Deterrence of Cyber Attacks and U.S. National Security

Charles L. Glaser
Professor of Political Science and International Affairs
Elliot School of International Affairs
The George Washington University

Report GW-CSPRI-2011-5

June 1, 2011

Abstract

This paper draws on deterrence theory to analyze the challenges that the United States faces in deterring cyber attacks. We begin by briefly reviewing the basic logic of deterrence theory and relating it to the challenge posed by cyber attacks. The following section explores what is commonly viewed as the key problem in deterring cyber attacks—the "attribution problem" arises when a state cannot determine who has attacked it and therefore cannot credibly threaten to respond. We suggest that this barrier to deterrence has been exaggerated, while acknowledging that it does create a number of dangers. The following two sections discuss deterrence of different types of cyber attacks—those designed to damage the U.S. economy and society, and those designed to weaken U.S. conventional military forces. The final section highlights a few points, including the need for the United States to design a clear declaratory policy that explains its cyber deterrence strategy and the importance of integrating deterrence into a multilayer policy designed to protect the United States from cyber attacks.

Work supported by the Office of the Vice President for Academic Affairs and the School of Engineering and Applied Science of the George Washington University

Deterrence of Cyber Attacks and U.S. National Security

Charles L. Glaser
Professor of Political Science and International Affairs
Elliot School of International Affairs
The George Washington University

Deterrence basics

In broad terms, we can envision protecting the United States with three separable, but complementary, layers of capability. The first layer is deterrence—capabilities and policies designed to convince an adversary not to launch a cyber attack. The second layer is defense—capabilities designed to reduce the effectiveness of the adversary's cyber attack. The third layer is reconstitution and robustness—capabilities designed to enable U.S. systems to continue functioning once they have suffered cyber damage and to enable the United States to restore and rebuild its cyber capabilities after they have been damaged.

These layers achieve their objectives in different ways. Deterrence influences the adversary's intentions, convincing an adversary not to attack; defense works against the adversary's capabilities, defeating attacks that the adversary launches; reconstitution and robustness reduce the implications of successful attacks by the adversary. The layers complement each other by making up for limitations in other layers. If deterrence were known to be perfect, defense and reconstitution would be unnecessary; similarly, if defense were perfect, deterrence and reconstitution would be unnecessary. But, when none of the layers is perfect, each contributes to a country's overall ability to protect itself. My paper focuses on deterrence, among other reasons because the effectiveness of the other layers hinges primarily on technical considerations.

Deterrence theory was developed in the 1950s and 1960s primarily to address the new strategic challenges posed by nuclear weapons. Since then scholars have explored deterrence of conventional attacks, the relationship between the credibility of various type of deterrence commitments, deterrence of terrorists, and a variety of additional extensions and applications.[1] Deterrence involves convincing an adversary not to take an action by leading the adversary to believe that the costs of pursuing the action will exceed its benefits. An attacker's basic deterrence calculus depends on four components: 1) the benefits of taking the action—the larger the benefits, the harder the adversary is to deter; 2) the probability of achieving the benefits—the higher the probability, the harder the adversary is to deter; 3) the costs the defender will impose if the adversary takes the action—the higher the costs, the more likely the adversary is to be deterred; and 4) the adversary's assessment of the probability that the defender will inflict these costs—the higher this probability, the more likely the adversary is to be deterred. This last factor—the probability that the defender will carry out its deterrent threat—is commonly termed the credibility of the threat and has often been one of the thorniest issues for strategists to deal with. When the expected costs of an attack exceed the expected benefits, an attacker will be deterred.

In terms of these four components, deterrence is frequently divided into two types—deterrence by punishment and deterrence by denial. When relying on a strategy of deterrence by punishment, the United States threatens to inflict costs in retaliation for being attacked. The effectiveness of deterrence by punishment depends on both the size of the costs being threatened and the credibility of the threat. Credibility depends on both the ability to retaliate and the will to retaliate. The credibility of its nuclear threats was a major concern for the United States during the Cold War because the United States was defending allies—which it valued less than its own country and, therefore, was willing to run only smaller risks to protect—and was highly vulnerable to Soviet nuclear escalation. While there was no doubt about U.S. ability to inflict massive retaliatory damage, many U.S. leaders worried about the effectiveness of the U.S. nuclear deterrent due to doubts about its credibility.

For analyzing a deterrence-by-punishment strategy for dealing with cyber attacks we will need to assess the credibility of U.S. threats for responding to cyber attacks. Here we flag three issues. First, the most commonly cited barrier to deterring cyber attacks is the "attribution problem": most analysts believe that the United States will have great difficulty determining who launched a cyber attack; if the United States is not confident about who launched an attack, then it may be unwilling to retaliate, and an attacker that recognizes this problem will doubt the credibility of U.S. threats. Second, the credibility of U.S. threats will require the attacker to believe that the United States has the ability to retaliate. This could pose different challenges in the cyber realm than in the kinetic realm. The United States can demonstrate its conventional and nuclear capabilities by buying forces, testing these systems, and engaging in training and exercises, all of which are observable (to varying degrees) by its adversaries. In contrast, U.S. offensive cyber capabilities may be entirely invisible. In addition, they may be untested against adversary systems, leaving the adversary with some doubt about the effectiveness of U.S. capabilities. Third, the United States could threaten traditional kinetic attacks in response to a cyber attack, but this would likely raise different doubts about U.S. credibility, reflecting among other things concerns about the appropriateness of escalating from cyber to kinetic attacks and concerns about the risks to the United States because this escalation might lead the adversary to escalate to still higher levels of conflict.

Deterrence by denial works by a different logic: in this approach, the United States deploys capabilities to convince its adversary that the probability of its attack succeeding are low; this reduces the expected benefits of the attack and can therefore result in successful deterrence. We see here a close relationship between the defense layer and the deterrence layer: defensive cyber capabilities that the adversary believed would be effective can convince the adversary not to attack in the first place. Pure denial strategies have limitations: even if an adversary believes that its attack is unlikely to succeed, he may not be deterred if the costs of attacking are low. For example, some scholars have expressed concern about conventional military strategies that emphasize deterrence-by-denial, because the key costs for the adversary of launching an attack are limited to the potential loss of soldiers and military material. This criticism was leveled at NATO's conventional strategy during the Cold War.[2] The problem is almost certainly worse for deterrence of cyber attacks because attacking would be essentially costless.[3] A partial "solution" is to integrate denial and punishment strategies, combining the ability to defeat attack with the threat to retaliate.

Cyber deterrence and the attribution problem

Many experts are quite pessimistic about the feasibility of attribution. For example, William Lynn, the U.S. Deputy Secretary of Defense recently wrote, "The forensic work necessary to identify an attacker may take months, if identification is possible at all."[4] Richard Clarke reports that a leading group of cyber experts concluded that it is "fruitless" to try to attribute the source of cyber attacks.[5] This view, however, may exaggerate the attribution problem by overlooking either the purposes of the attacker or the scenario in which the attack occurs.[6]

A state that launches a "countervalue" attack against the United States' economic infrastructure, economy and/or society is likely to have a political purpose. Possible purposes could include compelling the United States to make political concessions during a crisis before a war starts, compelling the United States to stop fighting a war, and reducing the U.S. ability to fight a war by weakening its economy and industrial infrastructure. For these compelling threats to be effective, the state would have to make demands and spell out its threat. In addition, it would have to provide the United States with some confidence that attacks would stop if the United States meets that attacker's demands. These communication requirements would largely eliminate the attribution problem. For the scenario of attacking to weaken the U.S. ability to fight, the country the United States was fighting would be immediately identified as the likely suspect; the possibility that the United States would likely come to this conclusion could be sufficient to deter the adversary's cyber attack. Alternatively, the attacker might not be deterred because the costs of U.S. retaliation were not large compared to the costs of the on-going war; but in this case the failure of deterrence would not result from the attribution problem but instead from the size of the retaliatory costs the United States was threatening.

Of course, actors that lack political objectives are not covered by this argument. Terrorist groups are therefore a natural concern, as they are often viewed as motivated simply by the desire to damage the United States. A very different perspective disagrees, however, arguing that terrorist groups, including al Qaeda, are motivated by political goals and use terror attacks as a means to achieve their political ends.[7] If this is the case, a terrorist group will find itself facing communication requirements that are not unlike those facing states. A terrorist group might be hard to deter by retaliation because there are no good targets to hit in retaliation, and almost certainly no important cyber targets, but again the difficulty of deterrence would not result from attribution problems, but the more familiar problem of threatening attacks that would inflict sufficiently high costs on a terrorist group. Another type of actor that might be of concern here are hackers who are motivated by the technical challenge of undermining U.S. cyber systems and not by political objectives.

The attribution issue for "counterforce" attacks—those directed against U.S. capabilities—is quite different, but may be even less of a problem than with counter value attacks launched by states. This type of attack is most likely to occur during a crisis or war, with the adversary employing the cyber attack to gain a military advantage. Attribution will likely not be a problem, because the United States will know which state it is involved within a conflict. This is not to say that deterring this type of attack will not be difficult; it might be for reasons other than attribution. This is a separate issue that we deal with briefly below.

All of this said, the difficulty of attribution does create a variety of potential dangers. One possibility is dangerous mischief: a third party—country, terrorist group, or hacker—could

launch a cyber attack against the United States while it was involved in a crisis or war with another state. Based on the logic sketched above, this could lead to misattribution, because the United States' first inclination would likely be to attribute the attack to the country it was already fighting. Consequently, the third party could use such an attack to generate escalation in the on-going conflict, with the goal of increasing the damage that the United States and/or its adversary would suffer. Another problem is that the inability to attribute attacks undermines the U.S. ability to deter (and otherwise respond) to much lower level cyber attacks, including data stealing, espionage, and disruption of commerce. At a minimum, attribution would enable the United States to try to deter these types of attacks by promising to pursue legal actions. But for the most part, these types of attacks do not threaten vital U.S. national security interests, so from a security perspective the attribution problem does not generate large risks.

Deterring coercive countervalue cyber attacks

A standard deterrent strategy for deterring countervalue attacks is to threaten similar damage in retaliation. In the nuclear realm, holding the adversary's cities hostage—that is, vulnerable to retaliation—is considered the basic requirement for deterring the attacks against one's own cities. The analogy in the cyber realm would be to threaten a cyber attack that would inflict comparable damage against the same type of targets that the adversary had attacked.

But this raises the question of whether the United States should rely on cyber retaliation to deter cyber attacks. Because deterrence works by threatening costs with sufficient credibility, not by threatening specific types of attacks, this type of retaliation-in-kind is not strictly necessary for deterrence to be effective. Instead, the United States could threaten to use conventional weapons to inflict damage in retaliation. If the United States wanted to make clear that it was attempting to inflict comparable damage (for example, to avoid further escalation), it could attack similar targets. For example, if the adversary's cyber attack had destroyed part of the U.S. electric grid, oil refineries, and/or pipelines, the United States could attack these infrastructure targets in retaliation. Alternatively, except when facing a major power, the United States could threaten to invade the attacker's country or impose a new regime, if the country launched a highly destructive cyber attack against the United States.[8] These costs would be very different from those imposed by the adversary's cyber attack, but there is no reason that the costs have to come in similar types for an adversary to be deterred.

Deciding whether to rely on cyber retaliation or alternative types of retaliation is a major project that is beyond the scope of this short paper. Here we offer a few brief comments that suggest directions for further analysis. First, traditional kinetic capabilities have the advantage of being relatively easy to demonstrate and observe. As noted above, this could add to the credibility of kinetic threats compared to cyber threats. Second, a related point is that the United States would likely have greater confidence in its kinetic capabilities than its offensive cyber capabilities, because it would have been unable to test the latter, at least not fully. Third, the impact of kinetic attacks is likely easier to anticipate than the impact of cyber attacks. If the United States wants to inflict a given amount of damage—to avoid inappropriate escalation or even to signal its willingness to deescalate—then it would see advantages in attacks that would result in damage that was relatively easy to estimate in advance and that would be easy to evaluate once they had occurred. Experts worry that cyber attacks could result in large uncertainties, leaving both the attacker and the attacked unsure about how much damage had been inflicted.[9]

But the case here is not entirely one sided—cyber deterrent threats could also have some advantages. First, cyber retaliatory attacks might provide a clearer means of tacit bargaining: the adversary is more likely to recognize a cyber attack as retaliation for its own cyber attack. Second, and related, cyber retaliation-in-kind would have benefits if cyber attacks were understood to represent a threshold between levels of violence. In this case, if the United States prefers that a cyber conflict not to escalate to conventional or nuclear war, respecting the cyber threshold would help to avoid escalation, while pursuing interwar deterrence.

Whatever type or types of attacks the United States decides should constitute its strategy for deterring countervalue cyber attacks, the United States should develop a declaratory policy that lays out how it will respond, and why. Deterrence depends on the adversary understanding the threatened consequences. Laying out ahead of time the type and spectrum of responses can help a state clarify its threats and develop its adversary's expectations. This will be especially important if the United States finds that it requires not only the ability to deter initial cyber attacks, but also a more complex deterrence strategy that would enable it to engage in limited cyber wars in which cyber attacks are used for bargaining. Developing a well designed declaratory policy will be particularly important if the United States decides to rely on non-cyber retaliation, or to complement cyber retaliation with conventional attacks.

Deterring counter-military cyber attacks

Deterring counter-military attacks presents a host of different issues. First, deterring cyber attacks in isolation is probably not the key to deterring this type of attack. Both the United States and its adversary are likely to consider counter-military cyber attacks to be part of their overall conventional fighting capability. Within types of weaponry and warfare, the United States has traditionally distinguished between conventional and nuclear warfare, and also made distinctions concerning chemical and biological weapons. In terms of counter-military attacks, cyber attacks may well not be considered a different type of warfare. Instead, counter-military cyber attacks are more likely to be viewed as a component of conventional warfare. This would be in line with current categorizations, which for example, include electronic warfare assets as an element of conventional capabilities. Similarly, imagine a cyber attack that damaged U.S. command and control capabilities. Why should the United States response to this attack, or its deterrent threat that is designed to prevent the attack, be different if the damage is done by a kinetic attack rather than by a cyber attack?

Second, if the preceding line of argument is correct, then the challenge the United States faces in deterring counter-military cyber attacks is to be able to deter the adversary's overall conventional attack, including the offensive cyber capabilities that would be a component of this attack. This overall deterrence will depend on relative U.S. cyber capabilities, including both its ability to defend against the adversary's cyber attacks and its ability to use offensive cyber attacks to weaken its adversary's overall conventional capability. But, deterrence will depend still more broadly on how U.S. conventional capabilities compare to its adversary's. The adversary could be deterred from launching a conventional attack, including its counter-military cyber component, if the United States has the ability to win a conventional conflict, even if its adversary enjoys a cyber advantage. And, more in line with standard worries, an adversary that enjoys a net advantage in counter-military cyber capabilities might not be deterred, even if U.S. conventional forces are otherwise clearly superior. In any event, the basic point here is that the

impact of cyber capabilities on deterrence has to be understood in terms of their net impact on U.S. overall conventional capabilities.

Third, counter-military cyber capabilities would likely increase states' uncertainty about their conventional capabilities, which could make failures of deterrence more likely. Theorists have argued that uncertainties about the outcome of a war are a fundamental source of bargaining and deterrence failures. Uncertainty about outcomes and, closely related, disagreements about the outcome of a war, can prevent states from reaching a political bargain that they prefer to war.[10] Therefore, if cyber capabilities are potent enough to significantly influence assessments of war outcomes, then the increased uncertainty they will introduce could make war more likely.

Concluding thoughts

Deterring cyber attacks may not be as difficult as the emerging conventional wisdom suggests. This is partly because the attribution problem may be less severe than is generally believed. Because states are driven by political motives, they will be unable to use countervalue cyber attacks to achieve their objectives without making known their identities. A state will also likely be able to identify the source of counter-military attacks because these attacks will be most important in the context of a conventional war.

To support its deterrence policy, the United States needs a clear declaratory policy that lays out its plans for responding to various types of attacks. If the United States plans to rely partly on kinetic attacks and conventional operations to deter certain categories of cyber attacks, this should be spelled out to increase the probability that adversaries appreciate the breadth of the United States' cyber deterrence strategy.

Finally, because even a well designed deterrence policy could fail, the United States must pay attention to the other layers that can contribute to protection from cyber attacks—both defense, and reconstitution and robustness undoubtedly have important roles to play and contributions to make. In addition to the direct protection this capability can provide, they can also contribute to the U.S. ability to deter cyber attacks because asymmetries in the ability to inflict cyber damage, especially countervalue damage, could provide a state with bargaining advantages. Evaluating the proper balance between these three layers of protection promises to be a highly complex, technical and imprecise enterprise. The brief evaluation presented in this paper suggests that cyber deterrent capabilities and strategy are sufficiently promising that they should not be the neglected as United States develops an integrated policy for reducing the danger posed by cyber attacks.

References

[1] Key early works include Glenn H. Snyder, *Deterrence and Defense: Toward a Theory of National Security* (Princeton: Princeton University Press, 1961); and Thomas C. Schelling, *Arms and Influence* (New Haven: Yale University Press, 1966). On deterrence before the nuclear age, see George H. Quester, *Deterrence Before Hiroshima* (New York: Wiley, 1966); on conventional deterrence see John J. Mearsheimer, *Conventional Deterrence* (Ithaca: Cornell University Press, 1983); for a thoughtful review, see Robert Jervis. "Deterrence Theory Revisited," *World Politics*, Vol. 31, No. 2 (Jan. 1979), pp. 289-324. On applying established deterrence concepts to cyber deterrence, see Patrick M. Morgan, "Applicability of Traditional Deterrence Concepts and Theory to the Cyber Realm," in Committee on Deterring Cyberattacks, National Research Council, *Proceedings of a Workshop on Deterring Cyber Attacks: Informing Strategies and Developing Options for U.S. Policy* (2010), at www.nap.edu/catalog/12997.html.

[2] Samuel P. Huntington, "Conventional Deterrence and Conventional Retaliation in Europe," *International Security*, Vol. 8, No. 3 (Winter 1983-84), pp. 32-56.

[3] A possible cost is that the defender will learn about the attacker's offensive cyber capabilities, resulting in a significant diminution of its future offensive cyber capabilities.

[4] William J. Lynn III, "Defending a New Domain," *Foreign Affairs*, Vol. 89, No. 5 (September/October 2010), pp. 97-198.

[5] Richard A. Clarke and Robert K. Knake, *Cyber War: The Next Threat to National Security and What to Do About It* (New York: HarperCollins, 2010), p. 132.

[6] For this perspective and some of the issues we raise below, see Richard L. Kugler, "Deterrence of Cyber Attacks," in Franklin D. Kramer, Stuart H. Starr, and Larry K. Wentz, *Cyberpower and National Security* (Washington, D.C.: National Defense University Press, 2009).

[7] See for example Robert Pape, *Dying to Win: The Strategic Logic of Suicide Terrorism* (New York: Random House, 2005).

[8] The possibility of relying on threats of different types of costs inflicted by different means has been identified as an option for the United States in deterring biological attacks; see for example, Victor A. Utgoff, "Nuclear Weapons and the Deterrence of Biological and Chemical Warfare," Occasional Paper No. 36 (Washington, D.C.: Henry L. Stimson Center, October 1997).

[9] On this point and many related ones, see Martin C. Libicki, *Cyberdeterrence and Cyberwar* (Arlington, VA: RAND, 2009), chap. Three.

[10] See for example, James D. Fearon, "Rationalist Explanations for War," *International Organization*, Vol. 39, No. 3 (1995), pp. 379-414; and Robert Powell, *In the Shadow of Power: States and Strategies in International Politics* (Princeton: Princeton University Press, 1999).