The Cyber Deterrence Problem

Aaron F. Brantly
Assistant Professor, Department of Political Science
Virginia Polytechnic and State University
United States
abraantly@vt.edu

Abstract: What is the role of deterrence in an age where adept hackers can credibly hold strategic assets at risk? Do conventional frameworks of deterrence maintain their applicability and meaning against state actors in cyberspace? Is it possible to demonstrate credibility with either in-domain or cross-domain signaling or is cyberspace fundamentally ill-suited to the application of deterrence frameworks? Building on concepts from both rational deterrence theory and cognitive theories of deterrence this work attempts to leverage relevant examples from both within and beyond cyberspace to examine applicability of deterrence in the digital age and for digital tools in an effort to shift the conversation from Atoms to Bits and Bytes.

Keywords: cyber, deterrence, denial, punishment

1. INTRODUCTION

The challenge of the digital era is not to define deterrence. Deterrence is a well-defined concept that has been studied and practiced throughout history and to an even greater depth following the advent of nuclear weapons. The present challenge it is to understand the role digital technologies play in the broader scope of interstate deterrence. Deterrence in one domain rarely if ever operates independently of other domains. Much of the literature on cyber deterrence focuses on within domain deterrence. Yet, this is a dangerous constraint that elevates risks and minimizes the probability of success. This paper seeks to draw out the literature on deterrence and identify its applicability within a newly delineated domain of interactions, cyberspace. The resultant analysis strives to encompass the complexity of deterrence and advance an argument beyond within domain modeling.
Classical deterrence centers on a potential adversary’s cost-benefit calculus to dissuade specific actions and differs from compellence by focusing on ex-ante behavior manipulation through a priori uses of force or other tools of state power. Both compellence and deterrence are forms of coercion, however, the former employs both hard and soft power both in the present and future with continued or escalated actions, while the latter threatens use of force (power) absent their employment. The focus below is on ex-ante actions by states and sub-state entities that threaten, but that do not use the tools of state against an adversary to manipulate their decision-making calculus. Additionally, actions undertaken independent of threats that can, ex-ante, reduce the benefits associated with a given attack are examined.

Focusing on classical deterrence and deterrence by denial helps illustrate the similarities and differences between deterrence in the pre- and post-delineation of cyberspace as a domain of military operations. Deterrence in cyberspace has been addressed by a variety of scholars across the subfields of International Relations. Many examinations of cyber deterrence rely on direct applications of IR theory absent robust technical understandings of how the domain functions. The development and application of classical deterrence theories to a domain necessarily requires an understanding of how state and non-state actors achieve, develop, and assess costs and benefits within this domain.

This work proceeds in three sections. First, it examines some of the relevant literature on deterrence and identifies some of the gaps within the field and provides a trajectory for the subsequent sections to examine a more dynamic theory of deterrence in cyberspace. The second section focuses on the technical, tactical, operational, and strategic aspects of the domain in an effort to identify those areas where deterrence can alter the costs-benefit analysis of adversaries. Third, the work concludes by providing a discussion on national strategy development for integrated cyber deterrence incorporating the lessons from the first two sections.

2. FROM ATOMS TO BITS AND BYTES

Deterrence is not a novel concept. The classical IR cannon on deterrence can be traced back to the Peloponnesian War and the threat of violence in response to adversary actions. Yet, more modern formulations of deterrence are largely rooted in the nuclear world following World War 2. The most common form of deterrence known as conventional deterrence was established by Bernard Brodie, Thomas Schelling and

---


others and focuses on the ex-ante dissuasion of adversaries through the threat of ex-post costs in response to potential adversary actions.

Robert Jervis identified three “waves” of deterrence theorizing to which a potential fourth wave has been added by Jeffery Knopf. First wave deterrence theory rested on the rise and consequences of nuclear weapons. Bernard Brodie et al. asserted that the use of nuclear weapons had almost no innate strategic or tactical value outside of being a threat against an adversary. The consequences of nuclear weapons use, even in limited strike situations, would quickly and dramatically escalate. This escalation made the limited use of such weapons untenable in all but the most extreme situations. Lawrence Freedman summarized the second wave as the realization that “total war could now only be threatened, but never fought”.

Second wave deterrence posited how nuclear weapons could be threatened and the dynamics of those threats. Thomas Schelling and others posited a series of conditions in which states could develop deterrence in the nuclear era. As Jervis noted, second wave theorizing became extremely popular because of its abstraction and logical structuring. Game theory and other rational models were used to illustrate rational costs and benefits, creating models suited to rigorous concepts of rationality. The second wave arose under stable bi-polar conditions in which it was assumed states engaged in rational decision-making in matters of foreign policy and national security. Schelling found deterrence largely dependent upon credibility and rationality. He illustrated that signaling potential costs to an adversary absent credibility creates deterrence failure. By using divergent game-theoretic structures from prisoner’s dilemma to chicken – theorists developed arguments about deterrence. Despite rigorous theory, this abstraction contained systemic flaws and gave rise to a third wave of deterrence.

The third wave of deterrence theory in the 1970s addressed challenges beyond game theoretic models, including the failing rationality. Irving Janis and Graham Allison, both, but with different perspectives, illustrated the weaknesses of rationality in decision-making. The third wave led to extensions into cognitive psychology and behavioral studies. Robert Jervis, Richard Ned Lebow, and Janis Stein provided insight into the general problems associated with parsimonious use of rationality through case analyses. Specifically, Jervis et al. identified the potential for over-valuation of

6 Ibid: 22.
Jeffrey Berejikian incorporated Daniel Kahneman and Amos Tversky’s analysis of prospect theory into the deterrence calculus and challenged parsimonious rational thought by illustrating cognitive dimensions associated with decision-making beyond groupthink and bureaucratic processes. His work highlighted issues related to risk in cognitive decision-making that undermine rationality. Concepts such as sunk costs or tying hands fit well within parsimonious deterrence theory, yet the mechanisms that made them effective were not well understood prior to the third wave.

Although modern deterrence theory encompasses a spectrum from pure rational modeling to cognitive models, the objective of deterrence as identified by John Mearsheimer remains the development of fear of the consequences (in particular of “military action”) or a “function of costs and risks”\(^\text{11}\) Developing shared knowledge about costs and risks for nuclear events differs from non-nuclear conflicts. Early deterrence models relied heavily on rationality and parsimony but did not underestimate the clarity provided by the use and subsequent impact of the weapons themselves. The generation of fear or knowledge of consequences to assess costs and risks loses clarity the as analyses shift away from nuclear weapons. Lawrence Freedman defines single weapon or type of warfare deterrence as “narrow deterrence”\(^\text{12}\). Narrow deterrence is less effective when expanded beyond single weapon or type warfare.

General or broad deterrence covers a range threatened actions to dissuade an adversary. Freedman writes: “broad deterrence involves deterring all war”\(^\text{13}\). Ted Hopf explains: within deterrence there is a need to expand deterrence beyond the scope of military tools to the entire range of options available to actors\(^\text{14}\). Extending analysis further, scholars also emphasize concepts of direct deterrence and extended deterrence. Direct deterrence is concerned with actions against “your” state and its immediate interests as opposed to extended deterrence – dissuasion of adversary actions against a third party or non-immediate interests. Delineating between these two types of deterrence in a globalized world is difficult. Cyberspace compounds the challenge of delineation because attacks on foreign infrastructure can and do have ramifications globally.

Concepts of the means to achieve deterrence or more simply how to deter are often contested. Threats can be narrowed to weapon type or category, or include

---


\(^{13}\) Ibid.

interdependent relationships such as diplomatic, informational, military and economic effects. Threats signaling a potential response to adversary action should provide clear, unambiguous consequences. The ex-ante threat should causally lead to an ex-post consequence; punishment.

Often left out of traditional international relations literature, deterrence by denial has seen a surge of interest in the years following the 9/11 terrorist attacks. Alex Wilner defines deterrence by denial as “reducing the perceived benefits an action is expected to provide a challenger”.¹⁵ Deterrence by denial in the physical world often includes hardening targets by building higher walls, adding security mechanisms, or other tactics to reduce the susceptibility of targets to attack. If the Strategic Defense Initiative (SDI – also known as Star Wars) had been successful, it would have been a deterrence by denial strategy to limit the effect of Soviet nuclear weapons. Commonly used forms of deterrence by denial in conflict zones include land mines, razor wire, surface to air missiles (SAMs) and fortifications.

Deterrence by punishment and denial are intended to manipulate the cost-benefit analysis of an adversary. To function they must both be credible. Credibility requires undertaking ex-ante costs by the deterrer. Threats absent ante impetum costs lack credibility. A state without nuclear weapons cannot credibly threaten nuclear retaliation. If a state wishes to deter it must provide demonstrable evidence that it is able to carry out its threat.

Likewise, deterrence by denial fails when it lacks the material capabilities to deny. The Maginot Line built by the French following World War I stands an example of failed deterrence by denial. The French system of fortifications on portions of their northern territory failed because the line itself only covered one vector of attack into France. The elevation of costs to a potential attacker must be complete and provide no reasonable alternatives to achieve the attacker’s intended utility. Both strategies require ex-ante costs by the defender to alter the ex-post perceived benefits of an attacker. Punishment strategies increase adversary costs after a violation and denial strategies increases adversary costs in advance of a violation.

Deterrence by denial is a successful strategy in many instances; SAMs effectively deter enemy aircraft. The relative costs of upgrading certain denial tools is comparatively less than the costs of surmounting them. In the case of SAMs, the United States spent billions of dollars to defeat the S-300 missile system (~$100 million/system).¹⁶ Following the development and use of stealth, S-300 designer Almaz upgraded its

---


systems to the S-400 variant with greater accuracy and anti-stealth technology.\textsuperscript{17} The cost ratio between the denial tool and offensive weapon system is approximately 1 to 1,000. The defensive and offensive capabilities, industrial, and financial resources of these two states exceed most other nations. Even with a $18.5 trillion GDP a $1 to $1,000 cost to benefit ratio is high and demonstrates how denial can be a remarkably effective strategy.

Deterrence by denial is not always successful as illustrated by the Israel – Hamas conflict. In response to Hamas’ use of Katyusha rockets, Israel developed the Iron Dome System. Iron Dome batteries cost $100 million and each rocket costs $50,000.\textsuperscript{18} To intercept an incoming Katyusha rocket, the Israelis launch 2 interceptor rockets.\textsuperscript{19} By contrast, Hamas spends between $500 and $1,000 per rocket launch.\textsuperscript{20} If the cost of the battery is ignored, the cost of deterrence by denial is still between 100 to 1 and 200 to 1.

Denial strategies are not passive. They require continuous modification relative to adversary capability development. Static denial strategies in cyberspace or in conventional conflict are likely to have limited credibility over time. Similarly, punishment strategies also require constant updating in relation to adversary capabilities and geopolitical considerations. In cyberspace, this involves adapting denial strategies to technological advances such as artificial intelligence, polymorphic malware and the Internet of Things, to name just a few.

Punishment strategies also require ex-ante costs. Below the nuclear threshold, threats of force are common, yet the credibility of these threats is difficult to establish. Alexander George and Richard Smoke identify three attributes important for signaling in conventional deterrence: “(1) the full formulation of one’s intent to protect a nation; (2) the acquisition and deployment of capacities to back up that intent; (3) the communication of intent to a potential aggressor”.\textsuperscript{21} These three aspects are also at times limited in their ability to convey commitment to fulfill the intent.\textsuperscript{22}

Charles Glaser, writing on cyber deterrence, established four components of basic deterrence:

\textsuperscript{19} Ibid.
\textsuperscript{20} Ibid.
\textsuperscript{22} Ibid: 558.
“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”.23

George and Smoke and Glaser acknowledge the challenge of establishing not just threats of punishment, but the credibility associated with carrying out that threat.

Creating material capability (i.e. weapon systems capable of carrying out a given threat) and clear signaling might occur and yet the utilization of this capability in response to an adversary’s action will lack credibility (fulfillment of commitment) unless it contains what James Fearon refers to as hand-tying within a sunk costs framework.24 Credibility and hand-tying are most closely associated with extended deterrence, yet when expanding deterrence to cyberspace it also finds relevance. The establishment of credibility through hand-tying establishes a forcing mechanism for decisions, indicating costs have already been incurred or are likely to occur. This subsequently alters the cost-benefit calculus of retaliation. The stationing of US forces in West Berlin serves as an example of hand-tying through prospective costs.25 An attack on West Berlin would have resulted in sunk costs and provided a strong inducement or “tripwire” to actuate US retaliatory threats. Nearly all forms of kinetic attacks against the direct interests of a nation implicitly include hand-tying. It is unclear how to effectively signal prospective costs within cyberspace to an adversary.

Charles Glaser identifies several problems associated with deterrence by punishment specific to cyberspace that extend beyond basic credibility issues. First, he notes that deterrence often relies on the attribution of an adversary’s actions.26 In cyberspace, this can be difficult and time-consuming.27 Although the attribution problem is decreasing as more data becomes available, it does not eliminate uncertainty.28 Second, hands-tying and other forms of credibility enhancing measures are likely lacking in cyberspace. Moreover, the ability to respond within domain simply might not be possible within certain conditions.29 Third, Glaser identifies potential spillovers...
in which limited within domain options result in cross-domain, kinetic responses.\textsuperscript{30} To date there is limited evidence of cross-domain responses and therefore lacks in credibility. Moreover, cross-domain retaliation alters the escalation framework from digital to kinetic or other and poses a challenge for states wishing to establish credibility while controlling potential escalatory behaviors.

Deterrence is more than simply threatening punishment. Deterrence requires substantial target relevant costs and the development of mechanisms to establish that further costs are credibly wagered to provide clarity for an adversary. The goal of this clarity is to establish within an adversary’s calculus that their expected gains are less than any potential losses incurred. Reassessments of rational modeling and the increasing importance of cognitive modeling increase the value of tailored deterrence strategies predicated on the uniqueness of conditions and actors. Paul notes that deterrence is complex and is most logically broken down into five ideal types:

“(1) deterrence among great powers; (2) deterrence among new nuclear states; (3) deterrence and extended deterrence involving great powers and regional powers armed with chemical, biological and nuclear weapons; (4) deterrence between nuclear states and non-state actors (5) deterrence by collective actors”.\textsuperscript{31}

It follows that tailored deterrence for cyber actors is also one potential avenue of exploration.

The potential for tailored deterrence strategies could be highlighted in numerous significant cyber incident cases. The 1998 cyber attack code-named SOLAR SUNRISE discovered by US Air Force Computer Emergency Response Team (AFCERT) stands as a prime example. The three-week hack affected more than 500 systems across the US Air Force, Navy, NASA, Lawrence Livermore Labs, MIT, Harvard, and UC Berkeley. The attack coincided with increased tensions between the United States and Iraq and resulted in high-level governmental meetings to identify a proper response action.\textsuperscript{32} At the time, the attack was believed to be state-sponsored cyber attack focused on degrading US military capabilities. Subsequently, it was discovered that the attack was conducted by two California teenagers with guidance from Israeli hacker Ehud Tenebaum. The incident is relevant to tailored deterrence because it highlights challenges faced in developing a deterrence strategy. The adversaries were domestic, yet foreign inspired and attacked the operational infrastructure of the Department of Defense. No form of deterrence by punishment delineated above could have appropriately accounted this challenge. The only realistic

\textsuperscript{30} Ibid.
deterrence frameworks for SOLAR SUNRISE would have been deterrence by denial or punishment in cooperation with allies.

Richard Kugler writes that a strategy or general framework for deterrence in cyberspace must necessarily be tailored to differing threats, situations, and objectives. The threats, situations, and objectives in cyberspace differ from the concerns addressed by first wave theorists. While the potential for physical damage through cyberspace has been demonstrated in tests such as the Aurora generator experiment that resulted in the destruction of a multi-ton diesel generator, or the Stuxnet attack that destroyed segments of a centrifuge cascade in Iran’s Natanz nuclear facility, many attacks do not have kinetic parallels. Building on Kugler, Jeffrey Cooper identifies three important factors that frame concepts on deterrence in cyberspace. First, there is a wide range of actors each with different capabilities and attributes as well as cost benefits structures; second, cyberspace is a unique operational domain that carries with vastly different concepts of risk and reward; third, to develop deterrence, models must be applicable to the virtual and physical aspects of the domain.

This section has provided a summary of a large and robust literature on deterrence. The concepts that need to be carried forward include, the type of deterrence, the credibility of that deterrence and the attributes of the environment in which deterrence occurs, and who and what actors and weapons are to be deterred. The next section builds on the literature above, with a specific emphasis on the technical, tactical, operational and strategic attributes of cyberspace.

3. ONE SIZE DOESN’T FIT ALL

To deter adversaries in cyberspace it is helpful to first define what cyberspace is and what types of actions and actors a state would like to deter. The US Department of Defense defines cyberspace in the following way:

“Cyberspace consists of many different and often overlapping networks, as well as the nodes (any device or logical location with an Internet protocol address or other analogous identifier) on those networks, and the system data (such as routing tables) that support them. Cyberspace can be described in terms of three

layers: physical network, logical network, and cyber-persona. The **physical network** layer of cyberspace is comprised of the geographic component and the physical network components. It is the medium where the data travel. The **logical network** layer consists of those elements of the network that are related to one another in a way that is abstracted from the physical network, i.e., the form or relationships are not tied to an individual, specific path, or node. A simple example is any Web site that is hosted on servers in multiple physical locations where all content can be accessed through a single uniform resource locator. The **cyber-persona** layer represents yet a higher level of abstraction of the logical network in cyberspace; it uses the rules that apply in the logical network layer to develop a digital representation of an individual or entity identity in cyberspace. The cyber-persona layer consists of the people actually on the network”.

The inclusion of the full definition illustrates the complexity within which defense strategists and operators in the various services engage. Because the domain spans the physical, logical, and persona layers, deterrence strategies can reasonably occur within and across all three. This fundamentally differs from the conceptualization of deterrence in physical domains of land, sea, air, and space. Physical domain deterrence might include physical and cognitive aspects analogous to the cyber persona and physical network layers, however, the logical layer is wholly absent. The cyber persona layer also diverges significantly from personas within the physical domain as individuals and states have the capacity to alter their attributes within the persona, logical, and network layers.

To construct a meaningful model of deterrence in cyberspace we must first ask what it is we wish to deter. Herein lies the largest distinction between deterrence in the physical world and in cyberspace. Whereas in the physical world deterrence is directed most commonly against physical attacks against specific assets or categories of assets that when attacked provide strong, largely non-repudiable forms of attribution, in cyberspace deterrence is directed against manipulations of the elements within the environment and the environment itself. Manipulation of elements of cyberspace and the environment itself can be examined in multiple ways. Simplifying cyberspace operations into three broad categories, there are cyber attacks, cyber espionage, and cyber theft. Despite simplification, it is important to note these categories are not entirely discrete in process or function. Cyber attacks are those acts in cyberspace that degrade, deny or destroy. Acts of cyber espionage steal information for state or corporate intelligence gain. Cyber theft is the stealing of information for financial gain with no direct state utility. Attacks, espionage, and theft occur across all levels

of actors from script kiddies to the military units of states – a problem which will be
examined more below. States are most commonly concerned with cyber attacks and
espionage at the national level, and theft at lower-jurisdictions.

Because attacks, espionage, and theft are perpetrated by a variety of actors against
almost any target in cyberspace, sending an overt signal from one state to another,
while still applicable, might not deter attacks at other levels that are of equal or greater
significance. Moreover, research by Shawn Lonergan and Erica Borghard indicate
a high prevalence of proxy\textsuperscript{37} usage by states to maintain plausible deniability.\textsuperscript{38}
Using proxies to engage in cyber acts against targets deflects deterrence by threats
of punishment unless sufficient evidence is present to indicate involvement by the
instigating state rather than the third-party proxy. The use of proxies to engage in
attacks, espionage and theft against target states outside of cyberspace has been the
practice of states since Katulaya and Sun Tzu.\textsuperscript{39} However, unlike the difficulties
of non-repudiability within conventional conflicts, cyber attacks are frequently
repudiable. Attackers might use Virtual Private Networks (VPNs), proxies or other
means by which to engage in an attack.

Additional problems in cyberspace not frequently encountered in conventional
physical domains are second and third order effects. As noted by Herbert Lin, the
results of a cyber attack itself might not be identifiable, rather it is second or third
order effects that generate an intended outcome.\textsuperscript{40} Classical deterrence and tailored
deterrence strategies used against terrorist organizations are unable to account for
disconnected action and reaction pairs commonly found in cyberspace. The time to
punish a violation can be weeks, months or years based on discovery and attribution
challenges, a problem not present in classical deterrence.

Cyber attacks are incidents occurring in or through cyberspace that degrade, deny
or destroy. Attacks in cyberspace can and are perpetrated by all levels of actors. The
differentiation between actors is most closely correlated with targets and outcomes
of attacks.\textsuperscript{41} For example, criminal actors may use phishing attacks to ingress into
a hospital’s computer systems to installCryptolocker or a similar ransomware
malware on the hospital’s systems. Cryptolocker is an attack that degrades civilian
critical infrastructure, denies user access and has the potential to destroy critical

\textsuperscript{37} Here proxy usage refers to the authority to represent someone else not the technical usage of the term in
information communications.

\textsuperscript{38} Borghard, Erica D, and Shawn W Lonergan. 2016. “Can States Calculate the Risks of Using Cyber

\textsuperscript{39} Kautalya and L. N. Rangarajan. 1992. \textit{The Arthashastra}. New Delhi: Penguin Books India; Griffith,

\textsuperscript{40} Lin, Herbert. “Operational Considerations in Cyber-attack and Cyber Exploitation”. In \textit{Cyberspace and
National Security Threats, Opportunities, and Power in a Virtual World}. Edited by Reveron, Derek S.

\textsuperscript{41} Brantly, Aaron F. 2015. “Aesop’s Wolves: The Deceptive Appearance of Espionage and Attacks in
Cyberspace”. \textit{Intelligence and National Security} 31(5): 674-685.
Very few states have national deterrence strategies aimed at sub-state actors, criminal organizations or individuals. State deterrence strategies aimed at non-terrorist sub-state actors are confined to criminological models of deterrence. Yet, if a soldier or spy from an adversary state walked into the server room at the same hospital and threatened to detonate a bomb and destroy all the files unless he was paid a ransom, the act would align more closely with a conventional deterrence framework of state-to-state deterrence by threats of punishment or tailored deterrence against terrorist actors.

Most scholars and practitioners are likely to contend that it is not the responsibility of the state to deter non-state actors (excepting terrorists), particularly criminals from cyber attacks against non-federal infrastructure outside of a criminological framework. Yet, the same tool used by a criminal is available to the state and presents the same challenges associated with attribution irrespective of the perpetrator. What actions could a state undertake to deter an adversary state actor from engaging in this behavior and would these actions have a measurable effect on non-state actors as well?

Examples of cyber attacks abound and include the destruction, denial or degradation of military or civilian communications platforms. Attacks such as the Mirai (malware) botnet attack in 2016 are capable of being directed at both critical and non-critical infrastructure by both state and non-state actors. A botnet using Mirai was able to generate in excess of 1Tbps of traffic and degrade dozens of websites in the United States on 20 September 2016. This same form of attack could be directed towards IP addresses of the FAA and emergency service providers or any number of Internet-enabled systems found on Shodan.io or similar services.

Although DDoS attacks are generally considered to be among the least complicated forms of cyber attacks they still challenge state and sub-state entities both public and private. DDoS attacks have been used against US government infrastructure, against Estonia in 2007 and the Republic of Georgia in 2008. To date, DDoS attacks against the US government or critical infrastructure have received little attention in discussions on deterrence in cyberspace. On 21 January 2016 a grand jury in the Southern District of New York indicted 7 Iranian Hackers in absentia for their involvement in DDoS

attacks against US financial sector interests and a variety of other US companies occurring from 2011-2013. These indictments are: (a) not deterrent threats or denials, but criminological deterrents; (b) temporally distant from the time of attack as to be ineffective at signaling deterrence; and (c) impose little to no costs on Iran or the individual perpetrators or organizers of the attack.

Beyond DDoS attacks, Russian attacks against Ukrainian electric infrastructure and US political organizations also resulted in no or weak responses that offer no indication that deterrence is making headway in cyberspace. In response to massive influence operations perpetrated by the Russian Federation against the United States and its two major political parties during the 2016 Presidential election the United States expelled 35 suspected Russian intelligence operatives and placed sanctions on Russia’s two leading intelligence services, the FSB and the GRU. The US response imposed insignificant costs in comparison to the utility achieved by the Russian Federation.

The latter case of Russian influence and hacking during the 2016 election cycle provides a case study for why deterrence by threat in cyberspace is so difficult to achieve. The first indications of Russian interference in the 2016 election were identified by the FBI in September 2015 more than a year before the election. The FBI phoned the DNC to try and alert them to a potential attack, but the call was not considered credible and was subsequently ignored by DNC staffers. The progression of hacking attempts against the DNC continued and President Obama was notified in the summer of 2016. Moreover, the “attack” against the DNC was not an attack, but espionage or theft and therefore falls outside conventionally defined deterrence frameworks. Yet the impact of the espionage and the later release of private DNC emails was substantial as indicated in a declassified report by the Office of the Director of National Intelligence (ODNI). The report assessed that information warfare conducted following the espionage campaign substantially degraded the DNC and engendered a loss of confidence in the US electoral system. Cyber deterrence has fundamental problems including the realization that the most valuable assets in cyberspace might not be destroyed or degraded, but rather stolen and used.

51 Ibid.
53 Ibid.
Even in instances where specific code is used to achieve damage such as Iranian efforts to hack a spillway dam\(^{54}\) or malware implants in critical infrastructure such as a German steel mill\(^ {55}\) there are no formal mechanisms by which to signal a threat within cyberspace or beyond other than by referencing responses to kinetic effects. Current deterrence by threat signaling for attacks occurring in or through cyberspace is ambiguous. Efforts by the NATO CCD COE through the production of the Tallinn Manuals have begun to outline the frameworks in which deterrence could legally take place, yet the application of threats is still uncertain\(^ {56}\).

Deterrence by threat within cyberspace is realistically only applicable to cyber operations that result in direct physical effects that are non-repudiable and attributed quickly. Using formal modeling in the *Decision to Attack: Military and Intelligence Cyber Decision-making* I found that most cyber attacks, with the notable exception of DDoS, operate under varying conditions of anonymity\(^ {57}\). The anonymity associated with attacks is usually necessary for attacks to be successful in bypassing deterrence by denial frameworks found in the perimeter defenses of networks such as intrusion detection and prevention systems found in the logical or physical network layers of cyberspace. Threats of punishment could impact the persona layer of cyberspace as well, but as will be examined below there are some fundamental challenges unique to cyberspace posed by anonymity.

**4. TECHNICAL CHALLENGES: THREATS OF PUNISHMENT WITHIN DOMAIN**

Punishing an adversary in cyberspace is not cheap or fast outside of pre-established botnets or damage done to physical infrastructure. Punishment in or across any of the layers cyberspace requires what the US Department of the Army refers to as intelligence preparation of the battlefield (IPB):

> “IPB is a systemic, continuous process of analyzing the threat and environment in a specific geographic area. It is designed to support staff estimates and military decision making”\(^ {58}\).

---


\(^{58}\) US Department of the Army. 1994. FM 34-130 Intelligence Preparation of the Battlefield. Washington, DC.
In response to a nuclear attack on a city in the US, the proportional response would be a counter attack on an adversary city. The city itself is geographically fixed and immovable both logically and physically. Threatening in-kind retaliation is both plausible and technically feasible with ballistic missiles or air assets. The same logic does not hold in cyberspace.

Why are in-kind retaliations or other forms of punishment not viable solutions for most retaliations in cyberspace? First, a state must fulfill the burden of proof in identifying the perpetrator of an action. All the above IPB and potential for retaliation still depends upon attribution of who, what, and potentially why an attack occurred.\textsuperscript{59} Retaliation absent strong evidence is likely to lead to misidentification and unnecessary escalation.

Second, a state must retaliate within a proximate temporal range. If state X does not have detailed intelligence on the asset it wishes to retaliate against, developing intelligence along with a cyber weapon to target it increases the time horizon of response such that it is days, weeks, months or even years out from the original attack for which it is retaliating. Due to this temporal disconnect, the threat to punish in response to a given action falls into a category of what economists refer to as hyperbolic discounting. The risk of punishment for an attack is possible but so temporally, distant as to be discounted to the point of irrelevance.

Third, deterrence by punishment requires proportionality. It is necessary to have comparable assets to punish to prevent escalation or violations of international law.\textsuperscript{60} Comparable assets are not a given within cyberspace and are often difficult to identify.\textsuperscript{61} To punish an asset within a domain requires pre-established access or knowledge of that asset beyond its location. Whereas a city is immovable and likely to be as susceptible today as it will be tomorrow to a missile or bomb, a computer system that is penetrated today for prepositioned access, might be patched, upgraded or taken offline tomorrow.

Fourth, a state must possess a specific cyber weapon system tailored to its target. If state X alerts state Y that it is going to punish an asset or state X uses a repeated cyber weapon to attack state Y’s system, it is likely to be ineffectual the longer it is used due to updated perimeter defenses, such as intrusion detection and prevention systems (IDPS), antivirus programs or a variety of other security measures. If state X wants to punish state Y it must have knowledge of the attributes of the asset it wishes to retaliate against and what the status of that asset is. State X must also develop new exploits to achieve effects or be confident that State Y has not accounted for previous exploits that have been used.

\textsuperscript{59} Brantly. 2016.
The challenges of signaling deterrence by punishment are numerous within cyberspace whether the conflict is contained within domain or crosses over domains. Advances in attribution within a timely manner and the availability and reasonable assumption that proportional assets of an adversary can be held at risk need to be improved to credibly threaten punishment. This is a challenge not isolated to within domain retaliation. While proportional target selection might be slightly easier in cross-domain retaliation, the first three issues raised above are still relevant.

Deterrence by punishment in cyberspace is possible, but it is not a reliable or credible option under most conditions absent sufficient and sustained intelligence. This assessment is not unique and is borne out in the analysis of Valeriano and Maness, who find that deterrence via punishment is generally ineffective and likely more dangerous than other means of preventing attacks. Moreover, sustained invasive intelligence into adversary networks creates its own unique problems, including a security dilemma. The more states engage in highly invasive intelligence via cyberspace, the more their actions are likely to be misinterpreted. Differentiating between various forms of cyber actions are difficult and can lead to miscalculation.

Figure 1 illustrates the relationship between attacking and defending forces and area where both forms of deterrence function.

FIGURE 1. TIMELINE OF CYBER ATTACKS AND DEFENSE

64 Brantly. 2016.
As seen in Figure 1, deterrence by threat of punishment and denial operate within the same temporal ranges, yet while attribution matters a great deal for threats of punishment they are generally unimportant for denial. In their initial stages both denial and punishment focus on ante-impetum means of dissuasion, yet deterrence by punishment necessarily needs post-impetum attribution for it to be used. Based on the technical realities of cyberspace and of international relations deterrence by threat of punishment is more complicated and difficult to effectively establish.

5. TECHNICAL CHALLENGES AND OPPORTUNITIES: DETERRENCE BY DENIAL

Both deterrence by denial and punishment require ante-impetum costs by the defender. The allocation of resources between denial and deterrence and the efficiency with which they deter adversaries differ. The establishment of credible deterrence by denial often starts with the allocation of financial capital to purchase technical resources and provide human capital sufficient to continually update, enhance, audit and manage complex network infrastructure.\textsuperscript{65} Network-based and host-based defenses such as intrusion detection and prevention systems, anti-virus products and similar systems are some of the variety of overlapping expenditures that can be undertaken to increasingly make the intrusion of adversaries into a given network more difficult.\textsuperscript{66}

In cyberspace, such expenditures are regularized and often included as overhead costs, however they are deterrent in nature.\textsuperscript{67} Although they are not glamorous, they substantially decrease the probability of penetration. The same types of deterrence strategies are used by stores in placing electronic tracking tags on their products and detectors at doors, by banks in the construction of vaults, silent alarms and dyed packets of money, by critical infrastructure in extending the perimeter of security outward to prevent vehicle-borne improvised explosive devices, increased numbers of security guards, cameras and the use of razor wire or other physical structures. These devices signal to adversaries both criminal and terrorist alike that the costs of successfully perpetrating an attack are high and that the likelihood of success is low, although both terrorist and criminal deterrence models include deterrence by punishment through criminal proceedings and potential lethal actions against terrorist they rely far more heavily on preventive measures that deny would be adversaries.

Sceptics might contend denial mechanisms are unlikely to deter a state, yet this is in and of itself not accurate. The vast majority of probes by states do not translate into successfully attacks. The US Department of Defense suffers from millions of probes

a day. Yet nearly 99.99% of them are unsuccessful. Moreover, in the face of a global onslaught of cyber attacks and espionage the United States re-architected much of its military network infrastructure. This restructuring allows the initial point of contact with adversaries to be chosen. In military parlance, it allowed the defenders to choose the terrain of the battle. While it did not obviate the need for denial mechanisms within the network infrastructure, it did signal increased cost imposition on adversaries and it did allow for more efficient resource allocation.

Unlike in any other battlespace, whether conventional kinetic terrorism, conventional kinetic or mass destruction military force, the opportunities for deterrence by denial are substantial in cyberspace and unique. While denial opportunities in land, sea, air, and even space are predicated on the control of a given geospatial area, the party establishing deterrence by denial has limited abilities to manipulate the nature of the domain itself. The same is not true within cyberspace. Every aspect of a defender’s cyberspace from the structure of the network, to the hardware, firmware, and software within a network, to the access of individuals within and external to that network is manipulable. At every stage of an attack an adversary is always attempting to operate on or against the defender’s cyberspace over which it has no control and has limited visibility.

For denial, the historical literature of deterrence theory remains relevant, in particular the second and third stages of deterrence which focused on rational game theoretic and cognitive modeling. While in conventional deterrence the emphasis was on punishment, here these same modeling techniques find applicability in deterrence by denial. Although the games might be the same, the payoffs in cyberspace manipulable and favor the defender. In few other applications of deterrence are the payoff matrices of deterrence so favorable to the defender. Despite the favorability of conditions, the ability to manipulate the potential payoff for attackers remains difficult. Although possible for defenders to reduce the probability of attack success, the potential payoff for a successful attack can remain large.

Despite conditions favoring defenders, the potential payoffs are often not affected by deterrence by denial. Minimizing the potential payoffs from attacks on data repositories requires disaggregation of data. These types of denial mechanisms come with efficiency or financial costs. Although denial offers more potential than punishment, it is not a silver bullet to the cyber deterrence problem. Denial decreases the probability of success for attackers and is likely to reduce classes of actors focused on certain targets. Despite efforts to signal through the purchase and implementation of various defensive measures, the re-architecting of network infrastructure, the cyber deterrence problem remains.

6. BEYOND THE DETERRENCE PROBLEM

If punishment and denial are unable to fully remediate the cyber deterrence problem, are there any meaningful solutions? The core debate remains, with no simple and readily apparent solutions. The search for a single solution is likely to remain fruitless for the foreseeable future. Deterrence has never been the single tool within the toolbox of the state to dissuade or shape adversary behavior. Rather, it has always been combined with efforts that extend beyond traditional concepts of deterrence to include geopolitical and technical practices including norm development, entanglement, cumulative deterrence, research and development, policies and laws, liability structures for software and hardware, training for users and human capital development within information technology and cybersecurity.69

Efficient and effective cyber deterrence should extend international politics and include fields such as criminology, immunology and public health.70 The capacity of states to punish criminals is high and the credibility of punishment actions in developed nations is strong. Despite a capacity to punish criminal behaviors, they still occur. Extending beyond punishment, states also focus on denying criminals opportunities to commit crimes. Yet crime still occurs. The root causes of crime are not simple nor isolatable to a single phenomenon. Likewise, states engage one another in cyberspace for a variety of reasons. Some reasons fit within conventional deterrence frameworks of denial and punishment and do not suffer from challenges with attribution. For instance, larger and more harmful attacks increase the probability of attribution. However, many states remain perturbed by the death by a thousand cuts phenomena which falls below thresholds and required to provide timely attribution.

Shifting the focus away from within domain deterrence focused solely on punishment and denial and changing the emphasis to a basket of strategies focused on reducing incentives, availability and anonymity fosters an environment less conducive both to hostile actions and potential malicious actors. The solution to the deterrence problem is not abandoning it, but expanding the range of alternative strategies not presently considered. By acknowledging the failures and inadequacies of deterrence strategies and the potential places where novel strategies found in other fields are applicable the intractable problem of cyber deterrence becomes more manageable.

REFERENCES


