Retaliatory Deterrence in Cyberspace

Eric Sterner

The view that deterrence is of little value in securing the nation's information infrastructure is based on a Cold War model of strategic nuclear deterrence. If one examines other approaches to preventing attack, however, deterrence may make significant contributions to US security in cyberspace. Success, however, will require a new mind-set and changed expectations.

Deterrence is ingrained in US national security posture. It dominated Cold War debates and thinking about preventing Soviet aggression against vital US national interests. The lack of a direct US–Soviet war seemed to confirm its utility. Indeed, with the collapse of Soviet communism, deterrence advocates continued to proclaim its primary value in preventing aggression by lesser threats. In 1996, then-secretary of defense William Perry asserted, "And if these powers [rogue states] should ever pose a threat, our ability to retaliate with an overwhelming nuclear response will serve as a deterrent. Deterrence has protected us from the established nuclear arsenals for decades, and it will continue to protect us."[1] Yet, more than two decades into the information age, US policymakers are still working through its applicability in cyberspace. This article first examines cyber vulnerabilities then moves to cyberdeterrence alternatives. Finally it proposes a cyberdeterrent posture and policy.

Cyberspace: Vulnerabilities and Conflict

For the better part of two decades, analysts have recognized, and feared, the new national vulnerabilities that the information revolution created

---

Eric Sterner is a fellow at the George C. Marshall Institute, Washington, DC. He held senior staff positions at the House Armed Services and Science Committees and served in the Office of the Secretary of Defense and as NASA's associate administrator for policy and planning. He has written for several journals, including *Comparative Strategy*, *Washington Quarterly*, and the *Journal of International Security Affairs*. He is a graduate of the American University and the George Washington University.

The author would like to thank the Critical Infrastructure Protection Program of George Mason University's School of Law for sponsoring earlier work on deterrence in cyberspace, and Tim Clancy, Michelle Van Cleave, and Jeff Kueter for their comments on earlier drafts.

for the United States. In 1991, a landmark National Research Council (NRC) study concluded:

> We are at risk. Increasingly, America depends on computers. They control power delivery, communications, aviation, and financial services. They are used to store vital information, from medical records to business plans to criminal records. Although we trust them, they are vulnerable—to the effects of poor design and insufficient quality control, to accident, and perhaps most alarmingly, to deliberate attack. The modern thief can steal more with a computer than with a gun. Tomorrow's terrorist may be able to do more damage with a keyboard than with a bomb.[2]

If anything, the NRC underestimated the scope of the vulnerability. Computers and the networks that link them have only become more crucial to the functions of a twenty-first-century economy. Systemic infrastructure failures have already been attributed to problems in information networks. The Commission to Assess the Threat to the United States from Electromagnetic Pulse (EMP) Attack began life to examine a unique kind of nuclear effect. In doing so, it identified a common vulnerability across multiple national infrastructures, namely the proliferation and integration of systems controlled by networked computer chips through the use of embedded supervisory control and data acquisition (SCADA) systems, digital control systems (DCS), and programmable logic controller (PLC) systems.[3] All are vulnerable to electromagnetic pulse in one way or another. More importantly, they represent nodes in cyberspace (not all networks are connected to one another, but the trend is toward greater interconnectivity). Collectively, they represent a massive national vulnerability.

As the NRC predicted, malicious actors ranging from criminals and miscreants to terrorists and nation-states have exploited cyberspace vulnerability for a wide range of purposes. Attacks on commercial systems are a daily occurrence, and it is rare for more than a few days to pass before some company announces it has been attacked. Of late, Google and the instant messaging service Twitter are among the most well-known victims, but their stories are common.[4] A recent survey by the security firm Symantec found that 75 percent of corporate respondents had been attacked in the prior 12 months, and 41 percent of those attacks had been somewhat or highly effective. One hundred percent of respondents admitted to experiencing cyber losses in 2009.[5] One estimate puts 2008 global losses from cyber crime at $1 trillion.[6]

There is a temptation to view such activities as private matters, more suitable for law enforcement than national security. After all, the victims

are commercial entities, and the losses inflicted are nominally private losses. Gross damage to a private entity may be comparable to operating in a known flood plain, hurricane zone, or earthquake-prone area during a natural disaster. In other words, attacks and losses are viewed as the "cost of doing business."

Unfortunately, the vulnerabilities go well beyond simple private losses. They have the potential to affect the entire country. Demonstrating the vulnerability of critical infrastructure to attacks through cyberspace, the US government tested the ability to attack the electrical grid and successfully destroyed an electric generator by hacking a replica of a power plant's control systems.[7] Press reports suggest that power grids in the United States and elsewhere have been penetrated by malicious foreign actors who have done real damage, causing blackouts in multiple cities.[8]

Indeed, the world is awash in cyber conflicts. At least three high-profile international conflicts have been reasonably well- and widely documented: the Israeli-Palestinian conflict of 2000, the Russo-Estonian conflict of 2007, and the Russo-Georgian conflict of 2008.[9] These are not isolated instances. Cyber attacks for ostensibly political purposes occur routinely. They may or may not involve governments. The United States and South Korea were both struck almost simultaneously by several waves of cyber attacks in the summer of 2009.[10] Attacks on Google's Chinese services clearly had political overtones, and Chinese-origin attacks are quite common around the world.[11] There are even signs of ongoing cyber conflicts between al-Qaeda and some of its Islamic opponents as well as a sectarian cyber conflict in the Persian Gulf.[12]

It does not come as a surprise that the United States, as the lone superpower, would find itself on the receiving end of such attacks. In 2007, the Department of Defense identified 43,880 malicious attacks against itself, rising to 54,640 in 2008, and 43,785 just through the first half of 2009.[13] The defense secretary's unclassified e-mail account was breached, and department officials report hundreds of thousands of cyber probes each day. Additionally, in 2007, NASA and the Departments of State, Homeland Security, and Commerce all reported major intrusions resulting in lost data and interrupted operations.[14]

Quite simply, the United States is already engaged in conflict in cyberspace and has been for years. Gen James Cartwright, then-commander of US Strategic Command, testified before the Senate Armed Services Committee:

> However, not unlike the targets of pirates or train robbers of the past, *America is under widespread attack in cyberspace*. Our freedom to use cyberspace is threatened by the actions of criminals, terrorists, and nations alike. Each seeks their own form of unique advantage, be it financial, political, or military, but together they threaten our freedom to embrace the opportunity offered by a globally connected and flattened world. The magnitude of cost, in terms of real dollars dedicated to defensive measures, lost intellectual capital and fraud cannot be overestimated, making these attacks a matter of great national interest. Unlike the air, land and sea domains, we lack dominance in cyberspace and could grow increasingly vulnerable if we do not fundamentally change how we view this battle-space (emphasis added).[15]

More recently, the former director of national intelligence, VADM Mike McConnell, who also served as director of the National Security Agency, stated quite bluntly, "The United States is fighting a cyber-war today, and we are losing."[16]

Deterrence and Cyberspace

The United States has responded to cyberspace as a national security domain in a variety of ways, primarily through improved defense and closer public-private cooperation and coordination. Nevertheless, as fundamental as deterrence is in US national security policy, it is not always clear how it relates to cyberspace. Many focus on the challenges of preventing attacks on or through cyberspace and are skeptical about the prospects for deterrence to contribute to this goal.

Their reasons are straightforward. It becomes quickly apparent that traditional models of deterrence have little relevance to cyberspace. Strategic nuclear deterrence theory, for example, largely presumes a stable bipolar relationship between nation-states of roughly equal power (made so by the possession of nuclear weapons) that share similar expectations and seek to avoid nuclear warfare at all costs, as it threatens each state's supreme interest in its own survival. Theoretically, these nation-states possess the perception and communication skills needed to manage a crisis successfully and avoid the worst possible outcomes. Acknowledging that reality fell well short of the abstract concept, Western policymakers sought to promote deterrence by addressing shortfalls in these key ingredients through force structure, arms control, improved decision making, and better communication links. Thus, deterrence was elevated from a tactic in international relations, to a strategy, to a means of cooperatively managing the superpower

relationship.[17] The concept was so well enshrined in Western strategic culture that some scholars even advocated—or at least argued for tolerating—the modest proliferation of nuclear weapons, whose destructive capabilities theoretically leveled the relative power imbalance, induced a particular clarity in decision making, and otherwise increased peaceful stability in the international system.[18]

Setting aside powerful critiques of strategic nuclear deterrence, none of the elements that purportedly made it successful are present in cyberspace. The number of actors possessing nuclear weapons has been historically low; only nation-states possessed the wherewithal to develop such capabilities. By contrast, the number of actors in cyberspace is astronomically high, growing rapidly, and constantly changing in character, thereby undermining stability, communication, and clarity.[19] Indeed, one might view cyber actors as a threat cloud, constantly evolving and changing shape.

Rather than symmetrical bipolar relationships, cyberspace is governed by a potentially infinite number of asymmetrical, multilateral, and bilateral relationships that are constantly in flux. Stakes, interests, power, and defenses all vary, while ambiguity will be prevalent before, during, and after engagements.

Perhaps the greatest problem encountered when applying strategic deterrence models to cyberspace is the difficulty of identifying the challenger and appropriate retaliatory targets. This was not a problem in traditional models of deterrence, whether nuclear or conventional. Theoretically, an attacker's identity would always be known; only nation-states possessed the capability of launching significant military attacks. Actors in cyberspace, however, are "created" in cyberspace. They may or may not correspond to the creator's identity in the real world. The legal, political, economic, and geographic characteristics that describe an actor in the physical world are not constraining in cyberspace. Worse, a cyberspace actor may not be persistent. It may be created and exist for the short time necessary to launch an attack, only to be quickly discarded after the fact. Thus, if one is to retaliate against a cyberspace actor in the physical domain—where retaliatory options historically lie—by legal, political, economic, or military means, one must first establish connections between the cyberspace actor and his or her physical-world counterpart. For many, this so-called attribution problem is insurmountable. Also, if the cyber attacker is not a nation-state, retaliation may involve impinging on the sovereignty of the country in which the cyber attacker is physically located or of the country(ies) through which the attack was launched. Thus, retaliation has

a high likelihood of collateral damage. In some cases, a challenger might launch an attack simply to provoke retaliation to advance some other political interest. In such cases, the threat of retaliation might actually invite the attack!

Alternatives to Deterrence

Left with few retaliatory options, the defender can only hope to ensure that its defenses are better than the challenger's offenses and take steps to manage the risks and consequences of losing the offense-defense interaction. Martin Libicki, who thoroughly analyzed cyberdeterrence and found it wanting, recommended an approach akin to safety engineering.[20] More recently, Greg Rattray noted parallels between public health and cyber security and suggested drawing from public health risk-management models to help secure cyberspace.[21] Because cyberspace is not defined nationally, it is necessary to improve its overall resistance to malicious behavior and, in so doing, improve the US defense posture. Rattray recommends improving partner security capacity and capabilities, engaging and supporting multi-stakeholder international organizations, and encouraging network operator groups to play active roles in making systems more resistant to attack. He concludes: "The United States should take lessons from public health efforts at national and global levels. Specifically, the federal government should support public-private collaboration that enables early warning of new threats, rapid response to contain the spread of malware, and long-term commitment to eradicating the malicious activity that often thrives in the cyber commons."[22]

Ultimately, resilience and flexibility become key for defense. It absolutely is necessary to improve the resiliency of cyberspace writ large, not to mention our own use of it, and our flexibility to deal with attacks—successful or otherwise—to improve our posture in an offense-defense interaction. Moreover, managing and minimizing the risks and consequences of an attack may dissuade some attackers by denying them the object of their attack. Deterrence by denial is a well-accepted posture. In the end, risk- and consequence-management policies will help allocate resources more efficiently than the ad hoc approach used now. In many ways, they will remain the main line of security in cyberspace after straightforward defense. Nevertheless, their limitation lies in the fact that they divorce cyber security from cyber conflict and the attack from the attacker.

Conflict involves interaction between conscious actors, each of which behaves in a way intended to defeat the other relative to the stakes of their conflict. Each will employ a strategy it thinks will advance its goals. Viruses, worms, safety flaws, and the like are not willful; they certainly do not employ conscious strategies for the purposes of defeating their victims. Rather, they are merely tools reflecting the intent and capabilities of an attacker or the vulnerabilities of a defender.

A risk/consequence–management policy framework pays less attention to threats as a function of intent and capability. Therefore, it may blind the defense to sudden changes in the nature of the threat, either in terms of general attitudes toward the United States or the particular goals and stakes of a specific engagement or campaign. It may also create a class of malicious behavior for which we are unprepared to hold actors responsible, with a new set of tools to employ against US national interests in conjunction with more traditional geopolitical maneuvers. Separating the attack from the intent of the attacker begins to break down the fundamental ingredients of a successful campaign. Defense shifts from an interaction between belligerents to an interaction of weapons. Of course, one cannot prevail in a cyber conflict any more than a conflict in other domains if one only thinks about it at this level.

This is a crucial challenge. Dominant modes of analysis seek to segment threats into a variety of categories based on a mix of factors, usually including the actor's physical description (criminal, nation-state, corporation), motives (criminal, harassment, political, strategic), target, and consequences of the attack.[23] Risk management ultimately focuses on the highest-risk challenges and may pay less attention to lower-level threats, such as criminal activity, or those primarily affecting private persons. Unfortunately, ambiguity in cyberspace creates incentives and opportunities for one kind of attacker to disguise itself and its motive for an attack. It also means that what appears to be one kind of attack may, in fact, be a particular tactic in another. For example, states may use front groups to assemble botnets which they rent out for criminal activity and use to launch distributed denial-of-service attacks as a distraction for something more decisive elsewhere. An activist may simply find itself the covert recipient of sufficient government funds to "rent" cyberspace weapons and launch harassment attacks. In other words, intentions and capabilities are subject to rapid change. Yesterday's criminal threat is tomorrow's strategic attack. With that in mind, it behooves a defender to pay excruciatingly

close attention to such dynamics lest it miss the suddenness with which a cyberspace threat to its security might arise.

Finally, too heavy an emphasis on a risk-management approach largely cedes the initiative to a challenger. Because it is focused on reducing vulnerabilities and minimizing consequences, it is largely reactive to a specific attack or campaign. A conflict typically involves both defense and offense, even if the offense is limited to counterattacks. Without imposing the consequences of a counterattack—strategic, operational, or tactical—on an attacker, the defender is merely taking a beating.

An Alternative Model

The limits of risk management and the offense-defense interaction return one to the discussion of deterrence. Its perceived limitations, however, are drawn from analysis of our Cold War experience with strategic nuclear warfare. As it turns out, much of this analysis used the wrong deterrence model.

Many treated Cold War strategic deterrence as a binary switch: deterrence prevents conflict; if a conflict breaks out, deterrence has failed. In 2001, Secretary of Defense Donald Rumsfeld restated the point in his forward to the *Quadrennial Defense Review Report*: "The strategy that results is built around four key goals . . . [including] decisively defeating any adversary if deterrence fails."[24] This view may be a relic from theories associated with nuclear weapons. Taking state survival as paramount, theorists concluded that nuclear war was always unacceptable and, therefore, to be avoided at all costs.

As discussed earlier, cyberspace, and American interests in it, are already under attack. Conflicts within cyberspace are continual, with relative peaks and valleys in the intensity of their connection to politics. A deterrence model that focuses on the prevention of armed conflict will thus fall short—the conflict is already underway. In the end, it may not be that deterrence falls short in cyberspace; merely that the deterrence model against which most analysts measure the cyber conflict problem falls short.

The chairman of the Joint Chiefs of Staff, ADM Michael Mullen, noted that US deterrence theory had not appreciably improved in 20 years and concluded, "We need a new model for deterrence theory, and we need it now. . . . We need to be ready—actually and completely—to deter a wide range of new threats. It is not just about cleaning someone else's clock

anymore. We need a new model of deterrence that helps us bring our own clock up to speed with the pace and the scope of the challenges of this new century."[25] Indeed, more than one model will be necessary. The need is particularly acute in cyberspace.

The role deterrence can play in shaping, containing, or even preventing a continuation of ongoing conflict is intuitive but often ignored in analyses of deterrence in cyberspace. During Operation Desert Storm, for example, US policymakers signaled clearly enough to Iraqi leaders that the United States could respond to Iraq's use of weapons of mass destruction by escalating its war aims to include regime change.[26]

More generally, deterrence threats can be used to affect a challenger's choices of means and aims in a conflict. Throughout its history, Israel has sought to deter attacks from nonstate actors by changing the nature of its conflict with those actors. It countered cross-border Palestinian raids, for example, by threatening and conducting retaliatory attacks against Jordan and Egypt, each of which had greater reason to fear Israeli retaliatory threats and possessed capabilities to threaten and punish Palestinian raiders.[27] In other words, Israel combined threats and actions to change the nature of the conflict in an attempt to create a better situation for itself. This "active deterrence" reflected a combination of the actual use of force and threats of force to achieve its security goals. Doron Almog offers an updated concept, dubbing it "cumulative deterrence." For him, "cumulative deterrence is based on the simultaneous use of threats and military force over the course of an extended period of conflict."[28] Israel's readiness to change the strategic dynamic of a conflict if necessary by escalating it horizontally or vertically has established a deterrent posture that effectively prevents some attacks and contains the dynamics of conflicts within certain boundaries. Consequently, Israel is able to wage conflicts on more-favorable terms that have the potential to limit the conflict and, ideally, bring peace. Unlike nuclear deterrence, which focuses on preventing conflict, these concepts revolve around shaping it over time.

Might such a posture be more appropriate for cyberspace? Certainly it suggests there is less reason for despair about deterrence than some have assumed. Of course it involves changing expectations. Law enforcement accepts imperfect deterrence as the nature of the beast rather than dismissing the concept entirely. The same can be said for cyberspace. Resigned admonitions to avoid overwrought strategic metaphors for security in cyberspace and instead approach threats by ascribing to defense the more pedes-

trian status of "safety engineering" are well heeded, but they should not become an excuse for forgoing deterrent options. Instead, it will be necessary to view cyberspace attackers as thinking beings who engage in some form of cost-benefit calculus and then seek to change that estimation in their minds. Deterrence in cyberspace will be far from perfect, but it is also far from hopeless.

Toward a Cyberdeterrent Posture and Policy

In moving toward a cyberdeterrent posture, the United States will need to change the strategic dynamic of the conflict. It will not be effective simply to meet challengers on their terms, at the times and places of their choosing. Doing so cedes the initiative, gives them an opportunity to continually probe and identify vulnerabilities, and enables them in advance to lay out lines of retreat from an engagement should the offense-defense interaction go badly.

First and foremost, the United States must retaliate for malicious cyber behavior. Today, US officials often consider punishing cyber aggressors through domestic law enforcement, largely because those means are readily available. Such tools are entirely inadequate. Domestic statutes regarding cyber crimes typically: (1) require prosecutors to attribute a monetary value to the damage inflicted, which may be irrelevant or inappropriate for national security matters; (2) utilize high evidentiary standards associated with criminal prosecution and its presumption of innocence; and (3) assume that a criminal defendant can be made to stand trial.[29] As a practical matter, these tests cannot reliably be met in cyber attacks that cross territorial boundaries, they are inadequate for dealing with harassing attacks or those that share traits with espionage, and they are inappropriate for dealing with state-sponsored or state-sanctioned cyberspace attacks. Moreover, such retaliation is extraordinarily slow with an extremely low likelihood of execution. Indeed, successful prosecutions are still remarkable events, largely because they are so rare relative to the scale of attacks.

Other retaliatory options will be needed. Political, economic, and military means must be explored. While usually considered in the context of state-to-state relationships, these methods have been used against nonstate actors for a variety of purposes, including advancing nonproliferation agendas and fighting the global war on terror. In the case of political and economic retaliation, the threshold needed to justify imposing sanctions

should be lower, usually left to the discretion of the president once he is confident that certain conditions have been met.

Kinetic and cyber retaliation are more problematic, due in part to questions of proportionality, collateral damage, and attribution. Kinetic measures may be precise but generally not precise enough to get the NRC's proverbial terrorist-with-a-keyboard without doing considerable collateral damage. Moreover, it can be argued that the prospect of taking life in a kinetic attack far outweighs the damage one can commit with a cyber attack; that is, it is disproportional. Richard Harknett summed up the dilemma:

> At its core, deterrence theory rests on the principle of retaliation in kind, where the cost inflicted in retaliation will at least match the level of costs associated with the offensive attack. If an attack reduces no buildings to rubble and kills no one directly, but destroys information, what is the response? We tend to think about information as intangible, but the loss of information can have tangible personal, institutional, and societal costs. What credibly can be placed at risk that would dissuade a state from contemplating such an attack?[30]

The dilemma is more simply framed as a "bits-for-lives" trade-off, in which the value placed on the challenger's life is always higher than the value placed on the defender's bits. Presumably, the United States values lives more than bits, so any retaliatory threats are not credible. Framing the dilemma in this manner is too limiting.

The United States has employed military measures in cases where its values, interests, and international prerogatives were at stake but its national survival was not. In the 1980s, it used force in Grenada and Panama because US citizens were threatened. In the 1980s and 1990s, it used force against Libya in retaliation for terrorist attacks in Europe; in the Persian Gulf to preserve the global flow of oil; in Lebanon, Somalia, and Haiti for peacekeeping and humanitarian reasons; and in the Balkans to prevent ethnic cleansing. Thus, the threat of force in retaliation for cyber attacks that adversely affected vital national interests in some meaningful way seems eminently credible, the concern over trading lives for bits notwithstanding. Certainly, the United States possesses the ability and has demonstrated the will to use force in instances that fall well below the threshold of national survival. Thus, if—and this is a big "if"—the United States can identify an attacker with enough confidence to permit retaliation, military options should be available.

Questions of proportionality go well beyond a lives-for-bits trade-off. Traditionally, the concept is drawn from theories of justice, whether in

war or the legal system. The punishment should fit the crime, as it were, and every military provocation should not necessitate a massive response. That said, in and of itself, cyber conflict lies somewhere between the two. It may not rise to the level of warfare, but the legal system is often inadequate to deal with it as a strategic tool in international relations. Meanwhile, small attacks of modest intent may have immense consequences, even perhaps inadvertently, as they propagate through global networks. Conversely, massive attacks of aggressive intent may have modest consequences, particularly if they are poorly executed or the target has effectively defended against them and/or taken steps to minimize the damage. Thus, concepts of proportionality drawn from other domains are out of place. Policymakers will ultimately have to decide what constitutes a proportional response on a virtual case-by-case basis, taking into account a variety of factors ranging from the attacker's intent, consequences of the attack, and confidence levels in identifying responsible parties to the strategic situation, concerns about repeat attacks, and available retaliatory options. Many of these judgments will have to be incorporated into rules of engagement to enable the defenders of cyberspace engaged in the conflict to make decisions about counterattacks, just as police and soldiers in the field are trusted with judgments about the use of lethal force.

There appears to be an unwritten assumption that knowing the physical-world identity of a cyber attacker is a prerequisite to retaliation. This is eminently reasonable when one's primary retaliatory tools were designed for attackers in the physical world. But, the challenge of cyberspace—that it is not limited by the physical world (even if it does not exist independent of the physical world)—also represents an opportunity. Instead of trying to fit the square pegs of retaliatory options developed for the physical world into the round holes of cyberspace, the United States needs to develop and employ policies, doctrine, tools, deterrent models, and rules of engagement for cyber retaliation against actors in cyberspace. In other words, it needs the ability to retaliate against cyber attackers without necessarily knowing who they are in the physical domain.

The challenge of identifying retaliatory targets remains. Attribution, however, is not an insurmountable problem. Many factors come into play. First, technical tools for identifying sources of cyber aggression are constantly improving. In studying an attack or the creation of offensive cyber capabilities, it is often possible to identify e-mail accounts, Internet service providers, and even servers from which certain kinds of behavior emanate.

Joseph Menn recently documented the efforts of a private security expert working with British and Russian law enforcement to track the online behavior of criminal gangs and defeat their attacks on private web business. In particular, he noted the success of nongovernment groups and individuals in building thorough profiles of malicious cyber actors, sometimes even tying them to their counterparts in the physical domain.[31] According to public reports, researchers identified websites during the Russo-Georgian cyber conflict of 2008 hosting downloadable "weapons," traced activities to computers known to be controlled by Russian organized crime, and linked related Internet traffic to servers controlled by Russian telecommunications firms.[32] Islamist websites contain instructions and links to means of cyber attack.[33] One such site, Al-jinan.org, offered downloadable software to attack a preapproved list of Internet protocol addresses and a simple Windows interface that enables the visitors to conduct attacks at their leisure, based in part on the speed of their connection to the Internet.[34] Some ostensibly legitimate businesses are even selling "hacks" and other software vulnerabilities to the highest bidder.[35] In short, in some significant cases it is possible to identify specific sources of cyber attack.

Secondly, strategic context matters. The Russo-Estonian cyber conflict did not occur in a vacuum but in the context of an ethnic dispute inside Estonia, to which Russia became a party. Similarly, the Russo-Georgian cyber conflict occurred against the backdrop of a physical invasion of the latter. This is not to suggest that an underlying strategic situation will definitively identify an attacker. Indeed, criminals may be motivated to take advantage of international crises; states engaged in a type of attrition cyber attack may engage in most activity at relatively peaceful times so as not to exacerbate a political conflict; and, third parties may well seek to disguise their activities to create a political crisis between two other parties. Nevertheless, policymakers should consider the strategic situation both in assessing an attack and executing retaliatory options. That context will contribute to confidence levels in attributing an attack and selecting a particular means of punishing an aggressor.

Thirdly, the United States can hold third parties accountable commensurate with their role in enabling or allowing cyber attacks that do it harm. Unlike other conflict domains (sea, air, land, and space), cyberspace is a created medium. Someone owns the servers, nodes, transmission lines, and infrastructure that create cyberspace and enable it to function. Arguably,

today we have established a norm of irresponsibility that holds these owners and creators harmless for third-party damages done by, with, or through the things they create. Establishing a deterrent will require defenders to put cyberspace creators on notice that they will be held accountable for use of their creation. Such an approach need not be always adversarial. More often than not, the interests of the government in deterring attacks will coincide with the interests of cyberspace creators in preserving the value and utility of their creation. For example, in 2008, while investigating a web-hosting firm engaged in suspicious activity, reporters from the *Washington Post* approached the enterprise running the server farm on which the hosting company had based its business. Shortly thereafter, the server farm disconnected the web-hosting company from its servers, and security experts noted a significant drop in global spam activity.[36] Should cooperative efforts fail, however, escalating horizontally to the creators of cyberspace will change their interests such that they use the leverage they have over the users of their infrastructure to constrain attacks.

The United States might start down this path by putting cyberspace actors on notice that it will hold them accountable for how their creation is used, perhaps by creating blacklists of bad actors who consistently tolerate malicious cyber attacks over or through their infrastructure. Persistent toleration of such attacks may become sufficient grounds for some form of retaliation by political, economic, cyber, or kinetic means.

It will be tempting to draw "redlines" and clarify what kinds of malicious behavior one is attempting to deter. One might understandably focus on deterring some sort of cyber Pearl Harbor or other nightmare scenario that involves widespread economic damage. Of course, clear redlines signal that malicious activity falling below that threshold is of less concern, inviting attackers to continue their efforts there. Rather than drawing specific redlines, the United States needs to consider a range of retaliatory options to use against a range of threats that it may not be able to rank hierarchically, given the speed with which threats might change. Thus, cyber attacks should be no more tolerable than major attacks on strategic infrastructure. Neither gets a "pass," as it were. If there is a parallel in the physical domain, the concept of "broken window" law enforcement comes to mind. By stopping small infractions, one creates a cumulative effect that deters bad actors from escalating to more serious behavior.[37]

Over time, a commitment to retaliation for cyber attacks by a variety of means (political, economic, military, or cyber) and a willingness to hold

cyberspace creators accountable for their role in permitting or enabling attacks will create a deterrent posture. By no means will the United States be able to retaliate for every attack, but visible retaliation will create risk for potential attackers, affecting their cost-benefit analysis. Those cyberspace actors contemplating attacks on the United States will have to consider the potential punishment that such an attack might invite. Similarly, those who own and maintain the infrastructure of cyberspace will have to weigh the risks of allowing their infrastructure to be used at will by various cyberspace attackers. Presumably, at least a portion of them will improve their situational awareness and be more accommodating to cyberspace defenders, lest they become retaliatory targets themselves.

The United States cannot adopt such a posture tomorrow or simply through declaratory statements. It will require sophisticated rules of engagement, careful mapping of global cyber networks to better anticipate secondary or tertiary consequences, accelerated development of advanced forensic tools, and improved retaliatory capabilities, ranging from cyber weapons and limited war plans to presidential sanction authority and international cooperation to identify cyber attackers and the legal means of punishing them. Careful study of the potential unintended consequences will be necessary. Finally, it will take a series of visible retaliatory actions—political, economic, military, and cyber—over time to create a reasonable, if not certain, expectation of the risk of punishment for potential attackers. These specific measures go well beyond the scope of this article. Moreover, developing these tools may take years, while the cyber threat is here now.

Conclusion

Conflict in cyberspace does not fall squarely within the bounds of law enforcement or traditional warfare. As a unique environment with unique actors, power distributions, and interests, it represents something else entirely. With that in mind, it is necessary to develop new intellectual frameworks for understanding cyber conflict and securing US interests. Simply importing concepts and thought processes from other domains will prove entirely inadequate. Strategic nuclear deterrence is unique to a nuclear environment; indeed, it may well be unique to the Cold War.[38] It does not represent a useful posture for cyberspace. That does not mean deterrence has no value. A more forward-leaning posture that incorporates the realities of cyberspace is necessary.

To be sure, the deterrent posture laid out herein may be controversial. It should be. An immense amount of study, analysis, and additional work is needed to understand the dynamics of cyber conflict, how different retaliatory options might affect attackers, the most useful means of holding an attack's enablers accountable, escalatory ladders, authorities, roles, and missions. Moreover, Americans are reluctant to escalate conflicts vertically or horizontally. Although the United States has done so in the past, holding third parties responsible for their toleration or enabling of bad actors adds risk to any given conflict. Nevertheless, the alternatives are insufficient. Risk management, consequence management, and the offense-defense interaction create a policymaking framework that may cede the initiative to attackers. Given the stakes involved for the United States, policymakers must explore all measures available to improve US security. Deterrence in cyberspace will not become a first, second, or even third line of defense. Risk and consequence management and the improvement of defenses at the point of attack are likely to long dominate US security in cyberspace. But, deterrence may yet contribute to security by helping contain the severity and frequency of attacks and focusing attention on cyber conflict as the interaction of conscious actors whose decision-making processes can be influenced. SSQ

Notes

1. Quoted in Keith Payne, *The Fallacies of Cold War Deterrence and a New Direction* (Lexington: University Press of Kentucky, 2001), 82.

2. National Research Council, *Computers at Risk: Safe Computing in the Information Age* (Washington: National Academies Press, 1991), 7.

3. Commission to Assess the Threat to the United States from Electromagnetic Pulse (EMP) Attack, *Critical National Infrastructures Report*, April 2008, 1–16, http://www.empcommission.org/docs/A2473-EMP_Commission-7MB.pdf.

4. Jessica Vascellaro, "Hackers Briefly Bring Down Twitter," *Wall Street Journal*, 18 December 2009; Nik Cubrilovic, "The Anatomy Of The Twitter Attack," *TechCrunch.com*, 19 July 2009; and Siobhan Gorman and Jessica Vascellaro, "Google Attack Linked to Asian Hackers," *Wall Street Journal*, 22 February 2010, http://online.wsj.com/article/SB10001424052748704751304575080362745174130.html?mod=WSJ_hpp_MIDDLTopStories.

5. Symantec, *State of Enterprise Security, 2010*, http://www.symantec.com/content/en/us/about/presskits/SES_report_Feb2010.pdf. Another security company, McAfee, sponsored a Center for Strategic and International Studies (CSIS) survey report, *In the Crossfire: Critical Infrastructure in the Age of Cyber War*, which revealed similar findings.

6. David Z. Bodenheimer, *Statement Before the House Armed Service Committee's Subcommittee on Terrorism, Unconventional Threats and Capabilities Concerning Private Sector Perspectives on Department of Defense Information Technology and Cybersecurity Activities,* 10 February 2010, 4.

7. Jeanne Meserve, "Sources: Staged cyber attack reveals vulnerability in power grid," *CNN*, 26 September 2007, http://www.cnn.com/2007/US/09/26/power.at.risk/index.html; and Glenn Derene, "How Vulnerable is U.S. Infrastructure to a Major Cyber Attack?" *Popular Mechanics*, October 2009, http://www.popularmechanics.com/print-this/4307521.

8. Siobhan Gorman, "Electricity Grid in U.S. Penetrated by Spies," *Wall Street Journal*, 8 April 2009; Michael Mylrea, "Brazil's Next Battlefield: Cyberspace," *Foreign Policy Journal*, 15 November 2009, http://www.foreignpolicyjournal.com/2009/11/15/brazils-next-battlefield-cyberspace; and Tom Espiner, "CIA: Cyber attack caused multiple-city blackout," *c/net news*.com, 22 January 2007, http://www.news.com/CIA-Cyber attack-caused-multiple-city-blackout/2100-7349_3-6227090.html.

9. For a discussion, see Col Patrick Allen and Lt Col Chris Demchak, "The Palestinian-Israeli Cyberwar," *Military Review*, March–April 2003; Brian Krebs, "Report: Russian Hacker Forums Fueled Georgia Cyber Attacks," *Washington Post*, 17 October 2008, http://voices.washingtonpost.com/securityfix/2008/10/report_russian_hacker_forums_f.html?hpid=sec-tech; Stephen W. Korns and Joshua E. Kastenberg, "Georgia's Cyber Left Hook," *Parameters* 38, no. 4 (Winter 2008/2009): 60–76; Eneken Tikk et al., *Cyber Attacks against Georgia: Legal Lessons Identified* (Talinn, Estonia: NATO Cooperative Cyber Defence Centre of Excellence, August 2008); Joshua Davis, "Hackers Take Down the Most Wired County in Europe," *Wired Magazine*, issue 15.09 (21 August 2007); and Richard A. Clarke and Robert K. Knake, *Cyber War* (New York: HarperCollins, 2010). Given the strategic context and manner in which the attacks unfolded, circumstantial evidence suggests that governments may have colluded in the attacks, although such collusion may not have been necessary for attacks to take place.

10. John Sudworth, "New 'cyber attacks' hit S. Korea," *BBC News*, 9 July 2009, http://news.bbc.co.uk/2/hi/asia-pacific/8142282.stm; and James A. Lewis, "The 'Korean' Cyber Attacks and their Implications for Cyber Conflict," Center for Strategic and International Studies, October 2009.

11. Bryan Krekel et al., "Capability of the People's Republic of China to Conduct Cyber Warfare and Computer Network Exploitation," prepared for US–China Economic and Security Review Commission, 9 October 2009.

12. Ellen Knickmeyer, "Al-Qaeda Web Forums Abruptly Taken Offline," *Washington Post*, 18 October 2008, A-01.

13. US–China Economic and Security Review Commission, *2009 Report to Congress* (Washington: US–China Economic and Security Review Commission, November 2009), 68.

14. CSIS, *Securing Cyberspace for the 44th Presidency* (Washington: CSIS, December 2008), 12–13.

15. Gen James Cartwright, USMC, *Statement Before the Strategic Forces Subcommittee, Senate Armed Services Committee*, 28 March 2007, 4–5.

16. Mike McConnell, "To win the cyber-war, look to the Cold War," *Washington Post*, 28 February 2010, B-01.

17. See, for example, Patrick Morgan, *Deterrence Now* (Cambridge, UK: Cambridge University Press, 2003), chap. 1. The logic of deterrence, as well as its flaws, is considerably more complex than this. Other factors such as psychology, cognition, certainty of retaliation, offense-dominance, and the like come into play. Developed in the context of strategic nuclear deterrence, each of these factors also, arguably, limits the utility of traditional concepts of deterrence in cyberspace.

18. See Kenneth Waltz, "More May Be Better," in *The Spread of Nuclear Weapons: A Debate*, eds. Scott D. Sagan and Kenneth Waltz (New York: W. W. Norton & Company, 1995); Martin van Creveld, *Nuclear Proliferation and the Future of Conflict* (New York: Free Press, 1993); Devin Hagerty, "Nuclear Deterrence in South Asia: The 1990 Indo-Pakistani Crisis," *International Security* 20, no. 3 (Winter 1995/96); and John Mearsheimer, "Back to the Future: Instability in Europe after the Cold War," in *The Perils of Anarchy: Contemporary Realism and International Security*, eds. Michael Brown, Sean Lynn-Jones, and Steven Miller (Cambridge, MA: MIT Press, 1995).

19. Some estimates indicate the number of Internet users rose from roughly 360 million in 2000 to 1.8 billion by 2010. See http://www.internetworldstats.com/stats.htm.

20. Martin Libicki, *Defending Cyberspace and Other Metaphors*, (Washington: National Defense University [NDU] Press, 1997), 41, fn. 1, 107–8. Libicki urges policymakers to adopt a philosophy that will lead "information warfare . . . to acquire the pedestrian status of safety engineering." For a more recent and thorough analysis, see Libicki, *Cyberdeterrence and Cyberwar* (Santa Monica, CA: RAND, 2009).

21. Greg Rattray, Chris Evans, and Jason Healey, "American Security in the Cyber Commons," in *Contested Commons: The Future of American Power in a Multipolar World*, eds., Abraham Denmark and James Mulvenon (Washington: Center for a New American Security, 2010), chap. 5.

22. Ibid., 167.

23. See David S. Alberts, *Defensive Information Warfare* (Washington: NDU Press, 1996), 22–39; Irving Lachow, "Cyber Terrorism: Menace or Myth," in *Cyberpower and National Security*, eds., Franklin Kramer, Stuart Starr, and Larry Wentz (Washington: NDU Press, 2009), 437–64. Alberts' analysis, while dated, has held up well in subsequent studies.

24. *Quadrennial Defense Review Report* (Washington: DoD, 2001), iii–iv.

25. ADM Michael Mullen, "From the Chairman: It's Time for a New Deterrence Model," *Joint Force Quarterly*, issue 51 (4th Quarter, 2008): 2–3.

26. See James A. Baker III, *The Politics of Diplomacy: Revolution, War & Peace*, 1989–1992 (New York: G. P. Putnam's Sons, 1995), 359. The role of Baker's threat in Iraq's nonuse of weapons of mass destruction remains debated. See, for example, Avigdor Haselkorn, *The Continuing Storm: Iraq, Poisonous Weapons, and Deterrence* (New Haven, CT: Yale University Press, 1999).

27. Jonathan Shimshoni, *Israel and Conventional Deterrence: Border Warfare from 1953 to 1970* (Ithaca, NY: Cornell University Press, 1988).

28. Doron Almog, "Cumulative Deterrence and the War on Terrorism," *Parameters* 34, no. 4 (Winter 2004/2005): 8.

29. See, for example, 18 USC § 1030, "Fraud and Related Activity in Connection with Computers," available from the Department of Justice Computer Crime and Intellectual Property Section, along with relevant sentencing guidelines tied to damages, at http://www.justice.gov/criminal/cybercrime/cclaws.html#fedcode. There is a robust and rapidly growing body of literature on laws relating to cyber conflict, including relevant bodies of international law, the Laws of Armed Conflict, international law, espionage statutes, and criminal and civil codes. For a useful summary discussion of the issues, see *Technology, Policy, Law, and Ethics Regarding U.S. Acquisition and Use of Cyber Attack Capabilities* (Washington: National Academies Press, 2009), chap. 7.

30. Richard J. Harknett, "Information Warfare and Deterrence," *Parameters* 26, no. 4 (Autumn 1996): 93–107. Not all deterrence concepts are limited by the need to retaliate in-kind or proportionally. Instead, at its core, deterrence rests on the defender's threat to impose costs on a challenger that exceed the challenger's willingness to pay, at least in the challenger's mind.

31. Joseph Menn, *Fatal System Error* (New York: Public Affairs, 2010).

32. John Markoff, "Before the Gunfire, Cyber Attacks," *New York Times*, 13 August 2008, http://nytimes.com/2008/08/13/technology/13cyberhtml?_r=1&pagewanted=print; Brian Krebs, "Report: Russian Hacker Forums Fueled Georgia Cyber Attacks," *Washington Post*, 17 October 2008, http://voices.washingtonpost.com/securityfix/2008/10/report_russian_hacker_forums_f.html?hpid=sec-tech. See also Tikk et al., *Cyber Attacks against Georgia*.

33. Steve Coll and Susan B. Glasser, "Terrorists Turn to the Web as Base of Operations," *Washington Post*, 7 August 2005, A-01.

34. Larry Greenemeier, "'Electronic Jihad' App Offers Cyberterrorism for the Masses," *InformationWeek*, 2 July 2007.

35. Brian Krebs, "Auction of Software Flaws Stirs Concerns," *Washington Post*, 13 July 2007, D-01, http://www.washingtonpost.com/wp-dyn/content/article/2007/07/12/AR2007071202070_pf.html. As of 14 October 2008, the website for the company was still active and advertising for additional researchers.

36. Brian Krebs, "Major Source of Internet Spam Yanked Offline," *Washington Post*, 12 November 2008, http://www.washingtonpost.com/wp-dyn/content/article/2008/11/12/AR2008111200658_pf.html.

37. George Kelling and James Q. Wilson, "Broken Windows," *Atlantic Monthly*, March 1982, http://www.theatlantic.com/magazine/archive/1982/03/broken-windows/4465/; and Kelling, "'Broken Windows' Works," *Forbes.com*, 16 July 2009, http://www.manhattan-institute.org/html/miarticle.htm?id=5091.

38. For that matter, strategic nuclear deterrence may be unique to the United States. There is evidence that the Soviet Union did not share the concept.