The European Journal of International Law Vol. 24 no. 1

How Cyber Changes the Laws of War

Jack Goldsmith*

Abstract

Michael Walzer's Just and Unjust Wars *anticipated many problems and developments in the laws of war, but it understandably did not anticipate how the Internet and associated computer and telecommunications revolutions would change war or the laws that govern it. This article seeks to assess, in general terms, the ways that the rise of cyber exploitation and cyber attacks challenge prevailing conceptions of the laws of war.*

Michael Walzer's *Just and Unjust Wars* anticipated many problems and developments in the laws of war. But it understandably did not anticipate how the Internet and associated computer and telecommunications revolutions would change war or the laws that govern it. In 1977, the year Walzer's book was published, Arpanet, the Internet's precursor, had been operating in practical secrecy with a crude packet-switching system for just eight years; the first email system was five years old; and Vinton Cerf and Bob Kahn had used the term 'Internet' for the first time in a paper on the Transmission Control Protocol just three years earlier. No one at this time, or for more than a decade, would worry much about the internal security vulnerabilities of this developing communications system. The wake-up call for security, such as it was, came in 1988, when Robert Tappan Morris, a graduate student at Cornell, introduced a worm on the Internet that was designed to determine the Internet's size but that inadvertently shut down about 10 per cent of the 60,000 computers then connected to it.[1] This event startled the Defense Advanced Research Projects Agency (DARPA), the futuristic Department of Defense (DOD) research wing. DARPA had given its financial support to what became the Internet to ensure that military communications could withstand nuclear attack. But suddenly its young creation seemed vulnerable from within.

These vulnerabilities would grow in the next two decades because, despite growing cyber-security concerns, the military and the society it defends would become ever more reliant on ever more vulnerable computer and telecommunication systems. This article explains these vulnerabilities, sketches how they affect the laws of war and

* Henry Shattuck Professor of Law, Harvard Law School. Email: jgoldsmith@law.harvard.edu.

1 Orman, 'The Morris Worm: A Fifteen Year Perspective', 1 *IEEE Security and Privacy* (2003) 35.

 doi: 10.1093/ejil/cht004

conjectures that international norms to regulate and temper attacks on these vulnerabilities are unlikely to develop.

1 Characteristics of Cyber

Many factors make computer systems vulnerable, but the most fundamental factor is their extraordinary complexity.[2] Most computers connected to the Internet are general purpose machines designed to perform multiple tasks. The operating-system software that manages these tasks, as well as the computer's relationship to the user, typically has tens of millions, and sometimes more than 100 million, lines of operating instructions, or code. It is practically impossible to identify and to analyse all the different ways these lines of code can interact or might fail to operate as expected. And when the operating-system software interfaces with computer processors, various software applications, Web browsers and the endless and endlessly complex pieces of hardware and software that constitute the computer and telecommunications networks that make up the Internet, the potential for unforeseen mistakes or failures becomes unfathomably large.

The complexity of computer systems often leads to accidental mistakes or failures. We have all suffered computer crashes, and sometimes these crashes cause serious problems. In 2009, the Internet in Germany and Sweden went down for several hours due to errors in the domain name system, which identifies computers on the Internet. A few years ago, a software problem in the Pentagon's global positioning system network prevented the Air Force from locking onto satellite signals on which they depend for many tasks. The accident on the Washington Metro in 2010, which killed nine people and injured dozens, was probably caused by a malfunction in the computer system that controls train movements. Several years ago, six stealth F-22 Raptor jets on their maiden flights were barely able to return to base when their onboard computers crashed.

The same complexity that leads to such malfunctions also creates vulnerabilities that human agents can use to make computer systems operate in unintended ways. Such cyber threats come in two forms. A cyber attack is an act that alters, degrades or destroys adversary computer systems or the information in or transiting through those systems. Cyber attacks are disruptive activities. Examples include the manipulation of a computer system to take over an electricity grid or to block military communications or to scramble or erase banking data. Cyber exploitations, by contrast, involve no disruption, but refer to merely monitoring and related espionage on computer systems, as well as the copying of data that is on those systems. Examples include the theft of credit card information, trade secrets, health records or weapons software and the interception of vital business, military and intelligence communications.

Both cyber attacks and cyber exploitations are very hard to defend against. 'The aggressor has to find only one crucial weakness; the defender has to find all of them,

[2] This part draws on Goldsmith, 'The New Vulnerability', *The New Republic* 7 June 2010, available at www.tnr.com/article/books-and-arts/75262/the-new-vulnerability.

and in advance', wrote Herman Kahn in his famous 1960 book, *On Thermonuclear War*.[3] This generally true proposition about defence systems has special salience for computer networks. Even if (as is often not the case) those trying to find and patch computer vulnerabilities outnumber those trying to find and exploit the vulnerabilities, the attacker often still has an advantage. Under the Kahn principle, in some fraction of the time the attacker will discover a vulnerability that the defender missed. And she need only find one, or a few, vulnerabilities to get in the system and cause trouble.

Once a vulnerability is identified, an attack or exploitation is relatively easy to disguise, because the operation of a computer is almost entirely hidden from the user. Malware can be embedded in a computer system without the user's knowledge, either remotely (when the user downloads an infected program, or when she visits an infected website) or at any point in the multi-country global supply chain that develops and produces most commercial software. And once it is embedded, malware can be used for any number of tasks, including data destruction, theft, taking over the computer for various purposes, recording keystrokes to discover passwords and much more. Many forms of malware are hard for engineers to find through diagnostic testing and are missed by anti-virus software. Computer users often do not discover malware before an attack makes clear that something has gone wrong. They often never discover malware that facilitates computer exploitations or (as happened in the China-Google kerfuffle), they discover it too late.

The inherent insecurity of computer systems is exacerbated by the number and incentives of actors around the globe who are empowered to take advantage of computer vulnerabilities. In real space, geography serves as a natural barrier to attack, theft and espionage: only if you get near the Pentagon can you attack it; only if you get near the Citibank branch in New York can you rob it. And if you are near these places in real space, American law enforcement and military authorities can exercise their full powers, within US sovereignty, to check or deter the attack. In cyberspace, geography matters much less because the Internet links computers globally with nearly instantaneous communication. As China's recent theft of information on Google's proprietary computers shows, someone sitting at a terminal in China can cause significant harm in the United States. And of course there are countless people around the globe with access to a computer who would like to do bad things inside the United States. To the extent that they are located outside the United States, American law enforcement authorities have much less effective power to stop or to deter them. The FBI must rely on law enforcement authorities in foreign countries who are often slow and uncooperative, giving bad cyber actors time to cover their tracks. And the American military cannot enter a foreign country unless the threat or attack rises to the level of war (a topic to which we will return).

Law enforcement and military authorities seeking to check malicious cyber activity face another fundamental challenge: the 'attribution problem' of identifying the author of a cyber attack or cyber exploitation. It is very difficult, and very

[3] H. Kahn, *On Thermonuclear War* (2007), at 535.

resource-intensive, and sometimes impossible, to trace with much certainty the computer origin of a professional cyber attack or cyber exploitation; it is even harder to do so in real time or even in the short-term. A thoughtful adversary can hide its tracks by routing attacks or exploitations through anonymizing computers around the globe. In 2009, a denial-of-service attack – a massive spam-like attack that clogs channels of communication – brought down some American and South Korean websites. Early reports said that the attack came from North Korea, but a few weeks later it was learned that the attack originated in Miami (and possibly, before Miami, elsewhere) and was routed through North Korea. It is still not known for sure who launched the attack, or from where.

Even if we can determine with some certainty which computer in the world is behind an attack or exploitation, that fact alone does not indicate who, or even which country, is responsible for the aggression. In 2009, a detailed study by the Information Warfare Monitor uncovered an extensive plot known as 'Ghostnet', which emanated from computers in China and infiltrated more than 1,000 sensitive government and commercial computer systems from over 100 countries. But the report could not determine whether the plot was controlled by the Chinese government or by private 'patriotic hackers' acting in the Chinese interest but without government involvement or by a criminal network in China. Nor could it rule out the possibility that 'a state other than China' was behind the plot, using agents to launch the operation from China in an attempt to 'deliberately mislead observers as to the true operator(s) and purpose of the *GhostNet* system'.[4] It is still not known who is behind the Conficker worm or the July 2009 denial-of-service attack on South Korea and the United States. Nor, more recently, do we know for sure who is behind the Stuxnet worm. Law enforcement and military officials are hobbled not only by geography, then, but also by their inability to know for sure where and by whom a cyber attack or exploitation originated.

To date, most harmful cyber operations have taken the form of exploitation – espionage, and massive theft of intellectual property, military secrets and the like. But cyber operations can also be attacks that potentially rise to the level of war, or that facilitate war-fighting. We saw glimpses of this when Russia (or groups in Russia) used denial-of-service attacks to shut down Estonian banks and government web sites in 2007 and cripple Georgian government web sites in 2008. Presidents Obama and Bush reportedly ordered covert computer attacks on computer systems related to Iran's nuclear weapons programme. Many experts believe that a cyber operation could shut down a stock exchange or destroy bank or money transfer records or operations, wreaking economic havoc. Or a cyber operation could corrupt or take over the computer systems (known as SCADA systems), which monitor and control infrastructure processes like the electrical grid and nuclear power plants, causing them to shut down or malfunction. The significance of the Stuxnet worm is that it seems designed to do just that.

[4] Information Warfare Monitor, *Tracking Ghostnet: Investigating a Cyberespionage Network* (2009), at 48–49.

2 Cyber and the Laws of War

Taken together, the factors outlined in Section 1 make it much easier than ever for people outside one country to commit very bad acts, possibly rising to the level of war, against computer systems and all that they support inside another country. This raises some well-studied (though not resolved) challenges to the laws of war, and some less obvious ones.

One challenge is to figure out when a cyber attack implicates *jus ad bellum*. The hard question is how to translate the UN Charter concepts of 'use of *force*' and '*armed* attack' into the cyber realm. The main answer that has emerged, drawing on Michael Schmitt's work, has been to focus on the scale of the kinetic effects of the cyber operation.[5] When the effects of a cyber operation are akin to the effects that would implicate the UN Charter terms, then cyber operations implicate the UN Charter. So, for example, a cyber attack that renders the electricity grid or air-traffic control system inoperable, and that as a result causes many deaths, would count as a use of force. But a cyber operation that merely involves espionage or that disrupts DOD computers conducting military research, likely would not be considered a use of force.

These cases are easy enough. But cyber operations introduce more challenging questions.[6] The challenges arise mainly because the Charter focuses its prohibitions on military means of inflicting damage on another state, but does not prohibit economic or political means of inflicting damage on another state. As a general matter, military means by one state that leads to deaths or physical destruction in another implicate the Charter, but political or economic sanctions that lead to deaths or physical destruction in another state do not. Cyber operations can cause havoc in a nation, including death and destruction, which might appear more like economic sanctions than a military use of force. Consider, for example, a cyber attack that corrupts data on a stock exchange and which in turn causes widespread economic harm but no direct physical damage. Is this more like a physical use of force or like economic sanctions? What about widespread economic harm caused by massive theft of digitalized intellectual property? Theft and espionage are not generally viewed as implicating the Charter, but the cyber context changes the scale and consequences of theft and espionage to a degree that can result in harm to the country at least as severe as a physical attack. Another difficulty with cyber operations is that, unlike many kinetic attacks, they can take place slowly and can be reversible. There is no settled answer to the question whether or when a slow disruptive

[5] See Schmitt, 'Computer Network Attack and Use of Force in International Law: Thoughts on a Normative Framework' 37 *Columbia Journal of Transnational Law* (1999) 885, at 914–916; see also Schmitt, 'Operations in International Law: The Use of Force, Collective Security, Self-Defense, and Armed Conflicts', in Cyber Computer Science and Telecommunications Board (ed.), *Deterring Cyberattacks: Informing Strategies and Developing Options for U.S. Policy* (2010) 151, available at www.nap.edu/openbook.php?record_id=12997&page=151.

[6] See generally NRC, *Technology, Policy, Law, and Ethics Regarding U.S. Acquisition and Use of Cyberattack Capabilities* (2009), ch. 7, available at www.nap.edu/catalog.php?record_id=12651#toc.; Waxman, 'Cyber-Attacks and the Use of Force: Back to the Future of Article 2(4)', 36 *Yale J Int'l L* (2011), at 421; Jason Barkham, 'Information Warfare and International Law on the Use of Force', *New York University International Law and Politics* (2001) 34, at 57–113.

cyber attack on critical infrastructure or an analogous system that gradually renders it sub-optimally operable becomes a use of force. Similarly, there is no settled answer to the question whether a temporary but reversible shut down of a computer system, lasting perhaps two days or two weeks, associated with a fighter-jet squadron or a reconnaissance-satellite system is a use of force. Nor is it clear whether 'mere' destruction of critical economic or military data, without any physical consequences, is a use of force.

Similar questions arise in trying to figure out which cyber attacks are 'armed attacks' under Article 51. In addition to conceptual problems analogous to those that arise with uses of force, the problem of attribution causes further complications. A thoughtful adversary can hide its tracks by routing attacks through anonymizing computers around the globe. Even if a nation knows which computer in the world is behind an attack, that fact alone does not indicate who, or even which country, is responsible for the aggression. Sometimes traceback and related forensic tools can provide pretty good attribution. And human and other forms of intelligence gathering can further help in attribution. But even taking into account these and other tools, the attribution of a sophisticated cyber attack is neither fast nor remotely certain. This makes it very hard for the nation responding to an armed attack to know which nation (if any) is responsible for it, and thus which nation it should use force against in self-defence. Opportunistic but plausible denials of responsibility for the armed attack will be frequent, attribution assessments will be probabilistic and mistakes in responding to cyber attacks will be inevitable. These problems create disincentives and uncertainty in responding to a cyber attack, and they lower *ex ante* disincentives to cyber attack in the first place. A related problem is that a cyber attack might start slowly and build over time, and waiting too long to respond to it might well make it harder to respond. But there is no way to know in advance whether an attack will grow in this way. Must a nation wait until the attack crosses a critical threshold when it might be too late, or can it respond earlier even though its prediction about the ultimate scale of attack might be faulty?

The attribution problem also underlies many of the problems in applying *jus in bello* principles. Cyber operations challenge both the principle of distinction and the principle of proportionality. In the cyber realm, it is often hard to know whether the computer system being attacked, or the nation associated with that computer system, is a military target. The mingling of civilian and military computer and telecommunication systems raises a similar problem. These examples suggest that a nation using cyber weapons often, and perhaps usually, cannot know with certainty whether it is attacking a military or civilian target. (Similar problems can arise, obviously, in the kinetic context, but the problem is much more pervasive when cyber weapons are employed.) In addition, a cyber attack can have unpredictable indirect and cascading effects on associated computer networks that make collateral damage very hard to calculate – much harder than the vast majority of kinetic targeting decisions. One reported reason why the Bush administration called off a planned attack on Saddam Hussein's financial network in 2003 was a worry about uncontrollable indirect effects on the global banking system.[7]

[7] J. Markoff and T. Shanker, 'Halted '03 Iraq Plan Illustrates U.S. Fear of Cyberwar Risk', *New York Times* 1 August 2009.

The laws of war are somewhat muddled when the source of an attack from one state is a non-state actor. The problem is exacerbated in the cyber realm. For one thing, a nation suffering an attack has a hard time knowing whether the attack comes from a state actor or a non-state actor. This can make it hard to know whether a military or law enforcement or some other response is appropriate. Assuming the actor is private, the criteria for state responsibility are unsettled. There are growing calls to deal with this cyber-attribution problem by making a nation responsible for all cyber attacks that emerge from within its borders, even if the attacks are not sponsored by that nation.[8] This would in theory ameliorate the attribution problem by eliminating the 'it wasn't us, it was private hackers' defence that Russia, China and other nations have invoked when criticized for cyber attacks from inside their borders. It is not clear whether technology permits nations (in a remotely cost-effective manner) to take the steps needed to control or arrest all malicious cyber agents from leaving their borders. But assuming away the technological hurdles, a strong state responsibility norm in this context would require extensive and intrusive governmental activity in the private network that at least for now is anathema in the United States and other western democracies.

A final problem is espionage. *Just and Unjust War* says practically nothing about espionage, perhaps because in both war and peace, international law hardly regulates it. Similarly, international law says little about state-sponsored theft of intellectual property and military secrets. It is unclear whether, in the cyber era, international law's non-regulation of spying and theft can continue. One reason is that the cyber realm makes possible massive theft of intellectual property and military and intelligence secrets that, in the aggregate, can (and many people believe now do) constitute a serious national security problem, a problem that could conceivably require a military response. Another reason is that the software agents that facilitate cyber espionage and those that facilitate cyber attacks are hard if not impossible (*ex ante*) to distinguish. This means that no nation can tell for sure whether the logic bombs and related agents it finds in its civilian infrastructure networks are agents of exploitation or attack – until, of course, they are used for attack. If these agents turn out to be used for attack, our complacency about the agents of exploitation – and about international law's non-regulation of digital spying and digital theft – will surely change.

3 Cyber and International Agreement

Even if we assume that some of the puzzles in Section 2 can be worked out, and that nations of the world can agree in theory on how *jus ad bellum* and *jus in bello* should apply to the cyber realm, a further hurdle stands in the way of developing true international norms to reflect this substantive agreement.[9] The main hurdle is verification, which is difficult in the cyber realm because attribution is challenging.

[8] See, e.g., R. Clarke and R. Knake, *CyberWar: The Next Threat to National Security and What to Do About It* (2010).

[9] For a more elaborate treatment of this issue, see Goldsmith, 'Cybersecurity Treaties: A Skeptical View', in P. Berkowitz (ed), *Future Challenges in National Security Law* (2010), available at http://media.hoover.org/sites/default/files/documents/FutureChallenges_Goldsmith.pdf.

If one sees the laws of war in instrumental terms, akin to solutions to a prisoners' dilemma, then it is doubtful that a rational government would forego using otherwise desired cyber-exploitation or cyber-attack capabilities in compliance with an international norm in exchange for mutual restraint by adversaries. The main reason is that a government cannot tell whether its adversary is complying with the norm, and thus it cannot know (until it is too late) whether it is receiving a benefit for its restraint. Moreover, cooperation in the prisoners' dilemma depends on credible retaliation when there is breach. Uncertainty in attribution makes retaliation for breach much harder for any president or general to order. ('Sir, we are 28 per cent sure the Chinese did it.') And this in turn makes retaliation less credible to some probably large degree, which in turn invites breach and unravels cooperation.

But perhaps the laws of war work more through normative influence. Once the norms are established and accepted, they will exert normative pressure on states to comply. I think this can work in institutions like the US Department of Defense, which have massive bureaucracies of relatively independent lawyers and compliance officers devoted to following the laws of war, and a strong culture of compliance. (These bureaucratic structures likely have an instrumental foundation, but I will set that aside for the moment.) And in fact there have been many reports that the DOD is deeply self-constrained – some say too constrained – by the laws of war in its use of offensive cyber weapons. But there are few militaries in the world with the type of self-constraining bureaucracy as the DOD. Can the laws of war have a normative influence on compliance with these other countries?

I am sceptical. There are many reasons for scepticism, but the main one is the problem of attribution. The laws of war would not be nearly as efficacious or have the same level of normative salience, if the nations that violated the laws could not be identified and publicly shamed. Norms cannot get much purchase in a world without serious attribution; anonymity is a norm destroyer. That, unfortunately, is the situation in the cyber realm. The problem is exacerbated by the fact that, even if a nation has perfect attribution, it often cannot publicly reveal the evidence of attribution because doing so would disclose espionage and attribution capabilities and render them less useful. To the extent that this is so, it makes the public shaming aspects of a verification regime, and thus the operation of norms, less robust.

In this world of anonymity, it is unlikely for the laws of war to have much normative purchase to constrain nations that lack robust bureaucratic commitments to compliance with the laws of war and that are otherwise inclined to use cyber weapons. Stewart Baker imagines the differences between the DOD and most other military bureaucracies like this:

> The Pentagon would be exquisitely sensitive to arguable violations of international law in carrying out operations in cyberspace. Our guys would sit with their fingers poised over the 'return' button for hours while the JAGs were trying to figure out whether the Belarussian remarks in committee were a consensus or an individual interpretation of article 42bis. And nobody else would give a damn what the treaty said, because they wouldn't expect to get

> caught and because even implausible deniability can't be rebutted with the certainty needed to make a legal case, let alone send missiles in response.[10]

Baker is exaggerating for effect, but his essential insight about the relatively robust compliance commitment by the United States and its allies as compared to other nations, and the resultant opportunities for mischief and opportunism in the cyber realm as a result, is right.

The many hurdles to developing international norms do not necessarily mean that the growing stockpiles of cyber weapons will lead to cyber war. Above I noted that norms would not have much influence on nations '*otherwise inclined to use cyber weapons*.' The moderately good news, I think, is that, although many nations have the capabilities to engage in large-scale cyber war, they are not using the weapons because doing so would be self-defeating. Even with the cloak of relative anonymity, the potential catastrophic costs to the globally integrated computer and telecommunications infrastructure from a large-scale cyber war creates powerful disincentives for a nation to engage in such war. One can, if one likes, call mutual restraint of this sort a 'norm'. This norm is nothing more than a behavioural regularity resulting from the coincidence of uncoordinated self-interest among nations. But if every nation continues to have such independent incentives, cyber war might not emerge.

To see the point, consider Richard Clarke's claim in his book *Cyber War* that China will in the near future engage in cyber attacks on the United States.[11] It is true that China has significant offensive cyber capacities that could in theory cause enormous destruction, and that it is stockpiling cyber weapons and planning for cyber war. But the same is true of the United States. What Clarke never adequately explains is why China or other nations would use these weapons in this way. Capacities and contingency plans, taken alone, do not add up to a serious threat. There must also be a plausible scenario in which a nation has the motivation to use these weapons.

Clarke addresses this issue briefly, in trying to explain why China might destroy American infrastructure by means of a cyber attack even though 'China's dependence on U.S. markets for its manufactured goods and the trillions the country has invested in U.S. treasury bills mean that China would have a lot to lose.'[12] He says that the United States and China might nonetheless be drawn into a war over Taiwan or the oil-rich islands in the South China Sea. Perhaps, but it is hard to imagine that China would wipe out the New York Stock Exchange or the electrical grid of the East Coast unless it were in a total war over those islands — the sort of war that would also involve enormously destructive non-cyber weapons. Clarke is also right that China's cyber weapons might (like China's conventional forces) deter the United States from intervening against China in a Pacific Rim contest. But he should also acknowledge

[10] Stewart Baker, *Going Wobbly on Russia's Cybersecurity Disarmament Proposal?*, The Volokh Conspiracy, June 6, 2010, http://volokh.com/2010/06/06/going-wobbly-on-russias-cybersecurity-disarmament-proposal/?utm_source=feedburner&utm_medium=feed&utm_campaign=Feed:+volokh/mainfeed+(The+Volokh+Conspiracy)&utm_content=Google+Reader; see also Stewart A. Baker, *Skating on Stilts: Why We Aren't Stopping Tomorrow's Terrorism* (2010), at 231.

[11] See *supra* note 8.

[12] *Ibid.*, at 61.

that this deterrent is weakened by China's dependency on a functioning American economy, which significantly reduces the credibility of its cyber threat.

I am not saying that there is no chance that a nation might want to use cyber weapons for attack, possibly rising to the level of war. We have already seen low-level cyber attacks related to war in Georgia and Lithuania. Many people think that Stuxnet is the first truly dangerous cyber weapon, and that it was designed by the Israelis to knock out the Iranian nuclear weapons programme. It is also possible that the stealth cyber-arms race, the difficulty of knowing for sure which nation is behind a cyber attack, and the general absence of effective norms to govern such attacks combine to create an unstable situation in which destructive cyber activities might escalate by accident. Finally, criminal groups have growing capabilities that could cause significant damage to nation states, and terrorists are now in the market for these capabilities. All of these developments are worth worrying about, and will present enormous challenges to, and likely require large changes in, our understanding of the laws of war.