SANDIA REPORT
SAND2020-5016
Printed Click to enter a date

Sandia National Laboratories

Why does cyber deterrence fail, and when might it succeed?

A framework for cyber scenario analysis

Eva C. Uribe
Benjamin J. Bonin
Michael F. Minner
Jason C. Reinhardt
Ann E. Hammer
Nerayo P. Teclemariam
Trisha H. Miller
Ruby E. Booth
Robert D. Forrest
Jeffrey J. Apolis
Lynn I. Yang

Prepared by
Sandia National Laboratories
Albuquerque, New Mexico
87185 and Livermore,
California 94550

Issued by Sandia National Laboratories, operated for the United States Department of Energy by National Technology & Engineering Solutions of Sandia, LLC.

Available to DOE and DOE contractors from
U.S. Department of Energy
Office of Scientific and Technical Information
P.O. Box 62
Oak Ridge, TN 37831

Telephone: (865) 576-8401
Facsimile: (865) 576-5728
E-Mail: reports@osti.gov
Online ordering: http://www.osti.gov/scitech

Available to the public from
U.S. Department of Commerce
National Technical Information Service
5301 Shawnee Rd
Alexandria, VA 22312

Telephone: (800) 553-6847
Facsimile: (703) 605-6900
E-Mail: orders@ntis.gov
Online order: https://classic.ntis.gov/help/order-methods/

ABSTRACT

Through cyberattacks on information technology and digital communications systems, antagonists have increasingly been able to alter the strategic balance in their favor without provoking serious consequences. Conflict within and through the cyber domain is inherently different from conflict in other domains that house our critical systems. These differences result in new challenges for defending and creating resilient systems, and for deterring those who would wish to disrupt or destroy them. The purpose of this paper is to further examine the question of whether or not *deterrence* can be an effective strategy in cyber conflict, given our broad and varied interests in cyberspace.

We define deterrence broadly as the creation of conditions that dissuade antagonists from taking unwanted actions because they believe that they will incur unacceptably high costs and/or receive insufficient benefits from taking that action. Deterrence may or may not be the most credible or effective strategy for achieving our desired end states in cybersecurity. Regardless of the answer here, however, it is important to consider why deterrence strategies might succeed under certain conditions, and to understand why deterrence is not effective within the myriad contexts that it appears fail. Deterrence remains a key component of U.S. cyber strategy, but there is little detail on how to operationalize or implement this policy, how to bring a whole-of-government and whole-of-private-sector approach to cyber deterrence, which types of antagonists can or should be deterred, and in which contexts. Moreover, discussion about how nations can and should respond to significant cyber incidents largely centers around whether or not the incident constitutes a "use of force," which would justify certain types of responses according to international law. However, we believe the "use of force" threshold is inadequate to describe the myriad interests and objectives of actors in cyberspace, both attackers and defenders.

In this paper, we propose an approach to further examine if deterrence is an effective strategy and under which conditions. Our approach includes systematic analysis of cyber incident scenarios using a framework to evaluate the effectiveness of various activities in influencing antagonist behavior. While we only examine a single scenario for this paper, we propose that additional work is needed to more fully understand how various alternative thresholds constrain or unleash options for actors to influence one another's behavior in the cyber domain.

ACKNOWLEDGEMENTS

We would like to thank all of the members of the Cyber Deterrence and Resilience Strategic Initiative, especially Richard Griffith and Heidi Ammerlahn, for their guidance and support. We also acknowledge the leadership of Division 8000, especially Any McIlroy and Dori Ellis, for their support of the strategic initiative. We acknowledge the following people for reviewing and commenting on this work: Carol Adkins (SNL), Marcus Chang (SNL), Susanna Gordon (SNL), Sheryl Hingorani (LLNL, formerly SNL), John P. Hinton (consultant to SNL), Margot Hutchins (SNL), Robert Hwang (SNL), Miriam "Mim" John (consultant to SNL), Karim Mahrous (SNL), Michael Nacht (UC Berkeley), Noël Nachtigal (SNL), James Novak (SNL), Ali Pinar (SNL), Anup Singh (SNL), Benn Tannenbaum (SNL), David White (SNL), and Evan Wolff (Crowell & Moring).

DISCLAIMER

The work described in this report was performed as part of the Cyber Deterrence and Resilience Strategic Initiative, an internally funded effort at Sandia National Laboratories. Strategic initiatives crosscut existing programs at the lab and fund exploratory studies that identify opportunities to inform the national dialogue on emerging national security threats and challenges. Contributing authors are from the Systems Research and Analysis Group, which engages multidisciplinary teams to investigate national security solutions in complex systems, through consideration of technology, policy, operations, and human factors. The views expressed within this publication are solely those of the authors and do not necessarily represent the views of Sandia National Laboratories or any other agency or sponsor. This paper is approved for unlimited release as SAND2020-5016.

Sandia National Laboratories is a multimission laboratory managed and operated by National Technology & Engineering Solutions of Sandia, LLC, a wholly owned subsidiary of Honeywell International Inc., for the U.S. Department of Energy's National Nuclear Security Administration under contract DE-NA0003525. This paper describes objective technical results and analysis. Any subjective views or opinions that might be expressed in the paper do not necessarily represent the views of the U.S. Department of Energy or the United States Government.

CONTENTS

LIST OF FIGURES

LIST OF TABLES

ACRONYMS AND DEFINITIONS

Abbreviation	Definition
APT	Advanced Persistent Threat
CAN	computer network attack
CNE	Computer Network Exploitation
CSS	Central Security Service
DEFCON	defense readiness condition
DNC	Democratic National Committee
ICJ	International Court of Justice
ICS	industrial control system
IDS	intrusion detection system
IPS	intrusion protection system
NPR	Nuclear Posture Review
NSA	National Security Agency
OPM	Office of Personnel Management
PPD-21	Presidential Policy Directive on Critical Infrastructure Security & Resilience
UN	United Nations
USCYBERCOM	United States Cyber Command
WMD	weapons of mass destruction

1. INTRODUCTION

As the people, governments, and infrastructure throughout the world become increasingly interconnected with and dependent on global information technology and digital communications networks, the question of how to protect them from intentional degradation and destruction becomes more urgent. Through cyberattacks and intrusions on these systems, antagonists[1] have increasingly been able to alter the strategic balance in their favor without provoking serious consequences. Conflict within and through the cyber domain is inherently different from conflict in other physical domains that house our critical systems.[2] The sheer number of highly capable cyber antagonists, and the speed and scale at which they are able to operate, is unprecedented in other types of conflict. It is relatively easy for cyber actors to conceal not only their identities, but also the effects of their actions, which complicates and delays the detection and attribution of cyber incidents. Nations have not yet had the opportunity to reach common understanding about what constitutes a use of armed force within cyberspace, and thus what justifies the use of force in self-defense, a process that took centuries of warring and bargaining to even partially negotiate for the physical domains. Moreover, "cyber" is not just another military domain.[3] Cyber tools and assets support and connect all of the other domains. Furthermore, they are the foundation upon which modern civilian society is built, which further ambiguates the civilian-military divide.

Together, these features generate new challenges for defending and creating resilient systems, and for deterring anyone who wants to disrupt or destroy them. They make it harder for actors to achieve their desired end states within the domain, whether that end-state is dominance of the cyberspace domain, stability[4] through mutual vulnerability, or some combination of both. The purpose of this paper is to further examine the question of whether or not *deterrence* can be an effective strategy in cyber conflict, given our broad and varied interests in cyberspace. We define deterrence broadly as the creation of conditions that dissuade antagonists from taking unwanted actions because they believe that they will incur unacceptably high costs and/or receive insufficient benefits from taking those actions. Deterrence may or may not be the most credible or effective strategy for achieving our desired end states in cybersecurity. Which end state we as a nation should be working toward is beyond the scope of this paper. But, whatever that end state is, it may well require a complex mixture of well-defended, resilient cyber infrastructures, judicious use of offensive cyber capabilities, and deterrence strategies to dissuade antagonists from taking unwanted actions in

[1] Throughout this paper, we will use the term *antagonist* to describe an actor that is considering to take an action in an attempt to upset or revise the status quo. We choose the term antagonist because it is broader than *adversary*, which may more narrowly imply the opponent in an active or hostile conflict. Antagonism, particularly in cyber contexts, includes actions below the threshold of active violence that still may have strategic effect. We use the term *protagonist* to describe an actor that wishes to maintain the status quo. A protagonist may seek to deter or dissuade an antagonist from attempting to revise the status quo. We also use the terms *Red* and *Blue* to identify actors in the context of specific scenarios. For the scenario described in this paper, Red is the antagonist and Blue is the protagonist, but of course in other contexts these roles could be reversed.

[2] Martin C Libicki, *Cyberdeterrence and Cyberwar* (Santa Monica: RAND Corporation, 2009)

[3] Schneider has aptly problematized the characterization of cyberspace as a *domain*, distinct from the physical military domains of land, air, sea, and space: "It might be administratively cohesive to think of cyberspace as a domain and deterrence, therefore, as across and through the cyberspace domain. However, the interpretation of cyberspace as a societal infrastructure that connects not only warfighting domains but also civilian networks and functions significantly complicates the deterrence discussion. Cyberspace in this understanding becomes a target we must deter others from attacking…Imagine, for example, examining a tank's ability to deter land, sea, and air conventional operations versus a highway's ability to deter those same operations." Jacquelyn Schneider, "Deterrence in and through Cyberspace," in *Cross-Domain Deterrence: Strategy in an Era of Complexity*, ed. Jon Lindsay and Erik Gartzke (Oxford University Press, 2019)

[4] Features that characterize conflict stability include: no potential party to conflict is incentivized to initiate conflict, no party is incentivized to escalate if conflict begins, off-ramps exist allowing either party to preserve their reputation, and there are transparent and formalized standards to prevent accidents or unintended consequences.

the first place. Regardless of the answer here, however, it is important to consider why deterrence strategies might succeed under certain conditions, and to understand why deterrence is not effective within the myriad contexts that it appears to be failing. This consideration is important primarily for two reasons.

The first reason is that perfect defense is effectively impossible. Cyber defenders may not be able to protect and defend all of their critical systems to the highest degree. Even for those systems that are well-protected, some may be so critical to national security that no amount of cybersecurity is truly adequate. Attackers only have to find one way in, but defenders must guard an infinite number of entry points, giving rise to a conflict dynamic dominated by offense. In 2018, the U.S. Department of Energy stated, "Anticipating and reacting to the latest cyber threat is a ceaseless endeavor that requires ever more resources and manpower. This approach to cybersecurity is not efficient, effective, nor sustainable… cyber threats continue to outpace our best defenses."[5] The Defense Science Board Task Force on Cyber Deterrence anticipates that this state will persist: "For at least the coming five to ten years, the offensive cyber capabilities of our most capable potential adversaries are likely to far exceed the United States' ability to defend and adequately strengthen the resilience of its critical infrastructures."[6] For this reason, even strategies centered on shifting cyberspace into a defense-dominated domain will still include options to deter and dissuade antagonists.

The second reason is that deterrence of cyber antagonists is currently the policy of the United States (U.S.). The U.S. has grappled with how to deter cyber actors for at least a decade. Serious consideration of developing a deterrence policy for cyberspace began in the Obama administration in 2009,[7] when the difficulty of attributing cyber events was largely thought to rule out most options for traditional deterrence by punishment.[8] The administration's 2009, 2011, and 2015 cyber strategies all advocated for deterring cyber adversaries using approaches that combined network resilience, denial of antagonist success, and attribution and response.[9] The 2015 strategy further stated that in conducting all cyber operations, the U.S. will follow a "doctrine of restraint" in accordance with the Law of Armed Conflict.[10] Over time, this approach was criticized for emphasizing restraint and resilience over actions that would impose more immediate and tangible costs on adversaries in retaliation for their actions.[11]

---

[5] Multiyear Plan for Energy Sector Cybersecurity, (Office of Electricity Delivery & Energy Reliability, 2018)

[6] *Task Force on Cyber Deterrence*, Department of Defense Defense Science Board (2017), https://www.acq.osd.mil/dsb/reports/2010s/DSB-cyberDeterrenceReport_02-28-17_Final.pdf

[7] Schneider, "Deterrence in and through Cyberspace,"

[8] William J. III Lynn, "Defending a New Domain: The Pentagon's Cyberstrategy," *Foreign Affairs*, no. September/October 2010 (2010), https://www.cybercom.mil/About/Mission-and-Vision/

[9] The Comprehensive National Cybersecurity Initiative, (Washington, D.C. 2009) ; International Strategy for Cyberspace: Prosperity, Security, and Openness in a Networked World, (Washington, D.C. 2011) ; The Department of Defense Cyber Strategy, (Washington, D.C. 2015)

[10] The Department of Defense Cyber Strategy, Short

[11] There are competing views about the effectiveness of cyber deterrence efforts in the years between 2011 and 2015. Nye describes a series of attacks that could all be considered "failures of deterrence," but all were considered low-threshold attacks with limited impact on national security (Joseph S. Jr. Nye, "Deterrence and Dissuasion in Cyberspace," *International Security* 41, no. 3 (2017), https://doi.org/10.1162/ISEC_a_00266). Schneider describes a series of responses the U.S. made to cyber events that strengthened its credibility for deterrence by punishment (Schneider, "Deterrence in and through Cyberspace,"), including Executive Order 13694, which allowed the Treasury Department to enact sanctions to respond to cyber attacks. During this time, the State Department released recommendations for norms in cyberspace, and Presidential Policy Directive 21, "Critical Infrastructure Security and Resilience" designated sixteen U.S. critical infrastructure sectors Presidential Policy Directive 21: Critical Infrastructure Security and Resilience, (Washington, D.C. 2013) to signal to potential adversaries that the U.S. wanted to deter attacks against these specific civilian targets.

A new administration brought renewed attention to the question of the role of deterrence in U.S. cyber strategy. The Trump administration's 2017 National Security Strategy called for the ability to deter cyber antagonists both by imposing "swift and costly consequences" and by building more resilient critical infrastructure.[12] An early Presidential Executive Order, "On Strengthening the Cyber Security of Federal Networks and Critical Infrastructure," required the new administration to assess strategic options for deterring cyber adversaries.[13] The unclassified summary report issued a year later claims "the United States remains in a strong position to deter attacks that would constitute a use of force because traditional tools of deterrence remain effective and potent," but notes "there are significant challenges in deterring the substantial increase in malicious state-sponsored cyber activity occurring below the threshold of the use of force."[14] Accordingly, the report defines two goals for U.S. cyber deterrence efforts: "A continued absence of cyberattacks that constitute a use of force against the U.S.," and "reduction in destructive, disruptive, or destabilizing cyber activities against U.S. interests below the threshold of the use of force."[15] The *use of force* deterrence threshold paradigm has been broadly adopted in policy and academia. A plausible theory has taken root that *below* the threshold of the use of force, deterrence is not a credible strategy for achieving U.S. interests in cyberspace, and that the fundamental nature of cyberspace calls for strategies based on constant operational contact and persistent engagement with cyber antagonists, as opposed to deterrence and operational restraint.[16] (Although this may prove to be an unsatisfactory dichotomy, as we shall discuss later.)

Deterrence remains a key component of U.S. cyber strategy, but there is little detail on how to operationalize or implement this policy, how to bring a whole-of-government and whole-of-private-sector approach to cyber deterrence, which types of antagonists can or should be deterred, and in which contexts. Additionally, there is reason to question the hierarchical breakdown of cyber incidents into those which are equivalent to the use of force and those that fall below the threshold of the use of force. It does not appear that nation state rivals are presently waging cyber war against each other. To the contrary, nation states appear to be using cyber methods specifically to produce strategic effects below the threshold of armed conflict, that is, below a level where a defender could justifiably resort to the use of force in self-defense, or to generate coercive options for themselves if conflict escalates above this threshold. The 2018 Command Vision of the U.S. Cyber Command (USCYBERCOM) recognizes that "adversaries operate continuously below the threshold of armed conflict to weaken our institutions and gain strategic advantages."[17] This statement implies that certain antagonists generate strategic effects without crossing the threshold of armed conflict. Therefore, there is great need to consider additional thresholds that may be implicitly guiding cyber conflict dynamics.

In this paper, we propose an approach to further examine whether deterrence is an effective strategy, and under which conditions. Our approach includes systematic analysis of cyber incident scenarios using a framework to evaluate the effectiveness of various activities in influencing antagonist behavior.[18] In the first part of the paper, we discuss the origins of the *use of force* threshold

---

[12] National Security Strategy of the United States of America, (Washington, D.C. 2017)

[13] Presidential Executive Order on Strengthening the Cybersecurity of Federal Networks and Critical Infrastructure, 13800 (2017)

[14] *Recommendations to the President on Deterring Adversaries and Better Protecting the American People from Cyber Threats*, (2018), https://www.state.gov/documents/organization/282253.pdf

[15] *Recommendations to the President on Deterring Adversaries and Better Protecting the American People from Cyber Threats*,

[16] Michael P. Fischerkeller and Richard J. Harknett, "Deterrence is Not a Credible Strategy for Cyberspace," *Orbis* 61, no. 3 (2017), https://doi.org/10.1016/j.orbis.2017.05.003

[17] Achieve and Maintain Cyberspace Superiority: Command Vision for US Cyber Command, (2018)

[18] Through the Office of the Secretary of Defense Minerva program, Robert Jervis and Jason Healey have conducted similar systematic study of the dynamics of cyber conflict (Robert Jervis and Jason Healey, *The Dynamics of Cyber Conflict*, School of

and propose an alternative framing. In the second part, we describe our cyber deterrence framework. And in the third, we demonstrate the use of this framework using a generic scenario involving an Advanced Persistent Threat against electric grid infrastructure. While we only examine one scenario for this paper, we propose that additional work is needed to understand how various alternative thresholds constrain or unleash options for actors to influence one another's behavior in the cyber domain.

International and Public Affairs, Columbia University (New York, NY, 2019),): "Since there are so many dynamics, and little work to structure and analyze the entire set, it has been easy for strategists and academics to, unconsciously or not, cherry pick some and ignore others."

2. THRESHOLDS AND DETERRENCE OF CYBER ADVERSARIES

2.1. Perspectives from Deterrence Literature

The concept of conflict "thresholds" was prominent in deterrence literature published during the Cold War. Thomas Schelling described thresholds as "finite steps in the enlargement of a war or a change in participation." Thresholds were especially salient in the analysis of conflict involving nuclear-armed powers; the point at which nuclear weapons might be employed represented an important dividing line between conventional conflict and something more devastating and uncertain. Indeed, for theorists like Bernard Brodie who believed that nuclear war was a losing proposition for all involved, deterrence was fundamentally aimed at preventing the crossing of that threshold.[19]

For others, like Herman Kahn or Henry Kissinger, there existed a broader spectrum of plausible escalatory thresholds. Lower rungs in the "escalation ladder" (a term coined by Kahn and subsequently widely adopted to describe both nuclear and non-nuclear conflict intensity) might not even involve overt employment of nuclear weapons, progressing through diplomatic statements and gestures, legal actions (e.g. sanctions), mobilization of conventional assets, and escalating levels of conventional warfare. If conflict reached a sufficient level of intensity (with neither combatant deterred from further escalation), mobilization and perhaps even demonstration of nuclear capabilities might take place to signal the potential for nuclear war (and to provide an opportunity for the opposing party to back down). This step might be followed by "local" or "limited" nuclear strikes in a battlefield context, then by extended and more widespread counterforce attacks on primarily military targets, and ultimately escalation to full-scale war involving the use of hundreds or thousands of nuclear weapons. Kahn also introduced the concept of "escalation dominance" to describe a suite of capabilities and strategies allowing one to prevail at any rung in the ladder; Kahn controversially argued for such a capability as the surest deterrent against Soviet aggression, and debates regarding the merits and limitations of escalation dominance would continue throughout the Cold War.[20]

While decisionmakers never adopted anything as complex as Kahn's own 44-rung escalation ladder, thresholds nonetheless played a prominent role in U.S. policy during the Cold War and continue to do so into the present day. The five-level defense readiness condition (DEFCON) scale is one of the most well-known examples of an explicitly defined set of thresholds, in this case intended to represent a condition of military readiness, with 5 indicating a state of normality (lowest readiness) and 1 indicating imminent nuclear war (maximum readiness). While the scale is ostensibly intended to inform U.S. military forces and allies of readiness requirements, a change in the DEFCON scale (to the extent it is broadcast openly) can also send an implicit message or warning to an antagonist that certain conflict thresholds have been (or are very close to being) crossed.[21]

Since 1945, it has been U.S. policy to maintain a certain level of ambiguity as to the exact declaratory threshold at which U.S. nuclear weapons might be used. This uncertainty was intended, in large part, to serve as a deterrent to Soviet aggression in western Europe; Soviet leaders considering even a limited conventional offensive could not be certain it would not invite a retaliatory nuclear strike.

---

[19] Thomas C. Schelling, *Arms and Influence* (New Haven, CT: Yale University Press, 1966); Bernard Brodie et al., *The Absolute Weapon: Atomic Power and World Order* (New York, NY: Harcourt, Brace and Co., 1946)

[20] Herman Kahn, *On Escalation: Metaphors and Scenarios* (Transaction Publishers, 1965); Henry Kissinger, *Nuclear Weapons and Foreign Policy* (New York, NY: Routledge, 1957)

[21] North American Air Defense Command and Continental Air Defense Command Historical Summary, (Directorate of Command History, Office of Information, Headquarters NORAD/CONAD, 1963)

More recently, U.S. Nuclear Posture Reviews (NPRs) since the George W. Bush administration have attempted to more clearly narrow the range of contingencies under which nuclear weapons might be used, going as far as to indicate they will not be used against non-nuclear weapons states in good standing with the Treaty on the Nonproliferation of Nuclear Weapons (the so-called "negative security assurance"), and that use will only be considered under extreme circumstances, such as in response to use of weapons of mass destruction (WMD) by an adversary, or, as articulated by the Trump administration, in the event of "Significant non-nuclear strategic attacks…on the U.S., allied, or partner civilian population or infrastructure, and attacks on U.S. or allied nuclear forces, their command and control, or warning and attack assessment capabilities."[22]

Thresholds may also be defined in response to specific crises. The basing of Soviet nuclear missiles in Cuba represented a crossed threshold for the Kennedy administration, which then set down a further (and more explicit) threshold through imposition of a naval blockade; attempts to run the blockade ostensibly would have been met with military action, to potentially include use of nuclear weapons, if the Soviets had not backed down.[23] A more recent example relates to the use of chemical weapons in the Syrian Civil War. Successive U.S. administrations (Obama and Trump) indicated that use of such weapons represented an unacceptable escalation of the conflict meriting a punitive response. In the case of the Obama administration, this response took the form of an international intervention to disarm Syria under the auspices of the Chemical Weapons Convention, while under Trump it took the form of military strikes.[24]

Schelling emphasized that, while some escalatory thresholds are explicitly defined, many (perhaps most) are implicit or arrived at through "tacit bargaining" in which combatants might demonstrate the existence of such thresholds through limits placed on their own actions (e.g., German non-use of chemical weapons against the Allied powers during World War II), or consistent patterns of behavior over time (e.g., U.S. non-use of nuclear weapons against conventional military aggressors in every regional conflict dating back to the Korean War). Schelling further recognized a normative dimension, noting that while there is often a "legalistic quality" to the existence and observance of thresholds, "[f]or the most part they are just 'there'; we don't make them or invent them, but only recognize them."[25]

2.2. Two Perspectives on Cyber Thresholds

2.2.1. Linear Escalation and the Use of Force

While thresholds are not necessarily the central focus of this study, the concept provides a useful construct for reflecting on the current state of cyber deterrence. *Figure 1* below illustrates five ostensible conflict thresholds – points at which an antagonist's actions cross a notional escalatory line in terms of the harm caused (or threatened) to national interests, and at which proportionally more intense retaliation might be considered. This simple escalation ladder is consistent with the linear approach to thresholds used by historical and contemporary scholars. It is inspired by literature previously cited, as well as the recent work of Nadiya Kostyuk et al., who define a more granular escalation ladder in relation to the 2008 Army Field Manual "Spectrum of Conflict."[26]

---

[22]Nuclear Posture Review Report, (Washington, DC: Office of the Secretary of Defense, 2010) ; Nuclear Posture Review, (Office of the Secretary of Defense, 2018)

[23] Graham Allison and Philip Zelikow, *Essence of Decision: Explaining the Cuban Missile Crisis* (New York, NY: Longman, 1999), p. 109-29

[24] Peter Baker, "For Obama, Syria Chemical Attack Shows Risk of 'Deals With Dictators'," *New York Times*, April 10 2017, A

[25] Schelling, *Arms and Influence*

[26] Nadiya Kostyuk, Scott Powell, and Matt Skach, "Determinants of the Cyber Escalation Ladder," *The Cyber Defense Review* 3, no. 1 (2018)

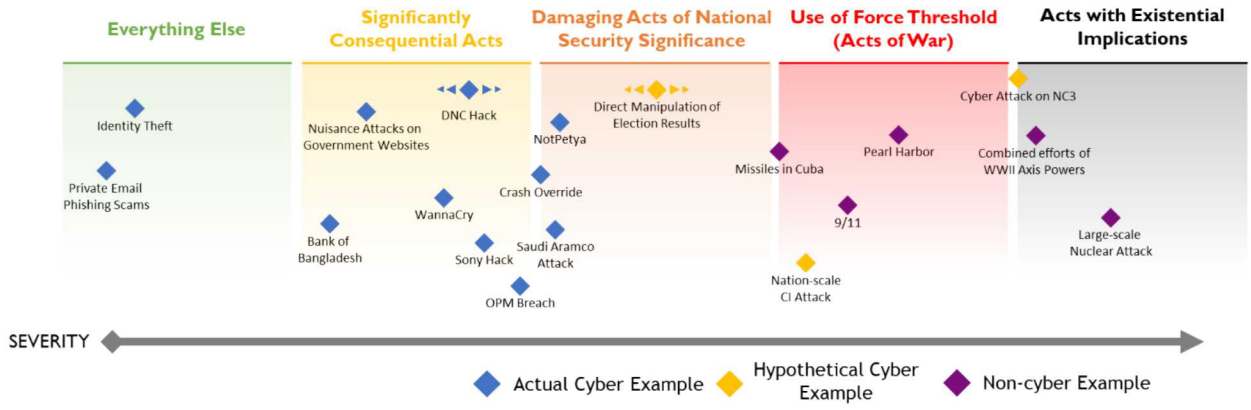

Figure 1: Incidents mapped against conflict thresholds

The left side of the figure begins with a catch-all category – "everything else" – comprising events and acts that, while not harmless, largely impact at the individual (rather than national) level. This includes lower-level criminal acts and disruptions that merit localized responses at best. Events in this category contribute to the day-to-day "cost of doing business" in cyberspace. Action in the second category – "significantly consequential acts" – do not arguably cause harm to national security, but their implications still resonate at a national level, causing minor but visible disruptions to the economy, infrastructure, or governance functions. Actions crossing this threshold may also include "routine" espionage activities carried out by means of human assets, signals collection, network penetration, and other methods (though such acts can bleed into the next category depending on the sensitivity of information acquired and the national security implications of its compromise). Responses to such activity may include legal and law enforcement actions and counterintelligence activities.

The middle category – "Damaging Acts of National Security Significance" – comprises actions which threaten serious harm to core national security interests but fall short of meriting an armed military response. They may result in significant but largely short-term disruption of the economy, governance functions, or critical infrastructure operations. This category may also include the theft of confidential information with severe implications for national security. Actions at this level might merit economic and diplomatic sanctions, counterintelligence operations, or in some extreme cases possibly even low-level, covert military action. The transition between this category of actions and the next represents a key inflection point in literature and discourse on international conflict; it is the point at which actions comprise a "use of force" that might justify an overt military response (i.e., war) under international law.

The body of international law governing armed conflict is not defined by a single document or protocol; rather, it has evolved through centuries of historical precedent established by state-to-state conduct in war and peacetime and, more recently, negotiated international agreements like the Geneva Conventions and the United Nations (U.N.) Charter. Unsurprisingly, interpretations and implementation can vary significantly from nation state to nation state, in accordance with domestic law, national interest, and political value systems. Even a concept as fundamental as the conditions under which states may justifiably resort to war is a controversial proposition, no more so than in

the present day, where states have a menu of unconventional offensive options available (including cyber) that test the historical boundaries of international law.

The U.N. Charter, considered by many international legal scholars to be authoritative on *jus ad bellum* (the right to war), makes it clear at the outset that members "shall refrain in their international relations from the threat or use of force against the territorial integrity or political independence of any state, or in any other manner inconsistent with the Purposes of the United Nations." It also makes it clear that when member states face such a threat or use of force, they are justified to act in self-defense. Michael Schmitt, a scholar of international law as it applies to cyber, notes that while adjudicating authorities like the International Court of Justice (ICJ) have "rejected a narrow interpretation of 'use of force' that limits the term to the employment of either kinetic force or non-kinetic operations generating comparable effects", the law is frustratingly vague as to precisely where the non-kinetic thresholds lie. The international group of experts that met to draft the 2013 *Tallinn Manual on the International Law Applicable to Cyber Warfare* agreed on a set of a factors or general criteria for judging whether a cyber incident might cross the use-of-force threshold, but only on the criteria of severity (i.e. damage) could they agree that anything more than *de minimis* (minimal or trivial) damage or injury could qualify an action for the designation. The group cited the damage done to Iranian nuclear facilities by the Stuxnet virus as an example of an incident that might meet the use-of-force threshold.[27]

A full treatment of the use-of-force threshold is beyond the scope of this study. Suffice to say, for the purposes of *Figure 1*, use-of-force actions threaten core national interests through large numbers of civilian and/or military casualties, damage or disruption to core military capabilities, large-scale disruption of critical infrastructure with devastating effect on national critical functions, loss of territory, severe economic disruption, or threats to core national interests abroad (including alliance commitments). A military response is conceivable under those circumstances. The final category – "acts with existential implications" – is an extension of the previous "acts of war" category, and comprises actions that threaten the very existence of the state through demographically transformative casualties, irrevocable collapse of governing structures, destruction of military capabilities or forces such that it can no longer provide for the national defense, large-scale loss of territorial integrity, or possibly even undermining the long-term viability of core economic systems. These acts are rare and often theoretical, with a large-scale nuclear attack being the archetypical example. Existential threats merit potentially extreme responses, including mobilization of national resources for "total war"; invasion, overthrow, and occupation of adversary nations; or the use of large numbers of nuclear weapons in an overwhelming first strike or counterstrike.

Various incidents are plotted against these thresholds in *Figure 1*; these include real-world cyber security incidents (represented by blue diamonds), hypothetical cyber incidents (orange diamonds), and non-cyber incidents (purple diamonds). The graphic is heuristic, so the placement of incidents should be seen in relative, rather than exact terms. It is suggestive of at least two important conclusions. First, deterrence of offensive cybers actions is plausibly taking place above certain conflict thresholds. Second, the linear escalation ladder approach to understanding conflict is imperfect at best, and possibly woefully inadequate, for understanding the nature of modern cyber conflict and evaluating deterrence successes and failures.

On the first point, there is good reason hypothetical and non-cyber incidents reside only on the right side of the graphic. These were included for illustrative purposes because – at least in the open

[27] Michael N. Schmitt, "The Law of Cyber Warfare: Quo Vadis?," *Stanford Law & Policy Review* 25, no. 2 (2014)

source literature – there exist few examples of cyber incidents crossing the "use-of-force" threshold, per most conventional interpretations. In fact, most publicized cyber security incidents – and particularly incidents directed at the United States and other major powers – fall well to the left of this threshold. This suggests any of three things: 1) antagonists are not attempting such attacks, 2) such attacks have been attempted but were not successful (and the attempts have gone unrecognized or unpublicized), and/or 3) successful attacks have in fact taken place but escaped the public eye. The last possibility seems unlikely, given that the effects of such an attack – if truly approaching use-of-force proportions – would almost certainly attract public attention. The second possibility is certainly plausible. However, if such attacks were happening on anything more than an infrequent basis, one would assume they would eventually attract public attention (either to the incident itself, or to the offended government's retributive response). The evidence is suggestive that attacks crossing the use-of-force threshold are, at the least, exceedingly rare.

If this is indeed the case, one must logically ask, why? Why are potential antagonists refraining from such actions in cyberspace? There are, again, three possible explanations: 1) potential antagonists lack the capability to carry out such attacks, 2) they lack sufficient motivation or incentive, and/or 3) they are refraining because they believe such actions would entail unacceptable cost (i.e. they are being deterred). In an era in which cyber warfare capabilities are increasingly accessible and ubiquitous, even to second and third tier military powers like Iran or North Korea, the first possibility seems unlikely.

The second is more plausible; an antagonist is arguably only likely to carry out an act of war if it is willing to accept the eventuality of war. However, given the inherent attribution challenges associated with cyber-attacks, especially if carried out through proxies or other means facilitating plausible deniability, incentive exists for antagonists – especially asymmetrically disadvantaged adversaries – to carry out such attacks as part of a "gray zone" escalation strategy that aims to exact pain and influence actions without retributive consequences. Many observers suggest this is precisely the strategy Iran has employed to combat U.S. influence in the Middle East, leveraging non-state proxies in Lebanon, Syria, Iraq, and the Persian Gulf to carry out actions that in some cases have had destructive and even deadly consequences for the United States and its allies. North Korea has similarly employed such a strategy for decades, engaging in provocative acts that arguably constitute acts of war, but which are plausibly deniable, or close enough to the margins of that threshold that its adversaries are hesitant to escalate militarily in response. Through intervention in Ukraine's eastern insurgency, Russia has also demonstrated willingness to operate on these margins.[28]

All of this suggests that something else besides lack of capability or lack of incentive is dissuading some antagonists from carrying out more provocative acts in the cyber realm, staying well left of escalatory thresholds that might draw a proportional military and/or cyber response. It is notoriously challenging to prove deterrence success – i.e. that an event did not happen because an antagonist believed the likely costs to be exacted through retribution or denial were unacceptable, versus other disincentives. However, by process of elimination, the evidence appears highly suggestive that some level of cyber escalation is being deterred, namely at and around the threshold of armed conflict.

At the same time, certain incidents plotted on *Figure 1* point to inadequacies in a linear conception of conflict thresholds, especially if ordered by severity of damage inflicted. One example is the hack of

[28] *Iran's Priorities in a Turbulent Middle East*, International Crisis Group (Brussels, Belgium, April 13 2018), ; Ken Gause, *North Korea's Provocations and Escalation Calculus: Dealing with the Kim Jong-un Regime*, CNA Analysis & Solutions (Arlington, VA, 2015), ; Lawrence Freedman, "Ukraine and the Art of Limited War," *Survival* 56, no. 6 (2014)

Democratic National Committee (DNC) email accounts in 2016, ostensibly undertaken by actors seeking compromising information on DNC activities that might influence electoral outcomes if publicized. From a purely material perspective on severity, this incident was not especially costly, either in terms of money or human lives; the effects were primarily political, harming the reputation of the party and the presidential campaign of Hillary Clinton. Email hacking is a common occurrence, and it is possible to simply chalk the incident up to the costs of doing business in cyberspace. However, if one interprets the incident as part of a more ambitious strategy to undermine the fundamental legitimacy of the U.S. electoral process, with intent to not only influence electoral outcomes but potentially destabilize governance, then placement of this incident on the graphic becomes more challenging; some might argue it easily crosses the threshold of national security significance.[29]

The hypothetical manipulation of electoral results through cyber means (denoted by an orange diamond in *Figure 1*) raises the stakes further. Rather than simply attempting to influence voter behavior, the precise effects of which might be challenging to discern, such an attack would attempt to sway outcomes through direct manipulation of electoral returns in favor of a specific candidate or party. Even if such an attack were conducted at a relatively low level of interference, and even were it detected and mitigated, public awareness of such an attack might cast public doubt on the integrity and legitimacy of the election writ large. Again, material costs might prove to be minimal (assuming no violent response on the part of domestic actors questioning the legitimacy of results). However, the political harm done might have significant repercussions for the legitimacy of the political process and governing institutions. Depending on the scale of tampering, and more importantly its overall effects on the continuity of governance, it is not hard to imagine some observers arguing that such an action at least flirts with the use-of-force threshold.

These challenges of interpretation and placement are not just limited to electoral examples. In the case of the 2017 NotPetya ransomware attacks, severe economic damage was done to numerous private entities, many of which were associated with Ukrainian interests. The Maersk shipping company, a Danish conglomerate, was also dealt a crippling blow that almost brought its global operations to an irreversible halt. While the attacks were carried out with no explicit political messaging, the economic effects of the attacks resonated at a national level (and in the case of Maersk, a multinational level).[30] How, then, should they be judged in terms of national security significance? In the case of Ukraine, which at the time faced simultaneous threats from an armed domestic insurgency and Russian occupation of the Crimea region, the nature and effects of the attacks might be interpreted as far more complex and insidious than can be represented easily on a linear threshold ladder.

2.2.2. An Alternative Approach to Conflict Escalation in Cyberspace

The ambiguities present in these cases suggest the need for a more nuanced conception of cyber conflict thresholds. *Figure 2* below represents an alternative, but complementary approach to thresholds that attempts to parse the problem across multiple dimensions. The major innovation is to disaggregate the *intended* political and *intended* material impacts of the various attacks. Cyber incidents can certainly result in impacts or consequences beyond what the antagonist intended, with potential implications for conflict escalation. However, this study is focused on strategies and

[29] Michael J. Assante, Robert M. Lee, and Tim Conway, *ICS Defense Use Case No. 6: Modular ICS Malware*, SANS Industrial Control Systems, Electricity Information Sharing and Analysis Center (Washington, D.C., 2017),

[30] Andy Greenbery, "The Untold Story of NotPetya, the Most Devastating Cyberattack in History," *Wired* (August 22 2018). https://www.wired.com/story/notpetya-cyberattack-ukraine-russia-code-crashed-the-world/

capabilities for deterrence, which necessarily involve rational choices on the part of both protagonist and antagonist, including calculated assumptions regarding the likely (intended) consequences of their actions.

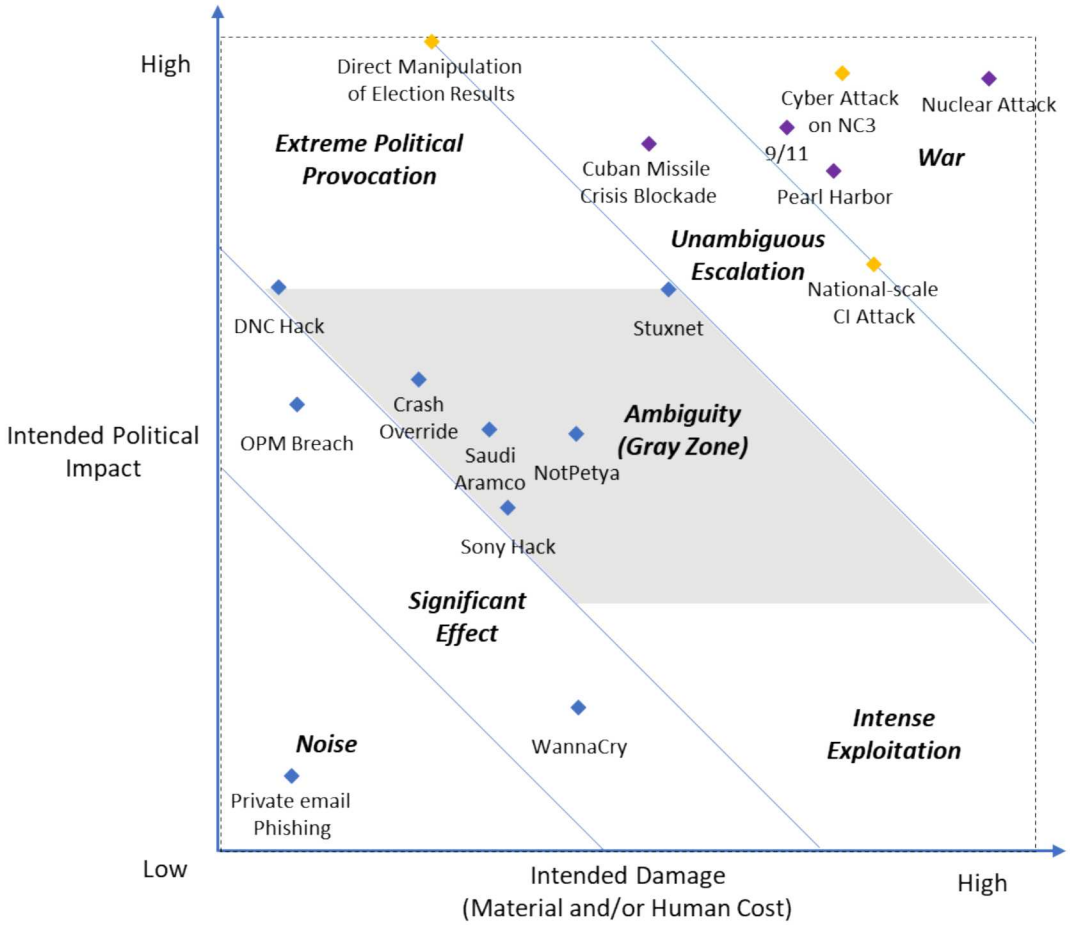

Figure 2: Alternative double-axis threshold

The x-axis of *Figure 2* indicates the level of intended material and/or human damage resulting from a cyberattack. This might include economic losses, damage to infrastructure, loss of service or functionality, and in extreme cases, even human injury or death. The y-axis indicates the level of intended political impact resulting from the attack. This might include influencing political decision-making processes (both in peacetime and wartime), the direction of public opinion on key issues, electoral outcomes, or diplomatic and alliance relationships between states. It should be noted that these axes are not completely independent; material consequences can surely have political impacts, and vice versa. Again, the purpose of the figure is heuristic, and the placement of individual incidents is relative.

Disaggregating the two impacts, however imperfect, nonetheless allows for more nuance in visualizing how different incidents might be situated relative to one another in "zones" (versus levels) of conflict escalation. The lower left and upper right corners of the figure correlate with the two ends of the previous linear figure; incidents in the lower left constitute the "noise" or

"everything else" of doing business in cyberspace – criminal acts targeting individuals or smaller corporations, low level espionage, and short-term nuisance disruptions of infrastructure or operations. The consequences of these actions may be political or material (or a combination of the two), but largely fall below a relative threshold of national significance. By contrast, the upper right constitutes actions that clearly cross the use-of-force threshold as previously defined, and that unambiguously suggest a state of war between protagonist and antagonist. Actions in this zone carry heavy political and material weight. Existentially significant acts from the previous figure populate the upper right-most limits of this zone.

The figure becomes more interesting as incidents populate zones outside these two extremes. The zone of "significant effect" comprises incidents where material and/or political effects begin to register at a national level, exhibiting consequences that are significant to national security. Political impacts may include short-term disruptions of governance processes, embarrassment and reputational effects, or compromise of low-level sensitive information. Material damage may include significant (but not systemically consequential) economic losses and shorter-term disruptions of critical infrastructure and corporate operations. The 2015 Office of Personnel Management (OPM) breach is an example of such an attack. Material damage was mostly limited to resources spent on investigating and mitigating the breach, but it registered significant political consequences, including reputational effects for OPM and its leadership, extending to the Obama administration itself. It remains to be seen if or how the stolen information might be leveraged by adversaries to achieve further damaging effects.[31]

Directly opposite, just before war, is a zone of "unambiguous escalation". The use-of-force threshold exists here, albeit ambiguously situated. Actions undertaken in this zone come with unmistakably significant material and political consequence, generally intended by their instigators to send an explicit message that an adversary has crossed red lines, and/or that war is potentially imminent – though importantly, there may still be opportunity to walk back from the brink. The previously cited U.S. blockade of Cuba in 1962 is an example of such an action.

The middle segments of *Figure 2* represent a more complicated and potentially nebulous conflict space, where many of the challenging incidents previously cited (e.g., the DNC hack and NotPetya) reside. On the upper left corner is a zone of "extreme political provocation"; attacks taking place here are unambiguously aimed at generating severe political effects with relatively minor material consequences. Attacks on electoral processes fall in this zone, particularly those aimed at influencing results in a specific direction. Attacks intended to influence the decisions of government officials in a direction that favors of the antagonist might also fall into this zone. Arguably, international governance systems and legal mechanisms are not well-equipped for calibrating or enabling responses in the zone of extreme political provocation, given that such actions probably do not involve overt militarized aggression, destruction of property, seizure of assets, or other measurable material damage. Moreover, perspectives may radically differ on what constitutes a threat to governance and political stability, especially among states with widely contrasting systems of governance (and possibly among interest groups within a targeted state, depending on whether they have been harmed or aided by an incident).

At the opposite extreme of the middle band resides a zone of "intense exploitation." Actions in this zone seek to exact as much material consequence as possible, but with little in the way of overt political intent. Large-scale economic heists or ransomware attacks, driven almost exclusively by

[31] The OPM Data Breach: How the Government Jeopardized Our National Security for More than a Generation, (Washington, DC 2016)

profit motives, reside in this space; the 2017 WannaCry ransomware attack was an event that at least bordered on this zone.[32] While it is tempting to assume that the zone of intense exploitation is primarily the realm of non-state actors, it is plausible that state actors seeking alternative sources of income might – perhaps because of international isolation and sanctions – operate in this space (or leverage non-state proxies to do so). Activities in this zone arguably fall under the purview of law enforcement, though the scale of consequences involved may very well elevate the salience of such events to higher levels of national and international authority.

At the intersection of the two middle conflict zones is a zone of intersectional activity, roughly analogous to what is often euphemistically referred to as "gray zone" conflict. One report defines such activity as "competitive interactions among and within state and non-state actors that fall between the traditional war and peace duality, are characterized by ambiguity about the nature of the conflict, opacity of the parties involved, or uncertainty about the relevant policy and legal frameworks."[33] Consequences in this zone, both political and material, are serious but similarly continue to fall below use-of-force thresholds. More importantly, the blending of consequences makes it more challenging to precisely divine intentions, especially if the antagonist does not explicitly broadcast them – as is often the case. NotPetya is an example of gray zone activity. So is the 2012 Saudi Aramco attack, an action for which no responsibility was formally claimed, but which struck at the heart of the Saudi economy during a time of high regional tension.[34] Indeed, many high-profile cyber incidents reside in this area. Cyberspace – which is subject to comparatively lax international regulation and oversight and which affords actors a high degree of potential anonymity (or at least ample mechanisms for obfuscating identity) and plausible deniability – is arguably well-suited to gray zone conflict actions.

Figure 2 strongly highlights the fact that the use-of-force threshold is not, by itself, a particularly helpful point of departure for understanding the nature of conflict and escalation in cyberspace; on one side of that line is a relatively small area of conflict (war), where few incidents are currently taking place. On the other side is a vast area of activity that itself can be broken down along multiple dimensions; only two were highlighted in this graphic (intended political and material consequences), but it is conceivable that many more exist. Even the "gray zone", which is already the subject of attention in academic and policy communities, is itself only a piece of the bigger picture.

Furthermore, *Figure 2* also highlights the need for more rigorous, systematic, and nuanced approaches to analyzing strategies of influence in cyberspace – including deterrence strategies. There is reason to suspect that strategies effective in one zone may not be relevant in another. As has already been suggested, deterrence may in fact be succeeding above the use-of-force threshold, somewhere in the zone of unambiguous escalation; however, this figure further suggests that deterrence may be failing in the gray zone, at least in the most high-profile instances. Deterrence may also be succeeding relative to the most extreme examples of political provocation and exploitation (e.g., overt tampering with electoral results). The problem is that analysts and scholars currently lack rigorous explanations for why this might be case, and thus fall back on broad-brush characterizations. There is a lack of both methodologically structured analysis and tools enabling such analysis. The next section of this report attempts to address this gap, proposing a framework for carrying out analysis of deterrence strategies in the context of a specific threat scenario.

---

[32] Investigation: Wannacry Cyber Attack and the NHS, (London, United Kingdom: National Audit Office, 2018)
[33] Philip Kapusta, "The Gray Zone," *Special Warfare Magazine* 28, no. 4 (2015)
[34] Nicole Perleroth, "In Cyberattack on Saudi Firm, U.S. Sees Iran Firing Back," *New York Times*, October 23 2012, https://www.nytimes.com/2012/10/24/business/global/cyberattack-on-saudi-oil-firm-disquiets-us.html

3. A FRAMEWORK TO ANALYZE DETERRENCE EFFECTIVENESS

Cyber operations span the entire spectrum of conflict, from harmless to nuisance to existential threat. In the earliest stages of conducting an operation with existential implications, an advanced persistent threat would pursue activities that appear to be indistinguishable from a very large volume of innocuous activities. In trying to develop a strategy to deter such an actor, which activities should we seek to deter? Which types of deterrence mechanisms are most effective at various stages of an operation or a campaign of multiple operations?

The purpose of the Cyber Deterrence Framework (*Figure 3*) is to analyze various deterrence options in a standardized way, in order to understand when (and why) deterrence will fail and when (and why) it may be more likely to succeed. It is clear from the outset that we cannot deter every possible actor from taking every potentially unwanted action, whether we are in the cyber domain or in any other domain. The purpose of the framework is therefore not to be prescriptive, but revelatory. For example, if we find through analysis of many cases that viable deterrence strategies are limited to a very small fraction of actors and to a very small fraction of their actions, then we would want to know the following:

1) What are the common conditions that may contribute to deterrence?
2) Can these conditions be preserved if conflict escalates?
3) Can these conditions be created in other cases where deterrence does not seem to be viable?

Our framework follows the layered structure of the Technical Cyber Threat Framework crafted by the Central Security Service of the National Security Agency (NSA/CSS).[35] The NSA/CSS framework categorizes antagonist activity along various threat stages and across a standard set of antagonist (Red) objectives for each threat stage. Our framework directly adopts the first two layers, threat stage (Layer 1) and antagonist objectives (Layer 2), and builds in additional layers for protagonist (Blue) deterrence objectives (Layer 3) and protagonist deterrence actions (Layer 4). A fifth and final layer evaluates each deterrence action across a common set of criteria.

[35] *NSA/CSS Technical Cyber Threat Framework v2,* Cybersecurity Operations The Cybersecurity Products and Sharing Division, U.S. National Security Agency (2018),

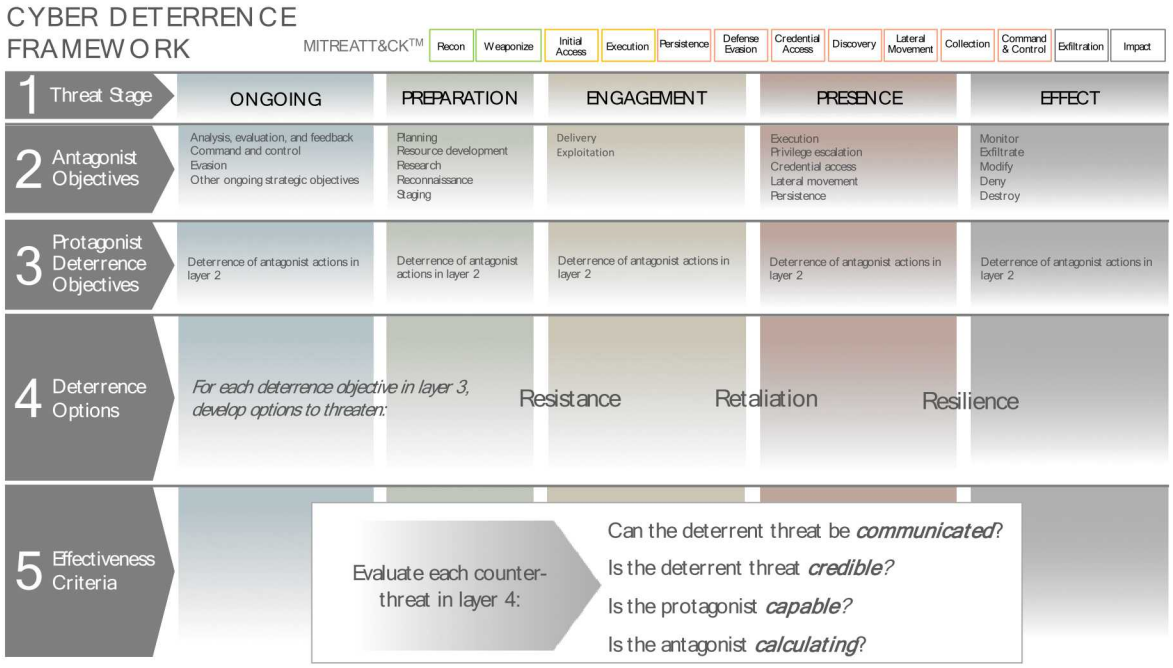

Figure 3: Overview of the Cyber Deterrence Framework.

First, we will describe the conceptual underpinnings of the framework in greater depth. Then, the following sections of this paper will provide examples of how the framework can be used to characterize deterrence strategies under specific scenario conditions.

3.1. Layers 1-3: Threat Stages, Red Objectives, and Blue Deterrence Objectives

The primary purpose of Layers 1-3 is to specify which of Red's activities are being examined. For example, are we concerned with deterring Red from planning an attack or from executing the attack to produce an effect? Alternatively, are we attempting to deter Red from achieving some ongoing strategic objective, or from achieving a specific, immediate and tactical objective? Obviously, the implications for deterrence policy and operations differ for all of these.

The NSA/CCS framework breaks the threat timeline into five stages, including ongoing, administration/preparation, engagement, presence, and effect. Our framework directly adopts this terminology as an example, but any similar threat sequence or kill chain could be used here (e.g., MITRE ATT&CK™), as long as one can arrive at specific deterrence objectives for Blue in Layer 3. The precise identity of Red and Blue is a complex question – they can include the governments of nation states, state-sponsored groups, private entities like corporations, organized criminal groups, groups of allies or international organizations, individuals, and many others. All of this nuance should be documented within the first three layers of the framework. The specific actors responsible for various stages of the threat may change, and their precise objectives certainly will change. There are probably multiple sub-actors within Red and Blue whose activities overlap across threat stages. Therefore, in the first three layers one must specify the actors involved, define their role in carrying out or deterring a particular stage in the threat sequence, and determine what their objectives are at that particular stage.

It should be noted that the quality of the analysis in Layers 4 and 5 depends on the depth of information available for Layers 1-3. Red's specific tactical objectives at various threat stages are linked by their broader strategic objectives, which are based on Red's own understanding of what is truly important to them. Without intimate knowledge of all of this, Blue's ability to deter Red at any stage is necessarily limited.[36]

3.2. Layer 4: Blue's Deterrence Counter-threat Options

The purpose of Layer 4 is to map the landscape of deterrence actions available to Blue for each different threat stage considered in Layers 1-3.[37] We conceive of deterrence broadly as creating conditions to dissuade an actor from taking an action because they perceive that they will be worse off taking the action than refraining from action. We generally adopt Nye's convention of using the terms *deterrence* and *dissuasion* interchangeably.[38] While this broad definition of deterrence encompasses manipulating Red's perception of the relative costs and benefits[39] of action *and* inaction, for this paper we focus more narrowly on manipulating Red's perception of the costs and benefits of action.[40]

Taxonomical structures can facilitate or hinder how broadly we are able to think about all of the ways that Red can be influenced. For example, the use of the term *deterrence* alone often implies threats of punishment, or the threat of imposing unacceptable costs or consequences. In contrast, *deterrence by denial* opens up space to consider Blue's capability to "deny the other party any gains"[41] from their actions. However, neither of these concepts includes consequences that either party might experience from passive mechanisms, such as entanglement, or from third parties, such as norms and taboo.[42] Additionally, we can consider conditions that Blue and others create that may have a deterrent effect, even if they are not specifically intended to deter.

We have chosen to adopt a taxonomy of deterrence mechanisms centered on the time that costs are imposed, or benefits denied, relative to the phase of the action Red is considering. These include both intentional deterrence strategies as well as conditions that Blue and others create, which may have a deterrent effect, even if they are not specifically intended to deter Red. In this taxonomy, deterrence mechanisms generally fall into three categories: resistance, retribution, and resilience (*Figure 4*). We use the term *deterrent threat* to describe specific actions in each of these categories, defined as a signal received by the antagonist that may dissuade them from taking a proscribed action because of the prospect of costs incurred or benefits denied. Such a "threat" can take many

---

[36] Blue's attempts to deter Red may include persuading them to pursue their broader or strategic objectives through alternative means. Ideally, these means would be less harmful to Blue, although calibrating that effect in order to avoid unintentionally incentivizing further escalation requires that Blue have a sophisticated understanding of the conflict dynamics at play.

[37] The temptation during Layer 4 is to limit one's thinking based on preconceived notions of which actions might be effective, plausible, or possible. This naturally limits the emergence of creative or unconventional ideas. Within reason, the analyst should resist this temptation and leave that determination for the analysis in layer 5.

[38] Nye, "Deterrence and Dissuasion in Cyberspace"

[39] Using the metrics of *cost* and *benefit* to Red and Blue to characterize deterrence interactions is problematic for several reasons. First, almost any action Blue takes to impose costs on Red also decreases the benefit that Red could obtain from taking that action. Similarly, almost any action Blue takes to decrease the benefit to Red of Red's actions will require that Red spend additional resources to gain back the original benefit, thus raising the costs of Red's actions. Secondly, the cost/benefit terminology invokes the idea that all actors are or must be rational if deterrence is to be useful at all. We will address this in the following section. Despite these shortcomings, we still adopt the cost/benefit paradigm as a useful heuristic, and for lack of a suitably concise alternative.

[40] Our approach, therefore, does *not* include actions that Blue could take to make the status quo more tolerable for Red, for example, by offering inducements or promises to alter the status quo in a manner favorable to Red.

[41] Glenn Snyder, "Deterrence and Power," *Journal of Conflict Resolution* 4, no. 2 (1960)

[42] Nye, "Deterrence and Dissuasion in Cyberspace"

forms. It can be an overt or implicit signal that Blue will issue some kind of response. It can also signal that Blue has already taken actions that will result in Red incurring additional costs beyond what Red originally calculated. It can also include passive mechanisms that do not involve any action by Blue, such as economic entanglement of Blue and Red. It should be understood that for each of the categories in *Figure 4*, it is the *prospect* of increased costs or insufficient benefits that might deter Red. Once Red actually experiences such costs or denial, we have ventured from the realm of deterrence into the realm of actual defense, war-fighting, or recovery.

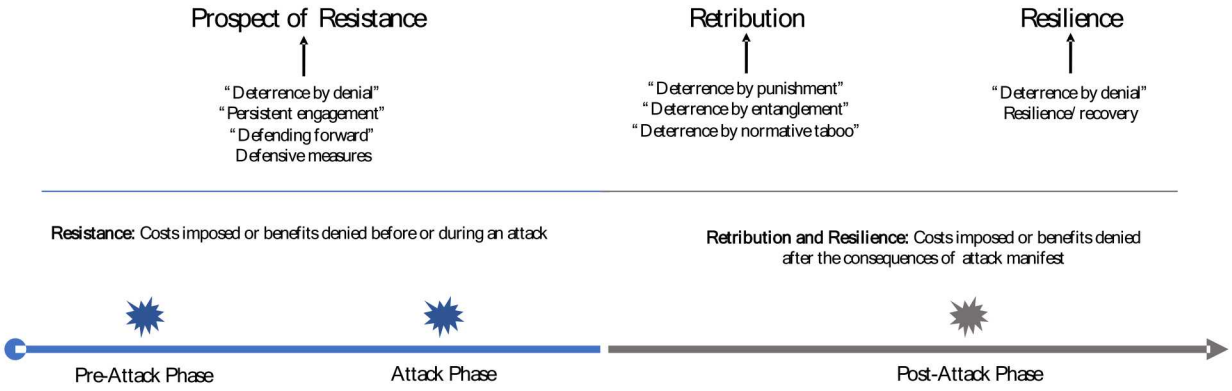

Figure 4: A breakdown of various deterrence mechanisms by time of cost imposition or denial of benefits relative to the attack phase.

It is important to note that for deterrence actions within each category above, Blue can and should develop, demonstrate, and signal their capabilities well in advance of any actions by Red. What distinguishes these categories is not necessarily when Blue acts, but when costs are imposed on Red or benefits are denied to Red by the sum total of Blue's cumulative actions and commitments up to that point. The following sections describe these categories in greater detail, and concrete examples of each within the context of a specific scenario will be described in Section 4.5.

The Prospect of Resistance

To deter by the prospect of resistance, Blue must dissuade Red from conducting a cyber operation because Red perceives that they will face defenses or other obstacles *during* their activities that will deny their ability to carry out the operation, or that will be too costly to overcome. This mechanism includes actions that Blue can take to prevent and degrade Red's ability or willingness to take the proscribed action. These types of actions raise costs for Red, because Red would have to spend additional resources to overcome defenses or recover capabilities. In the event that Blue takes resistance actions that completely eliminate Red's capability, the effect is not deterrence, but rather prevention or outright destruction (which may subsequently serve to persuade third party actors not to take similar actions against Blue in the future). When resistance is accompanied by Blue signaling the prospect of increased costs or denied benefits, this may have the effect of deterring Red from its original course of action. In general, deterrence by resistance makes Red's course of action appear less attractive or less likely to be successful, and thus raises the prospect of a prolonged or more expensive offensive for Red.

We consider resistance to include classical deterrence by denial, defined by Snyder in 1960 as "the capability to deny the other party any gains from the move which is to be deterred."[43] Even before the nuclear era, deterrence by denial was often used in conventional contexts to bolster defenses in

[43] Snyder, "Deterrence and Power"

allies' territory, often by the placement of troops, to raise the cost of invasion by an antagonist into those territories.[44] The most notable example from the Cold War is the placement of U.S. troops in Europe to act as a tripwire for Soviet expansion, and to raise the costs of conventional conflict in the region for *both* sides as a matter of course, rather than as coercion through the prospect of retribution.

We think the analogy to deterrence by denial in conventional contexts is important for understanding deterrence of cyber antagonists. Naturally nations do not always have jurisdiction over territories that are strategically important to them. So, they form alliances and take steps to bolster the defensive capabilities of their allies. As Gerson remarks, the purpose of doing so is not to completely remove the ability of the antagonist to attack, but rather to convince the antagonist that "it cannot achieve its objectives rapidly or efficiently…that his only alternative is a protracted war"[45] – in other words, to deny adversaries the prospect of quick victory without heavy losses. Therefore, deterrence by denial *still* involves manipulating the antagonist's perception of the costs and benefits of their actions. This is no less true when Blue makes commitments up front. Thus, a commitment on Blue's part (e.g., stationing troops or fortifying infrastructure) could be interpreted by Red as a threat of resistance by raising the magnitude and/or likelihood of unacceptable costs, if conditions are correct.

In cyberspace, nations similarly do not have jurisdiction over all networks and systems that are strategically important to them. Many of these are controlled, operated by, and within the physical territory of other nations or the private sector. Therefore, nations take measures to bolster the defenses of their allies and partners. We do not argue that mere possession of defenses by Blue or its allies constitutes *de facto* deterrence. Blue likely has compelling and immediate defense requirements, and they will commit resources and infrastructure to meet these immediate needs perhaps more so than to influence potential adversaries. But there is also the potential to deter by the same mechanisms discussed by Gerson. In terms of influencing Red's operational costs, we see no fundamental difference between stationing troops to raise the cost of Red's expansion and the emplacement of capable cybersecurity personnel who can find vulnerabilities and burn Red's hard-won exploits, for example.

Resistance is also an important mechanism of cyber deterrence, because in some cases it does not rely on attributing cyber activities to the true identities of the individuals responsible,[46] although it certainly may help to know something about Red's history and tradecraft. Additionally, much has been written about the potential utility of deterrence by denial to deter even the most highly motivated actors.[47] For example, even terrorists seek to minimize operational risk, and therefore may be deterred by the prospect of being apprehended or killed before completing their attack.[48] Terrorists "may be willing to give their lives, but not in futile attacks," note Davis and Jenkins.[49] The dynamics of cyber conflict are consistent with this perspective. Once vulnerabilities are discovered, they can be patched. The ability to manipulate cyberspace in favor of the defender makes it difficult for attackers to obtain the full potential payoff, yielding the advantage to the defender.[50] Thus, Red

[44] Michael S. Gerson, "Conventional Deterrence in the Second Nuclear Age," *Parameters*, no. Autumn (2009); A. Wess Mitchell, "The Case for Deterrence by Denial," *The American Interest* (2015). https://www.the-american-interest.com/2015/08/12/the-case-for-deterrence-by-denial/

[45] Gerson, "Conventional Deterrence in the Second Nuclear Age"

[46] Aaron F. Brantley, "The Cyber Deterrence Problem" (10th International Conference on Cyber Conflict, Tallinn, Estonia, NATO CCD COE Publications, 2018)

[47] Jeffrey W. Knopf, "The Fourth Wave in Deterrence Research," *Contemporary Security Policy* 31, no. 1 (2010)

[48] Robert W. Anthony, *Deterrence and the 9-11 Terrorists*, Institute for Defense Analyses (Alexandria, VA, 2003),

[49] Paul K. Davis and Brian Michael Jenkins, *Deterrence and Influence in Counterterrorism: A Component in the War on al Qaeda* (RAND National Defense Research Institute, 2002), https://www.rand.org/pubs/monograph_reports/MR1619.html

actors may go to great lengths to ensure their tools are not discovered without some benefit, and this provides Blue with opportunities to create perceptions of operational risk and uncertainty. If by resistance Blue can raise the uncertainty of a fast and easy victory for Red, then Red may choose alternative ways and means to achieve their ends, and this is deterrence.

Nye's concept of "deterrence by norms and taboo"[51] contains elements of resistance and retribution, depending on the mechanism through which the norms and taboo function. If violation of a norm results in punishment or consequences imposed on Red by a third party, then the normative deterrence mechanism is more properly categorized as part of deterrence by prospect of retribution. However, if Blue and others work toward the development of norms against the types of actions Blue would like to prevent Red from taking, such that actors within Red start to adopt these norms of behaviors or value systems themselves, Red may refrain from those activities, not because they fear punishment or shame, but rather because these actions have become contrary to their own identity. In this case, norms serve an important resistance function also, because they can serve to degrade an antagonist's willingness take undesirable actions in the first place.

Finally, we argue that what has come to be called *persistent engagement* or *cyberspace persistence* constitutes another example of resistance. The 2018 USCYBERCOM Vision defines cyberspace persistence as "the continuous ability to anticipate the adversary's vulnerabilities, and formulate and execute cyberspace operations to contest adversary courses of action under determined conditions."[52] A critical aspect of USCYBERCOM's persistent engagement strategy is the ability to "defend forward" to "counter and contest dangerous antagonist activity before it impairs our national power."[53] The 2018 Vision states that the U.S. government and military should be able to act with a unified, cross-agency response *before* adversaries have breached our networks or systems. In pursuit of persistent engagement and forward defense, cyber forces would maneuver "seamlessly between defense and offense" and "as close as possible to adversaries and their operations."[54]

There is considerable debate about whether persistent engagement does or could serve a deterrent purpose, or whether it constitutes a fundamentally different kind of cyberspace strategy. Many argue that persistent engagement seeks to leverage the inherent nature of cyberspace as an operationally rich environment in which myriad actors are in constant contact, whereas deterrence seeks to promote restraint in an operationally poor environment.[55] For example, General Nakasone recently commented that "[u]nlike the nuclear realm, where our strategic advantage or power comes from possessing a capability or weapons system, in cyberspace it's the *use* of cyber capabilities that is strategically consequential. The *threat* of using something in cyberspace is not as powerful as *actually* using it because that's what our adversaries are doing to us...So advantage is gained by those who maintain a continual state of action" [emphasis in original].[56]

While we think strategic advantage in nuclear deterrence is built upon far more complex mechanisms than simple possession of capability, the more important point to argue here is that persistent engagement could deter adversaries as well as continually contest their operational advantages in cyberspace. USCYBERCOM's 2018 Vision states that cyber persistence is intended to influence adversary behavior, and that "[t]hrough persistent action...we can influence the calculations

---

[50] Brantley, "Short The Cyber Deterrence Problem."
[51] Nye, "Deterrence and Dissuasion in Cyberspace"
[52] Achieve and Maintain Cyberspace Superiority: Command Vision for US Cyber Command, Short
[53] Achieve and Maintain Cyberspace Superiority: Command Vision for US Cyber Command, Short
[54] Achieve and Maintain Cyberspace Superiority: Command Vision for US Cyber Command, Short
[55] Fischerkeller and Harknett, "Deterrence is Not a Credible Strategy for Cyberspace"
[56] Paul M. Nakasone, "An Interview with Paul M. Nakasone," *Joint Force Quarterly* 92, no. 1st Quarter (2019)

of our adversaries, deter aggression, and clarify the distinction between acceptable and unacceptable behavior in cyberspace."[57] This intention is supported by recent statements made by former National Security Advisor John Bolton, who commented that as part of the Trump administration's new cyber strategy, offensive cyber operations have been authorized "precisely to create the structures of deterrence that will demonstrate to adversaries that the cost of their engaging in operations against us is higher than they want to bear."[58] Finally, even fierce critics of deterrence as a credible cyberspace strategy argue that a strategy of persistent engagement is more likely to create norms that persuade adversaries to change their courses of action. For example, Fischerkeller and Harknett argue that "[g]lobal norms of responsible behavior cannot take root if the universe of 'like-minded' states is a small proportion of salient cyber actors… those who operationally dominate the domain will be in the strongest position to argue for norms supporting their positions."[59] We consider this type of norm creation to be part of a strategy of deterrence.

The Prospect of Retribution

Deterrence by the prospect of retribution involves creating the perception that Red will incur unacceptable costs as a result of pursuing a course of action, that is, *after* the effects of a cyber operation occur. There are several mechanisms through which to cultivate the perception of unacceptable costs to Red. The most obvious is the threat of punishment by Blue, which can take the form of government instruments of power (use of military force, law enforcement responses, economic or diplomatic sanctions, etc.) or private sector vigilantism.

One potential major disadvantage is that a credible threat of punishment from Blue requires, first, that Blue be able to detect unwanted cyber activity, and second, that Blue be able to attribute this activity to the actors responsible in a manner visible to those Blue wishes to influence. Therefore, credibility breaks down if a) Blue lacks the ability to detect or attribute an attack, or b) Blue cannot or chooses not to make the fact or the details of detection and attribution known to actors it wishes to influence – for example, out of a desire to keep those capabilities secret. Even if the technical challenges of rapid, high-confidence attribution can be overcome, domestic and international law may preclude assigning responsibility or making attribution public.[60] Therefore, delayed detection and uncertain attribution pose significant challenges to strategies based on deterrence by punishment. Since most actors choose, for a variety of reasons, not to threaten punishment indiscriminately,[61] then deterrence by punishment is often not a credible strategy in cases where detection and attribution are very difficult.

---

[57] Achieve and Maintain Cyberspace Superiority: Command Vision for US Cyber Command, Short

[58] John Bolton, "National Security Advisor John Bolton on Cyber Strategy," news release, September 20, 2018, 2018

[59] Fischerkeller and Harknett, "Deterrence is Not a Credible Strategy for Cyberspace"

[60] The precise definition of who is responsible for a cyber operation is not always clear. See, for example, Herb Lin, "Attribution of Malicious Cyber Incidents: From Soup to Nuts," Aegis Paper Series No. 1607, Hoover Institute, Stanford University, 2016. Even in cases where it is known who sponsored or ordered an attack, who carried out the attack, the precise computers or networks from which the attack originated, and the precise location of those computers or networks, which may or may not be in the same country or jurisdiction as those who conducted or ordered the attack, at this time it is still not obvious, from a legal perspective, who can or should be held responsible, and in which ways. Additionally, Nye identifies three relevant audiences for attribution: 1) the defending government, who wants to avoid escalation, 2) the attackers, who can attempt to deny involvement if the case for attribution is not strong enough, and 3) domestic and international publics who may protest unjust retaliation. See Joseph S. Nye Jr., "Deterrence and Dissuasion in Cyberspace," *International Security*, 41, **2016**, 44-71.

[61] There may be cases where Blue could credibly threaten punishment without high-confidence or transparent attribution. In such a case, Blue would have to avoid creating the perception that an adversary would be punished regardless of whether they act or refrain from action, as this perception could provoke attack rather than deter attack. Therefore, Blue may have to take additional steps to assure innocent parties that if they refrain from action, they will not be punished. This would obviously be a very difficult, but perhaps not impossible, balance to strike.

In addition to punishment, Nye discusses two other mechanisms that we classify as part of deterrence by the prospect of retribution: entanglement and norms.[62] Nye defines deterrence by entanglement as "the existence of various interdependencies that make a successful attack simultaneously impose serious costs on the attacker as well as the victim."[63] Stated another way, Red may be deterred from taking an action if the action disrupts a status quo that benefits both Red and Blue. The advantage of deterrence by entanglement is that it does not rely on retaliatory threats and is thus indifferent to high-confidence attribution. It also does not require robust defenses of highly resilient systems. However, not all actors are entangled. Rogue, isolated nation states possessing few ties with the outside world will not be as dissuaded by entanglement as nations that are highly connected to the international world order and the global economy. Another disadvantage is that entanglement may be hard to observe or quantify, and thus difficult for Blue to credibly signal. Entanglements and interdependencies between actors often do not reveal themselves until an action has taken place and the antagonist feels the negative effects of their miscalculation.

Deterrence by norms describes a situation in which Red refrains from actions based on the perception that the damage to their reputation will outweigh the perceived benefits of the action. As stated previously, when actors within Red begin to adopt certain norms and become "self-deterred"[64] from violating them, we label this as deterrence by the prospect of resistance. In contrast, when Red does not share norms of behavior with Blue and is therefore not self-deterred, but instead refrains from actions from fear of potential damage those actions would have to their reputation, they are being deterred by the prospect of retribution. Red may perceive a threat of reputation costs from external third parties or from domestic audiences. We consider entanglement and norms to be part of retributive deterrence, because the costs they impose on an antagonist are not felt until *after* the attack has taken place; thus, they are a form of retribution, even if the retribution does not directly stem from actions taken by Blue.

The Prospect of Resilience

The final category is deterrence by resilience, which also imposes costs and denies benefits to the antagonist after an attack has occurred. Resiliency is a term that encompasses many different concepts, depending on the context. The Presidential Policy Directive on Critical Infrastructure Security and Resilience (PPD-21) defines resilience as "the ability to prepare for and adapt to changing conditions and withstand and recover rapidly from disruptions."[65] A demonstrated ability to mitigate and recover quickly from attacks has been discussed as an important mechanism for deterring terrorists.[66] Clarke and Knake adopt paradigms of resilience from social psychology, where "resilience is not about returning to a previous state after an individual experiences trauma, but about adapting to that trauma."[67] This concept of resilience includes the idea that resilient people (and systems) can *grow and become stronger* after a disruptive or destructive event. These definitions are all consistent with our framing of resilience. We define resilience as having systems in place that minimize the consequences or impact of an attack, that sustain operations throughout and after an attack, and that recover and adapt to new conditions after an attack has occurred. If an antagonist believes that a potential target is resilient in these ways, they face the prospect of fewer gains from

[62] Nye, "Deterrence and Dissuasion in Cyberspace"
[63] Nye, "Deterrence and Dissuasion in Cyberspace"
[64] Robert Jervis, "Deterrence and Perception," *International Security* 7, no. 3 (1982)
[65] Presidential Policy Directive 21: Critical Infrastructure Security and Resilience, Short
[66] Knopf, "The Fourth Wave in Deterrence Research"; Davis and Jenkins, *Deterrence and Influence in Counterterrorism: A Component in the War on al Qaeda*,
[67] Richard A. Clarke and Robert K. Knake, *The Fifth Domain: Defending Our Country, Our Companies, and Ourselves in the Age of Cyber Threats* (New York, NY: Penguin Press, 2019)

attacking that target and would thus probably need to expend more resources to achieve the desired effect than would be required absent such resilience. All else equal, a calculating antagonist would probably choose to target a relatively less resilient system.

Deterrence by resilience also encompasses actions that decrease the attractiveness of targets through intentional degradation or destruction of the target by Blue. If Blue can demonstrate, through actual, degradation of its own assets, that it can maintain and sustain operations despite such degradation, then they signal that they are resilient. A historical example is the use of scorched earth tactics by a defending army to slow the advance of an invading army. An example from the electric grid scenario described in this paper is the ability to rapidly switch to manual override operations. Once Blue demonstrates that they have preserved or developed the ability to maintain full power generation and distribution without digital, networked information technology systems and industrial control systems, they have effectively removed the value of those systems to Red, although they likely do so at some reduced efficiency and increased cost to themselves.

It is important to note that there are many actions Blue could take that will operate through both mechanisms of resistance and resilience. For example, network segmentation can serve a resistance purpose, that is, by raising the antagonist's costs for attacking a system because they must successfully compromise each segment individually. However, it can also create the prospect of resilience – an attack deployed on a segmented network will not spread as easily to other elements of the network, which will continue to operate normally, such that the intended effects are not as broadly distributed as the antagonist originally intended. While there may be some overlap between these two categories, we choose to distinguish them in order to facilitate analysts in thinking as broadly as possible about the contribution to deterrence of both defensive tools that raise antagonists' costs during their attack, and of resilience tools that decrease the impact and facilitate recovery once an attack has occurred.

3.3. Layer 5: Evaluation of Blue's Deterrent Threat Options

The requirements for deterrence are often invoked when analyzing threats that could persuade an opponent not act.[68] Over the decades the concepts of credibility, willingness, and capability have typically been invoked in pairs to describe what must be true for a deterrent threat to have gravity. In a 1954 assessment of Dulles' nascent doctrine of massive retaliation, Kauffman laid out a foundation for evaluating the requirements of deterrence.[69] These included clearly communicating the threat and taking appropriate measures to ensure that there is sufficient credibility in the threat. Credibility was argued to be a function of the protagonist having sufficient capability to carry out the threat, that sufficient costs can be incurred by the antagonist if the threat was carried out, and that the threat is believed by the antagonist to be something that would be carried out by the protagonist. Scholars and strategists subsequently refined and expanded Kauffman's concepts[70] and debated about alternate strategies and possible challenges for fulfilling the requirements.[71]

---

[68] Or, in the case of compellence, to act. See reference Schelling, *Arms and Influence*

[69] William W. Kauffman, *The Requirements of Deterrence*, Center of International Studies, Princeton Univeristy (November 1954 1954),

[70] R. D. Luce and Howard Raiffa, *Games and Decisions: Introduction and Critical Survey* (Oxford, England: Wiley, 1975); Richard Selten, "Reexamination of the perfectness concept for equilibrium points in extensive games," *International Journal of Game Theory* 4, no. 1 (1975); Frank C. Zagare and D. Marc Kilgour, *Perfect Deterrence*, Cambridge Studies in International Relations, (Cambridge: Cambridge University Press, 2000)

[71] Thomas C. Schelling, *The Strategy of Conflict* (Cambridge, MA: Harvard University Press, 1960); Schelling, *Arms and Influence*; Herman Kahn, *On Escalation: Metaphors and Scenarios* (1968); Lawrence Freedman, *The Evolution of Nuclear Strategy*, 3rd ed. (New York, NY: Palgrave Macmillan, 2003)

We propose an evaluative set of requirements for deterrence by beginning with Kauffman's definition and integrating forward through the subsequent literature. This framework considers four categorical considerations for deterrence effectiveness, two of which are each further decomposed into a pair of components. The resulting set provides a systematic diagnostic for assessing a proposed deterrent threat, assessing its effectiveness, and identifying challenges that must be overcome for it to be successful. These requirements are further discussed below and form a logical structure in that all of the requirements must be met, logical AND, for a deterrent threat to be effective. The likelihood of any particular requirement being met, and what is required to meet it, are precisely the discussions that the frameworks developed in this paper are meant to generate.

Communicated

A deterrent threat must be communicated by the protagonist to the antagonist in order to be effective. This requirement appears simple enough at first blush but can be quite involved. In the clearest cases, the protagonist might make an overt statement, publicly or privately, directly to the antagonist: "don't do X or I will do Y." In other cases, it may be better to make the communication implicit in the protagonist's actions, such as raising an alert level, preparing and deploying cyber defense tools, or stationing troops preemptively. For deterrence by resistance, communication may occur when Red discovers a defensive measure that Blue previously implemented, which dissuades Red from continuing operations as planned. Ensuring that a deterrent threat is accurately communicated and is understood by the antagonist as the protagonist intends is a complicated proposition, and history has shown that critical misperceptions are more common than we might like.

Credible

A deterrent threat made by the protagonist must be viewed by the antagonist as credible, meaning that the antagonist believes that unacceptable costs will be imposed or insufficient benefits yielded, if the antagonist choses to take the proscribed path. *Credibility* can be further decomposed into two considerations:

Principled: A deterrent threat from the protagonist must be viewed as aligned with the protagonist's values and principles, in order to be credible. That is, a deterrent threat that includes a protagonist action that is considered abhorrent to the protagonist's value system would likely be seen as less credible than an action that is consistent with the antagonist's view of what the protagonist would be willing to do. For example, if a pacifist protagonist made a deterrent threat that involved unbridled violent action, it would be difficult for the antagonist to view that threat as credible.

Rational: A deterrent threat from the protagonist must be viewed by the antagonist as something that the protagonist would choose to do, in the event that the antagonist takes the proscribed action. The antagonist must believe that, for whatever reason, the protagonist would prefer to carry out the threat, either because the protagonist would be better off (i.e., through an evaluation of instrumental rationality[72]), or is so committed that they have no choice but to invoke the threatened action (i.e., through some irrevocable commitment[73]). Costs of action incurred by the protagonist should be interpreted broadly and might include both immediate capital costs of action, as well as any anticipated costs due to further escalation, or reputational costs paid as a result of following through on the attack.

[72] Selten, "Reexamination of the perfectness concept for equilibrium points in extensive games"
[73] Schelling, *The Strategy of Conflict*

The sub-requirements that a deterrent threat be both principled and rational must be met in order for that threat to be considered credible, again following the logical AND convention. This decomposition may provide a more precise definition of credibility through which to judge proposed deterrent threats. In cases where the protagonist has less immediate agency over executing the deterrent threat, such as entanglement, or where the protagonist has already made commitments, such as erecting defensive infrastructure, the antagonist must believe that the protagonist will allow and sustain such conditions, that they align with the protagonist's principles and rational calculations, or that the protagonist has no agency or control over such conditions.

Capable

The protagonist must be viewed by the antagonist as being capable of carrying out the deterrent threat. Sufficient capability is a common focus of deterrence analysis, as possession of the means of carrying out the deterrent threat is generally seen as a powerful signal in and of itself. In fact, Zagare and Kilgour argue that this is the one absolute requirement for deterrence to be effective, dividing capability into two parts: "the physical and the psychological."[74] We frame these subcomponents as two considerations:

Executable: The antagonist must believe that the protagonist is able to execute, or carry out, a deterrent threat. This means that the technical capability exists for the protagonist to initiate the deterrent action, and that the action will result in the intended cost effect on the antagonist. For example, if a protagonist threatens a retaliatory missile strike as part of a deterrent threat, then the antagonist must believe that the protagonist possesses a missile capability that will launch on the protagonist's command, successfully travel to the target, detonate, and destroy the target. In this example, antagonist missile defenses could call the capability of the protagonist threat into question. For resistance and resilience deterrence threats, the antagonist must perceive that the protagonist is capable of executing and sustaining measures that would either raise the antagonist's costs before or during an attack or that would deny the benefits of their attack.

Painful (Costly): The antagonist must believe that any deterrent threat carried out by the protagonist will be sufficiently painful to cause the antagonist to reconsider their own actions. A deterrent threat must create consequences for the antagonist that are worse than those the antagonist would experience if they chose not to act in the proscribed manner. For example, if the antagonist is facing an existential choice, whereby *not* acting would lead to the collapse of their regime, then a threat by the protagonist to destroy antagonist military capabilities if that action is taken is unlikely to sway the antagonist. In such a case, deterrence is likely to fail, because the threat of lost military capabilities due to action is less painful to the antagonist than regime collapse due to inaction. The protagonist may also deter the antagonist by convincing them that the material costs of an attack are too high given the expected gains. Antagonists may choose not to risk revealing their capabilities if they know they will encounter a well-defended, highly resilient network, and they may choose instead to target a different system.

A capable protagonist deterrent threat is one that is both executable by the protagonist and sufficiently painful to the antagonist such that the antagonist must understand that the implied consequences are both certain and grave. Just as in the case of credible, both sub-conditions must be met to satisfy the logical AND requirement.

[74] Zagare and Kilgour, *Perfect Deterrence*

Calculated

The final requirement is that the antagonist must weigh the costs and benefits of action in some way that considers the deterrent threat itself. In other words, the antagonist must both consider the possibility that additional costs will be imposed or insufficient benefits will be gained, and also compare them to the costs and benefits of not acting prior to committing to a course of action. This is consistent with the concept of instrumental rationality as described by Raiffa.[75] A stronger condition would be Verba's definition of procedural rationality, where actors must consider all possible actions, under accurate assessments of the implications, and make choices consistent with clearly articulated preference functions.[76]

Traditional models for decision-making that were strictly based on a utilitarian calculation of costs and benefits was challenged by Kahneman and Tversky, who empirically demonstrated the important role that bias and heuristics play in making decisions. Their work launched the field of behavioral economics and a new concept of *bounded rationality*, in which rationality in human decision-making is limited by the information one has, the time available to make the decision, and one's cognitive abilities. Similarly, Tor divides *strategic rationality* into two parts: *instrumental rationality*, a strict mathematical calculation of costs and benefits, and *normative rationality*, which considers the value specific actors assign to various elements of the cost-benefit calculation.[77] Equally important when considering the nature of rational thought in human decision-making is asking how well that rationality is translated to organizational levels.

For the scenario analysis discussed in this paper, we assume that the antagonist displays some form of bounded rationality, that they will consider costs and benefits of action and inaction, but that this consideration is limited in the ways described above. Given enough information, it is possible for the protagonist to influence these perceptions. Therefore, we assume that the antagonist is *calculating* in some way, and that there is not much the protagonist can readily do, within the scenario, to change or alter the way in which the antagonist calculates and interprets its calculation. In reality, actors may go to great lengths in order to arrive at a common understanding of each other's rationality. During the Cold War, the United States and the Soviet Union negotiated and bargained through various means for decades in order to better understand how the other side thought. It is reasonable to assume that opposing actors in a drawn-out cyber contest could similarly enhance their abilities to understand each others' strategic interests.

With the set of requirements articulated above, a proposed deterrent threat can be evaluated, along several dimensions. It should quickly become clear that there are tradeoffs and interactions between these requirements. For example, a demonstration of a capability by a protagonist (e.g., a successful test-firing of a new missile) would also have implications for the antagonist's perception of credibility. As another example, situations that create a significant asymmetry of stake (where the antagonist's vital interests are in play, and the protagonist only has strong preferences on the outcome) pose challenges to both rationality of the deterrent threat (the protagonist may not be better off making good on the threat) and the painfulness of the threat (the threat is insufficiently costly to the antagonist when compared to them not taking the proscribed action).

Table 1 summarizes how the conditions *communicated, credible, capable,* and *calculated* apply to the four categories of deterrence strategies previously discussed.

---

[75] Luce and Raiffa, *Games and Decisions: Introduction and Critical Survey*

[76] Sidney Verba, "Assumptions of Rationality and Non-Rationality in Models of the International System," *World Politics* 14, no. 1 (1961), https://doi.org/https://doi.org/10.2307/2009558

[77] Uri Tor, Cumulative Deterrence. Uri Tor, "'Cumulative Deterrence' as a New Paradigm for Cyber Deterrence," *Journal of Strategic Studies* 40, no. 1-2 (2015), https://doi.org/10.1080/01402390.2015.1115975

Table 1. Deterrence requirements for various types of deterrence strategies

	Communicated	Credible		Capable		Calculated
		Rational	Principled	Executable	Painful/Costly	
Resistance	Norms: "XYZ values are an inherent part of who you are. Taking this action violates your core identity." Self-deterrence	The antagonist perceives that holding these norms and being aligned with a like-minded COA is in their own best interests. (COA = community of actors)	The antagonist believes that these norms are fundamental to their identity and values.	The antagonist believes that taking the proscribed action incontrovertibly and undeniably violates the norms they hold dear.	The antagonist believes that taking the proscribed action is not worth losing their inherent sense of self.	The antagonist perceives that the protagonist believes that the antagonist is a rational actor, and that given enough information about the antagonist's interests, thresholds, and red lines, the protagonist can influence the antagonist's decisions.
	Persistent engagement: "The protagonist has ramped up an effort to engage with the antagonist before they reach the protagonist's network, to generate tactical friction and force the antagonist to focus on defense instead of offense."	The antagonist perceives that the protagonist believes that persistent engagement is in the protagonist's best interests, that it is not too expensive and that it will not provoke escalation or retaliation.	The antagonist perceives that the protagonist believes persistent engagement with cyber adversaries is aligned with the protagonist's values and principles.	The antagonist perceives that the protagonist is able to engage with the antagonist persistently (technical capability)	The antagonist perceives that persistent engagement by the protagonist within or around the antagonist's networks will raise the antagonist's operational costs to unacceptable levels	
	Defense: "The protagonist has implemented sufficient measures to diminish the likelihood that the antagonist's attack will achieve the desired effect."	The antagonist perceives that the protagonist believes resistance measures are in its own best interests to create and implement (e.g. not too expensive).	The antagonist perceives that the protagonist believes resistance measures are in line with the protagonist's principles (e.g. do not violate certain rights or freedoms of citizens).	The antagonist has sufficient visibility into the protagonist's security to believe their attack would be ineffective. The antagonist believes that the protagonist's resistance measures are as consistent and effective as the protagonist claims.	The antagonist believes it would require too many resources to overcome the protagonist's resistance measures.	
Retribution	Punishment: Overt threat or precedent: "If the antagonist does X, the protagonist will respond with Y, which will impose unacceptable costs on the antagonist."	The antagonist perceives that the protagonist believes it is in its own best interests to carry out punishment.	The antagonist perceives that the protagonist believes the retributive action is consistent with the protagonist's principles.	The antagonist believes that the protagonist can carry out the retributive action.	The antagonist believes the impacts of punishment would be unacceptably painful.	
	Entanglement: "The economies/ infrastructure/allies/etc. of the antagonist and protagonist are interdependent. Therefore, any action the antagonist takes against the protagonist may also impact the antagonist."	The antagonist believes that they are interdependent with the protagonist. The antagonist believes that the protagonist would allow/tolerate these interdependencies based on the protagonist's own best interests, or that they are unavoidable.	The antagonist believes that the protagonist would allow/tolerate these interdependencies based on the protagonist's principles or values, or that they are unavoidable.	The antagonist perceives that they are in fact interdependent with the protagonist in the way the protagonist claims.	The antagonist believes blowback/shared impacts of attack would be unacceptable.	
	Norms: "The global standard is XYZ. Violating this norm has unacceptable consequences." (COA = community of actors)	The antagonist perceives that the protagonist or COA believe that the norm is important to uphold for their own benefit/livelihood.	The antagonist perceives that the protagonist or COA believe that norm is consistent with their values and principles.	If attributed, the protagonist or COA can impose reputation costs on the antagonist.	Reputation damage will result in unacceptable financial, social, or political costs for the antagonist.	
Resilience	"The protagonist has previously demonstrated that the effects of the antagonist's attacks have been mitigated, or that they (the protagonist) have been able to recover promptly."	The antagonist perceives that the protagonist believes resilience measures are in its own best interests to create and implement (e.g. not too expensive).	The antagonist perceives that the protagonist believes resilience measures are consistent with the protagonist's principles (e.g. do not violate certain rights or freedoms of citizens).	The antagonist has sufficient visibility into the protagonist's resilience to believe their attack would be ineffective. The antagonist believes that the protagonist's resilience measures are as consistent and effective as the protagonist claims.	The antagonist believes it would require too many resources to overcome the protagonist's resilience measures.	

4. CHARACTERIZING CYBER DETERRENCE UNDER SPECIFIC CONDITIONS

4.1. Scenario Description

This scenario describes a notional cyberattack on an electric power grid that we developed in order to exercise the Cyber Deterrence Framework in an actionable way. The details are inspired (but not restricted) by historical events[78] and abstracted for the purposes of analyzing the effectiveness of deterrence options in Sections 4.6 and 4.7. This notional scenario is intended to identify what can be done to deter in a such a situation. What options would be available? Which of these options have the potential to effect change and which options are likely to fail at achieving a deterrent effect?

In this scenario, the term *Blue* may be used, based on context, to refer to the country of Blue, the government of Blue, the population of Blue, the specific target within Blue that Red intends to disrupt via cyber means, or some combination of these. Red is primarily used to describe the specific antagonist (individual or group) conducting the offensive cyber operation against Blue, though Red may also be used, based on context, to refer to the country, government, population, etc., that the antagonist (individual or group) belongs to. This ambiguity is intentional and reflects the real ambiguity that exists in interactions between actors in cyber contexts.

The following subsection explains in detail a notional attack on Blue's power grid using an advanced malware platform designed for use in Blue's power grid control systems. This malware would be deployed in a transmission substation serving Blue's capital city, and, when triggered, would take down key transformers and wipe data within the control systems. This would cause a blackout in a portion of the capital city lasting for approximately one hour as the substation would be brought back online via manual operations. This scenario assumes that there would be very little in the way of physical damage, beyond the substation's transmission systems requiring manual point-of-contact restart and OT systems requiring re-imaging and reprogramming due to a data wipe of configuration files. Partial recovery would take only hours, but total recovery would require weeks to months. Consequences would be difficult to estimate due to the limited nature of the event. A post-event analysis of the malware would reveal that the threat actor had the potential to cause more damage but chose not to pursue this course of action. The novel nature of the malware's advanced capabilities and the timing of the event to roughly coincide with the passing of one year since Blue previously suffered a power outage due to cyber means, would likely instill a sense of fear and panic in Blue's public and lead to a loss of confidence by the public in Blue's governance institutions.

4.2. Scenario Summary

The following *profiles* provide a snapshot summary of the notional scenario broken down by Red actor, Blue target, attack details, and consequences.

[78] Assante, Lee, and Conway, *ICS Defense Use Case No. 6: Modular ICS Malware*, ; *CRASHOVERRIDE: Analysis of the Threat to Electrical Grid Operations*, DRAGOS (2017), ; Anton Cherepanov, *WIN32/INDUSTROYER: A New Threat for Industrial Control System*, ESET (2017),

Red Profile

Antagonist Type	Generic Advanced Persistent Threat (APT) antagonist with suspected backing of nation state antagonistic to Blue.
Ideology	Pro-hacktivists.
Motivation	Extreme political beliefs. Believe ICS attack will further their cause and demonstrate resolve.
Objective	Cause electric outages by taking a single Blue transmission-level substation offline. Sabotage ICS systems at the station through malicious operation of control units and deletion of files to delay recovery.
Effect	Disrupt operations, test capabilities, spread fear, send political message, destabilize region.
Technical Personnel	Tens of technical staff, working over about a year.
Cyber Proficiency	High IT proficiency with additional ICS SME level knowledge.
Resources	Significant but not exorbitant financial resources
Preparation Time	Months to a year. Previous event was likely precursor.
Stealth and Acting Through Proxy	Group responsible may be acting on behalf of a state actor, low expectation of anonymity for group which conducts attack.
Commitment	Highly committed.

Blue Profile

Protagonist Type	Critical infrastructure entity with strong partnership with Blue government.
Ideology	Maintain reputation and quality of service.
Motivation	Provide critical resource, electricity, upon which so many sectors depend, to the capital city of Blue.
Objective	Ensure successful, resilient operations of the substation, and prevent, detect, and respond to malicious cyber operations intending to disrupt these operations.
Security and Resilience Posture	Robust and continuing to improve.
Values, Norms, and Incentives	Previous cyber event targeting other portions of Blue's electric grid have highlighted the need for cyber defense and resilience at critical infrastructure entities across the nation.
Punishment Capabilities	Blue substation is unable to punish the APT group. Blue's government has options available at various levels, from name and shame and sanctions coordinated with allies to kinetic attacks.
Ability to Attribute	Additional resources available and intelligence available to Blue stakeholders will likely enable Blue's government and allies to attribute the event correctly to the APT group and to the suspected nation state sponsor.

Attack Profile	
Target(s)	Single transmission level substation.
Target Defenses	Appropriate protections in place, however, vulnerabilities exist. Networks are ostensibly air-gapped but antagonist will identify access points.
Geographic Reach	One substation serving a portion of Blue's capital city.
Attack Vector	Advanced Malware - modular software platform.
Access	Initially phishing allows access, then pivoting to ICS control systems.
Duration	Preparation and staging take months to a year.
Sophistication	The attack is relatively complex, requiring knowledge of power transmission systems and access to resources for testing.
Novelty	Highly novel.

Consequences	
Casualties	No loss of life, approximately 200,000 residents lose power for about one to two hours in the late evening and early morning.
Physical Damage	Transformers impacted and configuration files wiped.
Economic Impact	Difficult to estimate due to time of day and quick recovery, however, recovery requires manual operations. Loss of the transmission of several hundred megawatts of energy due to disruption in operations.
Displacement	No one is displaced.
Social Impact	Impacts society primarily via panic and distrust of government and institutions.
Reputation Impact	Loss of confidence in electric grid, loss of confidence in government to protect domestic infrastructure, large political impact with international attention paid to Blue's government and ICS as well as Red's suspected government backers.
Recovery Timeline	1 to 2 hours for partial recovery, manual operations continue for an extended but unspecified duration (likely months) until complete recovery.

4.3. Scenario Context

The Blue target in this scenario is a transmission substation serving Blue's capital city. Approximately one year prior, a separate portion of Blue's power grid suffered what was, at the time, considered to be the first known effective cyberattack on a power grid and commonly referred to as a practice run for the alleged but not confirmed antagonist Red. Blue's government has had a long, not always amiable history with its neighbors but is highly entangled with the suspected

antagonist. The Blue target's objective is to ensure successful, resilient operations of the substation, and prevent, detect, and respond to malicious cyber operations intending to disrupt these operations.

The antagonist Red is described here only in general terms: a generic, pro-hacktivist Advanced Persistent Threat (APT) group with extreme political beliefs. The group is highly committed with an intent to cause panic and undermine confidence in Blue's institutions causing Blue's population to feel vulnerable. Specifically, by violating loosely held norms about cyber-physical attacks, disrupting the grid, and risking real world harm which will demonstrate the seriousness of their resolve. The Red actor's objective is to cause electric outages by taking a Blue transmission-level substation offline through sabotage of the station's industrial control systems (ICS) and deletion of files to delay recovery. The APT group is believed, but not confirmed, to have the backing of a rival nation state to Blue's government. The government of Blue is aware of this connection and has seen other malicious cyber incidents target different sectors of Blue's economy and government in recent months. This attack would likely take tens of experts working for about a year. The antagonist Red would probably have access to real power system ICS hardware for testing. This would take significant, but not exorbitant, financial resources.

The malware used in the present scenario is a modular software platform and is based on the descriptions detailed in several public security firms.[79] The malware will be referred to simply as Advanced Malware.

4.4. Attack Description

This section will describe in detail the cyber kill chain of the scenario.

Preparation, Engagement, and Presence

Red will conduct reconnaissance operations to identify information about the Blue substation's operations, technology, and potential points of entry. Vulnerabilities in the facility that control power transmission and transformer hardware will have been found initially. These initial entrances through corporate firewalls are generally accomplished by traditional email phishing methods so that Red establishes a sustained presence on the internal corporate network. The antagonist Red will then pivot within the network over a timeframe of several months to establish additional access and backdoors and acquire deeper understanding of the ICS systems controlling the facility. Although corporate networks have industry standard defenses in place, time and persistence will allow the Red to establish a sustained presence. Additionally, Red will have power system subject matter experts to understand the power grid's transformer substation and the ICS control systems in the facility.

Red will take time to map the ICS environment and, with the necessary information in hand, identify an appropriate location to position the Advanced Malware for carrying out the attack. The Advanced Malware is modular platform that can launch various payloads to achieve Red's desired outcomes. The modules are tailored to affect the substation's operations, for example by continuously switching the state of operational control units between "open" and "closed" states to create instability and failure. A data wiper module will also execute after a set period of time from the launching of the Advanced Malware to obfuscate evidence of its usage and overwrite

[79] Assante, Lee, and Conway, *ICS Defense Use Case No. 6: Modular ICS Malware*, ; *CRASHOVERRIDE: Analysis of the Threat to Electrical Grid Operations*,

configuration files to impede recovery. One of Red's previously established backdoors will enable command and control to launch the Advanced Malware at the time of Red's choosing.

Effects

The Advanced Malware is deployed to execute in the late evening. The included modules will render much, if not all of the substation's transmission capabilities inoperable and unresponsive. This results in a loss of power equivalent to about 200MW, to approximately one fifth of the population of Blue's capital city, or roughly 200,000 residents. Operators at the facility will quickly discover that their control systems have been compromised. Attempts to restart these systems will prove unsuccessful. The physical units that transform voltage levels for further distribution will require physical presence of trained personnel to restart, and sustaining operations will have to occur manually. Returning the substation to partial state of operations through this process will take minutes to hours. However, the substation will require additional time and technical expertise for complete recovery and to identify the source of the outage, due in part to the data wiping module of the Advanced Malware. Reconstructing the timeline of events and performing attribution will involve multiple stakeholders from the substation, Blue's government, Blue technical personnel, and cooperation with allies of Blue. Ensuring that the substation's systems no longer contain hidden implants of the Advanced Malware or other backdoors will entail extensive analysis that may never be completed and, thus, result in a newly accepted risk for the facility.

Because of the duration of the event, and that fact that the event happened in late evening and early morning, there will be confusion in the larger population the following day. The details of what exactly transpired will not be not immediately clear. However, given the proximity to the one-year anniversary of the previous cyberattack on Blue's electric grid, theories and rumors will quickly spread. Much of Blue's larger population will be unaffected by the outage directly, but uncertainty about the future (near and far-term) will cause some panic and a loss of confidence by the public in Blue's government and authorities for not detecting or preventing the event. The substation will release information periodically about the event indicating a malfunction in operations and identifying some of the actions taken in response, such as manual operations and forensic analysis of the impacted systems. During the outage itself, certain interdependent infrastructure and other sectors in the affected area will suffer minor impacts. This includes but is not limited to: clean water distribution, sewage treatment, medical services, communications (e.g., cell service), transportation, and financial industries. However, these entities will likely have back-up energy sources to provide power through the outage further limiting consequences.

Additional details of the response of Blue's government and the substation are not included here, as the purpose of the framework is to analyze deterrence options for Blue (substation, government, some combination of the two, or another entity) at any point in the sequence of events described above. The tactical objectives of all actors involved evolve in each threat stage, so naturally certain points in the kill chain may provide different sets of deterrence options. The effectiveness of various deterrence mechanisms may also change depending on which specific kill chain activities they target. Therefore, careful selection of an analysis point within the kill chain, coupled with thoughtful definition of Blue, Red, and their respective objectives, informs which deterrence actions are available to Blue (Layer 4) and how effective these actions may be (Layer 5). In the following sections, we examine two case studies of the scenario described above for different points in the kill chain. We will demonstrate how the framework is used to analyze deterrence options for each case.

4.5. Two Case Studies

Figure 5 shows a graphical representation of the events in the scenario kill chain mapped to the threat stages described in the NSA/CCS framework, with cases A and B depicting the two stages we will analyze. In principle, we can examine any stage of the kill chain for deterrence effectiveness. Depending on the details of the scenario, various levels of granularity may be important to scrutinize. For example, are the deterrence options Blue might have to deter an antagonist from pivoting horizontally within their system substantively different than those available to deter an antagonist from privilege escalation, or are they similar? The cases selected for this paper should be considered prototypes or exemplars, rather than comprehensive truths.

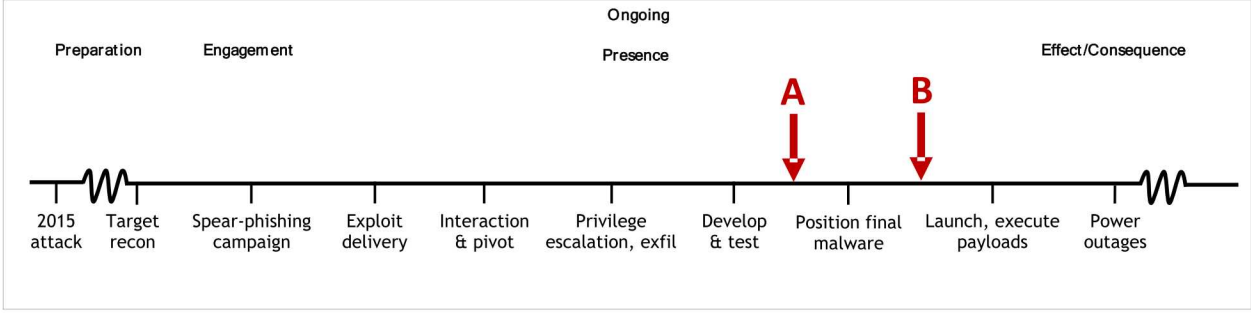

Figure 5: Kill-chain sequence for hypothetical scenario.

For Case A, Blue has observed that Red has gained access to the networks running their substation. Blue observes that Red has been able to escalate privileges, conduct detailed scans of their networks, query resources, and generally probe their system to gain an understanding of the substation's operations. Blue has no evidence that the substation's IT networks have been altered in any way that compromises Blue's core objective of being able to reliably deliver power to customers. However, based on the pattern and combination of Red's activities, the level of access they have gained, and the types of information they have sought, Blue suspects that Red's activities go beyond data collection and indicate Red's intent to be able to disrupt the substation's ability to distribute power in the future. We assume that this level of situational awareness is not unreasonable, because, given their previous experience, Blue has stationed highly capable network security teams at their grid substations. Therefore, at this precise moment, Blue's deterrence objective is *to deter Red from positioning malware that could compromise successful, resilient operations of Blue's substation.*

We contrast this situation to Case B, in which Blue observes that the Advanced Malware has already been placed onto the IT systems of their substation, thereby gaining the ability to compromise the substation's operation. Blue does not know when Red is planning to launch the malware, and they do not know how severe and extensive the damage will be. Upon discovery of the Advanced Malware, Blue would likely initiate certain actions to contain, quarantine, or remove the malware from their systems. However, Blue may be uncertain about the extent of the infection and may not be confident in their ability to completely and rapidly purge their systems without adversely affecting the substation's operations. Additionally, Blue knows that Red is still observing them, and Blue would not want to provoke or incentivize Red to launch the malware if Red suspects that Blue has discovered it and is taking steps to remove it. Blue may also suspect that Red intends only to position the malware, not launching it immediately, but rather saving its activation for an opportune moment or simply holding Blue's critical assets at risk. For all of these reasons, we assume that, in addition to defending their networks and sustaining operations, Blue would also consider options to

deter Red from launching the malware to cause a power failure. In this case, Blue's deterrence objective is *to deter Red from launching*

malware that would compromise successful, resilient operations of Blue's substation. *Table 2* summarizes the first three layers of the framework for Cases A and B.

Table 2. Framework Layers 1-3

		Case A	Case B
L1	Stage	After Red has established presence, but before they have installed or positioned Advanced Malware on Blue's systems.	After Red has positioned Advanced Malware on Blue's systems, but before Red has launched the malware to achieve their intended effect (power outage).
L2	**Red Objectives** Red = individuals of the APT group	*Motivation* Extreme political beliefs. Believe ICS attack will further Red's cause and demonstrate resolve. *Objective* Obtain the ability to cause electrical outages by placing malware onto the IT networks of Blue's transmission-level substation that is capable of taking the station offline, sabotaging its ICS systems through malicious operation of control units, and deleting files to delay recovery.	*Motivation* Identical to case A. *Objective* Cause electrical outages by taking a single Blue transmission-level substation offline. Sabotage ICS systems at the station through malicious operation of control units and deletion of files to delay recovery.
L3	**Blue Deterrence Objectives** Blue = the government and utility company working in close coordination	*Motivation* Provide electricity to Blue's capital city by ensuring successful, resilient operations of the substation. *Objective* Deter Red from *positioning* malware that could compromise successful, resilient operations of Blue's substation.	*Motivation* Identical to case A. *Objective* Deter Red from *launching* malware that would compromise successful, resilient operations of Blue's substation.

In both Cases A and B, it is important to remember that neither Blue nor Red can see into the future. Blue does not know if and how Red will choose to execute its strike, nor what Red's intended effects are. Blue does not know the extent of damage Red intends to cause (i.e., how long the power

outage will last and how many people will be affected). Similarly, Red does not know all of Blue's capabilities. Assuming Blue has identified Red as the actor that attacked their power grid a year earlier, then both sides can make strongly educated estimates of the other's capabilities and intent. For example, Red may know that Blue has the ability to begin manual override operations of its substation if its networks are compromised, which was done following the 2015 attack. Additionally, substantial communication and signaling is happening between Red and Blue through Blue's networks, although this communication is not overt and therefore subject to uncertainty.

Table 3 summarizes some of the deterrence actions Blue may consider for both cases, across all four mechanisms discussed in Section 3.2. This constitutes Layer 4 of the framework, for each case. Given the close proximity of Case A and Case B in the kill chain, the contents of Layer 4 are the same for both. These options are non-exhaustive and only intended to highlight possibilities.

Table 3. Deterrent threat options across three categories

<table>
<tr><th colspan="2">Resistance</th></tr>
<tr><td colspan="2">• Degrade/disable/destroy machines from which Advanced Malware originates or routes using cyber tools
• Degrade/disable/destroy machines from which Advanced Malware originates or routes using kinetic tools
• Preemptive arrest or dissuasion of responsible parties
• Name and shame to build public awareness and break antagonists' anonymity
• Work with international partners to establish norms against targeting civilian critical infrastructure
• Establish a moving target defense
• Establish an air gap
• Up-to-date patches
• Reverse engineer Advanced Malware, then patch
• Implement application white listing
• Segment network into logical enclaves (defense in-depth)
• Intrusion detection (intrusion detection/protection systems, security event identification manager)
• Outsource security operations (security as a service: monitoring, red-teaming, training, collecting information for attribution)
• Train a security-conscious workforce
• Wipe the machine
• Implement multi-factor authentication
• Implement security command center
• Set up a honeypot
• Implement antivirus system</td></tr>
<tr><th>Retribution</th><th>Resilience</th></tr>
<tr><td>• Government threatens legal, economic, or diplomatic consequences against individuals in APT group
• Persuade coalition of allies/partners to threaten legal, economic, or diplomatic measures against suspected state sponsors of cyber attacks
• Government threatens kinetic military retribution against Red's suspected state sponsor
• Government threatens cyber military retribution against Red's suspected state sponsor
• Industry coalition threatens retaliation
• Threaten to release a press statement – threaten to name and shame individuals
• Build entangled electric grids with Red or Red's allies</td><td>• Re-route grid; leverage distribution and production from alternative providers
• Isolate controls (limit damage to fewer systems)
• Manual override/operations
• Redundant systems – ensure components are replaceable/swappable
• Crisis protocols</td></tr>
</table>

In Layer 5 of the framework, we can compare each of these deterrent threat options based on the criteria discussed in Section 3.3 by seeking to answer the following questions:

1. How can Blue communicate the deterrent threat to Red?
2. Is the deterrent threat credible to Red? (Is it consistent with Blue's principles, and is it rational for Blue to carry out?)
3. Is Blue capable of carrying out the deterrent threat? (Can Blue execute the deterrent threat, and is it sufficiently painful or costly to Red?)
4. Is Red calculating (i.e., Does Red weigh the potential costs and benefits of their actions considering Blue's deterrent threats)?

For this paper, we assume that Red possesses some form of rationality, and that they weigh the relative costs and benefits of various courses of action, even if such considerations are qualitative and based on constrained perspectives or flawed knowledge. Therefore, as previously stated, we assume Red is calculating for both cases. Thus, we will discuss each of the first three questions for Case A and Case B in the following sections.

4.6. Case A: Deter Red from Installing and Positioning Advanced Malware

How can Blue communicate the deterrent threat to Red?

In considering communication between Red and Blue, the deterrence actions listed in *Table 3* fall into two general types: they are either 1) actions that Blue *takes or has taken previously*, but that Red does not yet know about, or 2) actions that Blue threatens to take but *refrains from taking* until Red has attacked Blue. In the former, Blue yields final agency to Red, who must decide on a course of action, given a more complete understanding of the characteristics of its Blue target. In the latter, Blue retains some agency to act, and ideally the two sides engage in tacit bargaining to arrive at mutual understanding about the costs that both sides are willing to endure to pursue their interests.

The first type of deterrence action is exemplified by the security measures Blue has already implemented from the "resistance" category of *Table 3*. This include actions such as up-to-date patching, multi-factor authentication, and training a security-conscious workforce. Of course, if Red is aware of such measures, and chooses to act anyway, then they failed to deter Red. However, throughout its activities in Blue's networks, Red may encounter new or unknown obstacles, simply through trial and error, that change their perceived cost/benefit ratio. Additionally, in Case A, Blue may have sufficient time to implement new security measures, such as moving target defense[80] or network segmentation, that could significantly frustrate Red's ability to carry out their attack as planned. Any such actions would also be examples of the first type of deterrence action. Furthermore, many of the resilience deterrent threats will also be implemented by Blue in advance. For these options, the question is how much Blue should communicate about its ability to sustain and recover power grid operations, keeping in mind that some of the most credible communication Blue has done on this front are its actions and responses to the previous attack.

Most of the retributive deterrent threats represent the second type of deterrence action, in which Blue would *threaten* to respond to Red's activities but would refrain from taking that response until after Red attacks. Such threats by Blue can be explicit or implicit, and they can come in the form of precedents that Blue and others have set previously. Red may also deduce retributive threats from its

[80] Aswin Chidambaram Pappa, "Moving Target Defense for Securing Smart Grid Communications: Architectural Design, Implementation and Evaluation" (Master of Science Iowa State University, 2016)

general knowledge of Blue's capabilities, rather than from any overt threat Blue makes under specific conditions (i.e., simply the knowledge that Blue *can* respond in a certain way could be enough to signal to Red that Blue *may*).

Is the deterrent threat credible to Red?

Aspects of both credibility and capability are highly dependent on the actors involved. For a generic Red vs. Blue scenario, we lack crucial information about the complex history of the actors, what their relationship is and has been, and what their strategic interests and long-term goals are. Crucially, we also do not know where each actor stands on the world stage in terms of their ability to project power and influence other actors. Are the nations of Red and Blue superpowers, or is there an asymmetry of power at play? Where we make assumptions about Red and Blue, we will state it, but the more important exercise here is to map the decision space as broadly as possible for various types of Red and Blue actors. We leave it to others to fill in the details for specific actors.

Generally, our assumptions about Blue are as follows. The nation of Blue is relatively well-integrated with the international community, that they value international trade, have some form of regulated free trade economy, and highly value and adhere to international norms and laws surrounding human rights and justifiable uses of force. Blue has some form of representative government such that public opinion has power over the government. Blue has strategic interests and a robust military capability to defend and project those interests abroad, but we do not assume that Blue is a superpower as such. Blue, perhaps in conjunction with key allies, has competent cyber military force and intelligence capabilities. Finally, we assume that the government of Blue has strong partnerships with its private sector, particularly those aspects supporting infrastructure that is critical to daily life for the citizen of Blue and to military operations. (We also acknowledge the challenge of achieving these close partnerships in reality.)

In assessing the *credibility* of actions that Blue may take to deter Red, Red would consider if Blue's actions are consistent with Blue's internal principles and values, and if Blue's actions are rational, (i.e., will ultimately lead to a preferable outcome for Blue). Red would likely perceive that all of Blue's resilience deterrent threats are consistent with Blue's principles. The same is true of the resistance deterrent threats that involve Blue taking defensive actions within its own networks and systems. However, not all of these options may be rational for Blue. For example, building and implementing intrusion detection system (IDS)/intrusion protection system (IPS) infrastructure is complex and costly. If Blue's utility companies are privatized, they may not be adequately incentivized to build, maintain, and monitor costly cyber defense systems, which could severely cut into their profits. If the ability to re-route power distribution or switch into manual override operations is not already in place, constructing that additional infrastructure may also be costly. Red would probably seek to make informed decisions about which systems Blue has protected, and which ones may be vulnerable by deliberate risk acceptance or by circumstance. If we assume a strong partnership between Blue's government and private sector, and a strong commitment by all to the importance of investing in cybersecurity for the electric grid, then the rationality, and therefore credibility, of many of the resistance and resilience options improves.

For Case A, Red has established presence on the networks of one of Blue's power grid substations, but they have yet to alter Blue's systems in such a way that could compromise Blue's ability to execute its core mission, that is, delivery of electricity to the capital. This limits the credibility of many of the retributive deterrent threats and any resistance threats that involve a kinetic use of force, such as the destruction of those machines from which Advanced Malware originates using a missile strike. Thus far, all of Red's activities fall into the category of computer network exploitation

(CNE), or "enabling operations and intelligence collection capabilities conducted through the use of computer networks to gather data from target or antagonist information systems or networks."[81] Blue's competent cybersecurity teams have just begun to observe behavior that indicates Red's intent is very likely to go beyond CNE to execute some kind of computer network attack (CNA), or "actions taken through the use of computer networks to disrupt, deny, degrade, or destroy information resident in computers and computer networks, or the computers and networks themselves."[82] States have not historically responded to CNE activities using kinetic force, so those options are incredible by precedent. Blue would probably view responding to CNE activity with any kind of kinetic military force as disproportionate – if not contrary to their principles, then not worth the potential for escalation from a rational perspective. Non-kinetic deterrent threats, such as legal, economic, or diplomatic action against individuals, or the creation of coalitions to threaten state sponsors of cyberattacks, are slightly more credible, depending on Blue's standing in the world. Red would have to assess if Blue's use of offensive cyber operations to preempt or respond to Red's actions is consistent with Blue's principles, or whether Blue might consider such actions to be too escalatory. At this stage, any direct threat against Red's suspected state sponsor would also likely be viewed as highly escalatory and destabilizing, and therefore incredible.

For deterrent threats that Blue has already implemented, certain common-sense defensive or resiliency measures, for example, it would seem that the question of credibility and capability has already been answered. Blue has demonstrated, by taking an action, that they believe this action is principled and rational and that they are capable of acting in this way. Red's belief that Blue took the action is all that Red should need to prove Blue's credibility. However, for certain deterrent threats, we think Red may also consider whether it is principled and rational for Blue to *sustain* actual implementation of the deterrent threat. For example, outsourcing security operations may be so expensive that Blue can only do so for a limited time, or only at a fraction of their substation locations. Similarly, a public outcry among Blue's domestic audiences in opposition to preemptive kinetic strikes against APT groups, for example, could make Red perceive that Blue's threats along those lines as not principled and therefore incredible. Similarly, there may be reasons why Red perceives Blue's capability to change over time.

Is Blue capable of carrying out the deterrent threat?

For this scenario we have assumed that Blue has competent law enforcement, standing armed forces, and a cybersecurity-conscious workforce running their electric grid infrastructure. We assume that if the private sector lacks sufficient resources or expertise to implement certain defensive or resiliency measures that they will receive material aid from the government to do so. Therefore, Blue, perhaps in conjunction with key allies, is capable of carrying out the vast majority of options listed in *Table 3*. One major limitation on Blue's capability is time. Blue may lack sufficient time to effectively carry out some of the actions in the resistive and resilience deterrent threat categories. For these actions, Red would be more likely to be convinced of Blue's capability for Case A than for Case B, when even less time is available to Blue.

A second major limitation is Blue's ability to attribute activities it observes to Red or to any suspected state sponsors of Red. Without transparent and high-confidence attribution, Blue's ability to carry out any of the deterrent threats requiring positive identification of individuals responsible is severely hindered, given Blue's likely unwillingness to retaliate against potentially innocent parties.

[81] "Computer Security Resource Center Glossary," National Institute of Standards and Technology, accessed 9 September 2019, 2019, https://csrc.nist.gov/glossary

[82] "Computer Security Resource Center Glossary,"

Red may also be aware that Blue would not want to make attributions public, if doing so would burn intelligence sources or other secrets.

4.7. Case B: Deter Red from Launching Advanced Malware to Achieve Power Outage

How can Blue communicate the deterrent threat to Red?

The most significant difference between Case A and Case B for Blue is the time they have to react to the presence of Red in their networks. In Case B, Blue would react with urgency to address the immediate threat that could compromise the substation's operations. Whatever defensive cybersecurity measures that were in place failed to stop Red's progress and did not sufficiently deter Red. There may be limited time or utility in implementing new cybersecurity tools, with the possible exception of degrading or disabling Red's command and control of the Advanced Malware. Such measures would be communicated from Blue to Red through similar "trial and error" mechanisms discussed in Case A. Options involving forward defense or persistent engagement (e.g., using cyber tools to degrade or disrupt Red's machines) at this later stage are much more likely to be intended to prevent Red from continuing their attack, rather than to deter them. In the case where such an action resulted in dissuading Red from further action, the communication would occur through Red's direct observation of the actions that Blue has taken to degrade Red's abilities.

There may be time to implement and signal resilience threats to Red. Assuming infrastructure is already in place, Blue could leverage alternative sources of electricity production and distribution for its capital city, ensure readiness for switching to manual override operations, or emplace redundant systems and components. Credibly signaling that such measures have been taken is more difficult without revealing information Red might put to good use later. Blue would therefore probably rely on a combination of vague public statements and its history of adequate responses to power outages in the past. Blue's best option here would be to increase the operational uncertainty for Red, for example, by introducing doubt that their actions will produce the desired effect, or doubt that the effect will be worth burning heavy investments in exploit and malware development. This uncertainty could make Red change their course of action to one they perceive as less risky.

Communication is more straightforward for retributive deterrent threats, than for other categories of deterrence. Given the clear and present danger to Blue's electric grid, they would probably feel justified in overtly threatening members of Red, or perhaps even Red's suspected state sponsor, with punitive consequences if Red launches the Advanced Malware. Public statements by the government have the additional advantage that they can be done very quickly.

Is the deterrent threat credible to Red?

In Case B, Red has positioned the Advanced Malware onto the networks of Blue's substation, thereby gaining the ability to compromise Blue's electric grid operations. The effects this escalation has on Blue's deterrent threat options mainly stem from two sources. First, Blue may perceive that its critical assets are at immediate risk, leaving Blue very little time to act. If novel cyber defense tools and resilience measures were expensive to implement in Case A, they are even more expensive (and perhaps impossible) to implement in Case B, under the pressure of time. Whichever resilience measures Blue is ready to implement would likely be viewed as principled and rational, however. Therefore, in this case, the main challenge is how to signal such actions to Red in time. While Blue may still consider resistance and resilience deterrent threats to be principled, in this case it may not view them as entirely rational. In the event that it becomes impossible for Blue to implement these

measures because of limited time and resources, then it becomes a question of Blue's lack of *capability*.

Second, Red's escalatory action of holding Blue's critical infrastructure at risk means that some of the retributive deterrent threats perhaps are now more credible than they were before. Threatening legal action against individuals now certainly seems justified, because Red is threatening the lives and well-being of Blue's citizens. Offensive cyber operations, as either part of resistance or retribution, are also likely to be viewed as more principled and more rational than in Case A – more principled, because they are more proportionate; more rational because this would seem to be a clear case of self-defense, which is less likely to risk escalation. Threats of kinetic military retribution may still seem disproportionate, especially if Blue calculates that it can mitigate the effects of a power outage as it did in the previous attack. However, Blue may calculate that the international community would view Red's launching of the Advanced Malware to achieve a power outage as a violation of Blue's state sovereignty, justifying some form of military response against Red or even potentially Red's suspected state sponsors if they can make a public attribution. Whether or not Blue would actually respond militarily is another matter. This would probably depend on the extent of the consequences of the power outage, and Blue's standing in the world, including the military alliances it can rely on. Unless the effects of the power outage are *very* severe, causing the deaths of hundreds or more of Blue's citizens, for example, we do not think a threat of military retaliation from Blue is credible against Red or its suspected state sponsor. In cooperation with other nations, threats of retaliation against individuals identified as being responsible (e.g., arrest, travel restrictions, individual economic sanctions, etc.) would be more credible.

Is Blue capable of carrying out the deterrent threat?

As in Case A, the primary limitations on Blue's capability of carrying out the deterrent threats in *Table 3* are time and the ability to make transparent, high-confidence attribution of malicious activities on its networks. The severe time limitation for Case B would probably rule out the implementation of any novel defensive or resilience measures not previously in place. This aspect of the scenario seems to favor retribution and some persistent engagement options, assuming Blue can attribute the attack. Even if attribution cannot be determined immediately, we assume that Blue eventually will be able to trace the activity to Red, and potentially even the suspected state sponsor. Therefore, we believe in this case that it is very likely Blue would be capable of carrying out retributive deterrent threats at some point in time.

5. CONCLUSIONS

The electric grid scenario we examined in this paper falls somewhere in the middle of the "gray zone" area of our notional two-axis threshold diagram (*Figure 2*, reproduced in *Figure 6*). Compromising the ability of a nation's infrastructure to deliver essential services to its public is a serious act of aggression. But for this particular attack, there were no casualties or displacements, and even if there had been, they would have been indirect rather than direct, which raises doubt about whether or not Red intended or expected such casualties to occur. Blue was able to recover power lost within approximately an hour, and given their prior experience, they probably knew that they would be able to do so. Therefore, the scenario at large seems to fall short of a clear use-of-force threshold. However, Red intended their attack to signal their capabilities and gain notoriety. As described in the scenario, there was evidence that Red could have produced worse effects but chose not to do so. All of this indicates that Red wanted to broadcast their ability to carry out these types of attacks against Blue and its allies in the future. There are political ramifications at work here that cannot be measured simply through the summation of hours of power lost, dollars to repair damaged systems, and casualties or displacements.

Others have argued that deterrence is fundamentally not credible in this gray zone – that we cannot deter actors here, or perhaps that deterrence has already failed in this space. To examine this question further, we created a framework to systematically analyze scenarios of interest (both in and beyond the gray zone), to evaluate various deterrence strategies. We then demonstrated how to use the framework on a single scenario. Much more work, preferably using actual rather than hypothetical data, is needed to draw definitive conclusions. However, our preliminary discussion indicates that there are some deterrence strategies that appear to be more suitable than others for this scenario. In this case, international law does not give Blue many options for anything involving kinetic resistance or retribution. There is simply no (or *very* limited) precedent. The credibility of even non-kinetic retributive options depends on Blue's military capabilities, its power and influence in the world, and its asymmetry with Red's suspected state sponsor.

However, the same cannot be said of defensive resistance and resilience options, which appear to be more credible options. Here the challenge lies with communication – with convincing Red that selection of particular targets carries with it unacceptable operational risk, because of the likelihood that these operations will not succeed or that they will yield insufficient gains, combined with the probability that Red's efforts in developing the attack will go to waste. Deterrence by denial was not the main focus during the Cold War. Because of the near impossibility of defending against or recovering from large-scale nuclear attack, civil defense programs were generally abandoned before they matured. Cyberattacks are different. We believe these differences merit further examination of how resistive and resilience deterrence might play a role in dissuading adversaries from selecting particularly critical targets in Blue spaces. It is possible that to the extent that deterrence *is* possible in the gray zone, the main source may be from these mechanisms. Another major advantage is that, for the most part, resistance and resilience do not require rapid, high-confidence public attribution.

Some version of the thresholds diagram shown in *Figure 6* could help analysts and decisionmakers understand what types of thresholds they want to set, and what kinds of thresholds may have already been set implicitly. Whether or not they want to publicize those thresholds is a policy question. However, *Figure 6* not only helps us conceptualize what types of attacks and which types of consequences "cross the line"; it also suggests what types of responses may be justified. Our ultimate vision is to use this more nuanced conception of thresholds in combination with a standardized methodology for cyber deterrence scenario analysis, such as the framework discussed

in this paper, to more fully map various types of responses across multiple levels of conflict escalation. Through further in-depth analysis of case studies (both real and hypothetical) that reside in different zones throughout the thresholds figure, it may ultimately be possible to identify generalizable deterrence strategy profiles for regions within a given zone. For example, many have argued that conventional deterrence by punishment remains a valid strategy for dissuading antagonists from using cyberattacks equivalent to the use of force – this may correspond to the areas in the upper right corner of *Figure 6*. We have hypothesized with this work that in the gray zone, resilience and resistance may be more effective at dissuading antagonists, with retaliation perhaps playing a less significant, though still important, role.

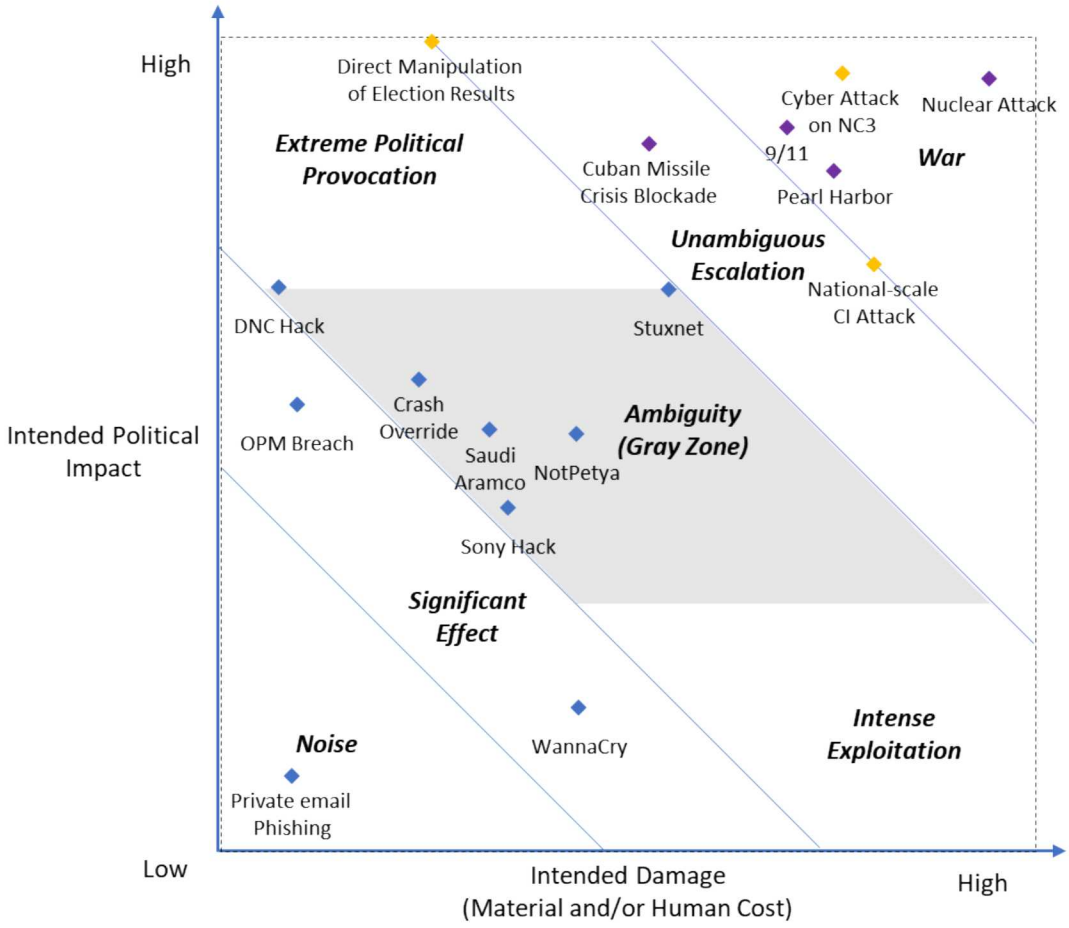

Figure 6: Strategy Profile Mapping

If one could construct a comprehensive mapping like this, it might serve as a decision support tool for cyber security implementers and policy makers as they weigh different options for influencing antagonists. Mapping an emerging threat scenario to the figure might not identify the exact strategy or strategy set that would be most effective, but it could potentially narrow the option space to a "most likely" set. Analysts might also be able to draw insight from comparable scenarios or case studies within the same zone or region. Analysis might further be expanded to understand the potential role of deterrence strategies in combination with other modes of influence, including compellence and coercion. The framework also allows us to examine the evolution of thresholds over time, as conditions change. Beyond that, we might also examine how to drive this evolution

through deliberate action. Can we move the lines in *Figure 6*? Can we make certain zones shrink or expand, or make them more obvious or clear?

We see two complementary paths forward for this work. The first is to use the framework to conduct an in-depth exploration of a few specific contexts of immediate policy relevance using real-world actors and data, and to explore the various thresholds involved for those specific actors within that context. Such work could inform policy on what thresholds may exist within that specific context of immediate interest. A second path forward is to use the framework to conduct a longitudinal study of many different types of scenarios to learn about the broader trends across the set. This would help drive our general understanding of how deterrence in theory fits into the grand strategies of nations and how they choose to conduct multi-domain conflict and competition. It may also help our understanding of how certain courses of action help or hinder the evolution of certain desired end-states for the domain (i.e., stability, superiority, or some optimized combination thereof). The framework described here is flexible and can be adapted to multiple use cases. Ultimately, we believe it is premature to abandon decades of learning we have gained from studying deterrence in other domains of conflict, and we argue that a more systematic examination of deterrence in cyber conflict scenarios is valuable and needed.

6. REFERENCES

Achieve and Maintain Cyberspace Superiority: Command Vision for Us Cyber Command. 2018.

"Alert (Ta17-163a) Crashoverride Malware." U.S. Department of Homeland Security, Updated 27 July 2017, 2017, accessed 6 September 2019, 2019, https://www.us-cert.gov/ncas/alerts/TA17-163A.

Allison, Graham, and Philip Zelikow. *Essence of Decision: Explaining the Cuban Missile Crisis*. New York, NY: Longman, 1999.

Anthony, Robert W. *Deterrence and the 9-11 Terrorists*. Institute for Defense Analyses (Alexandria, VA: 2003).

Assante, Michael J., Robert M. Lee, and Tim Conway. *Ics Defense Use Case No. 6: Modular Ics Malware*. SANS Industrial Control Systems, Electricity Information Sharing and Analysis Center (Washington, D.C.: 2017).

Baker, Peter. "For Obama, Syria Chemical Attack Shows Risk of 'Deals with Dictators'." *New York Times*, April 10 2017, A, 11.

Bolton, John. "National Security Advisor John Bolton on Cyber Strategy." news release, September 20, 2018, 2018.

Brantley, Aaron F. "The Cyber Deterrence Problem." 10th International Conference on Cyber Conflict, Tallinn, Estonia, NATO CCD COE Publications, 2018.

Brodie, Bernard, Frederick Sherwood Dunn, Arnold Wolfers, Percy Ellwood Corbett, and William T. R. Fox. *The Absolute Weapon: Atomic Power and World Order*. New York, NY: Harcourt, Brace and Co., 1946.

Cherepanov, Anton. *Win32/Industroyer: A New Threat for Industrial Control System*. ESET (2017).

Clarke, Richard A., and Robert K. Knake. *The Fifth Domain: Defending Our Country, Our Companies, and Ourselves in the Age of Cyber Threats*. New York, NY: Penguin Press, 2019.

The Comprehensive National Cybersecurity Initiative. Washington, D.C., 2009.

"Computer Security Resource Center Glossary." National Institute of Standards and Technology, accessed 9 September 2019, 2019, https://csrc.nist.gov/glossary.

Crashoverride: Analysis of the Threat to Electrical Grid Operations. DRAGOS (2017).

Davis, Paul K., and Brian Michael Jenkins. *Deterrence and Influence in Counterterrorism: A Component in the War on Al Qaeda*. (RAND National Defense Research Institute, 2002). https://www.rand.org/pubs/monograph_reports/MR1619.html.

The Department of Defense Cyber Strategy. Washington, D.C., 2015.

Fischerkeller, Michael P., and Richard J. Harknett. "Deterrence Is Not a Credible Strategy for Cyberspace." *Orbis* 61, no. 3 (2017): 381-93. https://doi.org/10.1016/j.orbis.2017.05.003.

Freedman, Lawrence. *The Evolution of Nuclear Strategy*. 3rd ed. New York, NY: Palgrave Macmillan, 2003.

———. "Ukraine and the Art of Limited War." *Survival* 56, no. 6 (2014): 7-38.

Gause, Ken. *North Korea's Provocations and Escalation Calculus: Dealing with the Kim Jong-Un Regime.* CNA Analysis & Solutions (Arlington, VA: 2015).

Gerson, Michael S. "Conventional Deterrence in the Second Nuclear Age." *Parameters*, no. Autumn (2009): 32-48.

Greenbery, Andy. "The Untold Story of Notpetya, the Most Devastating Cyberattack in History." *Wired.* (August 22 2018). Accessed September 20, 2019. https://www.wired.com/story/notpetya-cyberattack-ukraine-russia-code-crashed-the-world/.

International Strategy for Cyberspace: Prosperity, Security, and Openness in a Networked World. Washington, D.C., 2011.

Investigation: Wannacry Cyber Attack and the Nhs. London, United Kingdom: National Audit Office, 2018.

Iran's Priorities in a Turbulent Middle East. International Crisis Group (Brussels, Belgium: April 13 2018).

Jervis, Robert. "Deterrence and Perception." *International Security* 7, no. 3 (1982): 3-30.

Jervis, Robert, and Jason Healey. *The Dynamics of Cyber Conflict.* School of International and Public Affairs, Columbia University (New York, NY: 2019).

Kahn, Herman. *On Escalation: Metaphors and Scenarios.* Transaction Publishers, 1965.

———. *On Escalation: Metaphors and Scenarios.* 1968.

Kapusta, Philip. "The Gray Zone." *Special Warfare Magazine* 28, no. 4 (2015): 20.

Kauffman, William W. *The Requirements of Deterrence.* Center of International Studies, Princeton Univeristy (November 1954

1954).

Kissinger, Henry. *Nuclear Weapons and Foreign Policy.* New York, NY: Routledge, 1957.

Knopf, Jeffrey W. "The Fourth Wave in Deterrence Research." *Contemporary Security Policy* 31, no. 1 (2010): 1-33.

Kostyuk, Nadiya, Scott Powell, and Matt Skach. "Determinants of the Cyber Escalation Ladder." *The Cyber Defense Review* 3, no. 1 (2018): 123-34.

Libicki, Martin C. *Cyberdeterrence and Cyberwar.* Santa Monica: RAND Corporation, 2009.

Luce, R. D., and Howard Raiffa. *Games and Decisions: Introduction and Critical Survey.* Oxford, England: Wiley, 1975.

Lynn, William J. III. "Defending a New Domain: The Pentagon's Cyberstrategy." *Foreign Affairs*, no. September/October 2010 (2010). https://www.cybercom.mil/About/Mission-and-Vision/.

Mitchell, A. Wess. "The Case for Deterrence by Denial." *The American Interest.* (2015). https://www.the-american-interest.com/2015/08/12/the-case-for-deterrence-by-denial/.

Multiyear Plan for Energy Sector Cybersecurity. Office of Electricity Delivery & Energy Reliability, 2018.

Nakasone, Paul M. "An Interview with Paul M. Nakasone." *Joint Force Quarterly* 92, no. 1st Quarter (2019): 4-9.

National Security Strategy of the United States of America. Washington, D.C., 2017.

North American Air Defense Command and Continental Air Defense Command Historical Summary. Directorate of Command History, Office of Information, Headquarters NORAD/CONAD, 1963.

Nsa/Css Technical Cyber Threat Framework V2. Cybersecurity Operations The Cybersecurity Products and Sharing Division, U.S. National Security Agency (2018).

Nuclear Posture Review. Office of the Secretary of Defense, 2018.

Nuclear Posture Review Report. Washington, DC: Office of the Secretary of Defense, 2010.

Nye, Joseph S. Jr. "Deterrence and Dissuasion in Cyberspace." *International Security* 41, no. 3 (2017): 44-71. https://doi.org/10.1162/ISEC_a_00266.

The Opm Data Breach: How the Government Jeopardized Our National Security for More Than a Generation. Washington, DC, 2016.

Pappa, Aswin Chidambaram. "Moving Target Defense for Securing Smart Grid Communications: Architectural Design, Implementation and Evaluation." Master of Science, Iowa State University, 2016.

Perleroth, Nicole. "In Cyberattack on Saudi Firm, U.S. Sees Iran Firing Back." *New York Times*, October 23 2012. https://www.nytimes.com/2012/10/24/business/global/cyberattack-on-saudi-oil-firm-disquiets-us.html.

Presidential Executive Order on Strengthening the Cybersecurity of Federal Networks and Critical Infrastructure. 2017.

Presidential Policy Directive 21: Critical Infrastructure Security and Resilience. Washington, D.C., 2013.

Recommendations to the President on Deterring Adversaries and Better Protecting the American People from Cyber Threats. (2018). https://www.state.gov/documents/organization/282253.pdf.

Schelling, Thomas C. *Arms and Influence*. New Haven, CT: Yale University Press, 1966.

———. *The Strategy of Conflict*. Cambridge, MA: Harvard University Press, 1960.

Schmitt, Michael N. "The Law of Cyber Warfare: Quo Vadis?". *Stanford Law & Policy Review* 25, no. 2 (2014): 269-99.

Schneider, Jacquelyn. "Deterrence in and through Cyberspace." In *Cross-Domain Deterrence: Strategy in an Era of Complexity*, edited by Jon Lindsay and Erik Gartzke: Oxford University Press, 2019.

Selten, Richard. "Reexamination of the Perfectness Concept for Equilibrium Points in Extensive Games." *International Journal of Game Theory* 4, no. 1 (1975): 25-55.

Snyder, Glenn. "Deterrence and Power." *Journal of Conflict Resolution* 4, no. 2 (1960): 163-78.

Task Force on Cyber Deterrence. Department of Defense Defense Science Board (2017). https://www.acq.osd.mil/dsb/reports/2010s/DSB-cyberDeterrenceReport_02-28-17_Final.pdf.

Tor, Uri. "'Cumulative Deterrence' as a New Paradigm for Cyber Deterrence." *Journal of Strategic Studies* 40, no. 1-2 (2015): 92-117. https://doi.org/10.1080/01402390.2015.1115975.

Verba, Sidney. "Assumptions of Rationality and Non-Rationality in Models of the International System." *World Politics* 14, no. 1 (1961): 93-117. https://doi.org/https://doi.org/10.2307/2009558.

Zagare, Frank C., and D. Marc Kilgour. *Perfect Deterrence.* Cambridge Studies in International Relations. Cambridge: Cambridge University Press, 2000. doi:https://doi.org/10.1017/CBO9780511491788.

DISTRIBUTION

Email—Internal

Name	Org.	Sandia Email Address
Technical Library	01177	libref@sandia.gov

Email—External (encrypt for OUO)

Name	Company Email Address	Company Name

Hardcopy—Internal

Number of Copies	Name	Org.	Mailstop

Hardcopy—External

Number of Copies	Name	Company Name and Company Mailing Address

1

Sandia National Laboratories

Sandia National Laboratories is a multimission laboratory managed and operated by National Technology & Engineering Solutions of Sandia LLC, a wholly owned subsidiary of Honeywell International Inc. for the U.S. Department of Energy's National Nuclear Security Administration under contract DE-NA0003525.