

Cyberdeterrence between Nation-States

Plausible Strategy or a Pipe Dream?

Jonathan Solomon

CAN DETERRENCE strategies which have helped avert great-power war since the late 1940s similarly prevent attempts at corrupting, disrupting, or destroying a nation-state's vital information infrastructure? Is it even possible to deter state-executed or state-sponsored computer network attacks against the governmental, military, financial, industrial, and civil information systems upon which modern society relies so heavily? It seems intuitive that one nation-state could plausibly threaten another with debilitating punishment should the latter conduct a major "cyber attack" against the former. Examining computer network warfare characteristics, however, reduces one's optimism that even the most powerful states can deter strategic-level cyber attacks using the same methods that deter nuclear and conventional adventurism. While the threat of punishment certainly has a role in "cyberdeterrence" between nation-states, the defining variable in an aggressor's decision making will likely be whether or not the defending state can credibly retain use of its most potent tools of national power despite being subjected to an intense "cyber offensive."

This article examines the plausibility of potential US strategies for deterring such strategic cyber offensives, both during and outside of open war. We first study the characteristics of classical deterrence theory to derive the prerequisites for fielding a credible cyberdeterrent. We next explore the severe challenges decision makers face in attempting to attribute cyber attacks to other nation-states. We then look at where punitive cyberdeterrence thresholds might lie, how these would need to be coordinated with US nuclear and conventional military deterrence postures and assess the numerous efficacy questions surrounding punitive cyberdeterrence. Lastly, we suggest how cyberdeterrence by denial may actually be the stronger and more credible strategic path for the United States. We do not examine

Jonathan Solomon is a systems and technology analyst for Systems Planning and Analysis, Inc., in Alexandria, Virginia. He previously served as a surface warfare officer in the US Navy.

Disclaimer

The views and opinions expressed or implied in the SSQ are those of the authors and should not be construed as carrying the official sanction of the United States Air Force, the Department of Defense, Air Education and Training Command, Air University, or other agencies or departments of the US government.

questions of extended deterrence or deterring non-state-sponsored sub-state actors due to the substantially greater complexity of these issues. While these are valid questions for deterrence policy, their exploration is deferred for future analysis.

Theories of Deterrence

Deterrence is the art of convincing an enemy not to take a specific action by threatening it with intolerable punishment and/or unacceptable failure. In his classic work *Arms and Influence*, Nobel laureate and strategic theorist Thomas Schelling notes that successful deterrence using either punishment or denial methods depends upon effective communication between a state and the entity it wishes to deter. Though simple-sounding in theory, successfully signaling precise messages across cultural, linguistic, ideological, and strategic situational boundaries is extremely difficult in practice. Adversaries must first receive the messages communicating redlines and threats, recognize that these are indeed messages of deterrence, interpret the messages the way they were intended, and lastly, evaluate the declarations in the messages as credible.¹ In order to be effective, these messages must lead adversaries to conclude that the probable costs of taking proscribed actions outweigh desired gains.

Written toward the middle of the Cold War, Schelling's work largely focuses on deterrence by punishment, since it was widely presumed at the time that any direct conflict between the United States and the Soviet Union would have led to nuclear exchanges. This rendered deterrence by denial less credible as a primary strategy, given the capability of nuclear weapons to readily overwhelm passive defenses. Likewise, neither side possessed the capability to confidently conduct disarming first strikes against the other's nuclear arsenal. Under such circumstances, Schelling observes, deterrence can be well served by the appearance that decision makers may be unable to prevent or control their state's retaliatory response even if they want to do so, thereby creating a perceivable risk that aggressive actions will lead to an automatic response, irrevocably immolating both sides. He points out that the potential degree of self-inflicted pain makes little difference so long as the adversary state can be convinced that the defending state would willingly suffer this damage to gravely penalize the former for taking the proscribed action. If anything, Schelling suggests, credible expression of this extreme degree of will signals the defender's commitment to

detering the adversary's aggression. He stresses, though, a state must do more than merely threaten automated retaliation and signal its associated resolve. It must also overtly configure its deterrent forces so that allies and adversaries alike understand that the retaliatory threat can and will be carried out without fail. This can be done by laying a declared, visible, and unambiguous tripwire that if breached brings the two sides into direct contact and obliges the deterring state's response. This can also be achieved by making it clear the adversary's transgressions will ignite an inherently uncontrollable process that almost certainly ends in a conflagration neither side wants. Both of these approaches involve the defending state deliberately passing the initiative to its adversary in the form of a mutually recognized "last clear chance" to avoid collision.²

Another major aspect of deterrence by punishment is ambiguity. Schelling notes that in practice many deterrent threats are left deliberately ambiguous to provide the deterring state's decision makers with situational flexibility. Alternatively, deterrent threats might be left ambiguous because the complexity of a notional situation makes prior precise definition of deterrence thresholds impossible. This inevitably leads the adversary to conduct probes attempting to identify where the actual redlines lie, to erode the deterrent's credibility, and to restore some freedom of action. Schelling observes that the only way to parry these probes is to develop a reputation for unpredictable and potentially escalatory responses. One example might be by not cooperating with adversaries' attempts to extricate themselves from embarrassment when caught probing.³

Addressing deterrence by denial, Schelling notes that defense and deterrence combine when a deterrent threat cannot be made credible in advance. By demonstrating that the aggressor will pay a heavier than anticipated price for victory, the defender aims to deter its opponent from continuing to attack. Schelling distinguishes between a "pure, forcible defense," which aims not to deter the adversary but rather to prevent it from successfully achieving its objectives, and a "deterrent, coercive defense," which inflicts costs and pain on the adversary in hopes of convincing it not to proceed further.⁴ In a scenario where the attacker repeatedly penetrates the defense but is unable to effectively reduce the defender's resistance, the attacker may be unable to endure the mounting political and resource costs of continuing the offensive, not to mention the retaliatory pain inflicted by the defender. The attacker might therefore be pressured into halting its aggression well short of its objectives. It follows that if a

defender can communicate the probable resiliency of its defense to a potential aggressor, the defender might be able to influence the adversary's cost-benefit calculations and thereby deter an attack in the first place.

Under most conceivable circumstances, cyber warfare's lethality and capacity for permanent destructiveness is not equivalent to the nuclear and grand conventional warfare of which Schelling wrote. Yet, Schelling's outline contains many key factors which must be considered when crafting any potential strategies of cyberdeterrence by punishment. First, the list of proscribed activities and corresponding deterrent threats against them must be communicated to potential adversary states in ways which ensure the message will be recognized and interpreted correctly. Second, decision makers must understand that if they cannot clearly define redlines in advance due to complex circumstances or a desire for flexibility, they must communicate a credible willingness to act rashly or uncooperatively to protect the deterrent's ambiguity from erosion. Third, all redlines and threats must be made credible by decision makers either overtly demonstrating their resolve to act or creating visible mechanisms which would unquestionably force their hands when in extremis. Credibility demands that the defender's physical ability to carry out the retaliatory threat cannot be in doubt.

It is also important to note that for the deterrent to be considered credible, the retaliatory pain threatened by the defender against the aggressor must not be less than what the aggressor would inflict by taking the proscribed action against the defender. The defender's retaliatory threat does not necessarily have to be proportional in terms of potential inflicted damage so long as it would deeply harm things the aggressor is unwilling to trade off to attain its political objective. Nor does it necessarily have to be symmetrical to the aggressor's threatened attack. For instance, the defender can choose to threaten broad-based economic or kinetic retaliation against a specifically targeted cyber attack instead of or in addition to threatening a similarly targeted retaliatory cyber attack. The above questions of symmetry and proportionality are inherently political, as they must balance the deterrence-enforcing signals the defender wishes to transmit against the risks of escalation and diplomatic or domestic blowback if the retaliation must be executed.

Schelling's outline takes one critical aspect of deterrence by punishment for granted. In nuclear or grand conventional warfare, it is axiomatic that the defending state can attribute an attack to a specific adversary

state. A ballistic missile in flight can be detected and tracked by satellites or terrestrial radar, and its trajectory can be calculated back to its launch point. An armored brigade or an aircraft sortie crossing a border will eventually be detected and identified. Even if these weapons have no national markings or suppress their signatures, other unmistakable political or military signals soon reveal the attacker's identity with high confidence. A massive and sustained physical offensive leaves no attribution doubt. As potential aggressors are well aware, this ability to attribute responsibility with high confidence supports an adequately armed defending state's swift, harsh, and credible retaliation. Unless a cyber attack or cyber offensive is conducted in coordination with such physical attacks, cyber attribution will not enjoy the same level of clarity as the physical domain. This has major implications.

The Attribution Challenge

For cyberdeterrence by punishment to work, the defender must be able to identify the attacker with high confidence, and the prospective aggressor must believe that the defender will be able to achieve actionable attribution of the attack. Unfortunately, it is difficult to establish attribution through technical means alone. Ideally, technical attribution of a cyber attack will reveal the attacker's identity and location. This could be a name, a user account number, an alias, a physical location, or an Internet protocol (IP) address. The defender might also be able to match the attacker's exploit code, tools, or methodologies with tradecraft characteristics of specific countries, groups, or individuals. Unless the attacker uses sloppy tradecraft, however, the defender cannot hope for technical attribution to provide a "smoking gun" linking the people and machines involved to a specific country's government. Even if any of these data points are collected, a high level of confidence is hard to guarantee due to potential attacker countermeasures. These countermeasures can include falsifying end-user identity, encrypting attack flows, manipulating inadequately defended system logs or security protocols, bombarding the victim with forged reply packets from reflector host servers, laundering connection pathways through hijacked computers in a botnet, varying the speed of an attack to take advantage of intrusion detection system (IDS) protocol weaknesses, using parameter spoofing to decoy or evade IDS and firewall applications, or employing deceptive or misleading attack signatures

which implicate other parties. Technical attribution is further complicated by the fact that connection pathways typically cross network domain boundaries and national borders. Under such circumstances, forensic analysts may need cooperation from multiple network administrators as well as foreign law enforcement or security agencies. This cooperation can be difficult to secure, particularly when dealing with individuals or groups protective of their privacy, companies fearful of giving away trade secrets, or countries who either lack laws supporting pathway tracing or whose governments are antagonistic to those attempting to conduct the trace.⁵

It is important to appreciate that no single technical attribution technique provides a panacea. Most techniques focus on real-time and/or after-the-fact forensic analysis of an attacker's signatures and actions within a given system or network, not to mention within the upstream hosts used as stepping-stones in the attack. A separate class of techniques tags IP packets flowing through a network or modifies a network's terrain to localize and trace the attack pathways. Another grouping uses traditional intelligence collection methods such as covertly penetrating a suspected attacker's own systems to conduct surveillance, embedding beacons or tracing programs within data the attacker might exploit, or luring the attacker into providing traceable data. Some techniques even focus on increasing system or network security to reduce the number of attacks which need to be attributed.⁶ Most of these techniques must be used in parallel with each other if they are to overcome their individual shortcomings. Unfortunately, the network security community is still experimenting with how best to combine these techniques operationally, not to mention automate their use.⁷ Some of the more resource-intensive techniques may also impose significant cost or performance trade-offs on the systems and networks they protect. Attribution technologies offer the defender no "free lunch."

The challenges facing effective attribution are not impossible to overcome but are nevertheless extremely complex. Most attribution techniques require development and widespread implementation of standards governing protocol changes, processes, methodologies, and tools. Almost all require either implicit or explicit cooperation from upstream system administrators to enable intrusive forensic investigations, pervasive monitoring, and coordinated safeguarding of individual system activities. As trust cannot be surged in a crisis, successful attribution efforts require that system administrators develop relationships with each other, as well as with forensic investigators, prior to major attacks. This not only supports timely

information exchange and response coordination during an attack but also helps create a basis for authenticating individuals or groups requesting attribution data.

The private sector often resists paying the high costs of implementing attribution technologies and practices, as they largely view these as inherently military or law-enforcement responsibilities. Regulatory requirements may solve some of this, but enforcing compliance will be a challenge in its own right. The US government will most likely have to assume the primary role in funding technical attribution measures within US systems and networks, whether directly or through incentives. Considering how attribution efforts will often trace attacks through privately owned domestic systems and networks, the severe risk to personal privacy means that legal and technical protections as well as oversight measures must be implemented to defend against the risk of unconstitutional government surveillance.⁸ None of these measures will be accomplished quickly. It is not clear how advanced government efforts are in developing these policies and procedures nor whether a consensus even exists within the government regarding what these policies and procedures should be.

The United States will similarly need to work with foreign governments, both bilaterally and multilaterally, to encourage wider implementation of interoperable attribution technologies and methods, not to mention promote time-critical attribution information sharing. While some attribution methods can be coordinated via international institutions and industrial consortia, others will require close cooperation between militaries and intelligence agencies. It will not be easy for governments to assure each other that attribution cooperation will not be used for intelligence collection purposes or to unduly violate citizens' privacy. Successful cooperation may well compel governments to accept constraints on their cyber activities within or traversing each other's national network infrastructures. Such assurances will largely be built on trust and bounded transparency, as countries will understandably be reluctant to allow scrutiny of their most-sensitive network infrastructures and activities by even their closest allies. It follows that countries whose relationships are defined by the prevalence of strategic competition and hostility will be even less likely to grant the requisite degree of access. This also means it will prove difficult if not impossible to buttress deterrence by concluding cyber arms control agreements due to the inability to verify participants' compliance, and non-

binding political cyber reassurance will prove effective only amongst countries that already share a modicum of trust.

On the technology front, products which integrate and automate multiple, mutually reinforcing attribution techniques remain in their infancy. Developing effective automation tools will be particularly critical for tracing both high- and low-speed attacks. Still, attackers practicing disciplined tradecraft can make it extremely difficult to obtain actionable high-confidence traces. In fact, no standard methodology currently exists for establishing the degree of confidence in an attribution evidence set.⁹ Technical attribution data may end up being considered conclusive only when correlated with intelligence collected via other sources and methods.

Attribution can also be a double-edged sword. Deterrence by punishment cannot work unless the defender demonstrates some degree of effective attribution capabilities to potential adversaries. Conversely, disclosure of specific forensic methodologies may provide attackers with information on how to evade detection or attribution. Unless carefully sanitized, this information might even reveal exploitable details about defended systems and networks. These factors inevitably force policy decisions on whether revealing attribution capabilities and data as a means of establishing credibility or justifying retaliation might jeopardize the ability to defend against or trace potentially more-damaging future attacks. Further complicating matters, if an attacker is led to believe that the defender possesses advanced attribution capabilities and the defender fails to react to an obviously painful attack above the deterrent's stated threshold, the attacker may question the deterrent's credibility.¹⁰ This might provoke probing attacks designed to map the deterrent's ambiguities as well as test the defender's resolve. Should an attacker conclude that the defender's cyberdeterrent is not credible, it might question the credibility of the defender's deterrent postures in other warfare domains. This could prove extremely destabilizing, depending on strategic circumstances.

Decision makers must understand that successful technical attribution with high confidence may never be universally possible. Indeed, even the most technically sophisticated non governmental analyses of recent major attack campaigns did not provide incontrovertible attribution. The US Cyber Consequences Unit's examination of the August 2008 cyber offensive against Georgia revealed that even though the attacks were remarkably coordinated with Russia's military operations and many attack preparations were conducted far enough in advance of the cyber offensive to

suggest premeditation, no definitive Russian government or military participation could be found.¹¹ A June 2008–March 2009 joint study by the SecDev Group and the Citizen Lab at the University of Toronto’s Munk Centre for International Studies into exploitation of the Tibetan government-in-exile’s computer systems pointed to Chinese hackers but could not independently confirm whether the operation was conducted by a state-run strategic intelligence collection outfit, a private group or individuals without a political agenda collecting “interesting” information from random victims, a criminal group looking to profit from its thievery, or a false-flag group designed to lay blame on China.¹² Other analysts using data from the China study were able to exploit one of the suspected hacker’s tradecraft and identify his name and approximate geolocation. Nevertheless, they could not incontrovertibly prove his participation in the anti-Tibetan operation.¹³ Even the 2009–10 Aurora exploitation campaign against Google and other US companies has not been definitively attributed using forensics to the Chinese government. This is in spite of China’s obvious motives for exploiting these targets, the considerable amount of circumstantial evidence pointing to state-level involvement, and dataflow traces implicating specific sites and individuals within China.¹⁴ It is important to realize that the recognized experts who conducted these forensic investigations took many months to reach these limited conclusions despite using some of the most modern attribution technologies and methodologies. While it is possible that one or more of these investigations did achieve some degree of technical attribution but chose not to announce these results publicly cannot be completely discounted, the traditional independence and interconnectedness of the non-governmental computer security research community suggests that high confidence attribution conclusions could not be suppressed indefinitely.

Notwithstanding significant material, financial, and legal resource advantages over the above researchers, it is difficult to be sanguine about governmental chances of routinely achieving high-confidence technical attribution data within the foreseeable future. Although it is possible that some classified attribution technologies and techniques exist which offer very high-confidence results, one must remember that adversaries can and will adapt their tradecraft over time in response to their estimates of the defender’s attribution capabilities. Aggressive human intelligence operations cued by continuous technical exploitation of adversaries’ systems and networks might provide the only enduring methods for obtaining

actionable information linking specific individuals or machines with a particular foreign government.

A directly implicated government can nevertheless claim that the attackers were criminal entrepreneurs, uncontrollable “patriots,” or even terrorists. Even if it is revealed that the attackers were adversary government operatives, the implicated government could claim the attack team mistakenly thought they had authorization, were rogues, or accidentally caused the attack effects while exploiting a system for espionage purposes. Unlike command and control of physical military forces, there are fewer technical safeguards national leaders can use to establish positive control over state cyber warfare assets.¹⁵ This begs the question of whether defender attribution capabilities can confidently differentiate between unauthorized/nonsponsored and authorized/sponsored attacks. If the defender cannot make this distinction, one must wonder what level of retaliation should be threatened in advance, considering that the strategic risks of rashly acting on mistaken or inadequately substantiated impressions must be weighed against the need for repercussions appropriate to the damage sustained. Ill-considered impulsiveness is not the form of leadership “unpredictability” Schelling cites as a method for making deterrence more stable.

Ultimately, the defender’s decision makers can at best hope to collect massive amounts of potentially ambiguous, highly circumstantial attribution evidence. One analyst suggests adapting law enforcement’s classical model of establishing means, motive, and opportunity to support cyber attack attribution assessments.¹⁶ Under some circumstances, an outbreak of cyber attacks loosely attributed to a particular adversary state might appear remarkably well coordinated in both time and space with that state’s overtly conducted diplomatic, economic, or military initiatives. This might be enough to convince the defending state’s leaders of the adversary state’s guilt and thereby trigger as well as justify the deterrent’s threatened response. Linking cyber attacks to a particular adversary may be even more straightforward when the defender and the implicated attacker are already engaged in direct military hostilities. In this case, retaliation not only becomes part of the defender’s operational-strategic conduct of the war, but also can be configured as a “coercive deterrent” against continuation or escalation of the adversary’s aggression in general or cyber campaign in particular. Outside of war, however, unless the degree of attribution confidence is commensurate with the proposed retaliation’s severity, the

defending state's leaders may hesitate to strike back, and deterrence by punishment may be undermined.

Cyberdeterrence Thresholds, Retaliatory Methods, and their Implications for Nuclear and Conventional Deterrence

Cyber attacks can run the gamut from website defacement to packet-flooding a host to cut it off from the Internet; from covertly penetrating a network to manipulate or exfiltrate information to deliberately causing catastrophic failures of physical systems. Targets can include individual personal computers; private business internal networks; major inter-corporate networks, such as those used for money transfers between banks; or national security command, control, communications, and intelligence (C³I) networks. Inflicted damage can vary from the inconvenience of having to reboot or mirror a server to the inability to access critical services, theft of funds or sensitive information, loss of trust in potentially corrupted data and services, or physical neutralization or destruction of network-reliant systems. While none of these lists attempts to be all-inclusive, they do raise the question of where we draw the redlines which, once crossed, trigger a national response. Depending on how these triggers and any automatic trip wires are structured, a cyberdeterrence posture can either reinforce or considerably weaken a state's conventional and nuclear deterrence.

Schelling notes that failure to respond to an overt attack in accordance with a state's previously declared deterrent threat can erode the credibility of its other threats.¹⁷ A key factor is whether the deterrent threshold was drawn appropriately. Some of the cyber attacks described above are severely irritating but produce little or no lasting damage. If the defender believes that only specific retaliatory options can inflict enough pain on the aggressor to unambiguously signal the deterrent threat's fulfillment and the pain inflicted by these options will be disproportionate to the pain caused by the aggressor's original attack, the defender may be reluctant to retaliate out of fear of either appearing heavy-handed or provoking escalation. If this means the defender declines to fulfill its threat, the attacker may be encouraged to probe elsewhere and see if the defender is shaky on other deterrence commitments.

Another key factor somewhat unique to cyber attacks is whether the offending assault truly can be called “overt.” Unlike nuclear or conventional attacks, both the attacker and defender can hide a cyber attack to avoid triggering the deterrent response. The attacker may go to great lengths to conceal that an event was caused by a cyber attack, particularly if the attack was designed to covertly support diplomatic, economic, or military initiatives in other areas. The defender may realize an event was an attack but react as if it were only a system glitch if it cannot achieve actionable attribution, fears inciting further domestic chaos, abhors the effects of retaliating disproportionately, or wishes to block the adversary’s achievement of a particular attack objective. Similarly, the nature of a given attack may unintentionally conceal its overtness and thereby erode deterrence signaling. The attacker can never be sure that the defender will recognize an event as the result of a deliberate cyber attack as opposed to an accidental system failure. The defender may miss any signals the attacker deliberately sends to show hostile intent. All these ambiguities will almost certainly lead to probing in the cyber domain as well as potentially other deterrence domains by a particularly adventurous aggressor. These probes could prove highly destabilizing bilaterally, regionally, or even globally.

The concerns discussed above also argue against setting declared cyber-deterrence thresholds too low. The question of where to draw redlines may best be answered by identifying where the potential strategic costs of not retaliating almost certainly outweigh the potential costs of doing so. Since there are no international norms defining a cyber attack, it is useful to refer to the long-standing international norms defining an armed attack. If a notional cyber attack would cause the same level of damage as an armed attack, the defending state could threaten in advance that it would respond to such a cyber attack accordingly.¹⁸ Defacement, theft, and intelligence collection attacks would not rise to this level, but attacks which result in human casualties or physical damage to national critical infrastructure as defined in Homeland Security Presidential Directive (HSPD)-7 would.¹⁹ Nonlethal denial-of-service attacks against commercial or non-critical defense systems probably would not be considered armed attacks, but similar attacks against operational military command and control networks might, depending upon their effects. Discovery of attributable software “implants” within critical systems which, if triggered, would produce effects equivalent to an armed attack could arguably also qualify. Any attributable cyber attack against strategic command, control, and early

warning assets and networks would likely be considered an extremely hostile armed attack due to the safety and security implications of compromising these systems.

It is important to note that retaliation for a cyber attack may not necessarily use “cyber weapons.” Unlike kinetic military power, cyber attacks do not require specialized physical weapon systems. Considering that cyber attacks are conducted using commercial computing technology platforms, the real weapon in cyber warfare is creatively using knowledge of a particular vulnerability to develop an effective exploitation. Whereas defenders can use their nuclear and conventional forces to hold an adversary’s equivalent forces at risk, it is difficult to see how defenders can use cyber weaponry to symmetrically threaten the adversary’s cyber warfare capabilities. If a retaliatory cyber attack crashes the original attack platform, the attacker can simply relocate to another platform across the room on a different network which uses a different attack pathway. If a retaliatory cyber attack corrupts data or physically damages computing infrastructure, the original aggressor can restore the system using the last trusted data sets and system configurations as well as trusted backup or redundant hardware.²⁰ In short, symmetrical counterforce deterrence does not appear useful in the cyber domain.

Unless constrained by an overly specific deterrent threat, the defender could opt for diplomatic and/or economic retaliation if it believes either will send an adequate punitive signal. The defender could also attempt to publicly humiliate the attacker by sharing selective information about the cyber attack with independent computer security researchers and the international media. A wide international consensus regarding the aggressor’s probable guilt, even if based solely on circumstantial evidence, could prove extremely damaging to the aggressor state’s diplomatic and informational power. Such embarrassment might convince the attacker that the potential gain from similar future cyber attacks might not be worth widespread public diplomacy blowback. The trade-off, however, is revealing some information about one’s defenses and attribution capabilities to the attacker. Nonkinetic methods in general are probably best used for retaliating against relatively low-impact cyber attacks conducted outside of war, such as probes or harassment campaigns.

Nonkinetic retaliatory threats are even less useful for deterring an adversary’s potential use of deliberately lethal or physically destructive cyber attacks. This largely stems from the fact that nonkinetic first-order effects,

including those resulting from retaliatory cyber attacks, are neither as predictable nor necessarily as immediately painful as kinetic first-order effects. Threatened reprisals against high-impact cyber attacks might therefore include a conventional military response alongside or in lieu of retaliatory cyber attacks. This punishment could be directed against the adversary's conventional military forces and/or valued national institutions and infrastructure. At the extreme, a great power might threaten a potential aggressor with nuclear escalation to deter the latter from conducting catastrophically crippling cyber attacks against the assets and capabilities most vital to the former's national security. As Gen Kevin Chilton, commander of US Strategic Command, observed in May 2009, "You don't take any response options off the table from an attack on the United States of America."²¹ In fact, some analysts imply the United States should explicitly link cyber attacks on "critical infrastructure" with a nuclear response and that attacks conducted by "patriot volunteers" should be explicitly treated as attacks formally sponsored by the benefitting state.²²

It may be difficult, however, for the defender to make a public case that a given cyber attack merits the threat of a visibly harsh punishment, particularly if that punishment risks immolating both sides. Unless the defender's survival or vital interests are truly and incontrovertibly in extremis, the defender's leaders must justify the punishment to domestic and international audiences or risk losing their support. In the absence of obviously grave national danger or damage along with incontrovertible publicly-releasable attribution evidence, these audiences may not believe the threatened or implemented punishment fits the crime, nor might they even believe a crime was committed. Computer systems' technical complexity makes it difficult to publicly prove an incident was indeed caused by a cyber attack without disclosing detailed, sensitive information about the attack, the defenses, or the target itself.²³

Striking back at a country solely because it physically or virtually hosted a cyber attack conducted by patriot volunteers is even more problematic. The only foreseeable circumstance in which international norms make this kind of retaliation viable is if the implicated country refuses to cooperate in the investigation or with bringing the attackers to justice. Of course, this assumes the defender can show that the attackers or their enablers were truly citizens of or residents in the implicated country and that the attack was neither a false-flag operation tailored to smear that country nor an opportunistic use of that country's information infrastructure as the attack

platform. Considering that a small number of Internet service providers (ISP) in the United States regularly rank atop security researchers' global lists of hosts used most frequently by malicious actors, casting the first stone may be unwise from the perspective of establishing new international precedents for action.²⁴

A broad-based deterrent threat explicitly stating the defender is willing to ignite a conventional or nuclear war in response to a rather painful but nondebilitating cyber attack against unspecified "critical infrastructure" appears ill-advised. Unless the defender is willing to routinely demonstrate provocatively erratic decision making or otherwise create trip wires which automatically initiate escalatory processes, this kind of deterrent threat invites adversary probing and risks eroding the defender's conventional and nuclear deterrent credibility. While autocracies may be able to signal provocatively erratic behavior over long periods of time with few domestic repercussions, democratic electorates tend not to favor such qualities in their leaders. Although military computer systems could be publicly installed as trip wires on all upstream pathways from a protected nonmilitary system as a means of signaling to potential adversaries, it might be unwise to do so without great deliberation regarding the announced triggering thresholds as well as any domestic legal and privacy issues involved.

Efficacy of Punitive Cyberdeterrence

None of these arguments should be interpreted to mean there is no role for cyberoffensive capabilities in the US arsenal or for punitive cyberdeterrence postures which float the threat of US conventional or nuclear responses. Indeed, cyberoffensive capabilities are a new and extremely valuable element within current and future US combined arms operations in war. US cyberoffensive capabilities will be needed to disrupt, degrade, and exploit an adversary's C³I networks, integrated sensor and weapons systems, and military logistical systems. These capabilities may have similar utility against the adversary's domestic and international propaganda efforts, not to mention against the economic and industrial infrastructure supporting the adversary's military potential. The degree to which these capabilities are employed by either side will depend upon the respective political objectives at stake in the conflict. It follows that apparent US cyberoffensive capabilities and a high-threshold US cyberdeterrence

posture may be more effective at restraining overall enemy escalation at the start of and within conventional war than they will be at deterring an adversary's presumably less-damaging cyber aggression outside of war. As retired USAF general Eugene Habiger, former commander of US Strategic Command, observes, "The issue isn't whether deterrence and preemption should be part of our national security strategy to deal with cyber attacks; the issue is to what extent can we now rely primarily on these two doctrines to secure our nation against the sophisticated cyber threats we face each and every day."²⁵

It is important to distinguish between cyberdeterrent responses and a military unit's inherent legal right to self-defense. An infantry squad pinned down by fire from a building does not need to know who is shooting at them or who else might be inside that building before they can return fire in self-defense. Likewise, Soldiers under legal orders to defend a specific nonmilitary location do not need to request authorization from above to return fire if fired upon, even if they do not know the aggressor's identity. These forces may opt to hold fire for tactical or strategic reasons, but if their lives or the security of items they are legally charged with protecting are in danger, they do not need permission to use lethal force under standing rules of engagement (SROE). These responses are considered legal even if they result in innocent casualties or collateral damage, so long as they are deemed justifiable under SROE. It follows that a military system's operators or a military unit tasked with protecting a nonmilitary computer system can take actions to blunt the effects of a cyber attack. These actions might even be directed against the apparent source of the attack. If there are no passive measures that a military network defense unit can take to disrupt a cyber attack endangering the lives of US military members, the lives of those they are legally ordered to defend, or the security of entities they are legally ordered to defend, SROE can be interpreted to allow use of proportionate and discriminate "force" against even unknown aggressors to the extent needed to terminate the cyber attack.²⁶

Difficulties arise when the military is not or cannot legally be tasked with protecting a given entity or when a desired response goes beyond the minimum proportional and deliberate effort needed to break up an ongoing attack outside of open war. This enters the realm of cyber retaliation and cyberdeterrence. In situations where the military's inherent right of self-defense cannot be invoked, decision makers must answer several planning questions when developing deterrent postures or deciding upon retaliatory

responses. RAND cyber warfare analyst Martin Libicki notes that the defender must be able to know who committed the original attack, whether the attacker's assets can be held at risk, and whether these assets can be held at risk repeatedly during a long campaign or in the event of an escalatory spiral. He further states that decision makers must consider whether retaliation can disarm the attacker even if deterrence is unsuccessful, whether there is a risk that retaliation will bring third parties into the exchange, whether retaliation sends a desirable message to one's own side, whether there is an adequately defined credible threshold for response, whether retaliation will invite escalation, and what to do if the attacker has little worth striking.²⁷ Former Defense Intelligence Agency chief technology officer Bob Gourley recommends also considering whether attribution can be proven without unduly disclosing sources and methods, as well as whether retribution can be swiftly accomplished in an environment which impedes action.²⁸

While we have previously touched upon most of Libicki's and Gourley's observations, either directly or indirectly, there are three which we have not. First is the problem of third parties. These parties could be other countries, nonstate actor groups, or even individuals. Third parties might not be convinced the defender was truly the victim of a cyber attack. They might believe the defender's accusations against the original attacker are merely cover for aggression by the defender. They might have political, strategic, or ideological reasons for inserting themselves into the mix. They might even take the defender's side and initiate cyber attacks which others attribute to the defender. The threat of punishment might deter involvement by third-party countries under some circumstances, but deterring third-party groups and individuals is much harder. Furthermore, in a cyber warfare environment it will often be difficult to readily tell with confidence whether a follow-on cyber attack is being conducted by the original attacker or a third party unless the aggressor takes steps to unambiguously signal its identity. Even then, it may be impossible to know in near real time whether the original attacker and the third party are working together or independently. Either the original attacker or the third party could launder cyber attacks through the other to create plausible deniability or to falsely incriminate. All this blurs whom to hold responsible for a given cyber attack and creates considerable escalation spiral risks amidst the ensuing chaos.²⁹

Second is the problem of swift retribution. Former director of national intelligence Mike McConnell argues that the United States needs to be capable of detecting and attributing intrusions at an “evidentiary” level of confidence “in the milliseconds of network speeds.”³⁰ Gourley notes that “since attacks can hit us quickly and can do significant damage while underway, the response needs to be as swift as possible.”³¹ Habiger observes that retaliation must be timed to clearly link with the original attack to signal punishment, and that the ability to respond using cyber means depends upon a previously identified vulnerability in an adversary’s system(s) still being exploitable.³² As we have seen, attacker countermeasures, immature attribution technologies, and limited attribution cooperation mean the high level of attribution confidence needed for swift retaliation may simply not be possible for many years to come, if ever. While one can neither preclude development of advanced and highly reliable IDS and attribution technologies nor the emergence of broad public-private and transnational attribution cooperation, current deterrence doctrine cannot depend upon assumptions that we will someday possess specific deterrence capabilities. There must be a doctrinal balance between the defender’s desire to retaliate quickly and its ability to attribute correctly and confidently using contemporary means. Libicki suggests the ability to convince an attacker not to try again is more important than the retaliatory tempo. This means the defender’s speed of response must be aimed at the attacker’s human decision-making cycles rather than irrelevantly attempting the futile task of matching its “computing speeds.”³³ For the foreseeable future, other than taking actions to blunt an ongoing attack, the defender should not execute a retaliatory response which is explicitly linked to an earlier deterrent threat until decision makers believe they have sufficient confidence in what will likely be a highly circumstantial attribution case.

Third is the problem of holding the adversary’s assets repeatedly at risk. While nuclear and conventional weapons use brute force to demolish a target’s passive defenses, cyber weapons depend entirely on a computer system or network having exploitable vulnerabilities. Any given computing infrastructure element may have countless vulnerabilities, not all of which are known by the defender at any specific time. The defender, however, is constantly discovering and patching vulnerabilities. This means a vulnerability necessary for striking or exploiting a given system may no longer exist when needed. Upon experiencing a cyber attack, the defender begins examining the vulnerability and the exploit used. This analysis may teach

the defender new cyber warfare techniques or even give it the specifications for new cyber weapons of its own. In this light, an attacker must assume that a particular vulnerability and exploit pairing will be useable only once before it risks becoming a spent round, and any such usage risks being reverse-engineered for use against one's self or one's allies.³⁴ This further supports our earlier point that it may not be a wise idea for one's cyberdeterrent to rely primarily upon promoting the credibility of one's cyberoffensive capabilities.

Combined, all the factors we have examined argue that the punitive component of US cyberdeterrence posture should promote high-deterrent thresholds equivalent to those already established for discouraging armed attacks against vital national assets and capabilities, flexible retaliatory timelines, and flexible retaliatory methods against all but the most egregious and publicly obvious cyber attacks. The United States should openly emphasize the danger of cyber attacks against battlefield, theater, or strategic defense systems and networks during periods of extreme tensions. Such cyber attacks might be viewed as preludes to kinetic attacks if it appears the adversary's military deployments suggest imminent initiation of physical hostilities. US leadership cannot help but respond in this situation, and potential adversaries must be made to understand this. The price of such a high threshold for action may be an increased probability of adversary probes. Under such lesser circumstances it will be important to privately communicate to an attacker that a given use of national power against it is actually retaliation for its previous aggression, regardless of whether or not the public is aware of the original cyber attack or the nature of the retaliation. This will be particularly important when punishment must be delayed due to slow attribution. In other situations, though, it may be sufficient to covertly inflict unmistakable pain clearly linked with the original attack without necessarily taking national credit for the counterstrike. Potential aggressors must be made to realize that use of proxies and other plausibly deniable tradecraft to conduct damaging attacks cuts both ways.

Cyberdeterrence by Denial

Slow attribution means that the threat of punishment alone may not be enough to deter another country from conducting a crippling cyber attack or cyberoffensive campaign, especially if it believes it has little to lose by attacking or is particularly confident in its cyber attack skills. High impact

preparatory, distraction, or provocation cyber attacks against the defender's national security systems, civilian infrastructure necessary for military readiness, or general civilian infrastructure are particularly likely if an adversary has already decided to resort to conventional warfare as its means of attaining its political objectives. If cyber deterrence by punishment alone is insufficient, then the potential aggressor must be convinced that cyber attacks on the defender will not be able to achieve the desired effects or will fail outright. A painful attack may be irrelevant if it does not hurt enough to matter operationally or strategically. The embarrassment of failure, if recognized, could erode the credibility of the aggressor's own deterrence efforts.

Libicki argues that the goal of deterrence is to mitigate the risk that an adversary's actions will cause unacceptable amounts of pain. This risk can never be fully eliminated, but it can be reduced to a tolerable level. If improved defenses can reduce this risk more effectively than can retaliatory threats, it may make more sense to prioritize deterrence by denial over deterrence by punishment. Robust defenses must be built anyway, Libicki observes, since systems must be protected against criminals, terrorists, "hacktivists," and recreational hackers.³⁵

Cyberdeterrence by denial emphasizes hardening systems against penetration, designing redundancy and graceful degradation capabilities into systems, avoiding complete doctrinal or operational reliance upon individual systems or capabilities, and developing adaptation skills of system administrators and users through education and exercises. While defensive perimeters can always be improved, the impossibility of making them perfect means that systems and users must have fight-through resiliency to compensate for any losses of fighting potential on the margins.

The US government can do many things to implement its own cyber defenses as well as to incentivize private-sector and civil cyberdefense efforts. Network infrastructures as well as individual users' systems can be better secured through automated patch management, more-disciplined configuration management, thorough risk analysis, default disabling of insecure and extraneous functions or applications within systems, use of robust anti-malware and firewall/IDS applications, and routine network and system-defense testing by expert "red teams." User identity management and access control efforts can be improved to limit the ways an attacker can hijack an account as a ticket to access a given system and inflict widespread damage once inside. Access points to the US portion of the

Internet, not to mention US networks' Internet access points, can likewise be better protected. Government-sponsored education efforts can provide developers, users, and administrators with information about the severity of the threat, the nature of adversaries' tradecraft, and optimal security attributes and practices so they can better defend themselves and their systems.³⁶

The US government should also cultivate collaboration with allied governments, global academia, and the international computing industry to define and continuously update cyber security standards. This could include standardizing the following: specifications and interfaces for system and network product security features, methodologies for analyzing system or network security requirements, processes for developing and testing computing product security attributes, and practices for managing system and network security. Such collaboration can readily leverage existing international computer engineering standardization bodies. Even though standards development is often slow, it provides the benefit of extensive, continuous peer review. Standardization can also incentivize wide implementation, particularly if market-driving vendors and government customers are actively engaged in developing and adopting the standards.

Critical national security networks must be designed from inception as "integrated weapons systems" instead of mere data pipelines. This means a philosophical shift is necessary in the people, architectural concepts, processes, and technologies used to develop and maintain these networks throughout their life cycles. While technological commonality helps mitigate acquisition and life cycle costs, technological diversity across critical systems and networks can provide inherent redundancy as well as allow for graceful degradation. Critical government, military, and private-sector systems and networks must also be designed with computing performance capacities well in excess of anticipated normal needs, such as reserve data bandwidth to help resist potential denial-of-service attacks. It follows that information infrastructures can be designed and administrators can be trained to support rapid reconfiguration and capability restoration under periods of extreme network stress. Information assurance measures can be designed with improved human-system integration attributes to reduce users' temptations to bypass security procedures. Perhaps most importantly, vital systems and networks can implement a "war reserve" for use only in extremis which features special redundant systems isolated from

normal networks, special operating modes and configurations for front-line systems, and special “out of band” control channels.³⁷

Since not all critical systems and networks can be provided an equally high level of security, they should be designed from inception using trade-offs which favor only the confidentiality, integrity, or availability attributes most relevant to their particular missions and operating environments. These information infrastructure elements should also be designed to fail into hardened “safe modes,” which provide users with the minimum necessary capabilities to accomplish their missions. The most critical systems and functionality, particularly those needed to support safe modes, should avoid using hardware and software products not developed and manufactured within a government-vetted supply chain routinely scrutinized for counterintelligence purposes. Strict system and network design partitioning measures also must be implemented to protect critical capabilities. Nevertheless, it remains vital that users and administrators hedge against the risk of security or design lapses by striving to identify, plan for, and train against possible system and network failure states.³⁸ Leaders and operators must never fall into the trap of assuming that a given network, system, or capability will always be available, secure, and trustworthy. They must accept that safe-mode resiliency may require trading some degree of efficiency, net-centricity, and automation for older, less-integrated, and perhaps more manpower- and skill-intensive methods of communication and operation.

Protecting older national critical infrastructure elements, especially industrial and utility control systems, will be an even more complicated challenge. Many of these control systems lack inherent attribution-supporting capabilities. Anti-malware, firewall, and other software-centric defense solutions also heavily tax the limited computing resources of legacy control systems. Since this software was often custom designed using obsolescent programming languages and techniques, patch installation poses a higher risk of unforeseen, potentially crippling interoperability side effects. This means a more aggressive patch-testing regime is necessary if legacy control systems are to be adequately protected. Worse, companies are often tempted to integrate modern human-system interfaces, including wireless remote access technologies, with legacy control systems to satisfy user preferences. Unless companies conduct thorough risk assessments before taking such a step, they may open an untold number of new pathways for potential system penetration. In many cases, improving the security of legacy control systems will require a mix of determined government efforts


including tighter regulation, negotiating security standards with the private sector, and subsidizing a sizable portion of the private sector's critical infrastructure modernization costs.³⁹ Because environmental efficiency programs will drive infrastructure modernization over the coming decades, it is vital that the defense-enhancing systems engineering principles and fight-through measures mentioned earlier be implemented as key system redesign attributes.

Lastly, the government, military, and private sectors can do more to strengthen their incident response and cooperation skills through frequent exercises, improved information-sharing mechanisms, and coordinated development of contingency plans.⁴⁰ The government can also actively collect, exploit, sanitize, and disseminate intelligence about rest-of-world cyber-attack capabilities to inform public and private-sector systems engineering, defensive planning, and user training. The government can similarly monitor suspected adversary hackers and other potential aggressors to collect data which might later be used in identifying attack indications and warnings or conducting post-attack forensic analysis.⁴¹

Stronger cyber defenses give the defender more options than offensive measures alone. Defenses do not have to completely prevent penetration and exploitation. They only have to provide the resiliency needed for one's critical systems and networks to fight through the effects of persistent cyber attacks. If the defender actively communicates the resiliency of its defenses through tailored information disclosure, overt doctrine and exercises stressing fight-through capabilities and skills, obvious improvements in security measures, and aggressive vulnerability analysis and patching in response to probes, then potential adversaries may question the efficacy of their cyber arsenals. They can be led to believe their best attack tools may fail to achieve operationally or strategically desired effects when needed, perhaps spectacularly. They can be led to fear the risk that failed cyber attacks or campaigns might degrade the deterrent credibility of their other elements of national power. Combined with a cyberdeterrence-by-punishment posture featuring high thresholds and flexible retaliatory terms, this may be sufficient to dissuade adversaries against using cyber attacks as a means for capturing 'fait accompli' gains in other strategic domains or otherwise engaging in the most potentially catastrophic cyber attacks during situations short of unrestricted total war. Decision makers must accept that the United States will be unable to credibly deter adversary cyber attacks aimed at probing its defenses or exploiting its systems to collect intelligence.

In fact, it would be remiss if it were not covertly and selectively doing the same to potential adversaries' networks and systems.

Cyberdeterrence is certainly plausible, but it should not mirror the nuclear and grand conventional deterrence postures which have protected the United States and its interests since 1945. Given all the previously discussed considerations, the ideal US cyberdeterrence statement might very well be something similar to the following:

The United States will consider a cyber attack or cyberoffensive campaign conducted during a period of high international tensions and aimed at crippling the operational readiness of its military forces, strategic deterrence capabilities, or national security systems a prelude to war, and it will respond accordingly. The United States will additionally reserve the right to respond to cyber attacks upon its citizens, assets, institutions, and infrastructure at the time and place as well as using the means of its choosing. 

Notes

1. Thomas C. Schelling, *Arms and Influence* (New Haven, CT: Yale University Press, 1966), 54–55.
2. *Ibid.*, 36, 39–40, 42–48, 97–101.
3. *Ibid.*, 67.
4. *Ibid.*, 78–79.
5. David A. Wheeler and Gregory N. Larsen, *Techniques for Cyber Attack Attribution* (Alexandria, VA: Institute for Defense Analyses, 2003), 1–4.
6. The Wheeler and Larsen report describes the capabilities and limitations of 17 somewhat distinct attribution techniques. For a more thorough description of each individual technique, see *ibid.*, 11–41.
7. *Ibid.*, 42.
8. *Ibid.*, 43–47, 51–52.
9. *Ibid.*, 51–52.
10. Gen Eugene E. Habiger, USAF, retired, “Cyberwarfare and Cyberterrorism: The Need for a New US Strategic Approach,” Cyber Secure Institute white paper no. 1:2010, 1 February 2010, 25, <http://www.army-technology.com/downloads/whitepapers/computing-software/file1552/>.
11. John Bumgarner and Scott Borg, “Overview by the US–CCU of the Cyber Campaign against Georgia in August of 2008,” US Cyber Consequences Unit, August 2009, 2–5.
12. “Tracking Ghostnet: Investigating a Cyber Espionage Network,” *Information Warfare Monitor*, 29 March 2009, 15, 22, 48.
13. “Hunting the GhostNet Hacker,” *Dark Visitor blog*, 2 April 2009, <http://www.thedarkvisitor.com/2009/04/hunting-the-ghostnet-hacker/>.
14. John Markoff and David Barboza, “2 China Schools Said to Be Tied to Online Attacks,” *New York Times* online, 22 February 2010, <http://www.nytimes.com/2010/02/19/technology/19china.html>.

15. Martin C. Libicki, *Cyberdeterrence and Cyberwar* (Arlington, VA: RAND, 2009), 46–47, 78.
16. “The Aurora Operation—A Watershed Moment for Attribution,” *IntelFusion blog*, 20 January 2010, <http://intelfusion.net/wordpress/page/3/?s=attribution&task=search>.
17. Schelling, *Arms and Influence*, 55.
18. LTC Scott W. Beidleman, *Defining and Deterring Cyberwar* (Carlisle, PA: Army War College, 1 June 2009), 12–13, 15.
19. LTG Keith Alexander, “Advance Questions for Lieutenant General Keith Alexander, USA, Nominee for Commander, US Cyber Command,” Senate Armed Services Committee, 14 April 2010, 19.
20. Libicki, *Cyberdeterrence and Cyberwar*, xvi–xvii.
21. Joshua Pollack, “Is the Cyber Threat a Weapon of Mass Destruction?” *Bulletin of the Atomic Scientists* online, 20 January 2010, <http://www.thebulletin.org/web-edition/columnists/joshua-pollack/the-cyber-threat-weapon-of-mass-destruction>.
22. Ned Moran, “Achieving Cyber Deterrence,” *GroupIntel.com*, 17 May 2009, <http://www.groupintel.com/2009/05/17/achieving-cyber-deterrence/>.
23. Habiger, “Cyberwarfare and Cyberterrorism,” 34.
24. “Naming and Shaming Bad ISPs,” *Krebs on Security blog*, 19 March 2010, <http://krebsonsecurity.com/2010/03/naming-and-shaming-bad-isps/>.
25. Habiger, “Cyberwarfare and Cyberterrorism,” 2.
26. Alexander, “Advance Questions,” 11–12.
27. Libicki, *Cyberdeterrence and Cyberwar*, 39.
28. Bob Gourley, “Towards a Cyber Deterrent,” *CTOVision.com*, 29 May 2008, <http://ctovision.com/references/towards-a-cyber-deterrent/>.
29. Libicki, *Cyberdeterrence and Cyberwar*, 42.
30. VADM Mike McConnell, USN, retired, “We’re losing the Cyber-war. Here’s the Strategy to Win It,” *Washington Post*, 28 February 2010, <http://www.washingtonpost.com/wp-dyn/content/article/2010/02/25/AR2010022502493.html>.
31. Gourley, “Towards a Cyber Deterrent.”
32. Habiger, “Cyberwarfare and Cyberterrorism,” 32.
33. Libicki, *Cyberdeterrence and Cyberwar*, 62.
34. *Report of the Defense Science Board 2008 Summer Study on Capability Surprise, Volume II: Supporting Papers* (Washington: Office of the Undersecretary of Defense for Acquisition, Technology, and Logistics, January 2010), 37–38.
35. *Ibid.*, 32, 34.
36. Gourley, “Towards a Cyber Deterrent.”
37. *Report of the Defense Science Board*, 39, 41–42, 45–46.
38. Libicki, *Cyberdeterrence and Cyberwar*, 163–66.
39. “Joe Weiss, Crusader for Critical Infrastructure Security (Q&A),” *InSecurity Complex blog*, 10 May 2010, http://news.cnet.com/8301-27080_3-20004505-245.html?part=rss&tag=feed&subj=InSecurityComplex.
40. *Critical Infrastructure Protection: DHS Needs to Fully Address Lessons-Learned from its First Cyber Storm Exercise* (Washington: Government Accountability Office, September 2008), 12–14.
41. *Report of the Defense Science Board*, 47–48.