is difficult to say that neutralization is not an attack. There is a difference between neutralization as a first act, which is not enough to trigger armed conflict *per se*. But within the framework of armed conflict, it is different.

Another person responded that with respect to neutralization, Additional Protocol I: 49 says that an attack is violence against an adversary. It is repeated throughout the succeeding section. The *Commentary* also says that this is concerned with acts of violence. Neutralization may be inconvenient but it does not reach the level of a violent attack. LOAC does not address acts short of violence.

It was noted that speakers had emphasized cyber both as offering precision and as risking unforeseen consequences.

One person argued that obviously Stuxnet had spread all over world, but had activated only against its target. Cyber is potentially indiscriminate, but when you have programmers who are able to specify a certain set of conditions, it is possible to achieve two things at once. It can be sent out broadly, but defined as to when it will actually have an effect. Assuming that States can control what the tool will do, cyber tools are capable of precision effects, and can be the ultimate in discrimination.

One person then queried whether Stuxnet was representative of cyber warfare. In response, it was asserted that it was important to resist treating cyber attacks monolithically. Cyber is complex and always changing. Analogies and frameworks are hard to maintain. There is an understandable desire to generalize, but it is difficult to do so.

Another person wanted to emphasize a concern about using the word "attack" when talking about a violent act. If you look at cases from the International Court of Justice, an attack has to be significant to generate the right of self defense. There are proportionate measures permitted for lesser attacks. Most of so-called cyber attacks are exploitation. Most of the things complained about are efforts to get into computer systems, not to degrade them but to get information. Hacking fraud does not necessarily have any impact on a system. IHL has a very limited scope in cyber security. Most of it has nothing to do with the military.

One person agreed that he was enough of a legal formalist to adhere to that as well. Attack means different things in the *jus ad bellum* and the *jus in bello*. But, people call it a cyber "attack" for a reason. It is not all just the profit motive, as had been suggested in the previous discussion. "Attack" portrays more accurately what is at stake. This is again where

[11] Ibid.

analogies fail. When you lose data, or lose access to service – something more has been lost than the legal definition of attack suggests. It is just a normative point. If States are going to act like this is war, doesn’t LOW need to expand and acknowledge this reality? It would be better to see something more principled.