Philos. Technol. (2018) 31:379–407
DOI 10.1007/s13347-017-0271-5

RESEARCH ARTICLE

Why the World Needs an International Cyberwar Convention

Mette Eilstrup-Sangiovanni[1]

Received: 25 January 2017 / Accepted: 18 June 2017 / Published online: 21 July 2017
© The Author(s) 2017. This article is an open access publication

Abstract States' capacity for using modern information and communication technology to inflict grave harm on enemies has been amply demonstrated in recent years, with many countries reporting large-scale cyberattacks against their military defense systems, water supply, and other critical infrastructure. Currently, no agreed-upon international rules or norms exist to govern international conflict in cyberspace. Many governments prefer to keep it that way. They argue that difficulties of verifiability and challenges posed by rapid technological change rule out agreement on an international cyber convention. Instead, they prefer to rely on informal cooperation and strategic deterrence to limit direct conflict. In this article, I seek to rebut some of the main objections to seeking an international convention on the use of cyber weapons. While there are significant obstacles to achieving effective arms control in the cyber domain, historical experience from other areas of international arms control suggests that none of these obstacles are insurmountable. Furthermore, while most critics of cyberarms control assume that cyberspace favors offensive strategies, closer inspection reveals the dominance of cyber-defensive strategies. This in turn improves prospects for striking an effective international agreement on cyberarms control.

Keywords Cyberwar · Cyber conflict · Cyber deterrence · Arms racing · Strategic deterrence · Arms control · Offense/Defence balance · Attribution · Retaliation · Signalling

The capacity of states to use modern information and communication technology (ICT) to inflict grave economic, political, and material harm on enemies has been amply demonstrated. Before invading Iraq in March 2003, the USA 'pre-emptively' cut off Iraqi computer networks and internet grid to undermine the regime's political and military defenses (Saltzman 2013, 46). When Israeli fighter jets bombed a suspected nuclear site in Diaya-al-Sahir, Syria, in September 2007, they first hacked into Syrian air defense systems to blind them to the incoming attack (Lucas 2017). Later that year,

✉ Mette Eilstrup-Sangiovanni
mer29@cam.ac.uk

[1] Department of Politics and International Studies, University of Cambridge, Cambridge, UK

America and Israel used a sophisticated malware—*Stuxnet*—to shut down a nuclear reactor in Natanz, Iran. More recently, the Pentagon has used cyber-strikes against North Korea's missile program to sabotage test launches (Sanger and Broad 2017).

Meanwhile, American intelligence agencies have reported scores of attempted cyber attacks on critical infrastructure in the USA—including air traffic control systems, satellite systems, and national electricity grids.[1] The prospect of similar attacks on military air defenses, water supply systems, chemical and nuclear plants, or oil and gas pipelines has led political and military leaders to stress the need for stronger cyber defenses, and, in particular, stronger means of cyber deterrence. In 2009, the Pentagon created a "US Cyber Command" whose declared mission is to "conduct full spectrum military cyberspace operations…[to] ensure US and allied freedom of action in cyberspace and deny the same to our adversaries".[2] Similar commands have been established in many other countries. At least 29 governments now have military or intelligence units specifically dedicated to offensive cyber operations, and more than 60 countries are developing cyber-weapons according to a survey by the Wall Street Journal (Paletta et al. 2015).

We are witnessing the rapid unfolding of an international cyber-arms race. Just as the development of nuclear weapons revolutionized strategic thinking after World War II and sparked decades of nuclear arms racing that was only gradually brought under control through intense international diplomacy and formal arms control agreements, new cyber technologies have today sparked a frenzied contest to develop cyber offensive weapons and devise new strategies to defend against them (see Sanger et al. 2009). History suggests that arms races are best controlled through formal multilateral agreements, carefully crafted to reduce fears and tensions, increase transparency, and facilitate reciprocal arms reductions. From the invention of the expanding bullet and automated guns to the evolution of modern chemical, biological, and nuclear weapons, the introduction of new military technologies has generally led to efforts (some more successful than others) by the international community to control and limit their use.[3] Yet, despite the current race to weaponize cyberspace, calls for an international treaty to regulate cyber warfare have been roundly dismissed by international leaders as well as by academic experts (see, inter alia, Finnemore 2011; Goldsmith 2011; Singer and Friedman 2014, 127; Schmitt and Vihul 2014b, 21; Schmitt and Vihul 2016, 44; Lucas 2017, 75). Rather than advocating new multilateral instruments aimed at reducing uncertainty and lessening zero-sum competition, many scholars seem to accept the current dogma that in cyberspace, "the best defense is a strong offense."

[1] In 2009, reports surfaced that China and Russia had infiltrated the US electrical grid and left behind software that could be used to disrupt the system. See Gorman 2009, Clarke, June 2011. According to Saltzman (2013), Beijing has successfully hacked into American databases and critical infrastructure, such as nuclear and electric power plants and satellite systems. North Korea has sought to jam electronic signals for US guided missiles (Sanger and Broad 2017), and hackers linked to the Iranian government have hacked a small dam in New York and the networks of AT&T, Bank of America, and the New York Stock Exchange while Russia is suspected of supporting hackers who infiltrated the Democratic National Committees computers during the 2016 US Presidential election campaign. See Lior Div 2016, Lucas 2017.

[2] See https://www.stratcom.mil/factsheets/2/Cyber_Command/ (Accessed 10 November, 2016).

[3] The Geneva Conventions of 1864, 1906, 1929, and 1949 were negotiated by states to govern conduct during war. During the latter half of the twentieth century, a series of further arms control and international nonproliferation agreements were negotiated to prevent the use and spread of WMD. More recently, a majority of states in the international system have embraced the conventions governing anti-personnel landmines (1997) and cluster munitions (2010).

But, is strengthening cyber offensive capabilities with the aim to deter (and if necessary to wage) cyberwar really the best defense against cyberattack? In this article, I discuss the limitations of offense-based cyber deterrence and seek to build a case for an international convention limiting the development and use of cyber weapons. In doing so, I seek to rebut some of the main objections to a legally binding international treaty governing cyber warfare. I submit that while the distinctive features of cyberspace present significant obstacles to achieving effective arms control in this realm, none of these obstacles are insurmountable. I further contend that whereas most current political and military planners believe cyberspace favors *offensive strategies*, a closer examination of the political and material realities of cyberspace in fact suggest the dominance of *defensive* strategies. This in turn improves prospects for negotiating an effective multilateral agreement to decelerate the cyber arms race and thereby reduce risks of serious cyber conflict between states.

Let me be clear at the outset about what I am *not* arguing: I do not advocate that political and military leaders abandon current objectives of bolstering cyber deterrence in favor of a singular reliance on international diplomacy to manage cyber rivalry. Just like during the Cold War's nuclear standoff, international arms control and strategic deterrence must go hand-in-hand if we are to succeed in reducing the risk of global cyber conflict. However, I argue, creating an international legal framework to govern cyber warfare would abet efforts to achieve stable cyber deterrence. Indeed, many of the problems that currently undermine effective cyber deterrence (including difficulties of attribution, difficulties in distinguishing hostile attacks from innocent mistakes, lack of clarity about what constitutes an attack under international law, and—closely related to this—lacking credibility of retaliatory threats) would be greatly alleviated by the articulation of clear, binding international rules and norms that would both serve to distinguish lawful from unlawful behavior and facilitate punishment of cyber aggressors. My argument, in other words, is that international arms control and the desire to strengthen states' individual capacities to deter cyber aggressors are not antithetical goals but present mutually supportive strategies.

The discussion is organized as follows. Section 1 introduces the basic concepts relevant to my analysis. Section 2 documents the current rush by states to amass cyber offensive capabilities in order to bolster deterrence and analyzes the dangers associated with rapid build-ups of offensive capabilities. Section 3 presents the case for an International Cyberwar Convention (ICWC). This section begins by highlighting the limitations of offensive-based deterrence. Next, I outline the desired institutional functions of an ICWC. To be successful, an international treaty aimed at reducing risks of cyber warfare must fulfill (at least) four criteria: (1) it must offer sufficient positive incentives to ensure broad participation by states, (2) it must stipulate rules that effectively constrain behavior and that can be practically implemented given current technology, (3) it must provide sufficient credible information to reduce uncertainty about state interests and enable effective signaling, and (4) it must ensure significant costs to non-compliance.[4] Section 4 seeks to rebut some major objections to seeking a binding international treaty and discusses incentives for states to sign up to such a treaty. Section 5 concludes by pointing to the demand for strong international leadership in cyberspace by the most capable cyber-actors—in particular, the USA.

[4] For a discussion of the design of international treaties under uncertainty and the trade-offs between broad participation and constraining rules, see Keohane and Raustialla 2009, and Keohane and Victor 2015.

1 Definitions and Terms of Debate

There is considerable ambiguity about the meaning of terms such as "cyber warfare" and "cyber threats," and disagreement over what strategic goals can be achieved in cyberspace often boils down to analysts focusing on different problems or dimensions of threat. Before proceeding, it is therefore necessary to briefly define the key concepts of analysis.

To define the terms relevant to analyzing cyber conflict, we must first define "cyberspace." Simply stated, cyberspace is a man-made environment consisting of information/data and ICT control infrastructure (Tallinn Manual 2013). This environment is used every day by hundreds of millions of people to communicate, search for information, and conduct ordinary business transactions. At present, cyberspace is, however, being rapidly "weaponized." This weaponization takes two forms. First, cyberspace is being weaponized insofar as offensive arms are being introduced into the cyber environment that "are capable of destroying or damaging objects within that same environment" (Meyer 2016, 158). Second, cyberspace itself is increasingly viewed by states as a military asset. Contemporary weapons systems are often highly dependent on ICT infrastructure, and cyberspace therefore becomes a hotly contested military domain. The importance of cyberspace as a military domain is evidenced by the US Department of Defense (DOD) declaration that "DOD will conduct kinetic missions to preserve freedom of action and strategic advantage in cyberspace. Kinetic actions can be either offensive or defensive and used in conjunction with other mission areas to achieve optimal military effects".[5] A similar statement, but with greater emphasis of defense, was made by NATO leaders in July 2016 when they recognized cyberspace as "an operational domain in which the Alliance must defend itself as effectively as it does in the air, on land and at sea."[6]

One of the main contentions of this article is that the ongoing weaponization of cyberspace significantly heightens risks of international conflict, possibly rising to the level of *cyber warfare*. What precisely is meant by this term? In current literature, the term "cyberwar" is commonly used to refer to any hostile act that occurs in or through cyberspace. This is too broad. As Joseph Nye (2015a, b) observes, cyberspace harbors a wide range of threats to states and individuals—ranging from identity theft and cyber espionage to various forms of electronic crime and low-grade sabotage through denial-of-service attacks or defacement of internet sites. While these are hostile and criminal acts, few rise to the level of "acts of war." In this article, I therefore opt for a narrower definition of cyber warfare as the *deliberate and hostile use by a state of a cyber weapon with the intention of causing injury or death to persons, and/or to significantly disrupt, damage, or destroy another state's strategic assets or critical national infrastructure.*[7]

This definition rests on three criteria concerning the origin, means, and effects of hostile acts in cyberspace. The first criterion specifies that cyber warfare refers to hostile acts perpetrated by and directed at *states*. Some will take issue with this narrow focus on

[5] US DOD, National Military Strategy for Cyber Operations, Dec. 11, 2006, 15. http://nsarchive.gwu.edu/NSAEBB/NSAEBB424/docs/Cyber-023.pdf.

[6] http://www.nato.int/cps/en/natohq/topics_78170.htm

[7] This definition is similar to that offered in the Talinn Manual (2013) which defines cyberattacks as "cyber operations, whether offensive or defensive, that are reasonably expected to cause injury or death to persons or damage or destruction to objects." My definition differs in stipulating that to qualify as cyber *warfare*, cyberattacks must be perpetrated by state actors and aim to destroy or damage objects of strategic value to a country.

interstate conflict, objecting that many current cyber threats stem from non-state actors. This is true. But while many non-state actors are increasingly cyber-savvy, these actors generally lack the capacity to launch sophisticated cyberattacks against critical infrastructure on their own. As Saltzman (2013, 45) notes, there is a growing consensus among cyber experts that non-state actors such as "hacktivists" and "cyber militias" mostly operate in tandem with or under instruction by national governments (see also Nye 2010; Lucas 2017, 7). Furthermore, combatting cyber threats from non-state actors presents a different set of challenges for architects of international cooperation than governing interstate conflict. In this article, I therefore focus primarily on how to reduce the risks of interstate cyber conflict, although I also touch briefly on how an ICWC might improve states' individual and collective capacity to tackle threats from non-state cyber aggressors.

The second criterion specifies that cyber warfare refers to acts of war perpetrated by using *cyber weapons*. Cyber weapons, in turn, are defined as weapons built primarily of software and data, which are intended to cause targeted harm or widespread destruction in or through the cyber domain (see Clarke and Knake 2010; Lucas 2017, 7). It follows that a kinetic strike against a nation's ICT infrastructure would not constitute an act of cyber warfare (but might constitute an act of conventional war).

The third criterion stipulates that to qualify as cyber warfare, a cyberattack must be intended to cause *serious harm*. In addition to intending injury or death to persons, this may involve aiming to significantly damage or disrupt another state's strategic assets or critical national infrastructure. "Strategic assets" are usefully defined by Salzman (2013, 43–44) as "vital [] assets whose destruction may have a colossal effect on a state's national security and its capacity to operate normally". These may include a state's military constellations, defense industrial base, satellite communication, electrical power grid, internet connectivity, central banking system, stock market, and governmental agencies.

To sum up, cyber warfare refers to hostile acts that (a) are perpetrated by states, in the sense that they are directed by or controlled by state authorities,[8] (b) make use of cyber weapons, and (c) are intended to cause serious harm or damage to persons or strategic objects. By this definition, many threats commonly referred to as 'cyberwar' (including cyber intrusions conducted for purely economic benefit or for purposes of surveillance, as well as low-grade "information warfare" for political gain) fail to qualify as cyberwarfare and therefore would not fall under the purview of an international cyberwar convention. It is less straightforward, however, to determine what range of cyber intrusions fall clearly within our definition. As Lucas (2017, 20) notes, Struxnet is unique in offering the first example of a virtual weapon causing direct physical damage against a high-value strategic target in the real world, and its use is thus relatively easy to classify as an act of cyberwarfare.[9]Our definition does not, however, stipulate that cyberattacks must cause physical destruction to quality as acts of war. Virtual attacks against high-profile political targets (such as government agencies or military command structures) may have equally damaging effects if they succeed in disabling critical infrastructure or undermine trust in a political system. Determining whether and when a virtual attack crosses the threshold of being "equivalent to a use of armed force" in a conventional sense is difficult and is

[8] For a discussion of when states can be held responsible for cyberattacks carried, see Schmitt and Vihul 2014:62, Talinn Manual, Rule 7).

[9] As Lucas (2017, 20) notes, when the Struxnet worm was introduced into the SCDA system that controlled the nuclear centrifuges in Iran, it did exactly what conventional weapons launched by states often do: it targeted and destroyed military hardware that was seen to pose a direct threat to others.

presently one of the most hotly debated questions among international lawyers (see Schmitt and Vihul 2016, 41; Lucas 2017, 115–117). Yet, it is precisely this question (among others) on which international agreement is needed if escalation of cyber conflict is to be avoided. Drawing that line in international law will undoubtedly be very hard and require long and arduous negotiations with input from legal and technical experts. For now, I will simply observe that drawing some kind of clear line that states can orient themselves by is likely to be more important than where precisely the line is drawn.

2 The Cult of the (Cyber) Offensive

International security scholars and military officials widely agree that cyberspace favors offensive operations (see Nye 2015a, b; Lonergan 2016; Sheldon 2011; Libicki 2009, 32–33; Clarke and Knake 2010; Vanca 2013; Saltzman 2013, 43; Lindsay 2015, 52; Lucas 2017, 127). According to traditional offense/defense theory, when offense holds the advance it is relatively easier to 'move forward, destroy and conquer territory' than to protect and defend it (Jervis 1978; Evera and Stephen 1984; Glaser and Kaufman 1998).[10] Cyber experts list three main reasons why offensive strategies have an upper hand in cyberspace. First, as Sheldon observes, "attacks in cyberspace occur at great speed…putting defenses under immense pressure, as an attacker has to be successful only once, whereas the defender has to be successful all of the time" (Sheldon 2011, 98). Second, the prospect of launching attacks with relative anonymity (and therefore impunity) lowers the expected cost of offensive strategies in cyberspace (see Sheldon 2011, 98; Lindsay 2015). Third, physical distance is relatively inconsequential in the virtual world. Cyberattacks can emerge practically from anywhere, providing significant latitude for attackers to seize the initiative and catch defenders by surprise (Nye 2010; Sheldon 2011, 98). Another way of stating this point is that cyber technologies lead to great improvements in the mobility and reach of force—both factors held to increase offensive advantage (see Glaser and Kaufman 1998, 62).

The International relations (IR) theory points to four main strategic implications of offensive dominance. First, when offense is strong relative to defense, it becomes imperative for states to react quickly and resolutely to emerging threats, since even a small initial shift in the ratio of capabilities can cause a decisive shift in actors' ability to prevail in a conflict (van Evera 1984, 64). In a cyber context, the compulsion to act and react fast is further amplified by the changing vulnerability of most cyber targets. Cyber weapons consist of complex software designed to exploit vulnerabilities residing in other software, such as computer operations systems or industrial control systems (Lucas 2017, 7). Unlike most conventional targets, a given cyber target may therefore only be vulnerable—and a cyber weapon aimed at that target only potentially effective—until the point at which a string of code is fixed or the target is replaced (DOD 2006, 3). The implication, according to many cyber experts, is that, "if you do not act quickly, you may not be able to act at all" (Vanca 2013, 26. See also Buchanan 2016; Lonergan 2016; Lindsay 2015, 56).

[10] Traditional offense/defense theory focuses on conventional warfare aimed mainly at territorial conquest. Nonetheless, many of the theory's insights regarding the strategic implications of offensive advantage are also applicable to the cyber domain.

A second implication of offensive advantage is to encourage arms racing. As Glaser and Kaufman (1998, 48) explain, when offense is strong, states are likely to find that equally sized forces are inadequate to support a defensive strategy. Instead, they are likely to conclude that they require a substantial lead in military force to defend against attacks. This triggers a dynamic of competitive arms building, whereby even states that merely wish to defend the status quo strive to build up their military arsenals at a faster pace than their competitors in order to secure an adequate defense.

A third effect of offensive advantage is to increase rewards for striking first, thereby increasing the probability of preemptive or preventive attacks (Jervis 1978; van Evera 1984). In the cyber domain, incentives to attack enemies preemptively arise partly from the fact that actions in cyberspace move so fast that they leave little time for targeted states to mount a defense (Clarke and Knake 2010). The motivation to strike first may be further reinforced by the prospect of using carefully targeted cyber-strikes to neutralize an enemy's conventional defense systems and thereby limit its retaliatory capacity (Saltzman 2013, 44). Indeed, many military strategists believe that preemptive strikes in the cyber domain are likely to bring decisive advantages on the conventional battlefield (see, e.g., Clarke and Knake 2010).

Fourth, according to IR theory, when offense is strong relative to defense, the ability to *deter* attacks (as opposed to seeking to defend against them) becomes vital. Logically, in a strategic environment that is thought to favor offensive operations, the ability to deter aggressors hinges crucially on the ability to signal or demonstrate superior offensive capabilities. Hence, according to traditional offense/defense theory, the surest way to achieve effective strategic cyber deterrence is to develop strong offensive capabilities, which promise crushing retaliation against would-be attackers.

To sum up, conventional IR theory points to four strategic implications of cyber offensive advantage: (1) an emphasis on fast action, (2) a tendency towards arms-racing, (3) strong incentives for preemptive attacks, (4) a focus on boosting deterrence through enhancing offensive capabilities. These implications of cyber offensive advantage are not merely "theoretical." The importance of taking fact action is emphasized in leading national cyber security strategies. According to the US "National Military Strategy for Cyberspace,"

> Cyberspace affords commanders opportunities to make decisions rapidly, conduct operations, and deliver effects at speeds that were previously incomprehensible. In addition, increasing the speed of the policy and decision-making process potentially will yield greater effectiveness of cyberspace capabilities. (US DOD 2006, 4).

The perceived need to boost offensive capabilities as a means to strengthen deterrence is also clearly articulated in many strategic documents and in public statements by defense officials.[11] Admiral Michael S. Rogers, Commander of the US Military Cyber Command, has repeatedly pointed to the need to "increase our capacity on the offensive side to get to that point of deterrence."[12] Similarly, in a press conference in

[11] See, e.g., the Pentagon's updated Cyber Strategy of April 2015.

[12] Statement by Admiral Michael S. Rogers Commander US Cyber Command before the Senate Committee on Armed Services, 19 March 2015. See also Clapper 2015.

April 2015, US Air Force Chief of Staff Mark Welsh described the Pentagon's goal of developing cyber weapons that could inflict "blunt force trauma" on an enemy in order to deter aggressors (see Ewing 2015). The Trump Administration has also signaled an aggressive cyber position. During his presidential campaign, Donald Trump repeatedly vowed to expand America's offensive cyber capabilities—a commitment that has been followed up by plans for a 15% increase in military cyber-security spending in the DOD's 2017 budget (see Gady 2017).[13]

Other nations appear to be drawing similar conclusions regarding the need to adopt more offensive cyber postures. Russia, Iran, and North Korea all have military units specifically dedicated to offensive cyber operations (NY Times 2016). Since 2005, China has begun to incorporate offensive computer network operations into its military exercises, primarily in comprehensive first strikes against enemy networks (Salzman 2013, 51), and the Chinese have recently admitted the existence of purely offensive cyber units in the PLA (according to Raud 2016, 20–21). The germans have publicly disclosed that they are developing offensive cyber weapons (Limnéll 2013) as are Argentina and France. Even Denmark and the Netherlands have launched programs to develop cyber offensive capabilities (Paletta, Yadron and Valentino-Devries 2015).

In short, current political and military leaders focus overwhelmingly on improving their countries' cyber offensive capabilities and appear to premise national cyber security strategies on a firm belief that defensive strategies are insufficient to deter enemies in cyberspace. This strategic environment is strongly reminiscent of the "Cult of the Offensive" that swept Europe in the decades prior to World War I.[14] Much like political and military leaders prior to WWI, today's decision-makers appear to display a "highly exaggerated faith in the efficacy of offensive military strategies and tactics" (Van Evera 1984). And, much like political and military planners prior to WWI, they tend to discount the power of political factors (including the rule of international law) which may favor defenders.

Lessons of history suggest this is a dangerous approach. Just like "the Cult" has been found to have contributed to rapid crisis escalation in the run up to WWI, current beliefs in cyber offensive advantage pose a direct threat to international stability. A large literature in IR has found that offensive dominance increases volatility in international relations. As already discussed, one source of volatility is perceived first-mover advantages arising from offensive advantage, which imply that even states that wish to defend the status quo may be overcome by pressures to launch preemptive strikes (van Evera 1984; Glaser and Kaufman 1998). Some would argue this was the case with the "Olympic Games" attack on Iran's nuclear plant (see Lucas 2017, 59).

However, an equally serious danger arising from offensive advantage is that it heightens the risk of simple misunderstandings leading to inadvertent crisis escalation. Pressure to act fast in the cyber domain means that strategic and tactical decisions (for example, about how to respond to a suspected cyber intrusion or to a threat of intrusion) will often have to be made before actors are in possession of all the facts. This may lead to decisions being based on "worst-case scenario" assumptions about an intruder's

[13] In a campaign speech on 3 October 2016, Trump said "As a deterrent against attacks on our critical resources, the US must possess the unquestioned capacity to launch crippling cyber counterattacks." See Gady 2017.

[14] This term was coined by Stephen van Evera in his 1984 article which detailed how an exaggerated faith in offense advantage contributed to the outbreak of WWI.

intentions. Again, this danger is not merely theoretical. A tendency to "jump the gun" and respond to intrusions with aggressive counter-measures before even basic facts are known is evident in the growing embrace of so-called active cyber defenses (ACD). ACD refers to "proactive electronic measures" designed to "detect, analyze, and mitigate network security breaches in real time" and to take "aggressive, offensive countermeasures against an attacker's networks" (Dewar 2014, 6, 9). Effectively, ACD involves setting up systems to detect and automatically strike against attacking computer systems with the aim of shutting down cyberattacks midstream before the precise origin or nature of an attack are necessarily known (see Carr 2011, 46, 73; Maurer and Morgus 2014). While ACD may help to protect critical infrastructure, and might achieve a degree of deterrence by effectively 'relinquishing the initiative' (Schelling 1994, 137–8), this approach carries high risks of inadvertent crisis escalation. By striking against intruders indiscriminately, ACD widens the scope for mistakes. After all, an apparent cyberattack could be a simple accident. Thus, by automating retaliatory response, a targeted state runs the risk of accidentally targeting innocent systems and thereby escalating conflict (Libicki 2009, 28; Lindsay 2015, 57–58). What is more, ACD limits opportunities for reducing escalation by issuing clear warnings or by responding to intrusions with small but incrementally increasing countermeasures to persuade a challenger to back down, thereby limiting opportunities for active crisis management.

A third danger associated with offensive advantage is that it encourages bellicose diplomacy. According to Stephen van Evera (1984), the Cult prior to WWI pushed states to adopt a highly aggressive style of diplomacy, reliant on issuing ultimatums and threats rather than focusing on negotiation. A similar tendency can be observed in today's cyber domain. The US *International Strategy for Cyberspace* issued in May 2011 declares that, "we reserve the right to use all necessary means—diplomatic, informational, military, and economic…to defend our Nation, our allies, our partners, and our interests."[15] This implies a patent threat of escalation from cyber conflict to kinetic warfare, where the USA has a strong lead over most potential opponents. At the same time, perceived offensive advantage also prompts states to be more secretive about their military capabilities and plans (Glaser and Kaufman 1998). As van Evera explains, during WWI, belief in the advantage of launching a first strike created incentives to "conceal plans, demands, and grievances to avoid setting off such a strike by enemies" (Evera and Stephen 1984, 63–64). A similar tendency to shroud military plans in secrecy is visible in cyberspace, where concealment, stealth, and surprise are widely seen as integral to cyber security (after all, many cyber weapons and tactics would immediately lose their purpose if enemies knew of their concrete nature). This combination of bellicose and secretive diplomacy increases risks of conflict due to miscalculation of others' capabilities or interests and raises the specter of inadvertent escalation from an accidental cyberattack to full-scale kinetic warfare through a sequence of semi- or full-automatic counter-attacks.

To sum up, the Cult of the Cyber Offensive creates a volatile strategic environment with significant risks of rapid crisis escalation. Some would argue that this is simply an inescapable consequence of current technological developments. However, faith in the advantage of offensive strategies turned out to be ill founded prior to WWI, and the

[15] https://www.whitehouse.gov/sites/default/files/rss_viewer/International_Strategy_Cyberspace_Factsheet.pdf

misguided belief that offense would prevail during war is judged by many historians to have contributed directly to the outbreak of hostilities.[16] A similar scenario may be unfolding in today's cyber domain. As I will argue in the next section, there are in fact good reasons to believe that *defensive* cyber security strategies will prove stronger on balance than offensive ones. What is more, technology is rarely definitive in determining strategic choices. Technological momentum may play a large role in stoking current cyber conflict. Yet, as Salzman (2013, 42) reminds us, most instances of military competition and arms racing stem ultimately from political considerations and choices. It is therefore important not to lose sight of how political choices—including efforts to develop institutionalized arms control and conflict management—can alter the expected payoffs between offensive and defensive strategies and thereby reduce conflict potential. It is to these questions I now turn.

3 The Necessity of an International Cyberwar Convention

This section presents the case for negotiating an international convention to govern cyber conflict between states. Despite growing threats from actors using ICT for aggressive and illicit purposes, few treaties address international cyber security issues. Presently, the main international agreements governing cyber conduct are the 2001 Convention on Cybercrime (and its Additional Protocol, 2006) and the Shanghai Cooperation Organization's International Information Security Agreement (2009) (Vihul & Schmitt 2016, 39–40). Both agreements are severely limited, however, in terms of both their scope and membership.[17] Calls for negotiating a comprehensive international treaty to govern cyber conflict—a so-called E-neva Convention—have so far met with disapproval, especially from Western states (ibid.)[18] To date, the most elaborate discussions of cyber governance have thus focused on bringing existing international law to bear on cyber conflict. The *Tallinn Manual on the International Law Applicable to Cyber Warfare* (2013) is the most widely cited outcome of such discussions. This manual formulates a set of rules for how existing int ernational law—chiefly *jus ad bellum*, international humanitarian law (IHL) and laws of state responsibility—apply in a cyber context (Talinn Manual 2013; Schmitt and Vihul 2014a, b, 2016; Lucas 2016). But while the Tallinn Manual (hereafter "the Manual") provides a useful starting point for focusing attention on the need for stronger international cyber-governance, its practical value as an instrument to restrain cyber conflict is modest for (at least) three reasons. First, as Lucas points out (2017, 40), the Manual has generally failed to gain support outside the narrow group of NATO member-states that sponsored it. Second, the Manual does not propose or promulgate new international rules for the cyber domain but merely offers an interpretation of how existing laws may apply. So

[16] For a discussion and overview of the literature, see van Evera 1984.

[17] The Budapest Convention of Cybercrime has 53 state parties. Several important countries, including Russia, Brazil, and India have declined to adopt the Convention. The SCO Intl. Information Security Agreement has six state parties.

[18] In 2009, Russia proposed to the UN that use of cyber weapons by states be preemptively banned. This proposal was rejected by America. See Nye 2010, Clarke 2010. In Sept. 2011, Russia, China, Tajikistan, and Uzbekistan submitted a second Draft Intl. Code for of Conduct for Information Security to the UN General Assembly. This proposal too has failed to gain endorsement.

far, this interpretation has largely failed to persuade states to restrain their activities in the cyber domain (Lucas 2017, 17). What is more, the interpretation of many *jus ad bellum* and *jus in bello* treaty provisions as applied to cyber conflict remain unsettled (Schmitt and Vihul 2016, 43). Anticipating, as some of the leading legal scholars behind the Manual do, that "interpretative dilemmas concerning treaty law will be resolved through the recurrent practice of states in their application" (Ibid., 43–44) seems optimistic, especially when considering that the interpretation offered in the Manual has so far failed to influence state practice or opinio juris. Third, even if states were to agree on the general applicability of *jus ad bellum* and IHL in the cyber domain (which presently they do not), such agreement would hardly suffice. Each one of the international treaties and conventions currently governing the production, stockpiling, sale, transfer, and use of specific classes of armaments is testament to the states' recognition that the general international prohibition on the use of force and IHL are by themselves insufficient to prevent or restrain international armed conflict.

In this section, I therefore present the case for negotiating a separate international convention to govern cyber conflict. If appropriately designed, I argue, such a convention would serve to decelerate the cyber arms race and, at the same time, strengthen the defensive and deterrent capacities of individual state parties. Before we can evaluable the benefits of a formal system of international cyber arms control and conflict management, it is, however, first necessary to briefly consider the drawbacks of the main alternative; namely, a system of mutual deterrence based on strengthening individual capacities for retaliation via offensive counter-measures.

3.1 The Problem of (Offense-Based) Cyber Deterrence

Given the growing incidence of cyberattacks and the increasing vulnerably of many countries to such attacks, the goal of strengthening strategic cyber deterrence has gained increasing attention in recent years. Simply stated, the objective of strategic deterrence is to reduce the likelihood of attacks by ensuring that the anticipated cost of launching an attack exceeds the expected benefit. There are two ways of achieving this. The first is to build strong defenses. If a country's defenses are sufficiently strong to lower the likelihood that an attack will succeed, then the expected utility of attack naturally declines. This is known as "deterrence by denial." The second strategy is to develop the capacity to punish attackers. If an adversary anticipates that an attack will almost certainly trigger a reprisal, this may offset the expected benefits of attacking ("deterrence by retaliation").

Presently, most cyber experts reject deterrence by denial as simply too costly. According to Richard A. Clarke, former National Coordinator for Security, Infrastructure Protection and Counter-terrorism for the USA, the main reason why the US Military favors deterrence by retaliation is the belief that defending civilian networks is just too difficult. "It's almost as though the US military has thrown up its hands and said, 'Well we can defend the military network but we can't defend the civilian networks. And, since we can't do it effectively, we'll just go on the offense'" (Quoted in Lawson, June 2011).

Yet, deterrence by retaliation is wrought with difficulty in a cyber context. Offense-based deterrence works by convincing an adversary that attack is futile since any successful attack would trigger an immediate retaliation that would offset any benefits.

As many commentators have pointed out, it is difficult to credibly communicate such a stance in the cyber domain for (at least) three reasons. First, as many have argued, cyber conflict is characterized by a largely unknown balance of forces (see, e.g., Cirenza 2016). Classic deterrence theory assumes that stable deterrence is reached when a defending state is able to convincingly convey to a would-be attacker its willingness and capacity to respond to an attack by inflicting damage of a similar or greater magnitude on the attacker. But, whereas it is often possible to evaluate an opponent's capacity for inflicting damage by conventional military means with relative accuracy, it is more challenging to gauge relative cyber power (ibid.). Cyber weapons are often surreptitious and short-lived and thus cannot be accounted for the way one might count long-range missiles or nuclear warheads. Furthermore, unlike for a nuclear missile, there is high degree of uncertainty regarding the effects of employing cyber weapons (Libicki 2009, 52; Vanca 2013; Libicki 2009, 52–53). Insofar as cyber weapons depend on exploiting vulnerabilities in target software, the amount of damage that can be affected is often unpredictable (see Cirenza 2016). This inability to credibly threaten a specific magnitude of retaliation fundamentally undermines strategic deterrence (Lindsay 2015, 56).

A second limitation of cyber-deterrence by retaliation arises from the problem of *attribution*. The difficulty of correctly identifying the sources of cyberattacks and the possibility of routing attacks through foreign countries to conceal their origin implies that states often struggle to determine the specific identity of cyber attackers. This in turn makes it difficult to credibly threaten to punish aggressors (see Libicki 2009, 74; Goldsmith 2011; Limnéll 2013; Singer and Friedman 2014, 73; Div Lior 2016; Schmitt and Vihul 2016, 44). It is important to note, that while the attribution problem has been branded by many scholars as a major stumbling block, not only to effective deterrence but also to any form of international agreement governing cyber conflict (see, e.g., Schmitt and Vihul 2016, 44), the problem is often somewhat misconstrued insofar as it is treated principally as a problem of limitations to available technology (see Rid and Buchanan 2015). Yet, as some scholars have argued (and as I discuss in greater detail in the next section), given adequate time and resources, accurate attribution of cyberattacks is in many cases technically feasible—especially when it comes to major attacks against high-value national targets where the list of possible perpetrators tends to be limited and where incentives to invest the necessary time and funds in attribution are strong (Lindsay 2015, 53–4. 57; Rid and Buchanan 2015). The difficulty is that reliable cyber-attribution is often highly time demanding and resource intensive, making it unsuitable to underpin a strategy of offensive retaliation based on states taking immediate and forceful counteraction to halt on-going attacks or to deter further attacks. As Rid and Buchanan argue (2015, 32), analyzing a well-executed cyber operation in a narrow timeframe will be a significant challenge even for the most professional and best-resourced intelligence agencies. In serious cases, when decisions will often have to be made under time-pressure, the speed of political developments may therefore outpace the speed of the attribution process (ibid.).

In addition to widespread uncertainty about relative force ratios and about both the origins and effects of attacks, a third problem undermining cyber deterrence-by-retaliation is the need to conceal the means, aims, and tactics of cyber warfare. All cyber weapons depend fundamentally on exploiting vulnerabilities in target software. As Lindsay (2015, 55–56) observes, a cyber attacker therefore cannot reveal his knowledge of a target's software vulnerabilities and use this to issue a specific, credible threat ("I will attack this network with this effect unless you refrain from X"), since

revelation would prompt the defender to patch up weaknesses (see also Limnéll 2013). Unlike in the case of conventional or nuclear weapons, the only way to credibly reveal the possession of an offensive cyber weapon without at the same time rendering it worthless may be to use it. This inability to credibly signal the capacity to inflict specific harms dramatically weakens the prospect for offense-based cyber deterrence.

The above limitations suggest that offensive deterrence based on retaliation-in-kind is impractical in the cyber domain. Importantly, this does not rule out all forms of deterrence-by-punishment. However, it suggests that credible threats of reprisal must largely be based on the capacity and will of states to retaliate through non-cyber based means (so-called cross-domain deterrence). As discussed, several national cyber security strategies stress the option of kinetic responses to cyberattacks.[19] Assuming, however, that responding to a cyberattack with military force would be inappropriate and counterproductive in all but the most extreme cases, the deterrent effect of such threats is likely limited. This leaves two available routes for achieving stable cyber deterrence; deterrence-through-denial, or deterrence based on the promise of punishment via non-forceful means, such as political, economic, and normative sanctions. As I discuss in the next section, such cross-domain mechanisms of deterrence would in turn be strengthened by an international convention governing state conduct in the cyber security domain.

3.2 Institutional Functions of an International Cyberwar Convention

This section outlines the desired contours of a Cyberwar Convention from an institutional perspective. Solving the problem of impending cyber conflict will require careful institutional design. To be effective, an ICWC must fulfill the following goals: (1) secure broad participation from major cyber-faring states, (2) set out rules that effectively constrain state behavior, (3) provide sufficient credible information on actions in cyberspace to reduce uncertainty about state interests and allow effective signaling, and (4) ensure significant costs to non-compliance. Fulfilling these goals in tandem will be challenging. For example, IR-scholars often point to tradeoffs between the strictness of rules on one hand and participation on the other, since highly obligatory rules may reduce the willingness of reluctant states to be bound by an international agreement (see esp. Keohane and Raustialla 2009; Abbott and Snidal 2000; Finnemore 2011). This problem may be particularly pernicious in a cyber context given uncertainty about the long-term implications of adopting specific rules (Eilstrup-Sangiovanni 2009). A second challenge will be to devise an effective and legitimate system for monitoring states' cyber conduct (and thereby ensure costs to non-compliance) without encroaching unnecessarily on state sovereignty or creating disincentives for sharing information. Neither of these problems is insurmountable. Indeed, existing weapons conventions and arms control treaties provide useful templates for devising solutions. Against this backdrop, the remainder of this section outlines seven core functions of an effective ICWC, namely, i) supplying information, ii) spelling out clear rules of state conduct, iii) increasing transparency, iv) collective

[19] Since 2011, the USA has officially "reserved the right" to respond to cyber-based attacks with military force, and NATO has designating cyber attacks as events that could trigger Article 5 of the Washington Treaty, which seems to clearly leave open the door to responding with military force.

attribution, v) authorizing countermeasures, vi) clarifying state responsibility, and vii) offering support for compliance. For each function, I consider how the institutional design of an ICWC might be informed by lessons from existing international arms control agreements.

3.2.1 Information Provision

The most basic aim of all international arms control agreements (ACAs) is to provide information, thereby reducing uncertainty and enabling states to clarify their capabilities, interests and intentions so that misunderstandings can be avoided. One way ACAs provide information is by institutionalizing mechanisms for regular information exchange. Such exchange may take the form of "mutual assurance," whereby states provide information about their activities to reassure others about their motivations and objectives, or it may involve a measure of outside monitoring and verification (either carried out by individual states or by a central institution) (see Abbott 1993, 4). A second way ACAs provide information is by establishing formal rules that clearly distinguish between permitted and prohibited behavior. By declaring some forms of conduct illegal, ACAs serve to clarify state intention, since, as Clarke notes, if a government is prepared to violate an explicit and binding agreement, there is less ambiguity about its intentions (Clarke and Knake 2010. On "screening" and signaling effects of formal treaties, see, e.g., Simmons and Hopkins 2005). This function is particularly important in a cyber context, where states' capabilities and motivations often cannot be easily observed or inferred and where many tools that states typically use to signal their intent (such as, e.g., increases in armament expenditure, force mobilization, warnings shots, or military exercises) are unavailable or would be meaningless. By stipulating clear, binding rules of behavior, an ICWC would thus allow states to signal their intent clearly through either obedience or disobedience.[20]

3.2.2 Constraining Behavior by Establishing Clear Rules for (Im)/Permissible Actions

In addition to providing information about state interests and conduct, many ACAs place some form of restriction on the development, testing, deployment, or possession of specific categories of weapons. This would not be practical for cyber weapons. Given the dual-use nature of many cyber technologies, it would be difficult to agree a clear definition of what constitutes a cyber weapon prior to deployment.[21] An ICWC would therefore have to focus on restricting *use* of cyber weapons—for example by prohibiting first use of cyber weapons or by banning use of cyber weapons against certain targets.[22] In this regard, a crucial function of an ICWC would be to establish

[20] A large literature in IR focuses on the role of costly signals in communicating intent and avoiding misunderstandings leading to conflict. For example, Fearon 1997.

[21] The difficulty of banning certain classes of weapons is illustrated by the lukewarm reaction to the Wassenaar Group's 2013 agreement to restrict export of certain types of surveillance software to repressive regimes. Critics have objected that the definition of "surveillance software" is so broad it risks blocking export of legitimate technologies which are necessary to identify and repair vulnerabilities in software. See Zetter 2015.

[22] The UN Group of Governmental Experts on Information Security has recommended just this. See UN General Assembly 2015.

clear rules regarding *when* a cyber intrusion might be permissible and justified and when it would constitute an impermissible act of aggression. Some scholars insist this function is already fulfilled by existing international laws governing armed conflict. As discussed, the legal experts behind the Tallinn Manual argue forcefully that *jus in bello* and principles of IHL are directly applicable to cyberspace, thereby obviating the need for further treaties (see esp. Schmitt and Vihul 2014a, b, 2016; Lucas 2017). But, while the Tallinn experts agreed on the general applicability of international laws of armed conflict to cyberspace, their analysis revealed significant disagreement on how to apply specific principles of both codified and customary international law in cyberspace (Schmitt and Vihul 2014, 60–2, 2016, 43–44; Lucas 2017). Most crucially, the Manual fails to define the precise preconditions to treating a cyber operation as equivalent to an "armed attack" under international law, thereby allowing victims to respond forcefully in self-defense as well as legitimizing forceful response by third parties in collective defense (ibid., 41). The Talinn experts agreed that cyber operations that cause "significant injury, death, physical damage or destruction" amount to a "use-of-force" and thus qualify as an armed attack (Schmitt and Vihul 2014, 59). However, no consensus could be reached on the exact point at which a cyber activity crosses the threshold of use-of-force (see Tallinn Manual, Rule 1, Schmitt and Vihul 2016, 41). The experts also disagreed on whether cyber operations that cause severe *non-physical* harm could qualify as a use-of-force (Schmitt and Vihul 2014a, 59–65, 2014b, 17, 2016). The crucial question of when an offensive cyber-operation amounts to an armed attack thus remains unsettled.

But, even if the NATO-sponsored legal experts had managed to reach agreement on this crucial question, it is important to keep in mind that the Tallinn Manual does not create new, binding international law. Only governments can do that. At present, there is simply no international consensus on whether and how existing international laws of armed conflict apply in cyberspace.[23] As many commentators have argued, the lack of any codified international laws for cyberwarfare in turn means that there presently is no clear distinction—either in international law or in the declaratory policies of most governments—between criminal cyber intrusions (which may result in minor damage or economic losses) and major attacks against critical national targets or infrastructure, which may cause widespread destruction or loss of life. By extension, there is also no common understanding of what would constitute a proportional response to a cyberattack of either kind (see Kugler 2009; Himes 2015; Libicki 2009, 53). This lack of clarity greatly increases the risk of conflict. "Just cause," "proportionality," and "comparative justice" are key principles in international law, integral to justifying acts of war both politically and morally. Without adherence to such principles states may find themselves caught in endless spirals of opportunistic aggression and violent reprisals. However, without a clear common understanding of when a cyberattack amounts to an act of war, simply insisting that IHL applies in cyberspace is of little

[23] The Consensus Report of the UN Group of Governmental Experts on Information Security adopted in 2013 and updated in 2015 confirms that "international law, and in particular Charter of the UN, is applicable" to cyber activities. This implies that cyber operations that form part of an armed conflict or amount to a "use of armed force" would be subject to the same rules as kinetic warfare. However, the consensus report merely provides voluntary guidelines and offers no clear guidance as to when cyber operations amount to a use of armed force. Alledgedly, a draft provision that verbatim confirmed IHL's applicability was removed to secure unanimity (see Schmitt and Vihul 2014b).

use. A crucial function of an ICWC would therefore be able to provide an authoritative clarification of the conditions under which a cyberattack amounts to an act of war, which would serve to guide and constrain state behavior.

3.2.3 Increasing Transparency and Ensuring Effective Crisis Management

In addition to improving information and establishing clear rules of behavior, a third important function of an ICWC would be to lower the risk of accidental conflict by increasing transparency and by introducing mechanisms for crisis management. An important aspect of the nuclear arms control treaties negotiated during the Cold War was to institutionalize channels of communication between the superpowers and thereby lower the risk of accidents or misunderstandings triggering war. Many cyber experts assert that the risks of accidental conflict escalation are greater in cyberspace than in the nuclear realm (see, e.g., Libicki 2009, 28–39; Clarke and Knacke 2010; Lonergan 2016, 11). One reason is the difficulty of distinguishing cyber accidents from overt hostile attacks (Libicki 2009). As Vanca (2013) argues, a cyberattack on a nation's air defense network—even if by mistake—might plausibly lead that country's leadership to conclude that air attacks were imminent and inadvertently trigger a kinetic war. A second reason is the prospect of attacks by non-state actors. A cyberattack launched by a non-state actor from within another state's territory might lead to the mistaken conclusion that another government was responsible and trigger a forceful retaliation. Given these sources of instability, an ICWC must focus on reducing risks of accidental conflict by establishing strong channels of communication and by instituting obligatory early warning mechanisms whereby states agree to notify each other immediately in the event of the detection of any attacks from their territory (on this point see esp. Carr 2011; Nye 2015a, b).

3.2.4 Collective Attribution

An international cyberwar convention will not be effective without ensuring real costs to non-compliance. This brings us to the "attribution problem." No meaningful mechanism for enforcement and punishment can be devised without first solving the problem of how to identify transgressors. An important aspect of an effective ICWC would therefore be to institute some form of collective mechanism to facilitate faster and more reliable attribution of cyberattacks. As already discussed, whereas pessimism has long reigned regarding attribution in cyberspace, many experts now take the view that—given adequate time and resources—reliable attribution of cyber-attacks is generally possible, at least when it comes to large-scale attacks against critical infrastructure (for an insightful discussion, see Lindsay 2015). Nevertheless, reliable attribution is both time consuming and costly. Reliable attribution requires rapid emergency incident response and large-scale information and forensic evidence gathering. This may require states to train specialist "digital forensics forces" to retrieve physical evidence (including electronic equipment and data storage devices). Once collected, evidence and data must be analyzed for technical attribution. This may in turn require sophisticated tools for reconstruction, analysis, and interpretation and may depend on access to a variety of different sources of intelligence (see DOD Forensic Cyber Crime Center). Such technical capabilities are currently beyond the reach of many states. A

joint attribution mechanism would enable states to pool technical and financial resources and thereby make reliable attribution available to more states, at lower cost.[24] This would strengthen cyber deterrence in two ways. First, collective attribution would expand the group of states that are capable of systematically uncovering, blocking, and retaliating against malevolent cyberattacks. Second, as Lindsay argues (2015, 59), by lowering the general costs of attribution relative to the value of the systems being protected, joint attribution would increase the overall credibility of threats of retaliation (on benefits of collective attribution, see also Salzman 2013; Vink 2015).

When considering how to design a system for joint cyber attribution, a useful model is afforded by the Comprehensive Test Ban Treaty (CTBT) adopted by the UN General Assembly in September 1996, but not yet in force.[25] As Edward Ifft (2005) explains, the CTBT specifically seeks to address the problem of a highly uneven capacity by state parties to gather and interpret data regarding compliance. It does so by instituting a joint International Monitoring System (IMS) which—if completed—will have 337 facilities worldwide generating data from seismic, radionuclide, hydroacoustic, and infrasound sensors. This data could not be independently gathered or analyzed by most states, according to Ifft. However, an international data center in Vienna will supply individual state parties with expert technical analysis of IMS data and other relevant data that can be used to determine compliance free of charge (ibid).

Taking the joint IMS envisaged for the CTBT as a model, one could envisage a Joint Attribution Mechanism (JAM) which would assist states with cyber forensics, data gathering, and technical analysis at low cost.[26] Such a mechanism would clearly benefit countries that currently lack indigenous technological abilities to collect and analyze data for attribution, and might provide an important incentive for them to sign up to a cyber treaty. However, a JAM would also benefit countries that already possess sophisticated cyber forensic capacities by enhancing the credibility and legitimacy of the attribution process. As Ifft observes, unilateral attribution tends to not be as credible as collective attribution. "It is clear in the aftermath of not finding WMD in Iraq that, as far as most of the world is concerned, conclusions reached by respected international bodies will be more credible than those reached by any single country" (Ifft 2005, 3). In a cyber context, if a state were to retaliate against the suspected source of a cyberattack based purely on private attribution, this might trigger suspicion on the part of other countries regarding the sources and methods of attribution and could, in a worst-case scenario, be interpreted as an act of international aggression. For these reasons, attribution is best delegated to a neutral international agent tasked with supplying reliable data and analysis to states that require it (cf. Salzman 2013).

[24] A recent example of international collaboration on attribution and law enforcement in the cyber domain is Operation Shrouded Horizon. This joint operation, which involved national intelligence agencies from 20 different countries led by the FBI and Europol, succeeded in shutting down Darkode.com—an online cybercrime forum (see FBI 2015, Lucas 2017, 58).

[25] While it is not yet in force, the CTBT nonetheless provides a present example of a multilateral treaty with extensive monitoring and verification provisions, signed by 186 countries and ratified by 166 states.

[26] On the desirability of international collaboration on cyber forensics, see Clarke and Knake 2010, Nye 2015.

3.2.5 Authorizing Countermeasures

Once reliable attribution is ensured, steps must be taken to ensure that cyber aggressors are consistently punished. Here too, an international convention has an important role to fulfill. International law stipulates that counter-measures must fulfill the requirement of proportionality—that is, they "must be commensurate with the injury suffered" (for a fuller discussion, see Schmitt and Vihul 2014a). Presently there is, however, no international agreement on what would constitute a proportionate response to a cyberattack,[27] nor are there any institutional mechanisms for collective authorization of such a response.[28] This is a precarious situation. Just as individual attribution of cyberattacks is undesirable and insufficiently legitimate, unsanctioned retaliation against cyber transgressors risks escalating conflict and might open states to charges of opportunism and disproportionate response. By laying down clear rules for *when* punishment of cyber-intrusions is appropriate, and by specifying the nature and magnitude of retaliation warranted, an ICWC would help limit self-serving or overly harsh sanctions and thereby reduce the potential for conflict escalation. What is more, by specifying when and how cyber intrusions can be punished, an ICWC would serve to lower the political costs of retaliation and thereby strengthen deterrence by enhancing the credibility of retaliatory threats.

3.2.6 Clarifying State Responsibility in Cyberspace

So far, I have focused on institutional means to reduce uncertainty, increase transparency, and ensure effective monitoring and enforcement of rules regarding state conduct in cyberspace. While each institutional function would have to be tailored to the peculiar circumstances of cyberspace, such functions are part and parcel of most ACAs. To be effective, however, an ICWC would also have to address a further and distinct issue regarding how the law of state responsibility applies in cyberspace. Unlike the realm of nuclear conflict, which is squarely dominated by states, the cyber domain raises the specter of frequent attacks by non-state groups. This raises the question of when states can and should be held responsible for cyber activities by non-state actors. Schmitt and Vihul (2014a, 60) address this issue for a legal perspective observing that (a) the principle of state sovereignty empowers and obliges states to "exercise control over cyber infrastructure and activities within its territory," and (b) international law dictates that a state may not "allow knowingly its territory to be used for acts contrary to the rights of other states" (see also Tallinn Manual, Rule 5; UN Group of Cyber Security Experts 2015).[29] Together, they assert, these principles of international law imply that a state to which the cyber operations of a non-state actor are directly attributable (insofar as the non-state actor is acting under direct control or instruction

[27] Consider that most legal experts judge that a kinetic response to a cyber intrusion below the threshold of an armed attack would be illegitimate (see Vihul and Schmitt 2014a and Tallinn 2013). Nonetheless, many national cyber-security doctrines (including those of the US and UK) threaten just that.

[28] Except of course in the case of a cyberattack that were of such a grave nature and magnitude that forceful countermeasures might be mandated by the UNSC.

[29] The Consensus Report of the UN Group of Cyber Security Experts (2015) holds that the principle of state sovereignty and related principles apply to state conduct of ICT-related activities, and that states enjoy jurisdiction over ICT infrastructure within their territory.

of the states' government) can be held legally responsible and that an injured state can be justified in taking countermeasures against it (Schmitt and Vihul 2014a, 59, 62–64; Talinn Manual 2013, Rule 7; Tikk 2011, 123–4).

I agree with the Talinn experts that the principle of "direct state responsibility" is essential to prevent states from engaging in cyber conflict through proxies. The principle therefore ought to be formally confirmed in an ICWC. However, when it comes to enforcing the principle of direct state responsibility, attribution and countermeasures must be subject to collective oversight and mandate for reasons discussed in the previous sections. One might go further and insist that a state can also be held responsible for cyberattacks launched from within its jurisdiction even when the perpetrators *are not* acting under *direct* control of the government or its agencies. For example, Eneken Tikk (2011), 124) suggests a "Duty of Care Rule" whereby states would acquire "a duty to develop and implement reasonable levels of security standards for their ICT infrastructure and systems" to ensure they do not become havens for cyber militants and criminals. Such a rule would require states to take "reasonable measures" to prevent non-state actors from launching attacks against other states—for example by enacting and enforcing criminal laws against common cyberattacks (see Carr 2011).[30] Repeated failure by a state to act against groups exploiting its ICT infrastructure for criminal or militant purposes would result in it being declared a "sanctuary state," thereby opening it to countermeasures by other states (*opcit.*)—subject, in all cases, to collective authorization and implementation.

3.2.7 Support for Compliance

A final necessary feature of an ICWC is institutional support for compliance. Literature on compliance with international agreements has found that states are more likely to honor their commitments if they have access to funding and professional expertise to help them comply.[31] Such support is particularly germane in the cyber domain, where national capacities for compliance vary greatly. Compliance support would, as a minimum, include expert training and capacity-building measures aimed to improve national cyber defense systems and to enable national officials to cooperate with international monitors. Insofar as an ICWC would obligate states to take measures to forestall attacks by non-state actors from within their jurisdiction, support might also include legal, technical, and financial assistance to help improve supply chain security and prevent the proliferation of malicious code and other harmful hidden functions, as well as to improve states' capacity to assist other parties with investigations regarding cyber intrusions originating in or routed through their territories. Without such organizational support for compliance, many states would likely be unable to comply with the provisions of a cyber convention and therefore unwilling to sign on.

[30] The Budapest Convention on Cybercrime Art.2 and 11 requires parties to establish criminal offenses for many types of cyberattack under domestic laws. See Nye 2010. However, the limited scope of this Convention (which focuses mainly on Internet criminality and has only 53 state parties) makes it unsuited for addressing wider threats of cyber warfare.

[31] See, e.g., Chayes and Chayes 1998, Finnemore 2011, 92. Recent arms control treaties offer pertinent examples of institutionalized compliance support functions. For example, Art. 10, of the CWC instructs the technical secretariat of the Organization for the Prohibition of CWs to provide expert advice and assistance to states in identifying how they can best implement national control programs.

3.3 Benefits of an ICWC for Strategic Deterrence

In this section, I have outlined several reasons for and potential benefits of adopting an international convention to govern cyber conflict. As I alluded to in the introduction to this article, these reasons and benefits are contested by many scholars and practitioners. Before turning to address objections to an International Cyberwar Convention, however, it is worth briefly summarizing how an ICWC would serve to improve strategic cyber deterrence. At the beginning of this section, I argued that limitations of offensive cyber warfare dictate that strategic deterrence in cyberspace must be based as much on denying benefits to attackers (deterrence-by-denial) as on imposing costs via forceful retaliation (deterrence-by-retaliation) (see Lindsay 2015). By improving information-sharing, establishing best practices, and accelerating joint capacity building, an ICWC would help strengthen national cyber defenses and make national ICT systems more resilient to attack. Also, by instituting early warning mechanisms, an ICWC would lower the likelihood that cyberattacks will succeed. As such, a convention would serve to strengthen deterrence-by-denial (on collective cyber defense, see also Debar Dewar 2014, 14).

An ICWC would also improve deterrence-by-retaliation. As discussed, for cyber deterrence to work, states must be able to communicate clearly under what conditions a cyberattack will trigger a retaliatory response, and at what level. By clarifying what counts an act of cyber-aggression and what level of retaliation is deemed acceptable by the international community, an ICWC would thereby enhance states' capacity to adopt and communicate an effective deterrent posture (see esp. Carr 2011).

When considering the benefits of an ICWC with respect to strengthening deterrence-by-retaliation, it is important to appreciate that effective retaliatory deterrence need not rest on a promise of immediate or massive reprisals. Rather, what is required is a credible promise that attackers will be eventually identified and some form of punishment dispensed. By instituting joint management and oversight over the process of attribution and reprisal, an ICWC would serve to slow down the process of retaliation. This would have several benefits. First, a lengthier and more systematic process of attribution and retaliation would reduce risks of conflict escalation by leaving time for aggressors to either prove their innocence or agree to voluntary compensation measures. Second, by introducing joint authorization of countermeasures, an ICWC would lower the costs (both political and material) of retaliation, thereby making retaliatory threats more credible. Joint authorization of countermeasures might also encourage collective sanctions against transgressors and thereby introduce an element of "extended deterrence" among parties to an ICWC. Third, international authorization of reprisals would make it feasible to widen reprisals beyond narrow retaliation-in-kind (i.e., using offensive cyber capabilities to strike back against an attacker's information networks) to encompass a wider range of punitive measures: diplomatic, economic, and (in cases of grave assaults) military. Cross-domain retaliation is currently advocated by many cyber experts, since political and economic instruments are less sensitive to revelation than "in-kind" retaliation (Lindsay 2015, 58) and may be more potent against adversaries that lack sophisticated information networks to retaliate against (Kugler 2009). However, reliance on cross-domain retaliation would place high demands on the impartiality of attribution, lest vague cyber threats come to serve as political justification for economic sanctions (or even military strikes) against other

nations. This speaks strongly in favor of delegating attribution to an independent international body.

4 Objections to an ICWC

Critics have raised a number of objections to international cyber-arms control—ranging from the claim that *any* international agreement aimed at constraining state actions in cyberspace would be meaningless as it could not possibly be enforced, to the "softer" claim that the time is not yet ripe to negotiate legally binding rules regarding state conduct in cyberspace. In this section, I seek to address and rebut each of these criticisms.

Objection 1: The Impossibility of Monitoring and Enforcement The perhaps most frequent objection to an international treaty governing cyber conflict is that it would be impossible to verify compliance or enforce the terms agreement (see Nye 2010; Clarke and Knake 2010; Lonergan 2016, 7–8; Schmitt and Vihul 2016, 44–45). As discussed, cyber weapons are notoriously difficult to monitor, mainly due to their dual-use nature and the fact that they are easily concealed (Nye 2015b; Singer and Friedman 2014, 127). Since states would not wish to comply unless they were sure others were complying, critics argue, any international agreement seeking to restrain states' cyber conduct would quickly unravel.

When considering the force of this objection, the first thing to note is that problems of verification are hardly unique to the cyber domain. Take the example of chemical and biological weapons. Most precursors to chemical and bacteriological weapons have a range of domestic and industrial uses, and problems of "dual-use" (as well as a general inability to distinguish offensive from defensive R&D programs) have long plagued both the Chemical Weapons Convention (CWC) and the Biological and Toxins Weapons Convention (BCW) (see Ifft 2005). Nonetheless, these problems have for the most part been successfully addressed by a combination of highly detailed and strongly obligatory prohibitions, stringent domestic implementation and reporting requirements, intrusive international monitoring systems, and high penalties for cheating (for a discussion of the CWC, see Eilstrup-Sangiovanni 2009). So far, these institutional measures have deterred major transgressions. There is no a priori reason to assume a similar result could not be achieved in the cyber domain. In fact, the challenge of monitoring compliance may in some ways present a lesser obstacle to cooperation in the cyber domain than in the realm of "traditional" nuclear and bio-chem weapons. Consider that the need for reliable verification of compliance with a cyberwar convention arises not *ex ante* with respect to weapons development and possession, but rather *ex-post* with respect to weapons *use*. This makes an important difference. Due to the enormous and irreparable damage that can be wrought by a single nuclear strike, the bilateral treaties that dominated Cold War nuclear arms control focused primarily on restricting the development and possession of nuclear weapons and their means of delivery. By contrast, verifying and attributing the *use* of nuclear weapons was never a concern. Once a nuclear weapon is launched, there can be little doubt as to "who did it"; the only relevant question is

whether a victimized state has the capacity to retaliate. The ability for states to monitor build-ups of offensive capabilities is thus vital to nuclear arms control.

Cyber weapons present a different problem. Since a cyberattack is unlikely to cause as widespread or irreversible damage as a nuclear attack, and since the ability of a victimized state to retaliate is unlikely to depend on the speed with which a counter-attack can be launched (or even on the ability to retaliate in kind), the main concern is not to monitor other states' possession of offensive cyber weapons (which would admittedly be close to impossible). Instead, the thornier problem is to establish, *ex post*, who is responsible for an attack and decide how to respond. A joint attribution mechanism overseen by an international authority would greatly improve states' individual and collective ability to decide such questions and would thus go a long way towards solving the problem of monitoring and enforcement.

Objection 2: An ICWC Would Take Too Long to Negotiate A second objection to a formal treaty is that it would take too long to negotiate. According to Finnemore (2011, 93), "negotiating treaties can be a slow and cumbersome process, ill-suited to fast-changing issues like cyber security and Internet governance". Negotiating the UN Convention on the Law of the Sea took more than a decade, she observes. One might add that negotiations over the CWC, signed in 1993, spanned several decades.[32] It is important to note, however, that whereas negotiations over UNCLOS and the CWC took many years, in both cases, elements of norm gestation and informal cooperation kicked off early in the negotiation process as individual states sought to gradually bring their behavior into line with emergent international norms. Indeed, experience from these and other international treaty negotiations (such as the negotiations of the International Landmine Ban Treaty) suggest that while international norm creation and acceptance is generally a slow process, the process is often accelerated by formal treaty negotiations which add political weight and visibility to an issue. While the road to a binding international agreement on cybersecurity may be long, embarking on negotiations might therefore by itself have a positive impact.

An ICWC Would be Insufficiently Adaptable to Technological Change A third objection to an ICWC centers on the problem of rapid technological change, which implies that any treaty seeking to lay down detailed rules for the cyber domain would soon be outdated (Lonergan 2016, 8–9; Nye 2010). This objection too is often overstated. Virtually all spheres of international arms control must grapple with problems of unpredictable technological advances. Since the CWC's inception in 1993, new developments in the chemical industry have demanded periodic updating of the Conventions' verification annexes and lists of prohibited substantives (see Kelle 2003).[33] Consider also that most international arms control agreements provide for periodic review conferences that allow governments to update the terms of agreements. As an example, state parties to the Biological Weapons Convention have held seven review conferences since the Convention took force in 1975, most of which have

[32] Negotiations on a treaty to eliminate chemical weapons began in 1962 within the Conference of the Committee on Disarmament and continued within its successor institution, the Conference on Disarmament, before concluding in 1992.

[33] Art. XV, para. 5 of the CWV governs changes to the Annex on Chemicals.

focused on strengthening verification and reviewing the operation of the Convention "to take account of new scientific and technological developments."[34]

Undoubtedly, no international agreement will be perfect, and any agreement reached on cyber governance would likely require subsequent revision—both as result of learning experience and in order to adapt to technological change. However, the fact that an ICWC (as sketched here) would focus on prohibiting specific behaviors (such as, e.g., first use of cyber weapons or use of cyber weapons against civilian targets) rather than seeking a wholesale ban on the development or possession of an entire category of weapons technology implies that the problem of adapting to technological change may be less severe than critics allege. Consider that The Hague Conventions of 1899 and 1907 which ban the use of poisons and asphyxiating gasses and protect the rights of civilians and the wounded during war have largely stood the test of time despite enormous developments in weapons technology. The same is true of the Geneva Conventions of 1949, which regulate the conduct of armed conflict. These Conventions suggest that prohibitions on general forms of conduct may be relatively insensitive to technological change.

The Time is Not Ripe A fourth objection holds that it is simply too soon to negotiate an international treaty governing cyberwarfare. As Schmitt and Vihul argue (2016, 44), historically, treaties governing new weapons technologies have often been crafted only after the technologies have been in use for some time. Paradigmatic examples are the conventions governing anti-personnel landmines and cluster munitions. The reason, they argue, is that states are generally hesitant to restrict the use of weapons that may afford them an advantage on the battlefield until they have acquired sufficient experience to weigh the precise costs and benefits of doing so. A binding cyberwar agreement can therefore only be meaningfully pursued once states have become more familiar with emerging technologies and practices (Schmitt and Vihul 2014b, 18; Lucas 2017, 112).

Yet, this objection seems to be getting the problem backwards. It effectively argues that "ought" follows "is," and that the point of arms control agreements is simply to cement current state practice. Historically, states have often been a great deal more ambitious as the adoption of the Outer Space Treaty in 1967 shows. Along with the Environmental Modification Treaty of 1977, the Outer Space Treaty (which bans states from placing WMD in outer space) constitutes a farsighted bargain among states to forestall horrific threats to mankind made possible by new technologies *before* those technologies had fully matured or been put to use. A similar farsighted agreement is now required for cyberspace.

A Formal Treaty is Too Constraining A final objection to a treaty governing cyber conflict is aimed at its formal nature. International arms control agreements can take many forms. Some existing ACAs are highly specific and obligatory, while others amount to little more than loose gentlemen's agreements with no central monitoring, verification, or compliance apparatus. Most current advocates of closer international cooperation on cyber security favor lose norms and voluntary guidelines over binding, legal obligations. Indeed, some insist that a set of customary norms governing cyber conflict practices are

[34] See, e.g., proceedings of the 7th Review Conference of the BWC held in Geneva in 2011.

already emerging from state practice. Such a "soft law" approach, grounded in consensus and best practices, they argue, stand a greater chance of success than attempts at imposing formal legislation (Lucas 2016, 13; Lucas 2017, 112; Schmitt and Vihul 2016, 51).

To some, the appeal of a soft law approach rests mainly on the fact that cyber conflict is simply too novel—and the underlying technologies too fast evolving—for a binding international agreement to be possible (Finnemore 2011, 91; Nye 2010). Others argue that the cyber domain is characterized by strong conflicts of interest and highly fragmented power, which make formal agreement unattainable. These conditions favor decentralized cooperation based on flexible, non-binding commitment rather than a formal, centralized approach (see esp. Finnemore 2011; Schmitt and Vihul 2014a, b; Lucas 2017). In particular, starting with informal, voluntary rules would make it easier to persuade reluctant states to accept an agreement. Permissive rules could then be tightened down the line, when interests become better defined (Finnemore 2011. On the benefits of informal cooperation in conditions of uncertainty, see also Keohane and Raustialla 2009; Eilstrup-Sangiovanni 2009).

Contrary to these views, I insist on the need for a precisely codified and legally binding treaty to govern cyber conflict. Although there are clear benefits to informal cooperation in situations characterized by uncertainty and lopsided power-distributions (see Eilstrup-Sangiovanni 2009, 2014), four concrete features of cyber conflict speak strongly in favor of seeking a formal, obligatory treaty. The first is the large number of direct stakeholders in the cyber domain. As I have previously discussed, cyber weapons are within relatively easy reach of a range of different actors (state *and* non-state), and cyberattacks can be launched from—or routed through—almost any national jurisdiction. To avoid vast "havens" for cyber criminality, an international cyber agreement must therefore secure compliance by a majority of states in the system. Research on the design of international institutions has found that whenever the solution to a problem calls for cooperation among a large group of heterogeneous actors, a strong element of central authority is needed to coordinate action and monitor compliance. Informal, flexible agreements may suffice to elicit cooperation among smaller groups with relatively homogenous preferences (since uncooperative behavior among such groups is less likely and because effective peer-to-peer monitoring is more easily achieved among smaller numbers). However, the larger and more heterogeneous a group, the greater the demand for centralized coordination, monitoring, and verification to ensure cooperation (see Eilstrup-Sangiovanni 2009).

A second aspect of cyber conflict which favors a formal treaty is precisely the problem of verifiability which has been held up by some critics as a factor weighing against seeking an international convention. As discussed, difficulties of monitoring behavior in cyberspace means states will often be unsure about the actions taken by others, while difficulties of distinguishing benign from malign cyber activities may create uncertainty about others' preferences. According to rationalist cooperation theory, a standard response to uncertainty about behavior or preferences is to create highly detailed and obligatory agreements, which establish clear norms of prohibition and authorize strong sanctions against norm-breakers, and whose stringent implementation requirements can serve to expose uncommitted parties (See Lipson 1991, 50; Koremenos, Lipson & Snidal 2001, 787–8). An ICWC would establish a formal framework within which to anchor such norms.

A third, and closely related, argument in favor of a formal treaty is the need to foster robust norms of prohibition against cyberattacks (see Libicki 2009). Cyber weapons are both widely available and easy to conceal. Experience from other areas of international

arms control governing weapons with similar qualities (such as chemical and biological weapons) suggests a high value of establishing strong and unambiguous norms against use. Precisely because they are widely and cheaply available, it is crucial that the use of chemical and biological weapons is considered definitively out-of-bounds (see Price 1995).[35] The same is true of cyber-weapons. In an environment characterized by great uncertainty about the force postures and relative strength of potential adversaries, states must be able to trust that cyberattacks would meet with strong international condemnation and trigger severe sanctions, or else they will judge that the best strategy is to build-up their own armaments. This requires a formal treaty which spells out unambiguous rules of prohibition and establishes clear responsibility for responding to norm violations (Abbott and Snidal 2000).

A final argument in favor of a formal treaty is the need to ensure compliance at domestic level. Government officials who negotiate international norms and rules for the cyber domain may have difficulty persuading national legislatures to enact the necessary legislation to ensure effective implementation. IR-scholars have widely found that delegation to international organizations can serve as a potent tool for executives to strengthen their hand in domestic political battles. By accepting binding treaty obligations (as opposed to "voluntary norms and guidelines"), governments may thereby increase pressure on domestic law-makers to take adequate implementation measures.

Why Would States Sign Up? The perhaps strongest argument against seeking a formal treaty to govern international cyber conflict is that the insistence on binding rules will mean that only a small group of highly committed states are willing to sign on. Yet, if designed according to the principles and purposes outlined in this article, an ICWC would in fact offer a wide range of incentives for states to come on board. To begin, intelligence-sharing and a joint mechanism for attribution as outlined above would constitute important carrots for many states, as would access to technical assistance and funding to help improve national cyber defenses. One could imagine creating additional inducements to participation, such as joint disaster response. For example, a convention could oblige states to assist other parties if they came under severe cyberattack.[36] Similarly, one could envisage a multilateral fund for recovery and reconstruction which would provide financial assistance and expertise to help states repair and rebuild critical infrastructure after a cyberattack. In this regard, we can look to existing arms control treaties as a model. The CWC provides various forms of emergency assistance to state parties and establishes a voluntary fund for "Assistance and Protection against Chemical Weapons".[37] The Technical Secretariat also helps to improve national protection against chemical attacks by delivering detection equipment and alarm systems, protective equipment, and technical assistance to state parties.[38] Together with active support for compliance, these provisions have provided an

[35] The logic resembles that underlying domestic laws governing insurance fraud. Because insurance fraud is relatively easy to commit and difficult to expose, the penalty for being caught must be both severe and certain to deter fraud.

[36] Clarke and Knake 2010.

[37] Art. 8 of the CWC stipulates that the Technical Secretariat shall "coordinate the establishment and maintenance of permanent stockpiles of emergency and humanitarian assistance by States Parties."

[38] CWC 1997.

important incentive for states to participate. Similar provisions in the cyber domain could serve to encourage buy-in by reluctant states.

5 Conclusion

History teaches us that arms races are best tamed by formal international agreements carefully designed to reduce fear and uncertainty and to temper competitive dynamics through setting out clear rules for acceptable behavior. During the most hostile period of the Cold War, the ever-present risk of nuclear war was addressed by two main strategies: strategic nuclear deterrence based on a doctrine of mutually assured destruction and institutionalized arms control, manifested in a growing range of formal bilateral and multilateral treaty agreements. However, in today's cyber domain, only one strategy of conflict prevention is actively pursued, namely, strategic deterrence through retaliation.

To many observes, a one-sided reliance on strategic cyber deterrence is dictated by the current state of technology, which is said to favor offensive strategies and rule out negotiated solutions to conflict. Yet, there is in fact little basis for concluding that offensive strategies are strongly privileged in the cyber domain. Many sources of alleged offensive advantage (such as high speed of attack, anonymity, and a general irrelevance of physical distance or boundaries) introduce significant elements of uncertainty and mutual interdependencies which imply that offensive cyber strategies may risk backfiring. One must also keep in mind that although technological momentum can play a significant role in shaping international conflict, technological factors are rarely determinate (see Salzman 2013). As I have sought to demonstrate in this article, political choices can serve to increase the expected cost of aggression and lower the cost of defense, thereby altering the offense/defense balance in favor of more defensive strategies.

The arguments presented in this article should not be read as a statement of ingenuous optimism. Negotiating an ICWC will be fraught with difficulty given the diverse interests at stake, and the task will not be accomplished overnight. Nonetheless, I argue, most standard objections to embarking on a process of international negotiation (including difficulties of verifying compliance and problems of rapid technological change) fail on closer inspection. In the end, the main barrier to an international agreement on governing cyberconflict may be opposition by powerful (mainly) western states, whose desire to exploit current strategic advantages in the cyber-domain leads them to reject a treaty. For example, many observers judge that America's enduring hostility towards binding international rules for cyberspace is driven largely by its technological superiority in the realm of tactical electronic warfare, which provides a strong incentive to maintain maximum freedom of action in this domain (see Baruah 2013; Clarke and Knake 2010; Sanger 2015; Lindsay 2015, 46, 61–62; Sanger 2015; Goldsmith 2011). Not only would America be reluctant to bargain away its ability to exploit current tactical advantages, but as the world's strongest cyber power, American decision-makers may fear that by accepting binding international constraints on the conduct of cyber warfare, they would be trying their own hands while allowing other nations to rapidly catch up (see Singer and Friedman 2014).

Such reservations are too shortsighted. Given the current global rush to invest in offensive cyber power, America's unrivaled position in the cyber domain may well be

temporary. For now, however, America's leading position provides a unique opportunity to shape international rules and norms for cyber conflict. By assuming a global leadership role, Washington might succeed in encouraging other states to accept constraints on the use of cyber offensive power *before* these states acquire significant offensive capabilities of their own.[39] During the 1950s and 1960s, American leaders were sufficiently foresighted to see that they could exploit US superiority in the realm of nuclear technology to shape international rules regarding nuclear arms control. By promising self-restraint and by offering to share civilian nuclear technology, Washington succeeded in persuading other states to accept the norm of an international nuclear hierarchy and thereby prevented an uncontrolled global nuclear arms race. A similar opportunity is currently spurned in the cyber domain. The arguments presented in this article suggest it is time US policy-makers change course and give their support to negotiating an ICWC.

Open Access This article is distributed under the terms of the Creative Commons Attribution 4.0 International License (http://creativecommons.org/licenses/by/4.0/), which permits unrestricted use, distribution, and reproduction in any medium, provided you give appropriate credit to the original author(s) and the source, provide a link to the Creative Commons license, and indicate if changes were made.

Reference

Abbott, Kenneth W. (1993). "'Trust but verify:' the production of information in arms control treaties and other international agreements". *Cornell International Law Journal* 26.

Abbott, K., & Snidal, D. (2000). Hard and soft law in international governance. *International Organization, 54*(3), 421–456.

Baruah, Darshana M. (2013). "*Cyberspace governance: the American approach*", October 4. https://www.indiawrites.org/cyberspace-governance-the-american-approach/.

Buchanan, B. (2016). The life cycles of cyber threats. *Survival, 58*(1), 39–58.

Carr, Jeffrey. (2011). *Inside cyber warfare. Mapping the cyber underworld*, 2nd ed. (O'Reilly Media).

Cirenza, Patrick. (2016). "The flawed analogy between nuclear and cyber deterrence". *Bulletin of Atomic Scientists*, 22 February (http://thebulletin.org/flawed-analogy-between-nuclear-and-cyber-deterrence9179).

Clapper J.R. (2015) '*Statement for the record*: worldwide cyber threats, ' (http://www.dni.gov/index.php/newsroom/testimonies/209-congressional-testimonies-2015/1251-dni-clapper-statement-for-the-record,-wporldwide-cyber-threats-before-the-house-ptermament-select-committee-on-intelligence.

Clarke, R. A., & Knake, R. K. (2010). *Cyber war: the next threat to national security and what to do about it*. New York: Harper Collins.

Dewar, R. S. (2014). *'The "triptych of cyber security": a classification of active cyber defence.' 6th Intl conference on cyber conflict*. Tallinn: NATO CCD COE Publications.

Eilstrup-Sangiovanni, M. (2009). "Varieties of Cooperation: Government Networks in International Security." In *Networked Politics: Agency, Power, and Governance*, edited by Miles Kahler (Ithaca: Cornell University Press, 2009), chapter 10, pp. 194–227.

Eilstrup-Sangiovanni, M. (2014). "Network theory and security governance." In *Handbook of Governance and Security*, ed. James Sperling (Edward Elgar, Oct. 2014), chapter 3, pp. 41–62.

Evera, V., & Stephen. (1984). The Cult of the offensive and the origins of the first world war. *International Security, 9*(1), 58–107.

[39] Of course, the more advanced cyber powers would want to ensure that their current technological leads do not slip away through bootlegging and proliferation. This is best achieved by instituting informal export control regimes among smaller groups of advanced cyber states, much as has been the case for nuclear and chemical weapons. See Eilstrup-Sangiovanni 2009.

Ewing, Philip. (2015). "*The Pentagon's new cyber attack plan: 'blunt force trauma'*" Politico, 18 April (http://www.politico.com/story/2015/04/dod-hopes-cyber-can-create-blunt-force-trauma-117095) (Accessed, 12 January 2017).

Federal Bureau of Investigation (FBI). (2015). "*Cyber criminal forum taken down*". https://www.fbi.gov/news/stories/cyber-criminal-forum-taken-down/cyber-criminal-forum-taken-down.

Finnemore, Martha. (2011). "Cultivating international cyber norms." In *America's Cyber Future: Security and Prosperity in the Information Age,* eds. Kristin Lord and Travis Sharp, vol. II, pp. 89–100 (http://citizenlab.org/cybernorms2011/cultivating.pdf).

Gady, Franz-Stefan. (2017). "*Trump and offensive cyber warfare*". The Diplomat, January 16. http://thediplomat.com/2017/01/trump-and-offensive-cyber-warfare/ (**accessed** 14–03-17).

Glaser, C. L., & Kaufmann, C. (1998). What is the offense-defense balance and can we measure it? *International Security, 22*(4), 44–82.

Goldsmith J. (2011). "*Cybersecurity treaties: a skeptical view*" Hoover Institution (http://media.hoover.org/sites/default/files/documents/FutureChallenges_Goldsmith.pdf.

Gorman, Siobhan. (2009). "Electricity grid in US penetrated by spies". *The Wall St. Journal*. 8 April.

"Himes, Westmoreland (2015) members of cybersecurity subcommittee call for cyberwarfare rules", November 5, 2015, Issues: Intelligence, Science and Technology https://himes.house.gov/press-release/himes-westmoreland-members-cybersecurity-subcommittee-call-cyberwarfare-rules (Accessed 18 Dec, 2016).

Ifft, Edward. (2005). "*Witness for the prosecution: international organizations and arms control verification*". Arms Control Association, Nov.1. https://www.armscontrol.org/act/2005_11/NOV-Ifft.

Jervis, R. (1978). Cooperation under the security dilemma. *World Politics, 30*(2), 167–214.

Kelle, A. (2003). CWC report, The CWC after its first review conference: is the Glass Half Full or Half Empty? *The Acronym Institute, Disarmament Diplomacy*, issue no. 71, June-July 2003, http://www.acronym.org.uk.

Keohane, Robert O. and Kal Raustialla. (2009). Toward a post-Kyoto climate change architecture: a political analysis. In *Post-Kyoto International Climate Policy: Implementing Architectures for Agreement* (New York: Cambridge University Press).

Keohane, Robert O. and David G. Victor. (2015). "After the failure of top- down mandates: the role of experimental governance in climate change policy". In *Towards a Workable and Effective Climate Regime*, eds. Scott Barrett, Carlo Carraro, Jaime de Melo, pp. 201–212.

Koremenos, B., Lipson, C. & Snidal D. (2001). The rational design of international institutions. *Intenrational Organization, 55*(4), 761–99.

Kugler, R. L. (2009). "*Deterrence of cyberattacks*". *In Cyberpower and National Security, ed. Franklin D. Kramer et al*. Dulles: National Defense University Press and Potomac Books, Inc..

Lawson, Sean. (2011). "*Richard Clarke responds to administration cybersecurity proposals*" Forbes, June 3. (http://www.forbes.com/sites/seanlawson/2011/06/03/richard-clarke-responds-to-administration-cybersecurity-proposals/#42c5657ff80a).

Libicki, M. C. (2009). *Cyberdeterrence and cyberwar*. Santa Monica: RAND.

Limnéll, Jarno. 2013. "Offensive cyber capabilities are needed because of deterrence" In the fog of cyber defence, eds. Jari Rantapelkonen and Mirva Salminen (Helsinki: Juves Print), 200–208.

Lindsay, J. R. (2015). Tipping the scales: the attribution problem and the feasibility of deterrence against cyberattack. *Journal of Cybersecurity, 1*(1), 53–67.

Div Lior. (2016). "*Attack attribution does little to improve enterprise security*". Blog Post, Network World, July 29 (http://www.networkworld.com/article/3101727/security/attack-attribution-does-little-to-improve-enterprise-security.html). (Accessed, 10 March 2017).

Lipson, C (1991). Why are some international agreements informal? *International Organization, 45*(4), 495–538.

Lonergan, Shawn W. 2016. Cooperation under the cybersecurity dilemma" in Confronting inequality: wealth, rights, and power, eds. Hugh Liebert, Thomas Sherlock, and Cole Pinheiro (New York: Sloan, 2016).

Lucas, George R. Jr. (2016) "Emerging norms for cyberwarfare". In *Binary Bullets. The Ethics of Cyberwarfare*, eds, Fritz Allhoff, Adam Henschke and Bradley J. Strawser. Oxford University Press, pp. 13–33.

Lucas, G. (2017). *Ethics of cyber warfare. The quest for responsible security in the age of digital warfare*. Oxford: Oxford University Press.

Maurer, Tim and Robert Morgus. (2014). "*Compilation of existing cybersecurity and information security related definitions*". New America. https://www.newamerica.org/cybersecurity-initiative/policy-papers/compilation-of-existing-cybersecurity-and-information-security-related-definitions/ (Accessed 6 Nov. 2016).

Meyer, Paul. (2016). "Outer space and cyberspace: a tale of two security realms". In *International Cyber Norms: Legal Policy & Industry Perspectives*, eds. Anna-Maria Osula and Henry Rõigas (NATO CCD COE Publications), chapter 8, pp. 155–169.

Nye, Joseph S (2010). "*Cyber power*". Belfer Center for Science and International Affairs, Harvard Kennedy School, May 2010 http://belfercenter.ksg.harvard.edu/publication/20162/cyber_power.html.

Nye, Joseph S. (2015a). "*International norms in cyberspace*". Project Syndidate, May 11, (https://www.project-syndicate.org/commentary/international-norms-cyberspace-by-joseph-s–nye-2015-05?barrier=true).

Nye, Joseph S. (2015b). "*Opinions: the world needs new norms on cyberwarfare*". Washington Post, 1 October.

Paletta, Damian, Danny Yadron and Jennifer Valentino-Devries. (2015). "*Cyberwar ignites a new arms race*". Wall Street Journal.

Price, R. (1995). A genealogy of the chemical weapons taboo. *International Organization, 49*(1), 73–103.

Rid, T., & Buchanan, B. (2015). Attributing cyber attacks. *Journal of Strategic Studies, 38*, 4–37.

Saltzman, I. (2013). Cyber posturing and the offense-defense balance. *Contemporary Security Policy, 34*(1), 40–63.

Sanger, David E. (2015). "*U.S. and China seek arms deal for cyberspace*", September 19, 2015. (Washington D.C.).

Sanger, David E. and William J. Broad (2017) "Trump inherits a secret cyberwar against North Korean missiles". *New York Times,* March 4 . https://www.nytimes.com/2017/03/04/world/asia/north-korea-missile-program-sabotage.html.

Sanger, David E., John Markoff and Thom Shanker. (2009). "*U.S. steps up effort on digital defenses*". NY Times, April 27.

Schelling, T. C. (1994) *The strategy of conflict*. Cambridge, MA: Harvard University Press.

Schmitt, M. N., & Vihul, L. (2014a). Proxy wars in cyberspace: the evolving international law of attribution. *Fletcher Security Review, 1*(2), 55.

Schmitt, Michael N. and Liis Vihul. (2014b). "*The nature of international law cyber norms*, Tallinn Paper No. 5.

Schmitt, Michael N. and Liis Vihul: (2016) The emergence of international legal norms for cyberconflict. In Fritz Allhoff, Adam Henschke and Bradley J. Strawser (eds.), *Binary Bullets. The Ethics of Cyberwarfare* (pp. 34–55). Oxford University Press, 2016.

Sheldon, John B. (2011). "*Deciphering cyberpower strategic purpose in peace and war*". Strategic Studies Quarterly.

Simmons, B. A., & Hopkins, D. J. (2005). The constraining power of international treaties: theories and methods. *American Political Science Review, 99*(4), 623–631.

Singer, P. W., & Friedman, A. (2014). *Cybersecurity and cyberwar: what everyone needs to know?* Oxford: Oxford University Press.

Tallinn Manual on the International Law Applicable to Cyber Warfare. (2013). https://ccdcoe.org/tallinn-manual.html.

Tikk, E. (2011). Ten rules for cyber security. *Survival, 53*(3), 119–132.

United Nations, General Assembly. (2015). "*Group of governmental experts on development in the field of information and telecommunications in the context of international security*" *('UN Group of Cyber Security Experts')*. 70th session, A/70/171, 22 July 2015 (http://www.un.org/ga/search/view_doc.asp?symbol=A/70/174).

United States Department of Defense. (2006). "*National military strategy for cyber operations*", December 11, 2006, http://nsarchive.gwu.edu/NSAEBB/NSAEBB424/docs/Cyber-023.pdf.

Vanca, David. (2013). "Richard A. Clarke and Robert K. Knake's "Cyber war: the nextthreat to national security and what to do about it". Literature Reviews, *Georgetown Security Studies Review* 1(1).

Vink, Danny, (2015). "*America's secret arsenal*". Politico 9 December (http://www.politico.com/agenda/story/2015/12/defense-department-cyber-offense-strategy-000331).

Zetter, Kim. (2015). "*Why an arms control pact has security experts up in arms*". Wired. 24th July. (https://www.wired.com/2015/06/arms-control-pact-security-experts-arms/).