Journal of Military Ethics
Publication details, including instructions for authors and subscription information:
http://www.tandfonline.com/loi/smil20

The Ethics of Cyberwarfare

Randall R. Dipert [a]

[a] SUNY (State University of New York) at Buffalo, NY, USA

Version of record first published: 16 Dec 2010.

To cite this article: Randall R. Dipert (2010): The Ethics of Cyberwarfare, Journal of Military Ethics, 9:4, 384-410

To link to this article: http://dx.doi.org/10.1080/15027570.2010.536404

Journal of Military Ethics,
Vol. 9, No. 4, 384–410, 2010

Routledge
Taylor & Francis Group

The Ethics of Cyberwarfare

RANDALL R. DIPERT
SUNY (State University of New York) at Buffalo, NY, USA

ABSTRACT *The paper addresses several issues in the morality of cyberwar and cyberwarfare, defined as one nation's attacks on the governmental or civilian information systems of another nation. It sketches the diverse technical ways in which an attack may occur, including denial-of-service attacks and the insertion of various forms of malware. It argues that existing international law and widely discussed principles of Just War Theory do not straightforwardly apply to cyberwarfare, and many forms of cyberwarfare differ from previous forms of warfare in neither injuring nor killing human beings, nor causing lasting physical damage – but can nevertheless cause serious harm to a nation's vital interests. Another dissimilarity with traditional warfare is in the degree of knowledge of the identity of an attacker (the 'attribution problem'). The paper argues that cyberwarfare is not amenable to regulation by international pacts and that we can expect long periods of low-level, multilateral cyberwarfare, a Cyber Cold War, as a game-theoretic equilibrium is sought,. The paper laments the lack of a cyberwarfare policy, and concludes that it is only by applying game-theoretic principles that strategies can be discovered that are both moral and effective in suppressing overall harm to all parties in the long run.*

KEY WORDS: Ethics, cyberwarfare, cyberattack, cybersecurity, cyberespionage, just war theory

Introduction

'Cyberwarfare' is a recent term: the Oxford English Dictionary gives its first use as 1994 (OED 2010).[1] The topic of cyberwarfare has received widespread media attention only in 2009 and 2010.

The increasing discussion of cyberwarfare would come as no surprise to anyone who has been following the news in the last decade: cyberattacks on components of US defense cyber-infrastucture, and the probable strategic, military cyberattacks by Russia on Estonia, then on Georgia, (Beaumont 2009) and (USCCU 2009), and probable attacks by China, North Korea, and Iran on US defense and economic targets. Although organized attacks from China on Google and Gmail had a corporate target, the political implications for freedom of speech made them part of a larger cyberconflict and brought a warning from US Secretary of State Hillary Clinton (Clinton 2010) and (Markoff 2009b). In May 2010, the US Department of Defense formed a full command, Cyber Command, headed by a simultaneously promoted four-star general, General Keith Alexander, who also continues to serve as director of

Correspondence Address: Randall Dipert, Department of Philosophy, Park Hall 105, University at Buffalo, Buffalo, NY 14260, USA. E-mail: rdipert@buffalo.edu

1502-7570 Print/1502-7589 Online/10/040384–27 © 2010 Taylor & Francis
http://www.informaworld.com
DOI: 10.1080/15027570.2010.536404

the US National Security Agency (NSA). He had earlier noted that the Chinese government has hackers – cybersoldiers – organized into battalions and regiments (Alexander 2007).

Responding to intentional cyberharm ('cyberattacks') of all sorts, including by apolitical and anarchic ('black') hackers, as well as responding to economically motivated industrial cyberespionage, have all been a subject of discussion for decades. What is new is the acknowledgement of attacks that have been coordinated by the central commands of governments (or other political organizations) and that are directed at another country's governmental and military information systems, or at its commercial or infrastructure information systems for political purposes. Most of the public discussion has centered on *defense* against cyberattacks, that is, on governmental, military and economic cybersecurity.[2] What is still more recent is the open discussion of a need for an organized capacity for *offensive* cyberattack on hostile nations and political organizations by technologically advanced democracies.

Cyberwarfare has several unusual features. It is arguably the first major new form of warfare since the development of nuclear weapons and intercontinental missiles. This novelty has also meant that there is at present a virtual policy vacuum: there are no informed, open, public or political discussions of what an ethical and wise policy for the use of such weapons would be.

Second, it is very difficult to determine the source of cyberattacks: this is the 'attribution problem.' This fact would give many cyberattacks credible deniability – especially since in many cases nations can plausibly claim that the attacks may have originated from within their territory but their governments did not initiate them.

Third, many cyberattacks will not be lethal and will not even result in permanent damage to physical objects. This is of course extremely dissimilar from nuclear weapons and from virtually all traditional weapons of war.

Fourth, there are no exotic components to cyberweapons, again very unlike nuclear and other advanced technology weapons, and even unlike chemical or biological weapons. Any computer is a potential cyberweapon and anyone with advanced knowledge of information systems is a potential cybercombatant (Talbot 2010). This makes treaties that would ban cyberweapons virtually impossible from the outset (Libicki 2009).

One feature that cyberwarfare unfortunately shares with warfare using non-cyber weapons is that defense is expensive and liable to failure, while offense is comparably cheap: this parallels the well-known difficulties with nuclear civil defense, with anti-missile technologies, body armor, protection against improvised explosive devices (IEDs), and so on (Talbot 2010). Another feature that cyberwarfare shares with conventional warfare is the low degree of certainty of what will happen with an attack, or in a war, and that there are distant, deleterious side effects that cannot adequately be anticipated – disease, economic consequences, and so on.

My aim in this paper is to focus on the ethical issues surrounding cyberwarfare. I will examine how the widely accepted principles of the Just War Tradition might apply to cyberwarfare, but I will also consider deviations from or additions to this tradition that may be necessary either because of the inadequacy of these traditional principles or because of the sheer novelty of cyberwarfare. I have already signaled some of these hesitations. Traditional discussions of the morality of going to war, or morality in war, have been understandably motivated by the lethality and massive destructiveness of war. But cyberwarfare often will not be like that – although it *could be* both lethal and physically destructive.

I will begin by surveying some issues in the technology and tactics of cyberattacks, since these have ethical implications, and will also discuss issues of strategic policy. Second, I will turn to the separate domains of thinking about the morality of cyberwarfare and of how existing international law applies (or as I will argue, does not apply straightforwardly) to cyberwarfare. Then I will consider in more detail how a theory of morality in cyberwarfare needs to be augmented in two ways: In a *moral* dimension toward a general notion of harm, especially to national vital interests, and away from some of the key concepts in traditional Just War Theory, and in an *ontological* dimension, with a focus away from strictly injury to human beings and physical objects toward a notion of the (mal-)*functioning* of information systems, and the other systems (economic, communication, industrial production) that depend on them. Finally I will discuss what the future of cyberwarfare might be like and some possible remedies or ways we can mitigate its damage.

Cyberweapons: Technical Details and Policy Issues in their Use

Cyberattacks belong to a large genus of all kinds of attacks on information systems. Such attacks include traditional counterespionage and disinformation campaigns, old-fashioned destruction of telephone lines, jamming of radio signals, killing of carrier pigeons (FM 24-5, 1939: ch. 4), and so on. We can restrict ourselves here just to attacks on modern digital information systems, that is, on computers and computer systems: intentional damage to software, hardware, and the operations of information systems. Even the restriction to digital is not especially crucial. (Note that a restriction to 'electronic' information systems is also too narrow since both communication and computation may be performed by utilizing light, other electromagnetic radiation, or any phenomena that permits storage and switching.)

Although it has not been spelled out, most authors use 'cyberattack' to refer only to attacks on digital information systems via the Internet (or maybe, via any sort of network). The possibility exists of attacks on the software of modern information systems via other media (such as CDs of distributed operating systems, applications or data) or secretly incorporated in the BIOS of computers themselves that may be sold by one country to another. The BIOS is the Basic Input/Output System, part of most computers, a form of relatively simple firmware that survives powering down, and that any computer needs in order to know how to load an operating system from some volatile memory device. (Strictly speaking, the BIOS need not be stored in an actual hardware circuit, although most have been, but does need to be relatively non-volatile.)

There has long been discussion of incorporating unique identifying chips into the hardware of each individual computer, in order to reduce the secrecy-of-identity that has engendered much criminal and harmful activity on the Internet. A key cause of this activity is the 'Attribution Problem.' A chip, secretly built into the hardware of a computer or networked device, could also serve as a platform for cyberespionage or a cyberattack, that is, a hardware Trojan horse – see (Markoff 2009b). This approach has so far been widely rejected, but I will return later to the possibility that stronger, universal source identification may be necessary to limit the damage of cyberwarfare. It would be difficult now secretly to insert a chip because even brandname computers are manufactured in diverse countries; furthermore, large numbers of engineers have a hand in the design and testing of every chip and other component, and secrecy would be very difficult.

On the other hand, the worldwide dominance of several 'closed source' operating systems (OS), such as the Microsoft Windows and Apple Mac OS families, would theoretically be susceptible to the hidden incorporation by their creators of sections of the software that identify each copy, track usage, automatically send information to a company or security service, or allow external manipulation via the Internet (a 'back door'). Just such a concern motivated China in 2003 to sign with Microsoft an agreement to grant it access to all of the source code for the Windows operating system. The Chinese were concerned that the US security services, or the Microsoft Corporation, might have incorporated 'back doors' that allow secret access to any system that is running the operating system (Gao 2003). Such access had previously been given to the security agencies of NATO, Russia, and the UK. Although this might seem to be a positive step toward openness, sharing the source code for operating systems and complicated applications like Word, Acrobat, or Excel allows sophisticated analysts in a country like China, Russia, or the US to detect weaknesses in the code that can then be exploited via the Internet. One of the motivations for 'open source' operating systems, such as Linux, and for open source application programs, is to make a secret back door unlikely, since the many thousands of programmers worldwide with access to the source code are more likely to notice weaknesses and suggest fixes than are the hundreds that have access to a proprietary source code. In such large programs it is nevertheless inherently difficult to understand the purpose or exploitable weakness of every line or section of code.

Within the narrowed family of Internet attacks on modern digital information systems via the Internet, there are several very different forms that an attack can take. The classification of cyberattacks is complex, and one can classify cyberattacks along several dimensions. For example in Rowe and Crusty (2010) and Rowe (2006) we see an elaborate classification by the kind of deception that is perpetrated. Another dimension for classification might be the kind and amount of harm. One family of attacks may be termed 'unintrusive' cyberattacks, such as the Denial of Service (DoS) attack. In such an attack, an attacker does not actually gain access to the site. Instead, a site or a server (or even sections of the network) is bombarded by hundreds, or thousands or more, of spurious requests for information per second, so that it cannot successfully respond to them and thus is rendered inaccessible to its intended users. The attacker does not gain access to the internal working of software or hardware, and thus no damage is done to them.

The most successful such attacks in recent years have been Distributed Denial of Service (DDoS) attacks. A programmer designs a malicious piece of software (malware) that embeds itself in hundreds or thousands of computers all around the world. This turns the other computers into remotely controllable slaves, or botnets. Then, either at a pre-established time, or on command from the designer, the embedded malware bombards a targeted site with email or requests for responses. On the Internet, each stream or packet of information is identified by the unique address (the Internet Provider or IP address) from which it originates. In a simple Denial of Service attack, these pieces of information will have the same IP address (although they can be 'spoofed' – their real origin falsified and varied). In direct DoS attacks, the attacking message packets can be blocked, by the server or even at gateways and nodes in the Internet itself. In Distributed Denial of Service attacks, this technique will not work, since the requests to respond will be coming from many different IP addresses, often in different parts of the world.

Many of these techniques of DoS attacks and insertion of malware could be defeated by requiring a confirming dialogue between the source and target ('handshaking'), although the activity of filtering packets with suspicious IP addresses, and the initiation of confirming dialogues, may itself overwhelm the computational power of the target computer. Sites such as Google or sites within the US Department of Defense probably have the computational power to keep up with a huge number of such attacks in real time (for now). A smaller business website operating on a single server might not.

There are nevertheless different ways in which DDoS attacks too can be identified and blocked from disrupting legitimate Internet traffic, such as through telltale common features ('fingerprints' or 'DNA') of their contained messages. One can also track the path of the messages through the Internet, and identify the real source. These and other forensic methods of identifying the sources of DoS attacks and malware can be very expensive in terms of human and computer resources – and thus add to the cost for, and harm to, the attacked party.

The other major form of cyberattack is *intrusive*: malware gains access to sections of a computer's software or stored data through a site. Once there, the malware may alter various pieces of software or data, cause the system to crash or make certain pieces of software inoperable, erase hard-drives, send email messages pretending to be from the user, send information on the software and data back to the malware's author, and so on. If this intrusive malware can then be sent (perhaps in modified form) and infect still other computers, it is a self-replicating. If it is transmitted by attaching itself to another file, program, or email, it is a virus; if it is a free-standing program that can travel through information pathways, it is a worm. In some cases, such a virus does no harm; it is a prank, but its presence is still undesired. The ways a virus spreads are varied: some can, just like biological viruses, alter their own code, so that their outward appearance (equivalent to a biological virus's protein shell) is not immediately recognized as similar to the first versions of the virus. They can also be designed to 'blind' anti-virus software to their existence, or to prevent updates to anti-virus software.

An especially virulent but apparently non-destructive virus is Conficker; early attempts to eradicate it failed (Markoff 2009a). The damage viruses may do and the way they do it are highly varied. Intrusive malware that is not self-replicating and that is at first not easy to detect, or does not at first cause detectable damage (and might even be perceived to perform a useful function), is a 'Trojan (horse).'

Many forms of malware are strictly tools of espionage: they do not directly damage the targeted nation's information systems, or cause damage via these information systems. In this sense, they are not themselves cyberweapons. In such cases, it is only through the use of gathered information by an enemy that a nation's interest may be harmed. Traditionally, mere espionage has not been viewed as a *casus belli* (customary or legitimate reason for going to war), but may bring non-military retaliation, such as the expulsion of diplomats or the limiting of foreign aid or commerce. Consequently, I will not discuss pure cyberespionage further, even cyberespionage that is directed by one nation toward another, since it is not usually an activity that has been considered part of the moral considerations regarding going to war or conduct in war. The ethical considerations in espionage and other intelligence-gathering operations are but one of the several traditionally neglected aspects of the morality of war (although there has been growing interest in this field in the last few years).

Against a nation's military targets, there are essentially three categories for targets of cyberweapons.

First, the cyberweapons may target and impair a main function of military chains of command, namely command and control: communications and information gathering, as well as the communication of precise orders to maneuver, defend, or attack, may be harmed. Command and control data may be blocked, altered, or false reports and commands inserted. Most modern military organizations have hardened communication conduits (such as buried, shielded, or otherwise protected electrical or optical cables), elaborate schemes for encryption, and redundant systems.

Second, weapons or weapon systems can be rendered inoperable for a time or even physically sabotaged by faulty messages or intrusions into their controlling information systems; in the extreme case, they could be directed by an enemy to attack a false target. The most sophisticated computerized weapon systems, such as the Patriot and other anti-aircraft and anti-missile systems, and the Aegis system of the US Navy, could conceivably be rendered inoperable or directed to fire at false targets. These systems have already been proven fragile even in ordinary operation when they are not subject to cyberattacks: there have been friendly-fire and other targeting errors in the case of the Patriot anti-aircraft and anti-missile system in 1991 and 2003, and the disastrous shooting down of an Iranian civilian airliner by the Aegis Combat System (in 1988) aboard the US cruiser *Vincennes* that automates fire control of guns and missiles.

Finally, cyberweapons could target joint-use infrastructure, that is, systems and structures for both civilian and military uses, or even civilian targets with the goal of demoralizing, weakening, or confusing an enemy's military and civilian leadership. Joint-use infrastructure would include a wide variety of computer-guided systems, such as the satellite global positioning system (GPS) network, and energy, communication, water, or sanitation infrastructure, or petroleum and chemical refining systems. Targeting joint-use industries and systems in a lawful military conflict is generally permitted by international law and Just War Theory, but some authors have proposed additional criteria (Walzer 2006: 144–51). Targeting primarily civilian structures and networks is prohibited by international law and by almost all theories of morality in warfare.

The main aspects of chemical processing and energy production and transmission facilities, as well as communications networks, are now largely managed with the help of complex information systems using many modules of software. This software relays data and warnings to human operators, in some cases make decisions itself, and allows remote monitoring and control by engineers. Increasingly, at least in petroleum refining and chemical processing plants, key engineer supervisors are allowed access to the control systems via the Internet. Access by supervising engineers or military officers via the Internet would theoretically place near-instantaneous operational control of these facilities in the hands of the most informed and capable people, while they are at home or traveling. Access to the computer control chemical-processing facilities is typically highly protected, such as by layers of high-quality and frequently changed password protection, as well as accessibility only via a single, identifiable, company-owned computer. Likewise, access to secure military information – and presumably the ability to issue military commands – in the US Department of Defense is controlled by a user's Common Access Card (CAC), whose chip is encrypted and requires a computer to have a special piece of hardware with restricted distribution, the CAC Reader, in addition to the use of high-quality and frequently changed passwords. There have already been several generations of increasingly sophisticated hardware and encryption methods used in the CAC and CAC reader.

Nevertheless, in both industry and military, access via the Internet to this information and control is obviously susceptible to highly sophisticated hacking. The US Department of Defense, and each individual military service, has at least several layers of increasingly secure computer systems. Some of these systems are especially secure networks not immediately accessible via the Internet. The most sensitive such computers are constantly monitored for any unusual activity. This is termed cybersecurity 'defense in depth.' The danger to civilian banking and energy infrastructure has been widely noted in the media and in mass-distribution books, such as (Clarke and Knake 2010).

Public information, including General Keith Alexander's unclassified remarks before the Senate confirmation hearings of the Armed Services Subcommittee (US Senate 2010), indicate that there are hundreds of thousands of probes per hour of Department of Defense and NSA computer systems.

These probes include mostly harmless and unsuccessful attempts at getting past required passwords, or identifying a server's or system's operating system, metadata on its websites (such as the program with which the HTML code was written), etc. The percentage of those that probably originate with foreign governments is classified, as is what has been their level of success. (There are rumors that a majority of these attacks actually originate in the US). This does not include probes or attacks on other US government systems, such as in the Departments of Justice, State, and Homeland Security, or on other Western governments. Among the many officials I have heard or read, all, including General Alexander and the Senators and staff of the Armed Services Committee, are alarmed. (The one exception was a remark by Howard Schmidt (Singel 2009), but this was regarded as a major blunder.) There have been some spectacular recent successes at penetration, such as theft from a contractor of specifications, and thus weaknesses, of the F-35 Joint Strike Fighter.

Yet another peculiarity of cyberweapons is that, once intrusive malware is detected, there will usually be countermeasures and damage repair that can be executed in minutes, hours, or days by a technologically advanced user or country. The malware may be removed or quarantined, the system reloaded (reset, rebooted) with data and software from before the attack, and the incoming data stream can be filtered for this type of malware by looking for suspicious code sequences or IP source addresses. From the point of view of the offensive use of the cyberweapon, this makes many of them 'one-time use' weapons whose effectiveness will likely rapidly diminish. Few traditional weapons have had this characteristic. As an attacker, one will thus *not* want to release one's very best cyberweapon until there are occasions where risking the rapid development of enemy countermeasures might be worth it. An attacker will also probably want to develop offensive cyberweapons 'in depth,' that is, multiple cyberweapons (sometimes called cyber - or logical bombs) that are substantially different in the way they penetrate an enemy information system, and in their distinctive sequence of code, damage they cause, or the means by which they could be detected.

Another aspect of 'cyberweapons in depth' is that sustained, successful development and use of offensive cyberweapons might actually require three or more layers or kinds of attacking software:

1. Reconnaissance or espionage software to determine a system's nature and weaknesses, and to target it uniquely (rather than, say, also damaging civilian information systems).[3]
2. Damage-producing software that harmfully alters the behavior of the enemy's software or produces other undesired effects – possibly including deaths and damage to physical systems. These are the logic- or cyber-bombs themselves.
3. Damage-assessment espionage software that determines the effectiveness of the damage-producing software.

In some cases, reconnaissance and damage assessment might be achieved through traditional intelligence-gathering measures, although nothing would be as useful for future development or assessment of cyberweapons as details about how the damage-producing software actually affected the targeted information system. This is not usually going to be superficially apparent from human, electronic, or satellite surveillance.

Viewed from the point of view of limiting damage in the world, there are several positive and even unique features of cyberweapons. Most cyberweapons will probably fail to achieve their intended or expected degree of penetration and damage (Rowe 2009). In this they will resemble most mutations of biological microorganisms: bioweapons often lack the ability to spread, and harmful organisms are difficult to 'weaponize.'

Second, most cyberweapons will probably not be lethal. Even a massive electrical power failure would not immediately kill many (except those on life support and without backup power, for example).

Third, even if cyberweapons cause intended damage to military or even civilian information systems, they could be designed so that this damage could be reversible (ibid.). One could develop cyberweapons, together with countermeasures or ways of restoring operability or data, that one would eventually share with an enemy or release if an enemy's provocation ceases or if the cyberweapons accidentally cause damage to strictly civilian information systems. The only thing in traditional warfare that is remotely comparable to such reversible damage in cyberwarfare would be the joint production of chemical weapons and antidotes to them.

The Morality of Cyberwarfare: The Easy and Hard Cases

The moral questions of cyberwarfare divide themselves in this way:

1. Is a cyberattack ever morally justified in response to an enemy *conventional* attack?
2. Is a cyberattack ever morally justified in response to an enemy *cyberattack*?
3. Is a *conventional* attack ever morally justified by an enemy *cyberattack*?
4. Is a cyberattack ever morally justified in cases where the enemy has launched neither a cyber- nor a conventional attack? (With United Nations sanction, preemptively, preventively, or for some other reason.)
5. Once a war (cyber- or conventional) has begun what kinds of cyberattacks are morally justified?

Of the five questions, (1) and (2) would seem to be easy questions, subject to the restraints that might be posed by the issues in (5). A cyberattack by nation A on B would seem *prima facie* to justify morally a retaliatory cyberattack by B on A, and one at least as destructive as A's first strike. Likewise, a cyberattack in response to a conventional attack would seem to be at first sight justified, again subject to the issues raised in (5).[4] There may also be other *jus ad bellum* criteria that need to be considered, such as Proportionality and Likelihood of Success, but there do not seem to be any issues here that are peculiar to cyberwarfare.

Cases (3) and (4) are the hard cases. Case (4), such as the moral permissibility of a 'preventive' first-strike, shares many of the same issues and arguments as does a discussion about a non-retaliatory conventional attack. We have just gone through nearly a decade of discussion of the conventional case – the moral issues in the coalition's 'preemptive' attack on Iraq in 2003 – with remarkably little light being shed on the issue. The case of a cyber first-strike is, however, more problematic, because it need not involve widespread death and permanent destruction of physical installations. If a preventative 'attack' is ever justified, then it is likely to be something like a cyberattack, that can be free of the usual death and destruction of traditional forms of warfare.

Cases (2) and (3), being responses to cyberattacks, involve us necessarily in what was earlier called the Attribution Problem. More broadly these are *epistemic* problems that have been ignored by most theorists of the morality of war; namely, how much justification or evidence is necessary concerning the threshold conditions for morally going to war? We can distinguish here between the metaphysical conditions for a moral war, and the epistemic conditions: one is behaving morally if one objectively has extensive evidence that all the necessary conditions for going to war are held, even if it turns out one did not hold. This epistemic uncertainty is one of the peculiarities of responding to a cyberattack that makes it similar to preemptive and preventive war. In the case of a cyberattack, the problem is uncertainty about *who* attacked us, and in the case of preemptive and preventive war, the 'who' is known, but there is uncertainty about *whether* an enemy would indeed, in the future, attack us.

Can we do nothing at all if a cyberattack cripples our industrial, military, and governmental systems when there is some uncertainty about who did it? In another paper (Dipert 2006b) I argued, on game-theoretic grounds, that a preemptive attack can be morally justified if the evidence exceeds a certain threshold of objective likelihood (roughly 90 percent) and if there will be a high level of expected damage to us if we do not preemptively attack. I argued, for consequentialist reasons, that if every state followed such a policy, overall damage to all parties over the long run would be minimized because of a deterrent effect. A similar calculation would seem to hold in the case of a cyberattack when the attacker is not certain: if the expected damage is very large, and it is highly likely that we can identify the attacker, and can likely eliminate the threat or deter others in the future, then we might be morally justified even without certainty in our identification of the attacker. Admittedly, a great deal of moral weight lies on the extent to which we can objectively measure the likelihood of identifying an attacker and in objectively estimating the damage to us if we do not stop this attack, or deter others like it in the future.[5]

I believe Just War Theory, and much – but not all – recent theorizing about the morality of war relies upon rickety meta-ethical foundations. The origins of Just War Theory are to be found, in St Thomas Aquinas (1225–74) and Hugo Grotius (1583–1645), in natural law theory. Yet few contemporary philosophers have a metaphysics that would allow them to maintain such a view. It is also true that Just War Theory, having been endorsed by most industrial democracies and in international law, has also acquired the status of damage-minimizing convention. However, the increasing number of nations, especially non-Western ones, who show no serious effort to endorse or follow this convention – and the unwillingness of other nations to *force* compliance – means that the advantage of a widely accepted convention is lost; it merely handicaps nations with a developed public sense of morality and prevents them from moral intervention.

The only meta-ethical route left to many theorists is one based upon ‘intuitions,’ legitimized by the ‘reflective equilibrium’ of John Rawls (1921–2002) and by contemporary ‘experimental philosophy’; but it is unclear how these intuitions arose and whether they are reliable – whether, for example, they are not inter-individual ethical intuitions that are vaguely applied to states because of a lack of developed international-political modes of thinking, that is, Michael Walzer’s ‘domestic analogy.’

The meta-ethical view I am tentatively employing is realistic, diachronic, universalizable, enforceable consequentialism for states. It is ‘realistic’ not in the sense of dismissing moral thinking (Christopher 1999) but in recognizing that the condition of conflict between societies, as it has been for hundreds (if not thousands) of years, is exacerbated by communication and mobility, and is likely to remain for some time. It is universalizable in the sense that were a policy practiced by all states, then they would be markedly better off than they have been. It is enforceable or ‘compellable’ in that the practice of this policy would in fact compel other states to follow a similar policy, rather than encouraging free-riders to profit by disregarding it. Universalizability and what I am calling ‘enforceability’ are related to the desiderata that Robert Axelrod proposed as components of an optimizing strategy (which he proposed was achieved by tit-for-tat).[6] Finally, it is diachronic insofar as it does not look at what is best for this or that state right now, but what is attainable and optimizing in the long run (it is temporally universalizable).

Case (3) is a very hard case indeed: could a cyberattack, perhaps without deaths and permanent destruction, wreak such unjustified harm on a nation that it is morally permissible to launch a conventional attack, which will involve lethality and physical destruction? What if that is the only way of stopping this or future attacks? I will return to these cases.

Cyberwarfare and Traditional Just War Theory

As of the writing of this essay there are two papers mainly about the ethics of cyberwarfare: Arguilla (1999) and Rowe (2009), and the similar Rowe (2010); the writers were not experts in the morality of war, but experts in computer science. A third, book-length monograph was written in consultation with experts on the morality of war (Owens et al. 2009: xiii), but is largely a compendium of existing law. Two characteristics of these approaches are the following: (1) they regard Just War Theory as a settled body of opinion on the morality of war and (2) they accept or argue for the view that Just War

Theory is straightforwardly applicable to cyberwarfare, without the need for modification or addition. I argue that both of these claims need to be questioned.

It is a separate question whether existing national laws and international law are applicable to cyberwarfare, and whether they are sufficient. Laws (statutes, agreements, treaties, etc.) rarely coincide with permitting all of what is morally permissible, and outlawing all of what is morally impermissible. Although it is not my primary focus – nor is it my area of expertise – I hazard some guesses about the legal situation. Roughly, I believe that the legal status of cyberwarfare parallels the situation concerning the morality of cyberwarfare: most legal frameworks do not clearly apply to many instances of cyberwarfare, and cyberwarfare involves aspects of damage or harm that are typically not addressed by law, such as harm to the functioning of information and other systems that might not harm physical objects or persons.

The traditional theory of what is morally permissible in war, Just War Theory, is usually divided into two main questions. First, there is the question of when a country may morally take part in, or begin, a war: *jus ad bellum*. Second, once one finds oneself at war, there is the question about how one may morally fight that war: *jus in bello*. Excellent accounts of traditional theories can be found in Orend (2005), Fotion (2007), and Christopher (1999). The best available account in English of historical thinking on these issues is in Reichberg et al. (2006). The most informed account from the point of view of the recent history of warfare is Walzer (2006). Each of these two aspects of the morality of war have criteria for moral adequacy that are widely (although not universally) accepted by both major past and contemporary thinkers about moral issues of war.

For *jus ad bellum* the criteria are (1) Just Cause, (2) Last Resort, (3) Likelihood of Success, (4) Proportionality, (5) Proper Authority, and (6) Right Intention. Just Cause is regarded as by far the least controversial component of Just War Theory. A Just War Theory that abandoned or extensively modified the traditional understanding of *casus belli* would simply not be Just War Theory as we know it. Consequently, I will focus my remarks on Just Cause. I will make remarks about the other *jus ad bellum* and *jus in bello* criteria in passing.

As Michael Walzer notes in *Just and Unjust Wars* (Walzer 2006: 51–52) the key necessary condition of morally permissible war, Just Cause, is rather blandly and uniformly described as 'aggression' or 'attack' by an enemy. In the United Nations (UN) Charter, what is prohibited in order to maintain peace is broadly characterized as the 'threat or use of force against the territorial integrity or political independence of any state' (UN Charter 1945: Ch. 1, Article 2, Section 4). Later, an 'armed attack' is the only stated condition in the UN Charter under which a nation may justifiably defend itself before the Security Council takes action (UN Charter 1945: Ch. 7, Article 51). Observe that this seems, literally understood, to designate soldiers using 'arms,' roughly, as artifacts for inflicting injury, death, or causing physical destruction of objects.

It is thus a stretch to consider a cyberattack an 'armed' attack, since the artifact doing the damage is a computer (designed for other purposes), or, still more abstractly, an information-theoretic entity. The most prominent and large-scale work on the topic (Owens et al. 2009) considers a cyberattack to be a straightforward instance of an armed attack. This is not obvious. The paradigmatic historical form of aggression or attack is the invasion of the sovereign territory by armed, centrally commanded, enemy soldiers of another state who are prepared to use deadly force. Sometimes this invasion may be momentary, such as at Pearl Harbor in 1941, and the primary harm is extensive death or destruction.

This view of aggression and attack as invasion or destruction is retained by most writers who roughly follow the just war tradition, such as Fotion (2007). The language of 'attack' may be broad and vague enough to include non-paradigm cases of aggression I am about to discuss, although Fotion and most authors give examples only of violent, armed invasions. Yet 'customary' and accepted just causes of war have included a wider array of phenomena that are broadly categorized as *casus belli*: embargos, systematic attacks on and harassment of citizens and businesses abroad, blockades, blocking of necessary supply lines such as a pipeline, attacks on military or civilian ships in international waters, and so on. A cyberattack by one nation on another more resembles one of these. Martin Libicki (Libicki 2009: 179f) notes that NATO explicitly rejected the claim that Russia's cyberattack on Estonia was an act of war that would trigger mutual defense obligations. The National Research Council study (Owens et al. 2009: 257–258) somewhat unsatisfactorily separates the concepts of sanctions and blockade, the latter constituting 'just cause', the former not, regardless of the degree of harm: this makes too much of a legal (not a moral) distinction. Surely sanctions *could* be morally – but not legally – unjust, even if directed by the Security Council or by a regional international organization. Although it was clearly used as a pretext for large-scale war, the obstruction of the movements of goods and people between East Prussia and the main body of Germany in the 1930s (the Danzig Corridor) could reasonably have led to justified acts of war. Likewise the isolation of Berlin from the West by land routes (the Berlin Blockade) could have led to a justified war if the airlift had not been successful. In diplomatic language, and as a warning that these actions might risk war, such events are broadly conceived as intentional damage to *vital* interests of a state. Most discussions of *jus ad bellum* have taken scant notice of them. An interesting exception is Walzer (2006: xiv–xviii), who in the preface to the 4th edition of *Just and Unjust Wars*, calls them 'force-short-of-war,' and notes that it is force and thus morally prohibited nonetheless.

Attacks on another nation that kill or destroy buildings or other artifacts, but that do not include invasion by enemy soldiers, cover a broad range of activity, from small arms fire and artillery fired across a border, to the use of mortars and rockets. Cruise missile attacks or attacks by unmanned aircraft would seem to count as aggression that would, in certain circumstances, provide just cause for a wider counterattack. Even these actions still are traditional in the sense that they involve macro-physical objects that are intentionally directed to intrude into the airspace of another nation and cause physical destruction. One study (Owens et al. 2009: 239f) classifies these as kinetic projectile weapons, placing cyberattack in the category of non-kinetic weapons.

An unprovoked cyberattack by one nation on the civilian or military infrastructure of another nation is thus not very much like traditional, paradigmatic forms of aggression or attack. A cyberattack does not involve intrusions into the territory or airspace by soldiers or even by physical objects. A better analogy would be that a cyberattack is more like Electronic Warfare (EW), such as jamming the radio communications of another nation from outside its borders, or the intentional use of electromagnetic radiation, such as a laser, or an electromagnetic pulse (EMP) weapon to destroy or hinder the functioning of human beings, machinery, or infrastructure from beyond a nation's borders. Exactly this point was noted by General Keith Alexander (2007), who since has become the Commander of US Cybercommand (USCYBERCOM). The field manuals on electronic warfare and information operations (IO) of various military services of several countries have guidelines and policies that consider legal and ethical issues. In the case of radio jamming or a laser or maser attack, there is a nominal sense in which the photons 'invade' the airspace of another nation, but that in itself seems harmless; foreign radio waves constantly pass through nations' airspaces without complaint. In the case of EMP weapons, it is not the electromagnetic radiation, the photons themselves, that violate a nation's sovereignty and cause the ethical problem; it is the secondary production of magnetic fluctuations, and then induced electrical current, that is the problem.

Neither real historical examples nor Just War Theory gives us much direct help in thinking about such matters. Yet we can imagine a cyberattack that caused as much or more damage to a nation's infrastructure and institutions as would more conventional attacks by bombs or artillery. The further physical damage could have been intentional or not, and foreseeable or not. We can even imagine a disruption caused by a cyberattack killing large numbers of human beings – by damaging or maliciously 'invading' the software of a large nuclear reactor, medical information systems, navigation systems, or of passenger aircraft, for example.

The Search for a Moral Aspect of War beyond Violation of Sovereignty: Cyberharm

The broadest useful notion in the discussion of cyber-ethics is *intentional cyberharm*: this is intentional harm caused by an agent, *via* an informatics network such as the Internet, in which the functioning of a system (a person, a machine, software or an economy) is in some way impaired or degraded. An attack is intentional cyberharm to a specific system. A virus playfully set free on the Internet by mischievous hackers is not in this sense an attack, although it might do harm. An attack intentionally causes harm to a specific organization, system, etc. Another target of a cyberattack might be the functioning of an artifact, such as vehicle (say, by disabling it through corrupting the OnStar system), or the healthy functioning of a person (by disabling their pacemaker, and so on). The vehicle or person would be the indirect target of a cyberattack, since by its nature the direct attack is specifically on the functioning of software, etc., and it is the malfunctioning of these systems that causes the intended harm to a person, organism, or artifact.[7]

I focus on cyberattacks (intentional cyberharms) that are instigated or controlled by political organizations (or their military services) on other political organizations or military services. (The usual cases are of states, including nation-states, that have internationally acknowledged sovereignty over a territory.) I use the locution 'political organization' to include insurgent and rebel groups with political goals, as well as political organizations that are not localized to a territory – such as international terrorist organizations. Libicki (2006: 23) puts it this way: '*cyberattack*, for the purposes of this discussion, is the deliberate disruption or corruption by one state of a system of interest to another state.') These are *international cyberattacks.* If the attacks between political entities are sufficiently 'widespread' we might then speak of a cyber*war*. This modifies the useful definition of Brian Orend of 'war' as '*actual, intentional* and *widespread* armed conflict between political communities' (Orend 2005). I would further stipulate that war in its usual sense involves the intentional use of deadly force on human beings. A cyberwar might then not literally be a war in this stricter sense, unless death or severe injury of human beings was the further intended result.

So far as I can see, there are no serious concerns that restrict what a nation may morally take as strictly *defensive* measures to prevent nuisance harm, or to block cyberattacks. However, defensive cybersecurity efforts could violate the privacy or other civil rights of innocent non-state parties, or incidentally cause damage to one's own citizens, economy or computer systems. Defensive actions against cyberattacks could include blocking a system's or nation's Internet from certain foreign or hostile IP addresses and the physical severing of all information conduits (satellite, cable, radio, and so on), when this is possible. As for cyberespionage, there seems to be for now the general understanding that the only possible moral countermeasures are cyber-defenses, including retaliatory cyberespionage. This toleration may change as cyber-espionage and accompanying threats and harm become more serious, and might in the future include diplomatic, economic, and cyber-retaliation.[8]

Another way of categorizing cyberharm uses a (much debated) distinction in computer science, between *data* that an algorithm operates on, and the *algorithm*, a computer or network might be running. There are at least four kinds of entities: the data, and the algorithm (both information entities), the running of the algorithm on the data (an event or process), and the hardware in which this is taking place. *Theft* of data occurs when an unauthorized user gains access to private data – such as schematics of nuclear weapon design or social security numbers, which can then be used to harm the interests of the data's rightful owner. Government-on-government computer espionage falls into this category. These are sometimes casually described as 'attacks'[9]; because they are usually unsuccessful and do no harm; a better word would be that they are hostile probes. I do not consider such espionage a cyberattack, although the further use of information gained to commit cyberharm (or conventional harm) would be.

Martin Libicki (2006: 102f) lumps espionage into a category of mere 'computer network exploitation.' He considers at length whether a cyber-attack might be launched as retaliation for espionage and his answer is, for a variety of interesting reasons, no. Successful attacks require getting past IP-address confirmation exchanges, password protection, and encryption of the data (if any). Countermeasures would include adding further layers of security or, in the worst case, making the data inaccessible altogether from the Internet. *Corruption* of data is perhaps a worse problem, since the changes to the data might be subtle: imagine corruption of a list of the GPS coordinates of all of a hostile nation's fixed targets, or changing bank account data so that it skims pennies from a large number of bank accounts.

Another form of intentional cyberharm involves changing the programs – the algorithms – of a target computer system (or the algorithm of an application program, such as a spreadsheet or inference engine). This may involve appearing to make updates to software (most seriously in the operating system), which in fact causes the software to cease to function, or function in a way undesired by the regular users. This is the realm of the more usual computer viruses. Notice that the effect of changing a program may be to allow easier access to data, which is then stolen or corrupted and then falls into an earlier category.

Returning again briefly to international law, it would appear that these useful ideas of harm (to vital interests) and functioning of systems are poorly addressed. If we imagine that a real war has begun, the only traditionally permissible attacks that damage civilian infrastructure are those whose purpose is to impair an enemy's military operations, and which are militarily necessary to pursue the just goals of a war. This is usually taken to include key components of the electrical grid, for example if these are necessary to support the enemy's military communication and command and control. It may also include components of the transportation infrastructure, such as roads and bridges, airports, and so on, if these have likely military value.

The likely effects of damage to civilian infrastructure with regard to health is especially important, so that in addition to sites such as civilian or military hospitals, which have long been explicitly addressed by international law, this moral principle would seem also to protect water supplies. Indeed, in Protocol I (1977) to the Geneva Conventions, civilian water supply is explicitly protected.[10] The US, in the First Gulf War, and possibly in the Second (Operation 'Iraqi Freedom'), has been accused of intentionally damaging sections of the largely civilian water supply system of Iraq; the matter is complicated, however, since joint military and civilian use is not carefully addressed by the statute and the overall intent of the statute seems to be to prohibit such damage that is used to drive a population out of an area or destruction of objects that are 'indispensable to [human, civilian] survival' rather than merely conducive to good health.

The extensive Protocol I to the Geneva Conventions that covers this general type of damage, seems to be flawed in a way that one can describe as ontological, and thus renders it fundamentally inapplicable to cyberwarfare and certain software- or data- forms of damage to civilian infrastructure.

Namely, Chapter III (Articles 52–56) is concerned with what are there called 'civilian objects.' These are limited, in language and by examples, to material entities such as cultural and religious objects (Art. 53), the environment (Art. 55), the just discussed category (Art. 54) of 'objects indispensable to the survival of the civilian population,' and to '[public] works and installations containing dangerous forces' (Art. 56), such as nuclear (radiation) and dike and dam facilities (flood).

As an introduction to Chapter III there is a general ban on a residual class of non-military objects that are called 'civilian objects.' There are several limitations of this treaty that render it of little application to cyberwarfare, and indeed, to economic warfare. One is that it discusses only (material) *objects*, and not the *functioning* of these objects. This failure is ontological insofar as severe damage to software, data, and operating and control systems does not require the damage or destruction of objects in the usual sense. There are other flaws as well. First, there is an emphasis on prohibiting damage to these entities as military objectives, that is, of intentional attack; this leaves open a wide category of possibly severe incidental damage. Second, the key concept of targetable 'civilian object' is defined very narrowly as those objects that do not '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.'

Special Attributes of Cyberwarfare and Proposed *Jus ad cyber-bellum* Criteria

What I earlier called the 'hard [moral] cases' in cyberwarfare can be answered through analogical reasoning, comparing cyberwarfare with morally similar cases about which we have clearer thinking, and by looking to the ethical foundation that grounds what is wrong with one nation harming another nation in any way. My proposal for the jointly sufficient, and paradigmatic, conditions such that nation B morally may counterattack nation C with a cyber- or conventional attack, after a cyberattack by C, are:

1. The attack of C on B was unjust and substantial.
2. The source of the attack by C was, with overwhelming likelihood, ordered or permitted at the highest levels of a government.
3. Reasonable measures had been taken by nation B to defeat or minimize the cyberharm that a hostile nation or other non-state cyberattacker (black or gray hat hackers) might cause.
4. The expected damage to the enemy (C) is likely to be commensurate to the damage B has suffered, or is the minimum necessary to stop continuing cyberattacks.

Additional conditions would undoubtedly be needed, but they would be shared with whatever criteria one has for morally justified non-cyber wars. This is not to say that minor attacks never permit counterattack, nor that origination with a central governmental or military authority is absolutely necessary.

On (1): if the damage intentionally inflicted on B is primarily to its offensive capability, then this is for ethical purposes insubstantial damage. If damage is done to defensive capability or civilian infrastructure then it should be measured by the amount of increased vulnerability (in the case of harm to defensive installations, such as radar systems) or damage to the civilian population in which measurement takes place in this order: human lives, human well-being, material infrastructure. Thus a cyberattack on the Federal Aviation Authority flight data systems that foreseeably resulted in the crash of several airliners and deaths of hundreds of people, would probably exceed the threshold necessary to launch a conventional attack. Likewise, a massive cyberattack on defenses a nation has against physical attack (such as radar, spy satellites, command and control systems), would risk giving the attacked nation reason to believe that a conventional attack was imminent, and thus possibly trigger the conditions of justified preemptive war (in the technical sense of an attack being imminent and highly likely, see Walzer (2006) and Dipert (2006b)). It is often difficult sharply to distinguish defensive from offensive units, weapons, and installations – such as in command and control systems, including communication.

On (2): in conventional warfare, the identity of the attacker is not usually problematic. In cyberwarfare, identifying the attacker *is* especially problematic – due to problems of identification of IP addresses and the presence of diverse non-state agents with similar capabilities (especially clever black-hat hackers or groups of them). When information technology is concentrated or carefully controlled by a government, as it is in, say, North Korea, then if the IP nationality can be determined with great likelihood, the source is very likely to be the government itself. Likewise the presence of a large, indigenous community of black-hat hackers may give a nation's capacity for offensive military cyberattacks some sort of smokescreen. The enemy cannot determine if the source is governmental or military, or corporate or simply mischievous and anarchic.

On (3): this is an unusual condition, without clear analogy in the case of conventional warfare. We would not normally consider, for moral assessment, whether the attacked nation (B) had built enough bomb shelters or had failed to develop countermeasures to rocket or mortar attacks. Some condition like this nevertheless seems required in the case of cyberwarfare because civilians are in possession of tools that are as destructive as those nations can organize, namely computer programming skills and Internet access, and because one can expect occasional attacks or mischief from them. Civilians do not have tanks.

In my presentations, (3) has proven extremely controversial – and with some justification. My defense of it is analogical, following a mode of ethical reasoning described eloquently as 'Wittgensteinian ethics' in Stroll (1996 and 1998).

The Internet is, to use a quaint analogy, a very polluted stream. It is known to be very polluted with spam, unwanted popups, phishing scams, phony and defaced websites, and so on. We nevertheless continue to use it. We 'drink' from it informatically and 'swim' in it (using it for emails and our own posts), because it offers remarkable access to data and communication, as well as storage and computational power. Yet we are people – and government agencies – who drink from and swim in polluted water. It would be silly to drink large quantities of untreated water from a lake, surrounded by cottages, knowing many have faulty sewage systems that are slowly leaking into the lake, and then become sick, and sue the other cottage owners. A known level of danger requires us to take reasonable precautions – especially since our use of the Internet at all times is not necessary for survival. If our use of the water is especially critical, such as in giving water to infants, then we should take even more precautions: boil the water, etc. A relevant consideration in this analogy is that cyberwarfare is a case of someone deliberately polluting the water (so to speak) to harm us. In a civil suit against such an intentional polluter, I might indeed win a civil suit (or a criminal complaint), but the amount of the settlement would almost certainly be reduced by my failure to take reasonable precautions. This is radically unlike, say, the Katyusha rockets falling on Israeli territory from Gaza – there is not a widespread presumption of a falling rocket danger – and is more like acid rain or other background pollutants.

On (4): we have not yet witnessed (so far as I am aware) a case where sufficient damage was done in order to justify a large-scale conventional attack. It would likely take a large number of deaths or the irreversible crippling of an economy or vital economic sector. Although I hesitate to give terrorists a roadmap, such an attack might be on the control systems of airliners, of nuclear power plants, or of highly automated weapons, damaging large medical information networks, infecting large numbers of individual computer systems that control cars, medical instruments, and so on. Even then, the likelihood of stopping such a destructive cyberattack by a conventional attack may be small, since the source might not physically be located in a small number of sites. Although this is more a matter of the general theory of the justification of wars, I believe that the understanding of 'Likelihood of Success' is not sufficiently cognizant of game-theoretic considerations: goals may be the modification of this and other enemies' behavior in the future, through punishment and deterrence.[11] There is a contradiction between Just War Theory and wise geopolitical thought. The former, endorsed by a majority of the most intelligent moral thinkers, rejects disproportionate attacks (or threats thereof), and wars without a chance of success. Yet a majority of intelligent geopolitical experts, historians, and game theorists, found precisely this gambit – in a nuclear deterrent strategy – to be the best available option. Traditional Just War Theory thus has difficulties both with soft force and with the hardest of force.

A Coming Cyber Cold War?

It now appears that there were concrete proposals for a cyberattack on Iraq in 2003, before the physical attack began (Markoff & Shanker 2009). However, a more likely scenario for the future is that there will be cyberwarfare 'skirmishing' among the major players, perhaps lasting decades. This would take the form of aggressive Internet probing of military and industrial secrets, and perhaps numerous Denial-of-Service attacks, and corruption of data and software, especially if the origins of the attack can be somewhat disguised (China, the US) or if the government is not sensitive to world opinion (N. Korea). Any weakening of industrial or military power of other nations through cyberattacks would likely enhance a nation's own geopolitical power.

In short, what we are likely to see in the next years, perhaps decades, is something like the Cold War between the West and the Soviet Union. The espionage 'cat and mouse games' of the Cold War are well known, and there was also extensive probing of each other's territorial defenses, by the incursion of small numbers of air, sea, and ground forces, never giving sufficient reason to believe that a large-scale attack was imminent. What we are likely to see is the informal development of a similar 'equilibrium' in the accepted quantity and seriousness of cyberattacks.

It is relatively clear what the reasonable (and thus moral) constraints on a Cyber Cold War would be. There should be little targeting of strictly defensive computer control systems. There should be no attacks that disable or panic global financial or economic systems. There should be no interference in the vital economic and security interests of a major power, especially, as a practical matter, one with the power to attack with conventional physical force. (In Libicki (2006: 181) there is, a chart 'Ranking Various Forms of Harm in Cyberspace.' The top most harmful categories include casualties and, still more harmful, 'Interfere with Nuclear Systems' and 'Affect Military Operations.' This is far too unspecific, since disrupting a training operation, even seriously, is not on a par with blinding radar or rendering defensive electronic systems inoperative.)

Obviously, the most desirable ethical principle would be not ever to harm, through cyberattacks, the interests of other peoples. Since it is likely that many global players (industrial and governmental) will not abide by such a principle, then the 'containment' of harm seems to urge the development of defensive and even deterrent offensive cyberwarfare capacity by any nation whose information infrastructure is large and sensitive to harm. The offensive capacity is necessary in case – as is likely – defensive cybersecurity efforts are not sufficiently successful (or too expensive), and retaliatory cyberattacks are thus necessary to achieve some equilibrium.

There are some marked dissimilarities with the Cold War, however. First, a Cyber Cold War would be multilateral rather than bilateral: it would involve many nations, with different interests and not allied by treaty. Furthermore, the parties would include major non-governmental players such as private companies or even individuals or groups of individual hackers, perhaps with political interests. It is unlikely, in the more capitalistic and constitutionally free countries, that national governments can easily rein in these potential corporate and individual cyberattackers. Second, even if a nation's interests are attacked, it will often be difficult to determine immediately which country or organization is the culprit. For example, a harmful cyber-event may be the result of an organized attack by the government of Russia, by rogue elements in the Russian military, by groups of computer attackers tolerated by the government of Russia, by cyberattackers controlled by large criminal syndicates, by organizations supported by Russian oligarchs or corporations, or by individual hackers, political or apolitical. (The example of Russia is not chosen because I believe it to be especially nefarious, but it provides a wide and colorful cast of characters.)

Finally, computer technology is not dependent on any single controllable bit of technological knowledge (such as was possessed by a small group of nuclear physicists c. 1948), or on physical substances (such as U-235). It is remarkable, for example, how cryptological techniques and skill have spread to a wider public. Cyberattack technology is more like an idea than like a physical thing (or person). These facts would seem to make the creation of international treaties governing cyberattacks between governments, and laws within sovereign territories, extraordinarily difficult to develop, verify, and enforce. It is likely for a long time to be a brutish, if highly informed and non-physical, constrained combat between defensive and offensive cyberoperations. It will probably increasingly require greater and greater allocations of money and human resources.

How can the damage from cyberwarfare – and cyberattacks of all sorts – be eliminated or at least mitigated? The Holy Grail would undoubtedly be a solution to the 'Attribution Problem.' However, such a solution would run afoul of the relative anonymity and privacy that Internet access seems to offer, and to which users (in the West) have become accustomed. The alternative is certain identification of every user and every message. Even if users who have inflicted harm within one state could be identified, then if they are in another jurisdiction, and protected by another state, existing legal and diplomatic remedies will be neither quick nor easy. In the case of states with quite different juridical and moral systems, and distinct state interests, the problem is not now even remotely tractable. Progress is extraordinarily slow and impeded by political interests: witness European cybersecurity policy: 'The European Commission in [the] EU [European Union] is planning to establish a European rapid response system for cyber attacks, and may present an EU Internal Security Strategy in October 2010' (CybercrimeLaw 2010) and 'US asks EU to up cyber security' (Euroactive 2010), and this only describes political issues in the North Atlantic world.

Most domestic DDoS and malware attacks (perhaps working as bots for foreign agencies) could be blocked by requiring Internet Service Providers (ISPs) to require that users have recently employed malware-detection software before they are permitted to connect. This information, via an encrypted handshake with the ISP, would block DDoS attacks and severely limit foreign malware attacks. It could be strengthened if detection of foreign DDoS and malware (and messages through the domestic gateways as allowed by the US Foreign Intelligence Surveillance Act court) were reported to a central authority, with the legal proviso that this could not be used in domestic criminal investigations.

I would argue that this may eventually be required by the prerogatives of foreign and military policy within the US Executive Branch, and the foreign policies of other governments. It makes a mess of the distinction in the US between warmaking (Title 10), intelligence-gathering (Title 50), and law enforcment authorities; namely, in order to have an effective military and diplomatic policy, domestic law enforcement may need to identify and rein in domestic attacks on foreign information systems. If these attackers are sophisticated, or organized, then there is a risk that their efforts – whatever their goals – would be confused by foreign powers and corporations with centrally authorized attacks by the US government, and thus be grounds for counterattack and even war. The 'attribution problem' will thus have curious domestic consequences, not just hobbling our knowledge of our foreign attackers, but requiring us to identify and take action against domestic attackers on foreign entities. Even if there is not a suspicion of a government-directed and organized attack, foreign powers could still complain about – and retaliate against – our state toleration of these international cyber-outlaws and polluters.

Conclusions

On close examination, cyberwarfare appears to be almost entirely unaddressed by the traditional morality and laws of war. Because of a traditional emphasis on damage to human lives and material objects, there are not even clear restrictions on 'soft- or cyber-' damage that would leave wholly civilian targets, necessary for the well-being of the population, inoperable for long periods of time but not, in the strictest sense, damage them as objects. In some rare cases, the morally permissible response to a cyberattack may be a harsher retaliatory cyberattack, or even a conventional attack. As noted elsewhere (Owens et al. 1999: 41f) the requirement of proportionality in response to aggression does not require counterattack-in-kind, especially if it would not redress the harm done, diminish the threat, or deter future attack. This seems to open a window in which a morally permissible proper response to a damaging cyberattack is sometimes conventional, physical action.

The relevant entities in cyberwarfare are so unusual in comparison with the ordinary objects of daily life that the only useful way of thinking about them is by analogy. The relevant cyberentities include such things as the functioning of a system, software, and more broadly, information entities. Consequently, moral reasoning about cyberwarfare requires either the consideration of analogies with more traditional moral problems, or broader, less traditionally legalistic, and clearer moral theories capable of application to all possible ethical events and states of affairs.

It is perhaps understandable that the traditional morality of war, and laws of war, would not want to address such vague notions as the welfare or well-being of the civilian population, and harm to them. The most obvious and undebatable damage of war is on human beings as organisms: they die. Nevertheless, a cyberattack may do such extensive damage to the well-being of a populace, and to the functioning of a government, that it would satisfy the *casus belli* (just cause) requirement of reasonable criteria for morally going to war. Likewise, there would seem to be a need for additional moral reasoning, and additions to international law, such that militarily unnecessary damage to non-objects, namely, the functioning of civilian informatic systems, is limited in time of war.

There is widespread agreement (Libicki, Owens et al., and myself) that something like Cyber-arms Control treaties will not be forthcoming and perhaps even would be impossible to enforce. Others have advocated them (Clarke & Knake 2010), seemingly without a full understanding of the difficulties (including in verification).. Cyberarms are very unlike conventional, nuclear, biological, and chemical (NBC) arms (Owens et al. 2009: 293). Part of the problem is epistemic (attribution), but part is also due to the fact that cyberattack tools – unlike NBC weapons – are already held by diverse individuals and non-state actors. Libicki and I tend to believe that the use or threat of cyberattacks as deterrence is morally permissible and even desirable as policy. The National Academies of Science report (Owens et al. 2009: 41 and ch. 9) is less sanguine about the emergence of a deterrent effect.

In writing this paper, three serious worries have occurred to me. First, there is a large array of possible scenarios involving cyberattacks for which there do not exist obvious moral reasoning, or even straightforward analogies, that could guide us. Second, I am disturbed by the extent to which, through easy Internet access, much of our economic and defense informatics infrastructure is vulnerable to attack. I am myself engaged in research in developing systems that require regular access to open information sources (notably a variety of ontology and other web-based resources) that would not be available for long periods if a large-scale cyberwar erupted. I have long been disturbed by the departure from the relatively secure ARPANET (Advanced Research Projects Agency Network) for use in defense applications in the 1980s to a wide-open Internet, that does not have a single or identifiable handful of portals. Third, it is clear especially from General Alexander's comments (notably his 2007 paper) that serious thought is being devoted to the development of policy and strategy. To date it has remained largely shrouded in secrecy. This will be, and maybe even now may be, a serious problem, since the making public of many elements of policy are absolutely required for a deterrent effect.

Finally, it is interesting that, as far as I can see, traditional ethical and political theories – utilitarianism, Kantian theory, natural rights theory, etc. – cast so little light on this new, and difficult domain. Certainly no one of these theories seems at a special advantage in clarifying the issues. Instead, it is only the vaguest of *prima facie* notions that are most useful in elucidating the issues: (human) well-being, harm, intentional or knowing harm, political entities and their roles as agents, and essential function of states in promoting or protecting the well-being of their citizens. A suspicion of the role that highly developed ethical theories might usefully play for issues in the morality of war is one of the curious remarks in the excellent account by Fotion (2007: 1; see also Stroll (1996 and 1998)). It has also been my working assumption that fully understanding moral constraints on warfare requires understanding certain conclusions from game theory and working them into more traditional moral thinking about war. To date, there is virtually no effort in this, more mathematical and logical, direction.[12]

Acknowledgements

I thank Alan W. Dipert, Monika A.E. Dipert, George Lucas, William Mandrick, J. Jeremy Meyers, Richard Schoonhoven, and Catherine Ullman for comments and encouragement on earlier drafts; Ms Ullman, a philosophy graduate student and a cybersecurity expert at my institution provided me with a constant stream of technical articles on cybersecurity and cyberwarfare. At the interface of technical expertise and ethics, I am grateful to Neil Rowe (US Navy Postgraduate School) for having shared his essays with me. At the interface of defense policy and cyberwarfare, I am indebted, *unbekannterweise*, to the work of Martin Libicki. I especially thank Edward Barrett (Stockdale Center, US Naval Academy) and Steven Kershnar (SUNY Fredonia) for discussing with me at length a number of the issues in this paper; Dr Barrett additionally provided me with copies of much of the literature.

Notes

[1] I have worked as a contractor for the US Army, working on a Command and Control Ontology with funding from the Army Net-Centric Data Strategy Center of Excellence, Building 1209, Fort Monmouth, NJ 07703. I allude to this research in my discussion of the entities that are necessary to discuss ethical questions in warfare; any views expressed in this essay are strictly those of the author, and not of the US Army or the US Department of Defense.

[2] The exclusive concern with defensive cybersecurity operations continues in the May 2010 National Security Strategy of the United States (NSS 2010: 22). There is one oblique reference to offensive cyberoperations in the small section, 'Use of Force': '[Defense of the US requires] credibly underwriting US defense commitments with tailored approaches to deterrence and ensuring the US military continues to have the necessary capabilities across all domains – land, air, sea, space, and cyber.' This remarkable lack of a clear, declared strategy of response to cyberattacks by other nations is extremely problematic, I will argue below. There is considerable confusion in the media about the differences between cybersecurity and cyberwarfare, with the appointment of Cyber Security tsar Howard Schmidt seen as also covering cyberwarfare (it does not). Schmidt did not help things when he declared 'There is no Cyberwar,' when that is not in his portfolio' (Singel 2010).

[3] At the writing of this paper, the Stuxnet computer worm was identified. It targeted industrial control systems (marketed by Siemens) and was apparently intended to strike nuclear processing facilities in Iran. Its sophistication and knowledge of industrial control software suggests to many a degree of organization that only a state could develop. Nevertheless, its failure to do what it apparently was intended to do is clearly a targeting problem: it rapidly spread around the world but apparently caused little or no damage in Iran.

[4] There is a developed theory of *Prima facie* ethics developed by writers such as W. D. Ross and Louis Pojman. I do not here mean to refer to anything so sophisticated, but rather to what I think most reasonable, and historically and morally informed, people would hazard as an initial guess about moral permissibility. To be sure, its actual moral permissibility will hinge on the details.

[5] I am indebted to Steven Kershnar for pointing out the relevance of my earlier work for the Attribution Problem.

[6] In Axelrod (1984), anticipated in Schelling (1960) and various works on game-theoretic equilibria in the 1950s and 1960s. This is not to say that any single strategy, such as tit-for-tat, employed by a player against another with any other strategy is always superior, but only that a 'cooperative' optimizing strategy, in the context of conflict, would have these two features. Axelrod's own characterization of desirable features of a strategy when there is conflict are: being nice, retaliatory, forgiving, and clear (Axelrod 1984: 54).

[7] These notions of organization, person, artifact, and organism, as well as function and the event of processing have precise, and carefully interrelated, definitions in work in formal ontology for the military that my research group, NCOR, and others are developing. This begins with the Basic Formal

Ontology (Spear *et al*. 2006) including a nascent ontology for all data interchange in the federal government (UCORE-2).

[8] As Martin Libicki notes (2006: xvi and 41f), the threat of cyber-retaliation is less likely to deter than nuclear retaliation, since in the Cold War there would have been no epistemic problem of determining who launched a nuclear attack. In cyberattacks, identification of the attacker may not be immediately evident, and it might be problematic ever to determine the attacker and the attackers' exact affiliation with a state. However, with the possibility of low-level nuclear attacks by non-state organizations or proxy non-state actors, the epistemic parallel becomes much closer.

[9] 'According to the 2009 Annual Report to Congress of the US-China Economic and Security Review Commission, in 2008 the number of reported cyber attacks against the Department of Defense was 54,640. In 2009, from 1 January to 30 June the number was 43,785' (Zifcak 2009: 1).

[10] See Geneva Conventions (1949 and Additional Protocols), Protocol I (1977) to the Geneva Convention: 'Art 54. Protection of objects indispensable to the survival of the civilian population ¶1. Starvation of civilians as a method of warfare is prohibited. ¶2. It is prohibited to attack, destroy, remove or render useless objects indispensable to the survival of the civilian population, such as food-stuffs, agricultural areas for the production of food-stuffs, crops, livestock, drinking water installations and supplies and irrigation works, for the specific purpose of denying them for their sustenance value to the civilian population or to the adverse Party, whatever the motive, whether in order to starve out civilians, to cause them to move away, or for any other motive. ¶3. The prohibitions in paragraph 2 shall not apply to such of the objects covered by it as are used by an adverse Party: (a) as sustenance solely for the members of its armed forces; or (b) if not as sustenance, then in direct support of military action, provided, however, that in no event shall actions against these objects be taken which may be expected to leave the civilian population with such inadequate food or water as to cause its starvation or force its movement.'

[11] A game-theoretic dimension of the morality of war is discussed in the works of Thomas Schelling, especially Schelling 1960, as well as Dipert 2006a, Dipert 2006b, Dipert 2008, and Dipert 2010.

[12] A version of this paper was originally presented at the International Society of Military Ethics (ISME) Conference in San Diego, January 2010, and at the 2010 McCain Conference in April 2010 sponsored by the Stockdale Center at the US Naval Academy.

References

Alexander, Gen. K. B. (2007) Warfighting in Cyberspace, *Joint Forces Quarterly*, 46(3) (July), available at: http://www.military.com/forums/0,15240,143898,00.html; Internet.

Arguilla, J. (1999) Ethics and Information Warfare, in: Z. Khalizad, J. White & A. Marshall (Eds), *Strategic Appraisal: The Changing Role of Information in Warfare*, pp. 379–401 (Santa Monica, CA: RAND Corporation).

Axelrod, R. (1984) *The Evolution of Cooperation* (New York: Basic Books).

Beaumont, C. (2009) Russia 'Helped Co-ordinate' Attacks on Georgian Websites, *Daily Telegraph* (UK), 18 August 2009, accessed 25 October 2010, available at: http://www.telegraph.co.uk/technology/6048978/Russia-helped-co-ordinate-attacks-on-Georgian- websites.html; Internet.

Christopher, P. (1999) *The Ethics of War and Peace: An Introduction to Legal and Moral Issues* 2nd ed. (Upper Saddle River, NJ: Prentice Hall).

Clarke, R. & Knake, R. (2010) *Cyber War: The Next Threat to National Security and What to Do about It* (New York: HarperCollins).

Clinton, H. (2010) Remarks on Internet Freedom, 21 January 2010, accessed 13 September 2010, available at: http://www.state.gov/secretary/rm/2010/01/135519.htm; Internet.

CyberCrimeLaw (2010), available at: http://www.cybercrimelaw.net/Cybercrimelaw.html; Internet.

Dipert, R. R. (2006a) Strategies, Rationality, and Game Theory in the Philosophy of War, Paper, Joint Service Academy Conference on Professional Ethics (JSCOPE), Springfield, VA (January 2006).

Dipert, R. R. (2006b) Preventive War and the Epistemological Dimension of the Morality of War, *Journal of Military Ethics*, 5(1), pp. 32–54.

Dipert, R. R. (2008) Act, Policy, and Mathematical Calculation in the Philosophy of the Morality of War, International Society for Military Ethics Conference (ISME, formerly JSCOPE), University of San Diego, CA (January 2008).

Dipert, R. R. (2010) Game Theory and the Morality of War, Paper, Department of English and Philosophy (West Point, NY: US Military Academy).

Euroactive (2010) US Asks EU to Up Cyber Security, 17 September 2010, available at: http://www.euractiv.com/en/infosociety/us-asks-eu-cyber-security-news-497876; Internet.

FM 24-5 (1939) *Basic Field Manual: Signal Communications* (Washington, DC: War Department, United States).

Fotion, Nicholas (2007) *War and Ethics: A New Just War Theory* (New York: Continuum).

Gao, K.(2003) China to view Windows code, *CNET News*, 28 February 2003, available at: http://news.cnet.com/2100-1007-990526.html; Internet.

Geneva Conventions of 1949 and Additional Protocols, at http://www.icrc.org.

Libicki, M. (2009) *Cyberdeterrence and Cyberwar* (Santa Monica, CA: RAND Project Air Force, RAND Corporation).

Markoff, J. (2009a) Cyberwar: Defying Experts, Rogue Computer Code Still Lurks, *New York Times*, 27 August.

Markoff, J. (2009b) Cyberwar: Old Trick Threatens the Newest Weapons, *New York Times*, 27 October.

Markoff, J. & Shanker, T. (2009) Halted ’03 Iraq Plan Illustrates US Fear of Cyberwar Risk, *New York Times*, August 1.

NSS (2010) *National Security Strategy of the United States*, May 2010, accessed 13 September 2010, available at: http://www.whitehouse.gov/sites/default/files/rss_viewer/national_security_strategy.pdf; Internet.

OED (2010) *Oxford English Dictionary*, accessed 13 September 2010, at dictionary.oed.com (word = cyberwarfare); Internet.

Orend, B.(2005) War, in *The Stanford Internet Encyclopedia of Philosophy* (last revisions 2005), accessed 13 September 2001, available at: http://plato.stanford.edu/entries/war/; Internet.

Owens, W., Dam, K. & Lin, H. (Eds) (2009) *Technology, Law, and Ethics Regarding US Acquisition of Cyberattack Capabilities* (Washington, DC: National Research Council of the National Academies of Science).

Reichberg, G., Syse, H. & Begby, E. (2006) *The Ethics of War: Classic and Contemporary Readings* (Oxford: Blackwell).

Rowe, N. (2006) A Taxonomy of Deception in Cyberspace, International Conference on Information Warfare and Security, Princess Anne, MD, USA, March.

Rowe, N. (2009) The Ethics of Cyberweapons in Warfare, *International Journal of Cyberethics*, 1(1), pp. 20–31.

Rowe, N. (2010) Toward Reversible Cyberattacks, The 9th European Conference on Information Warfare and Security, Thessaloniki, Greece, July 2010, in: L. Janczewski & A. Colarik (Eds), *Warfare and Cyber Terrorism* (Hershey, PA: Information Science Reference).

Rowe, J. & Crusty, E. J. (2010) Ch. XII Deception in Cyber Attacks, in: L. Janczewski & A. Colarik (Eds), *Warfare and Cyber Terrorism* (Hershey, PA: Information Science Reference).

Schelling, T. (1960) *The Strategy of Conflict* (Cambridge, MA: Harvard University Press).

Singel, R. (2010) White House Cyber Czar: ‘There Is No Cyberwar,’ *Wired*, 4 March 2010, accessed 25 October 2010, available at: http://www.wired.com/threatlevel/2010/03/schmidt-cyberwar/; Internet.

Spear, A. (2006) Ontology for the Twenty First Century: An Introduction with Recommendations, accessed 25 October 2010, available at: http://www.ifomis.org/bfo/documents/manual.pdf; Internet.

Stroll, A. (1996) Ethics Without Principles, in: K. S. Johannesen & T. Nordenstam (Eds), *Wittgenstein and the Philosophy of Culture, Proceedings of the 18th International Wittgenstein Symposium*, pp. 310–320 (Vienna: Verlag Holder-Pichler-Tempsky).

Stroll, A. (1998) Ethics without Principles, *Topoi*, 17, pp. 133–147.

Talbot, D. (2010) Moore’s Outlaws, *Technology Review* July/August 2010, accessed 25 October 2010, available at: http://www.technologyreview.com/computing/25564/; Internet.

UN Charter (1945) *The Charter of the United Nations*, accessed 25 October 2010, available at: http://www.un.org/en/documents/charter/chapter7.shtml.

USCCU (United States Cyber Consequences Unit) (2009) Overview by the US-CCU of the Cyber Campaign against Georgia in August 2008, accessed 25 October 2010, available at: http://www.registan.net/wp-content/uploads/2009/08/US-CCU-Georgia-Cyber-Campaign-Overview.pdf; Internet.

US Senate (2010) Armed Forces Committee of the US Senate, Hearings on the Confirmation of Lieutenant General Keith B. Alexander, US Army, to be General and Director, National Security Agency/Chief, Central Security Service/Commander, United States Cyber Command, 14 April 2010, accessed 25 October 2010, transcript available at: http://armed-services.senate.gov/testimony.cfm?wit_id=9367&id=4505; Internet.

Walzer, M. (2006) *Just and Unjust Wars* 4th ed. (New York: Basic Books).

Zifcak, N. (2009) New Cyber Chief Faces Dynamic Challenges, *Epoch Times*, 12 October 2009, accessed 25 October 2010, available at: http://www.theepochtimes.com/n2/content/view/26958/; Internet.

Biography

Randall R. Dipert is Charles S. Peirce Professor of American Philosophy at the SUNY University at Buffalo, New York state. From 1995 to 2000 he taught at the United States Military Academy at West Point, NY and he has recently been a contractor (in applied formal ontology) for the US Army. He has published a book on artifacts and action theory, and has co-authored a book on logic. He has published numerous articles on Peirce, logic, artifacts, mathematics, the philosophy of mind, artificial intelligence, and ontology. He has written numerous pieces of software in Prolog and LISP for logic and logic instruction. He has published one paper in the *Journal of Military Ethics* and given presentations on the morality of preventive war and the application of game theory to the morality of war.