PNNL-26932

Pacific Northwest
NATIONAL LABORATORY

Proudly Operated by Battelle Since 1965

Cyber Deterrence and Stability

Assessing Cyber Weapon Analogues through Existing WMD Deterrence and Arms Control Regimes

September 2017

R Goychayev
GA Carr
RA Weise
DA Donnelly
SL Clements
JM Benz
KE Rodda
RA Bartholomew
AD McKinnon
RB Andres

U.S. DEPARTMENT OF ENERGY

Prepared for the U.S. Department of Energy
under Contract DE-AC05-76RL01830

DISCLAIMER

PACIFIC NORTHWEST NATIONAL LABORATORY
operated by
BATTELLE
for the
UNITED STATES DEPARTMENT OF ENERGY
under Contract DE-AC05-76RL01830

Printed in the United States of America

Available to DOE and DOE contractors from the
Office of Scientific and Technical Information,
P.O. Box 62, Oak Ridge, TN 37831-0062;
ph: (865) 576-8401
fax: (865) 576-5728
email: reports@adonis.osti.gov

Available to the public from the National Technical Information Service
5301 Shawnee Rd., Alexandria, VA 22312
ph: (800) 553-NTIS (6847)
email: orders@ntis.gov <http://www.ntis.gov/about/form.aspx>
Online ordering: http://www.ntis.gov


This document was printed on recycled paper.
(8/2010)

PNNL-26932

Cyber Deterrence and Stability

R Goychayev
GA Carr
RA Weise
DA Donnelly
SA Clements
JM Benz
KE Rodda
RA Bartholomew
AD McKinnon
RB Andres

September 2017

Prepared for
the U.S. Department of Energy
under Contract DE-AC05-76RL01830

Pacific Northwest National Laboratory
Richland, Washington 99352

Preface

In recent years, it has become increasingly clear that the situation in cyberspace is deteriorating. A short list of attacks by other States on U.S. interests includes Iran's none-too-subtle attack on U.S. banks and energy infrastructure in 2013,[1] North Korea's blatant attack on the Sony Corporation in 2014,[2] China's multi-year intellectual property piracy campaign from scores of private U.S. firms,[3] and Russia's attempt to manipulate the 2016 elections.[4]

These attacks and others like them affect America's strategic position. Where we have relied on a combination of economic might, technological know-how, and close relationships with allies to support a peaceful and prosperous world system, our adversaries strike directly at each of these via cyberspace. Attacks on private industry have restricted U.S. economic growth, infiltration of critical civilian and defense infrastructure have created significant military risks, and Russia's information operations in Europe and elsewhere have undermined America's relationships with some of its closest allies.

Over the last decade, a number of us have oscillated between cyber-threat-related committees at National Defense University, the White House, Fort Meade, and the Pentagon. Each year the warnings grow louder, the evidence grimmer, and our frustrating lack of ability to solve problems greater. Often, the conversation turns to the potential importance of diplomacy, deterrence, and arms control, but seldom does the discussion expand beyond a vague desire to somehow apply these tools to improve the situation.

This report represents a clear-eyed and systematic first step toward exploring how diplomacy and arms control can contribute to deterring cyber-attacks. This report describes the ways arms control agreements can help to deter conflict, defines key Cold War agreements and how their mechanisms might or might not apply to cyber conflict, and concludes by describing how an assortment of arms control and deterrence tools might be used to reduce the threat in cyberspace.

On the surface, this report assesses diplomatic means to diminish cyber threats. Less obviously, but equally important, this report also systematically walks readers through the characteristics of arms control agreements that do not apply well to cyber conflict. In this role, it guards against the unexamined wishful thinking that too often clouds the role of diplomacy in cyber policy. If diplomacy is to play a greater role in U.S. cyber policy, this report represents the kind of thinking that will make it work.

Richard Andres
National War College

---

[1] "Manhattan U.S. Attorney Announces Charges Against Seven Iranians For Conducting Coordinated Campaign Of Cyber Attacks Against U.S. Financial Sector On Behalf Of Islamic Revolutionary Guard Corps-Sponsored Entities," accessed August 31, 2017, https://www.justice.gov/usao-sdny/pr/manhattan-us-attorney-announces-charges-against-seven-iranians-conducting-coordinated.

[2] "Update in Sony Investigation," accessed August 31, 2017, https://www.justice.gov/opa/pr/update-sony-investigation.

[3] "U.S. Charges Five Chinese Military Hackers for Cyber Espionage Against U.S. Corporations and a Labor Organization for Commercial Advantage," accessed August 31, 2017, https://www.justice.gov/opa/pr/us-charges-five-chinese-military-hackers-cyber-espionage-against-us-corporations-and-labor.

[4] "Background to 'Assessing Russian Activities and Intentions in Recent US Elections': The Analytic Process and Cyber Incident Attribution," accessed August 31, 2017, https://www.dni.gov/files/documents/ICA_2017_01.pdf.

Executive Summary

Throughout the 20th and early 21st centuries, deterrence and arms control have been cornerstones of strategic stability between the superpowers. However, the weaponization of the cyber realm by State actors and the multipolar nature of cyber conflict now undermines that stability. Strategic stability is the state in which nations believe that if they act aggressively to undermine U.S. national interests and the post-World War II liberal democratic order, the consequences will outweigh the benefits. The sense of lawlessness and lack of consequences in the cyber realm embolden States to be more aggressive in taking actions that undermine stability. Accordingly, this paper examines 1) the role of deterrence and arms control in securing cyber stability, and 2) the limitations and challenges associated with these traditional national security paradigms as applied to this emerging threat domain. This paper demonstrates that many 20th-century deterrence and arms control concepts are not particularly applicable in the cyber realm. However, they are not entirely irrelevant. **The United States can distill lessons learned from this rich deterrence and arms control experience to develop and deploy a strategy to advance cyber stability**.

This paper presents a tool box for the United States to enhance cyber stability based on elements of deterrence and arms control that, for reasons discussed herein, are roughly analogous to the cyber domain. The tools are not mutually exclusive and can be selected as necessary or expedient. Furthermore, **those tools can be used to develop tailored deterrence strategies for every would-be U.S. state-level adversary, based on that adversary's motivations, tolerance for discomfort, and level of cyber infrastructure.**

Key findings of this report are as follows:

- Traditional deterrence approaches are often not applicable in the cyber realm given the difficulty in attribution of cyber-attacks, the secrecy surrounding States' cyber capabilities, and the difficulty of verifying with adequate confidence a States' adherence to many potential cyber norms and obligations, or international law.
- While not perfect analogues, certain elements of traditional arms control treaties hold promise for potential use to advance cyber stability, such as
  - states agreeing to limit destabilizing behavior, as in the Incidents at Sea Agreement;
  - mechanisms for international investigation of suspected State-authorized cyber-attacks, similar to mechanisms in the nuclear test ban treaties or the Chemical Weapons Convention (CWC);
  - frameworks for escalation avoidance and for communication and information sharing, as modeled in the Incidents at Sea Agreement and the Helsinki Accords;
  - confidence-building measures analogous to those associated with the CWC, the Biological Weapons Convention (BWC), and the Helsinki Accords; and
  - limitations on development, deployment, or use of dual-use technology based on indicators of State intent, as contained in the BWC and CWC.
- A suite of tools useful for traditional strategic stability holds potential to strengthen cyber stability, such as

- enacting international agreements that leverage lessons learned from arms control treaties that might apply in the cyber realm, as outlined above;
- taking measures to clarify signaling of U.S. intent, thresholds, and retaliation to potential adversaries;
- improving detection and attribution capabilities;
- tailoring specific deterrence strategies to different adversaries rather than adopting a "one-size fits all" approach;
- naming and shaming for deterrence purposes;
- enhancing coordinated use of national criminal laws among States;
- implementing sanctions, unilaterally or multilaterally, to censure offending states;
- retaliating against cyber-attacks that exceed acceptable thresholds with "loud" cyber acts or kinetic force; and
- investing additional resources in cyber resilience and reconstitution, including with international allies and partners.

Based on our research findings, our **key recommendations for further research are as follows**:

- Conduct more detailed research and development on certain cyber stability tools identified in this report, such as technical and institutional options for international cooperation on attribution of hostile cyber actions, reinvestment in international alliances and partnerships to amplify the U.S. cyber deterrence posture, and development of capabilities and policies for U.S. forces to conduct "loud" cyber retaliatory acts against offending State actors.
- Conduct a pilot study on tailored deterrence strategies targeted toward a selected near-peer rival and a specific cyber-enabled rogue. Comparing strategies, techniques and anticipated outcomes for such divergent threat actors may yield useful insights for further development.
- Work with USG partners to develop an agreed lexicon for signaling U.S. intent, thresholds of unacceptable damage, and threatened retaliation in the cyber domain. If possible, work with select international partners to develop a common lexicon for such concepts. During the Cold War, the U.S. and the Soviets were able to develop a "common language" for signaling in the nuclear domain the strengthened strategic stability despite enormous political differences. The same may be possible for cyber.
- This report may be responsive to Executive Order 13800[1] in which, among other needs, President Trump called for studies on cyber deterrence, but additional work is clearly needed.

This report is structured as follows:

- Sections 1 through 3 introduce the topics of cyber stability and deterrence and define key terms.
- Section 4 outlines the attributes of the cyber realm and its weaponization.

[1] 2017. Statement by President Donald J. Trump on the Elevation of Cyber Command. Washington, DC: Executive Office of the President of the United States. Accessed September 1, 2017. https://www.whitehouse.gov/the-press-office/2017/08/18/statement-donald-j-trump-elevation-cyber-command.

- Section 5 reviews a non-exhaustive list of arms control agreements that may have elements applicable in the cyber realm. This includes Strategic Arms Limitation and Reduction Treaties, nuclear test ban treaties, BWC, CWC, Incidents at Sea Agreement, the Helsinki Process, and international regulation of piracy to determine the extent to which those treaties or agreements could potentially be applied in the cyber realm, as highlighted above. **While none of these agreements are not perfect analogues for cyber, certain elements from each agreement may offer useful insights for development of a regime or strategy for enhancing cyber stability.**
- Sections 6 and 7 outline a model of cyber deterrence and conclude that **traditional deterrence approaches are often not applicable in the cyber realm given States' reticence to "display" their cyber weaponry and the difficulty of attributing malicious cyber acts to a State with sufficient confidence to merit retaliation.**
- Section 8 assesses how the United States might promote cyber stability with a variety of technical and policy options and tools. The options described are not mutually exclusive and different options may be best suited for different adversaries.

Weaponization of the cyber realm is unquestionably destabilizing. Though neither traditional deterrence paradigms nor traditional approaches to arms control are perfect analogues for the cyber realm, certain elements from deterrence and arms control may nonetheless offer a rich platform of experience from which to derive useful lessons. Using such tools and lessons, the United States should carefully consider its potential adversaries, its desired outcomes, and its available options to create a robust cyber stability strategy, tailored to deter specific states of greatest concern that contributes to deterring malicious actors from attacks that exceed acceptable thresholds and encourages international cooperation in supporting a robust cyber policy infrastructure.

Acknowledgments

We, the authors, would like to thank many people for helping make this publication possible. First, we would like to thank the State Department, especially Astrid Lewis for her faith in this concept and Grant Schneider and Jaisha Wray, subject matter experts at the State Department, for generously contributing their knowledge and experience to answer questions and provide insight.

Second, we are grateful to the National War College and the National Defense University for their collaboration and expertise. We feel extremely lucky to have worked with Rich Andres, who provided critical thought leadership at this project's inception and throughout its duration. We also benefited from the guidance and knowledge of Justin Anderson, Nima Gerami, and Shane Smith, who all had a hand in shaping the final product.

Finally, we would be remiss if we did not thank Laura Denlinger. She saw value in a fleeting thought and was a driver in formulating the initial concept for this paper. Laura was our indefatigable champion throughout the project and a wise sounding board whose advise dramatically improved the overall product.

Our sincere thanks to all!

R Goychayev
GA Carr
RA Weise
DA Donnelly
SA Clements
JM Benz
KE Rodda
RA Bartholomew
AD McKinnon

Acronyms and Abbreviations

ABM	Anti-Ballistic Missile
BWC	Biological Weapons Convention
CBM	Confidence-Building Measures
CERT	Computer Emergency Response Teams
CSCE	Conference for Security and Cooperation in Europe
CTBT	Comprehensive Nuclear-Test-Ban Treaty
CWC	Chemical Weapons Convention
DHS	Department of Homeland Security
E.O.	Executive Order
GGE	Groups of government experts
ICBM	intercontinental ballistic missiles
ICT	information and communication technologies
IMS	International Monitoring System
INF	Intermediate-Range Nuclear Forces
LTBT	Limited Test Ban Treaty
MAD	Mutual Assured Destruction
MIRV	multiple independently targeted reentry vehicles
OPCW	Organisation for the Prohibition of Chemical Weapons
OSCE	Organization for Security and Cooperation in Europe
OST	Outer Space Treaty
PPD	Presidential Policy Directive
PNET	Peaceful Nuclear Explosions Treaty
PNNL	Pacific Northwest National Laboratory
SALT	Strategic Arms Limitation and Reduction Treaties
START	SALT II agreements, the Strategic Arms Reduction Treaty
TTBT	Threshold Test Ban Treaty
UNCLOS	UN Convention on the Law of the Sea
UNSC	UN Security Council
UNSCR	UN Security Council Resolution
US-CERT	U.S. Computer Emergency Response Team
USSR	Union of Soviet Socialist Republics
WMD	weapons of mass destruction

Contents

Figures

Tables

1.0 Introduction and Objectives

This paper seeks to consider the elements of deterrence theory and practice, with a particular emphasis on the role of treaty regimes during the Cold War and beyond, that may generate relevant insights for establishing a new state of cyber stability and help deter State rivals from cyber-attacks with consequences the United States would deem unacceptable. The paper reviews classical deterrence theory and its application to a series of historical examples then discusses cyberspace attributes and the stages of a cyber-attack. To understand what analogues do or do not exist between kinetic weapons and cyber weapons, the paper analyzes specific arms control precedents and the extent to which those mechanisms may apply to cyberspace. Finally, the paper outlines options for the United States to reward or punish States' good and bad behavior in cyberspace.

"A cyber-attack perpetrated by Nation States or violent extremist groups could be as destructive as the terrorist attack on 9/11," stated then-Secretary of Defense Leon Panetta in 2012. "Such a destructive cyber terrorist attack could virtually paralyze the nation."[1] States developed treaties following costly and destabilizing arms races, and those treaties became integral elements of Cold War-era strategic balance, thereby reinforcing strategic stability. As time passes and technology advances, new weapons emerge and new domains undergo the destabilizing process of weaponization. Then, as States search for geopolitical equilibrium in these new domains, they seek to establish a new balance of deterrence, often relying on arms control agreements. Thomas Schelling described this process as follows:

> A "balance of deterrence" – a situation in which the incentives on both sides to initiate war are outweighed by the disincentives – is described as "stable" when it is reasonably secure against shock, alarms and perturbations. That is, it is "stable" when political events, internal or external to the country involved, technological change, accidents, false alarms, misunderstandings, crises, limited wars, or changes in the intelligence available to both sides, are unlikely to disturb the incentives sufficiently to make mutual deterrence fail. Arms control agreements are used as the tools to minimize the impact and/or likelihood of the events, tech change, accidents, false alarms, misunderstandings, crises, and intelligence available.[2]

Just as deterrence evolved and arms control emerged as a cornerstone of deterrence following the advent of nuclear weapons, weaponization of the cyber realm is challenging strategic stability for the United States today as it has been the target of multiple cyber-attacks, and new deterrence paradigms have been slow to emerge.

This paper focuses on State actors as the primary actors as both deterrence theory and international law (arms control agreements) operate at the State level. Non-State actors, such as criminal organizations, terrorist groups, or individuals with malicious intent, play a role in the cyber realm. However, unless otherwise noted, **for the purposes of this paper, to the extent that any non-State actor is operating without direction from or affiliation with a State, it will not be considered to impact cyber stability**

---

[1] 2012. Remarks by Secretary Panetta on Cybersecurity to the Business Executives for National Security, New York City. edited by Department of Defense. Washington, DC. Accessed 17 April 2017. http://archive.defense.gov/transcripts/transcript.aspx?transcriptid=5136

[2] Thomas Schelling, *Strategy and Arms Control* (Twentieth Century Fund, 1985), 50–51.

or the balance of deterrence because their actions are "unlikely to disturb the incentives sufficiently to make mutual deterrence fail," according to Schelling's definition.[1] Where non-State actors are acting on behalf of or under the direction of a State, those proxies are presumed to be State actors, as they would be presumed under international law, and are part of the scope of this paper.

To write this paper, the project team leveraged subject matter experts from Pacific Northwest National Laboratory (PNNL) and National Defense University in a variety of technical and non-technical fields. The multidisciplinary team brings together subject matter expertise on cyber issues, arms control, and verification policy and technology; weapons of mass destruction; and international law to address the multifaceted challenges associated with stability and deterrence in cyberspace.

2.0 Definitions

Standardized definitions related to the cyber realm and hostile actions in the cyber realm are limited. This section defines key terms necessary to discuss cyber stability consistently. Additional terms are defined in the Appendix A.

Cyber realm refers to a global domain within the information environment consisting of the interdependent network of information technology infrastructures and resident data, including the internet, telecommunications networks, computer systems, and embedded processors and controllers.

Cyber stability refers to a state of relations between States characterized by the absence of serious hostile cyber actions against one another, where the States have a sufficient common understanding of each other's capabilities and intentions so as to be inclined generally to avoid such actions, likely associated with a common belief that the costs of such conduct would outweigh the benefits.[2] From the U.S. perspective, cyber stability is undermined when actors behave aggressively in the cyber realm in ways that undermine U.S. national interests and the post-World War II liberal democratic order.

Cyber weapons refers to malicious code or exploitation of vulnerabilities in cyber networks or systems through various techniques intended to cause damage.

In addition to these key terms, Sections 3 and 4 define the concepts of traditional deterrence theory and the attributes of the cyber realm, which are crucial for understanding this paper's central assessment of whether deterrence and arms control apply in cyber space.

3.0 20th Century Deterrence Theory

This section outlines deterrence theory as it has traditionally been applied to nuclear weapons and serves as the baseline for determining if deterrence has applications in the cyber realm. For much of human history, wars were fought with the expectation that the costs of fighting would be outweighed by

[1] Ibid.
[2] Internal definition used by authors of this paper.

the spoils of victory. However, the advent of nuclear weapons fundamentally transformed that expectation. The development of nuclear bombers, and later missiles, that could reliably deliver a payload ultimately made deterring conflict the focus of serious study.[1,2] The rise of cyber weapons—weapons that can damage physical systems and critical infrastructures and against which no sure defense has yet been conceived—necessitates a renewed discussion of deterrence and lends the topic new urgency.

Henry Kissinger posited that deterrence (D) is the *product* of **Capability** (C) multiplied by **Resolve** (R) multiplied by **Belief** (B); if any of those is "zero," the whole result is "zero," also referred to as a deterrence failure.[3]

$$D = C \times R \times B$$

Capability refers to the technical ability to act—the possession of the weapons systems, delivery vehicles, and command-and-control infrastructure necessary to conduct first and retaliatory strikes.[4] ***Resolve*** is the willingness to carry out a threatened action.[5] ***Belief*** is in the mind of the adversary about both one's ability and willingness to take action.[6]

Intertwined with both Resolve and Belief is the idea of ***Signaling*****,** which can be either explicit or implicit. Explicit signaling refers to public statements or direct communication to an adversary about Capability or Resolve, while implicit signaling refers to tests and demonstrations of one's capabilities. Signaling is important for making capabilities and intentions clear, thereby reducing the risk of miscalculation. A well-known though fictional example can be found in *Dr. Strangelove*. In this classic film, the Soviet Union failed to communicate that it had created its "doomsday machine," knowledge of which might have persuaded General Ripper not to order the pre-emptive strike that is the focus of the film; hence, the deterrent effect of the machine that was presumably its primary purpose was not achieved.[7]

Scholars generally agree on two major approaches to deterrence: deterrence by denial and deterrence by punishment.[8] Deterrence by denial seeks to make the adversary doubt it can achieve its goals, while deterrence by punishment seeks to make the adversary believe that achieving its goals is not worth the impending retaliation. In the Cold War, deterrence by punishment was predicated on the notion that no conceivable active anti-air defense or passive civil defense schemes (deterrence by denial) could

---

[1] Brodie, Bernard. 1946. *War in the Atomic Age*. Edited by Bernard Brodie, *The Absolute Weapon: Atomic Power and World Order*. New Haven, CT: Yale Institute of International Studies.
[2] Thomas Schelling, Strategy and Arms Control (Twentieth Century Fund, 1985).
[3] Kissinger, Henry. 1961. The Necessity for Choice: Prospects of American Foreign Policy. Lansing, MI: Doubleday.
[4] Ibid.
[5] Ibid.
[6] Nye, Joseph S. 2017. "Deterrence and Dissuasion in Cyberspace." *International Security* 41 (3):28. doi: 0.1162/ISEC_a_00266. Accessed February 6, 2017. https://www.belfercenter.org/sites/default/files/files/publication/isec_a_00266.pdf
[7] In the film the eponymous main character, upon learning the existence of a Soviet doomsday machine, shouts "The whole point of the doomsday machine is lost if you keep it a secret! Why didn't you tell the world?" to which the Soviet Ambassador responds, "It was to be announced at the party congress on Monday."
[8] Snyder, Glenn H. 1960. "Deterrence and Power." *The Journal of Conflict Resolution* 4 (2):16. Accessed March 21, 2017. http://www.jstor.org/stable/172650

adequately protect against nuclear weapons.[1] Deterrence by punishment became the center of an offense-dominated posture.[2] Concepts such as Massive Retaliation, Flexible Response, and famously Mutual Assured Destruction (MAD) became the hallmarks of deterrence by punishment. Massive Retaliation was the Eisenhower Administration's attempt to deter without, as John Foster Dulles said, "exhausting ourselves"[3] and matching the Soviet Union soldier for soldier and tank for tank; any Soviet aggression would be met with nuclear retaliation. The Kennedy Administration's Flexible Response posture resulted from the realization that a massive nuclear response was not feasible—and therefore, not credible—for smaller-scale conflict.[4,5] MAD was the outgrowth of the idea that general nuclear war lay at the end of any escalation ladder; thus, to take aggressive action was to ensure one's own destruction.

The capabilities associated with deterrence by punishment and deterrence by denial did not manifest themselves in weapons alone. Concepts of operations were developed to support the use of nuclear weapons. For example, airborne alerts prevented bombers from being caught on the ground. Radar systems were supplemented by massive phased array sets watching for inbound missiles, and eventually joined by satellites watching for the thermal bloom of a missile engine.[6] Hardened communications capabilities and airborne command posts constantly aloft ensured no decapitating strikes on civilian or military leadership could prevent a retaliatory nuclear launch.[7] Under no circumstances would any aggression remain unanswered, as even a successful surprise attack would have to contend with second-strike capabilities in Submarine Launched Ballistic Missiles (SLBMs) controlled by these airborne command posts.[8,9] The cost of acting was nuclear retaliation.

Deterrence by punishment is predicated on two interconnected elements: 1) the capability to retaliate and 2) knowing against whom or what to retaliate. Nuclear weapons and their supporting systems provided the ability to retaliate, while radar systems and space-based surveillance satellites provided the necessary physical evidence for attribution. Attribution is simplified by the exclusivity of the set of potential nuclear threat actors. If a nuclear-armed intercontinental ballistic missile (ICBM) were launched at Washington, the Union of Soviet Socialist Republics (USSR) or (possibly) China would have been nearly the only likely culprits, based on their known capabilities and postures. During the Cold War, nuclear weapons and their delivery systems were routinely and visibly tested—making clear the ability to

---

[1] Wohlstetter, Albert. 1959. "The Delicate Balance of Terror." *Foreign Affairs* 37 (2):25. Accessed January 24, 2017.
[2] Brodie, Bernard. 1946. *War in the Atomic Age*. Edited by Bernard Brodie, *The Absolute Weapon: Atomic Power and World Order*. New Haven, CT: Yale Institute of International Studies.
[3] Dulles, John Foster, "The Evolution of Foreign Policy," Before the Council of Foreign Relations, New York, N.Y., *Department of State, Press Release No. 81* (January 12, 1954).
[4] Ibid.
[5] Snyder, Glenn H. 1962. "Deterrence, Defense, and Disengagement." *World Politics* 14 (2):11. Accessed March 28, 2017. http://www.jstor.org/stable/2009305
[6] Lippold, Kirk S. 1989. "U.S. and Soviet Strategic Command and Control: Implications for a Protracted Nuclear War." Master of Science in Systems Technology (Command, Control, and Communications), Naval Postgraduate School (39).
[7] Ibid.
[8] Ibid.
[9] Nye, Joseph S. 2017. "Deterrence and Dissuasion in Cyberspace." *International Security* 41 (3):28. doi: 0.1162/ISEC_a_00266. Accessed February 6, 2017. https://www.belfercenter.org/sites/default/files/files/publication/isec_a_00266.pdf

retaliate. Radars and satellites could help determine the "return address" for any incoming nuclear attacks; thus, attribution during the Cold War did not present any particular challenge to deterrence.

However, while deterrence is generally credited with having helped to prevent a major conventional or nuclear war, nuclear deterrence did not, and was not designed to, stop all levels of conflict. The Cold War was characterized by numerous "peripheral" conflicts under which competition between superpowers played out through regional proxies. The superpowers sought to assert power and achieve advantage in third countries with reduced likelihood of direct hostility against each other and of conflict escalation. As Joseph Nye writes, deterrence "[…] is complex and involves more than just retaliation […] While second-strike capability and [MAD] may have been enough to prevent attacks on the homeland, they were never credible for issues at the low end of the spectrum of interests."[1] However, the balance of deterrence was stable enough to remain, as Schelling described, because the lower levels of conflict were "unlikely to disturb the incentives sufficiently to make mutual deterrence fail."[2]

Bipolar deterrence as a paradigm for defense and international relations may have prevented nuclear holocaust, but the concept had considerable shortcomings, including Schelling's assumption that both sides correctly understood the incentives and disincentives to their actions. Balanced deterrence was predicated on presumptions of mutual comprehension, effective control, and attribution, which were, at best, highly imperfect. Accident, mistake, miscalculation, and miscommunication could and did occur. Schlosser and Hoffman, for example, offer alarming histories of accidents, incidents, and misperceptions in the superpowers' respective weapons complexes and decision-making structures during the Cold War.[3]

Under deterrence theory, tools of signaling and strategic communications are used to influence behavior. Unfortunately signaling can be imperfect, particularly when multiple threat actors are reading (and misreading) the signals or missing them altogether, resulting in flawed conclusions about how an adversary will act. Uncertainty about the size, structure, and capabilities of the adversary's nuclear forces ultimately proved destabilizing, and the consequent risk of a quick slide into nuclear Armageddon incentivized the world's two nuclear superpowers to enter into treaties and agreements. Such agreements limited or froze capabilities, clarified resolve, verified compliance, detailed red lines necessary for accurate signaling and credible belief, and prevented misperceptions and miscalculations from turning into nuclear conflict.[4]

4.0 Attributes of the Cyber Realm

To apply deterrence concepts to the modern cyber threat landscape, common concepts and terms describe the cyber realm. The cyber realm is a superset of multiple components. For purposes of this paper, the cyber realm comprises four distinct categories:

---

[1] Nye – Ibid.
[2] Schelling, Thomas. 1960. *The Strategy of Conflict*. Cambridge, MA: Harvard University.
[3] Hoffman, David E. 2009. *The Dead Hand*. New York, NY: Anchor Books.; Schlosser, Eric. 2013. *Command and Control: Nuclear Weapons, the Damascus Accident, and the Illusion of Safety*. New York, NY: Penguin Books.
[4] See, e.g., Agreement Between the Government of The United States of America and the Government of The Soviet Socialist Republics on the Prevention of Incidents On and Over the High Seas. 1972. https://www.state.gov/t/isn/4791.htm.

1. The cyber domain
2. Cyber weaponry
3. Attacking the cyber domain
4. Abusing the cyber domain.

Together, these attributes make up the cyber realm as referenced throughout this paper.

4.1 The Cyber Domain

Rattray and Healey characterize the cyber domain as having five key attributes.[1] The cyber domain is

1. logical but physical;
2. usually used, owned, and controlled predominately by the private sector;
3. tactically fast but operationally slow;
4. a domain in which the offense generally dominates defense; and
5. fraught with uncertainty.

Logical but physical. Hoffman writes of the cyber domain, "Unlike the land, sea, air and space where the laws of physics do not change, cyberspace is a man-made creation that continually changes and evolves."[2] The cyber domain features routers, switches, wires, and other digital components that are physical devices generally located within States' borders and subject to State jurisdiction. The cyber domain relies on standardized protocols that permit information to transit through the physical equipment. Additionally, certain aspects of the cyber domain, such as open-source code, have no clear ownership. Virtually any computing device that can store or process digital information can be considered to be part of the cyber domain. While the internet and its associated infrastructure and protocols have become a major medium for connectivity and transmission of information among such devices, it does not fully delimit the cyber domain. Any computing or digital device, regardless of its connectivity to other devices, can be considered as within the domain and can be subject to attack or accessed in an unauthorized manner.

Usually used, owned, and controlled predominately by the private sector but fall within the jurisdiction of one State or another. Most physical infrastructure that makes up the cyber realm is privately owned, used predominantly by private citizens or corporations, and non-State entities play a significant role in setting the standards that define the underlying protocols. While much of the infrastructure is owned by the private sector, Nation States nonetheless exercise their sovereignty over the physical elements of the cyber realm (i.e., servers, computers, and computer users) that are within their jurisdiction. Just as most buildings in a country can be privately owned, States can still regulate those

[1] Rattray, George, and Jason Healey. Categorizing and Understanding Offensive Cyber Capabilities and Their Use. National Academy of Sciences: Proceedings of a Workshop on Deterring Cyberattacks: Informing Strategies and Developing Options for U.S. Policy.
[2] Department of the Air Force. Cornerstones of Information Warfare, Department of Defense. Washington DC, 1995 available at http://www.iwar.org.uk/iwar/resources/usaf/iw/corner.html, accessed May 30, 2017.

buildings and permit or prohibit activities occurring within the buildings.[1] Moreover, while non-governmental organizations may be developing standards for operations in the cyber realm, the cyber realm is not a global commons like the high seas. When in the high seas, no State has jurisdiction. However, in cyberspace, typically at least one State—and often multiple States—may be able to assert its law over particular acts or persons based on the nationalities or locations of persons, instrumentalities (e.g., servers, Ethernet lines, or computers that originate content, to name a few), or effects related to the actions in question.

Tactically fast but operationally slow. The visible strike of an attack in the cyber domain is fast and the effects can be sudden; however, the planning and organization needed to create a precision effect demand significant time and resources. For example, the initial phishing email that established a toehold on the Ukrainian electric utilities networks came in the spring of 2015; the attackers then proceeded to map the networks, elevate access, and stage their payloads before finally executing the true attack of shutting off the power in December of 2015.[2]

The offense generally dominates the defense. The advantage in the cyber domain, as in the nuclear domain, favors the offense. A defender must block all attacks, while the attacker only needs to be successful once. At the same time, unlike in the nuclear domain where the arsenal is finite and costly and each weapon can be used only once, there is little to deter an attacker from mounting a continual barrage of cyber-attacks given the low cost of doing so.

Fraught with uncertainty. Cyberspace is a very complex environment prone to rapid change, adaptation, and unpredictability. At its core, cyberspace is an amalgamation of globally common technologies, protocols, and interconnections engineered to enable robust communication channels that are resistant to disruptions. Interestingly, cyberspace has its initial foundations in U.S. efforts to develop a command-and-control communications network that would still function in the event of nuclear war. Today, the global internet and similar networks function by separating information into small data packets that are routed from source to destination without necessarily following the same path, then reassembled for processing at the end. This architecture provides an inherent resiliency to disruption, outages, and misconfigurations. At the same time the extreme complexity, frequent change, and unpredictability of the cyber domain make it extremely difficult to fully understand the state of the environment.

4.2 Cyber Weaponry

Defining a cyber weapon is not as straightforward as defining a nuclear weapon—while nuclear weapons have one purpose, many cyber "weapons" have legitimate dual-use purposes. **Examining the anatomy of a cyber-attack through the lens of the Cyber Kill Chain[3] provides illustrative examples**.

---

[1] "Group of Governmental Experts on Developments in the Field of Information and Telecommunications in the Context of International Security," June 24, 2013, http://www.unidir.org/files/medias/pdfs/developments-in-the-field-of-information-and-telecommunications-in-the-context-of-international-security-2012-2013-a-68-98-eng-0-518.pdf.

[2] "Analysis of the Cyber Attack on the Ukrainian Power Grid," March 18, 2016, https://ics.sans.org/media/E-ISAC_SANS_Ukraine_DUC_5.pdf.

[3] Hutchins, Eric, Michael Cloppert, and Rohan Amin. "Intelligence-Driven Computer Network Defense Informed by Analysis of Adversary Campaigns and Intrusion Kill Chains," accessed August 31, 2017, http://www.lockheedmartin.com/content/dam/lockheed/data/corporate/documents/LM-White-Paper-Intel-Driven-Defense.pdf.

The kill chain concept delineates the series of events that must occur prior to a desired action. The concept can be used either to structure an attack or plan a defense. There are **seven stages to a cyber-attack** as defined in the Cyber Kill Chain (see Figure 1.

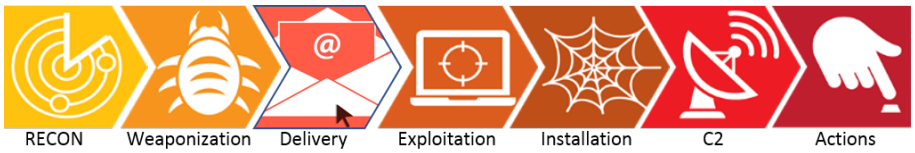

Figure 1. Stages to a Cyberattack (as defined in the Cyber Kill Chain)

- Stage 1 **Reconnaissance**: Similar to conventional warfare, adversaries must learn about their target prior to attack. While reconnaissance techniques in the cyber realm may differ, adversaries will still use any available information at their disposal to prepare for an attack. In the physical world, latitude and longitude coordinates, physical protection systems used, number and height of fences, and other defensive details may be of use. In the cyber realm, the corollary would be domain names, IP addresses, networking equipment providers, software and version, trusted partners, make, model and version number of digital hardware, and cyber defensive tool providers. This information might be gleaned from directly scanning the target's IP space, its public web presence, employee résumés, job postings, social media, reports on mergers and acquisitions, environmental impact studies, reports on regulatory compliance, search engines, and marketing material from companies that have worked at the target location. In the United States, **sunshine laws** often require the public availability of information that is very valuable to attackers.[1]
- Stage 2 **Weaponization:** Knowing what systems are running and the architecture of the target, an attacker will craft malicious software to exploit a vulnerability or otherwise help an attacker gain initial access into a target environment. Development and testing of this malware (i.e., cyber weapon) is most often done on the attacker's systems and is undetectable to the victim.
- Stage 3 **Delivery:** Like nuclear warheads, cyber "payloads" must be delivered. A variety of common methods can deliver a cyber weapon (i.e., to get the malware onto the target's system or device) such as direct attacks on target infrastructure, malicious emails/phishing, USB thumb drives, social media links, or compromised websites. The delivery stage is the first opportunity for the target's defenders (e.g., virus scanning software, or cyber security team) to truly engage with attackers and observe their tools, tactics, and procedures. The target can learn about its attackers at this stage by analyzing its logs and the adversarial targeting activities.
- Stage 4 **Exploitation:** Exploitation occurs when the malware's code is triggered, which then begins to exploit vulnerabilities in the target. In the cyber realm, not all weapons detonate on contact. Most are only successful against a very small subset of the devices or software in the cyber realm. Successful execution of the weaponized software establishes the initial access to the victim's environment.
- Stage 5 **Installation:** This stage is also referred to as Persistence or Lateral Movement**.** The malware installs an access point ("backdoor") for attackers to move throughout the environment, seeking specific information or looking to attain persistent access.

[1] Government in the Sunshine Act 5 U.S.C. 552b …every portion of every meeting of an agency shall be open to the public http://accessreports.com/statutes/sunshine.htm

- Stage 6 **Command and Control**: At this stage, the malware enables the attackers to have persistent access to the target's network or device. The uncertainty of operating in the cyber environment requires that most cyber weapons contain some sort of communication mechanism to interact with the attacker.
- Stage 7 **Actions on Objectives/Target:** The specific activities of the final stage can vary widely, but this stage is when the attackers attempt to achieve their goals. Traditional cyber-attack objectives might be to quietly exfiltrate information over long periods of time or run a "smash and grab," taking whatever can be found but not hiding the fact that intruders are present. Other actions may be website defacement to establish credibility or to embarrass the target. Destruction of software or hardware, staging for future activities, or creating real world kinetic effects by manipulating control systems are also potential actions.

Numerous high-profile attacks illustrate the kill chain process. For example, the Office of Personnel Management breach exposed the background check material on millions of U.S. citizens,[1] and the hack on the Ukrainian power grid knocked out power to over 200,000 Ukrainians in December 2015,[2] both illustrate the multi-stage approach to compromising and performing varying actions on target.

Many attempts have been made to categorize malicious cyber activities. For example, the Open Web Application Security Project's categorizes attacks based on the types of techniques used,[3] while the Department of Defense categorizes attacks based on the effect or intent of the operation.[4]

Two broad categories merit additional discussion in defining hostile actions and attacker may take to achieve a given effect or purpose: 1) attacking the cyber domain and 2) abusing the cyber domain.

4.3 Attacking the Cyber Realm

Attacks on the cyber domain are the most commonly referred to cyber-attacks. **Attacks on the cyber domain are activities that take advantage of weaknesses, vulnerabilities, or misconfigurations of hardware or software.** These attacks are used to steal information, destroy data, or compromise the confidentiality, integrity, or availability of the digital systems.

4.4 Abusing the Cyber Realm

Cyber-attacks that abuse the cyber domain do not technically compromise the underlying infrastructure, nor do they use unauthorized access (though they may violate terms of service agreements or de facto norms of behavior). Instead, **these attacks abuse the systems to achieve their objectives of**

---

[1] "The-OPM-Data-Breach-How-the-Government-Jeopardized-Our-National-Security-for-More-than-a-Generation.pdf," accessed August 31, 2017, https://oversight.house.gov/wp-content/uploads/2016/09/The-OPM-Data-Breach-How-the-Government-Jeopardized-Our-National-Security-for-More-than-a-Generation.pdf.
[2] "Analysis of the Cyber Attack on the Ukrainian Power Grid," March 18, 2016, https://ics.sans.org/media/E-ISAC_SANS_Ukraine_DUC_5.pdf
[3] "Category:Attack - OWASP," accessed August 31, 2017, https://www.owasp.org/index.php/Category:Attack.
[4] "Department of Defense Dictionary of Military and Associated Terms," accessed January 23, 2017, http://www.dtic.mil/doctrine/new_pubs/jp1_02.pdf

misinformation, fake news, or other propaganda by preying upon human nature and trusted relationships. The broad acceptance and use of social media has given the age-old tactic of information warfare new legs and an interesting twist.

In the cyber realm, there is an ongoing conflict as States both attack and abuse the cyber realm. The next sections consider whether arms control or traditional deterrence concepts can dissuade State from such behavior.

5.0 20th Century Arms Control Mechanisms and Their Applicability to Cyber Deterrence and Stability

Numerous bilateral and multilateral agreements have been adopted outside the cyber realm to strengthen strategic stability by limiting and balancing certain military capabilities, improving clarity in signaling, reducing miscommunications, and minimizing risks of unnecessary escalation. **This section looks at several of these agreements to illuminate which elements of those agreements were successful and to what extent the same principles might be relevant to the cyber realm**. The authors did not strive to conduct an exhaustive review of historical international arms control instruments, which would extend back many hundreds of years and could cover dozens of treaties.[1] Rather, this report focuses on treaties relevant to the analysis of how a user might promote strategic cyber stability. Specifically, treaties selected for this report either

1. are considered to have "worked" in terms of stabilizing nuclear arms races;
2. dealt with categories of arms (such as chemical and biological weapons or piracy) that seemed analogous in some regard to cyber weaponry in their verification and monitoring challenges;
3. regulated issues relevant to hostile cyber activity, such as incident prevention, escalation management, and reducing disinformation; and/or
4. regulated State behavior in ungoverned or poorly governed spaces.

This section outlines the context and rationale for adopting each selected treaty, summarizes the key operating principles for that treaty, and assesses the extent to which those principles are relevant or could be cross-applied to the cyber realm. In addition, the authors reviewed certain additional treaty regimes, such as the Intermediate-Range Nuclear Forces Treaty, the Outer Space Treaty, the UN Convention on the Law of the Sea (UNCLOS), and UN Security Council Resolution 1540, which were deemed to have negligible applicability to cyber deterrence and stability. These included in detail in Appendix B.

Figure 2 **provides a high-level summary of the analysis of the authors by identifying major attributes of specific treaties or agreements and their potential relevance as a model for cyber stability**. A more thorough definition of the agreement attributes is available in Appendix C and the analysis of cyber relevance or lack thereof can be found in the subsections below.

[1] Thomas Graham, *Disarmament Sketches: Three Decades of Arms Control and International Law* (Seattle: Institute for Global and Regional Security Studies [u.a.], 2002), 34.

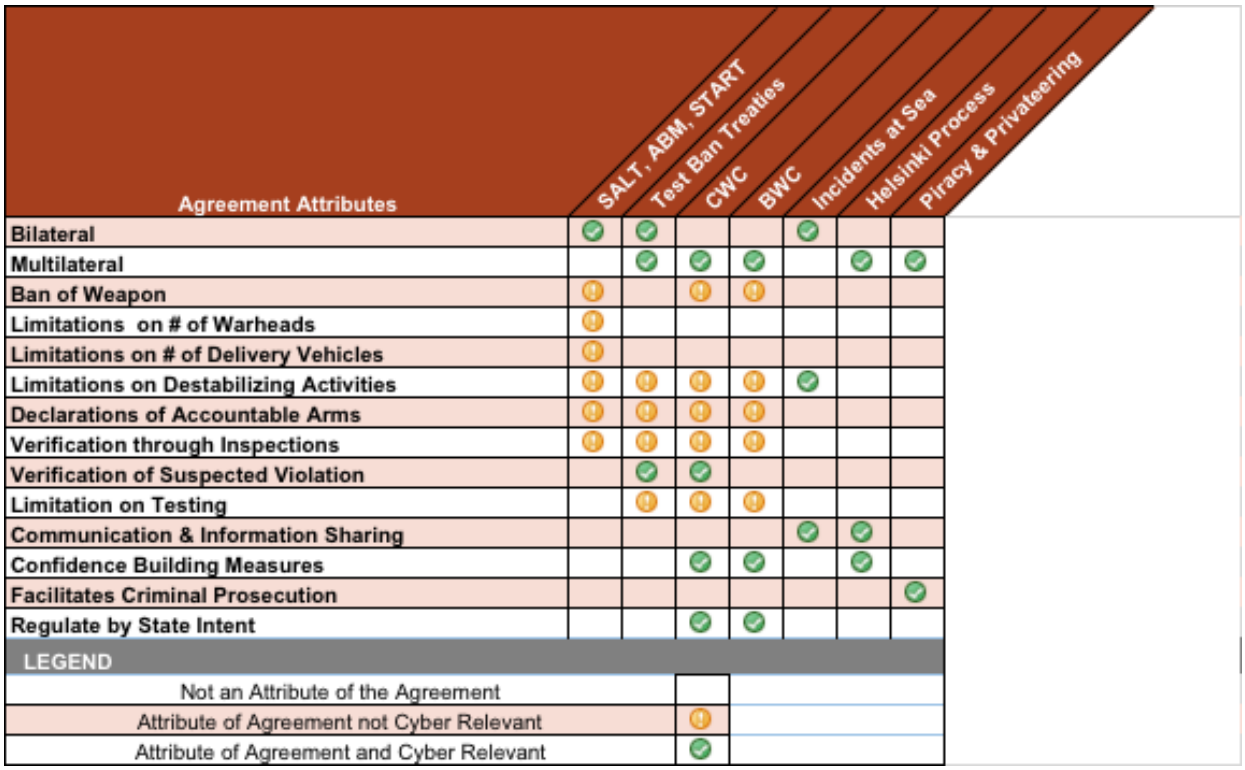

Figure 2. Summary of Section 6 Analysis

5.1 Strategic Arms Limitation and Reduction Treaties: SALT, ABM, START

Through the 1950s and 1960s, the international community grew progressively more alarmed over the sharply escalating U.S.-USSR nuclear arms race, which seemed to be producing far larger nuclear arsenals than appeared reasonably necessary for deterring or waging war. This spiraling growth in stockpiles was dangerously destabilizing, elevating the likelihood of loss of command and control, or of accident, misunderstanding, or miscalculation. The United States and Soviet Union had also experienced multiple serious confrontations and escalations with a nuclear dimension, such as the Suez, Berlin, and Cuban Missile crises. In 1969, with the impending entry into force of the Nuclear Non-Proliferation Treaty, the United States and Soviet Union would shortly be obligated to undertake good faith negotiations to cease the nuclear arms race and begin the process of nuclear disarmament. These and other factors led both countries to formally begin the Strategic Arms Limitation Talks (SALT).

The SALT process led to several agreements—including the Anti-Ballistic Missile (ABM) Treaty, the SALT I Interim Agreement, SALT II agreements, the Strategic Arms Reduction Treaty (START), START II, and New START, and the Strategic Offensive Reduction Treaty[1]—that were intended to limit numbers of armaments and specific capabilities that both sides agreed were potentially destabilizing.

[1] Interim Agreement Between The United States of America and The USSR on Certain Measures With Respect to the Limitation of Strategic Offensive Arms (Interim Agreement) (signed May 26, 1972); Treaty Between The United States of America and the USSR on The Limitation of ABM (ABM Treaty) (signed May 26, 1972); Treaty Between The United States of America and the Union of Soviet Socialist Republics on the Limitation of Strategic

5.1.1 Principles of the Agreement

Bilateral format. Cold War competition between the United States and USSR (later Russia), while dangerous, took place in a relatively stable bipolar framework, meaning each side could focus the bulk of its deterrent strategy on one primary adversary that was using similar tools for signaling and defense. A bilateral format for negotiations provided a workable and effective path for achieving significant progress on strategic arms limitation and reduction.

Numeric limitations on delivery vehicles. Both the SALT and START treaties relied, and in the case of New START rely today, upon agreed conventions for counting warheads based upon the number and type of deployed delivery vehicles, since warheads are too technically challenging and sensitive for the other side to count directly. Delivery vehicles (e.g., ICBMs, SLBMs, heavy bombers) can be verified without compromising significant sensitive information on design, performance and technology (particularly relating to the highly sensitive nuclear explosives that they carry). These delivery vehicles are large enough to be counted from some distance, or even using so-called "national technical means" (space-based surveillance). Reduced to a general principle, SALT's success was in part due to an emphasis on means of delivery of a weapon payload that are relatively less sensitive than counting warheads.

Numeric limitations on warheads. Even though SALT and START focused on counting deployed strategic delivery vehicles for verification purposes, the treaties have also limited the total number of deployed strategic nuclear warheads.[1] Even today, the United States and Russia continue to view direct counting of warheads (as opposed to delivery vehicles) as posing too great a risk for the disclosure of sensitive warhead design information to be utilized for current verification purposes. For this reason, warhead verification is measured indirectly through verification of numbers of deployed strategic delivery vehicles.

Limitations on destabilizing capabilities. The ABM and the START treaties imposed limits on the number and/or type of strategic offensive and defensive systems the United States and USSR could develop and deploy that could undermine the strategic balance. The ABM Treaty prevented both the United States and Soviet Union from fielding comprehensive defenses against strategic nuclear attack that might allow one side to execute a first strike and successfully defend against an adversary's retaliation. START I included certain restrictions on mobile ICBMs and multiple independently targeted reentry vehicles (MIRVs), which were thought to reduce the deterrent value of the adversary's forces or

---

Offensive Arms (SALT II) (signed June 18, 1979); Treaty Between the United States of America and the USSR on the Reduction and Limitation of Strategic Offensive Arms (START) (signed July 31, 1991); Treaty Between the United States of America and the Russian Federation on Further Reduction and Limitation of Strategic Offensive Arms (START II) (signed Jan. 3, 1993); Treaty Between the United States of America and the Russian Federation On Strategic Offensive Reductions (Moscow Treaty) (signed May 24, 2002); Treaty between the United States of America and the Russian Federation on Measures for the Further Reduction and Limitation of Strategic Offensive Arms (New START) (signed Feb. 5, 2011).

[1] START I and Strategic Offensive Reduction Treaty imposed certain numeric limits with no verification measures. START II (if enacted) would have required the sides to designate the number of warheads assigned to individual missiles but did not authorize counting of warheads on all missiles subject to inspection. The sides may conduct ten random checks per year of either one missile or three bombers to verify on-vehicle warhead counts. This was intended to be a visual inspection showing only the presence or absence of warhead housing, with no technical means at that time of verifying whether the housing contained an actual warhead.

destabilize the strategic balance in a system that relied on restrictions on quantities of delivery vehicles. START II[1] continued that trend and was designed to "de-MIRV" both countries' nuclear forces.[2] In its preambular text, New START[3] at the Russian Federation's insistence acknowledges a connection between strategic offense and strategic defense, highlighting that defenses will likely be a continued subject of negotiation between the United States and Russia.

Mutual declarations of accountable arms or other systems. An integral element of many arms control regimes is data declarations that require each party to demonstrate a degree of transparency, i.e., they must declare an inventory that can then (in most cases) be verified. The completeness and accuracy of the declaration are the key concerns for the parties to be confident of compliance.

Mutual verification of declarations through inspections. An integral element of the arms limitation and reduction treaties was verifying that balanced reliability with protection of sensitive information. Inspection under mutually agreed modalities proved a viable mechanism, allowing a prescribed number of on-site inspections or verification visits to relevant sites to confirm that countries' declarations were accurate and within the levels permitted by the treaty. The modalities of inspections were designed to protect potentially sensitive information that would not affect the ability of the inspection team to arrive at a correct assessment of treaty compliance. Certain frequently discussed techniques that are a focus of research for future strategic arms limitation include "managed access," through which the access to physical spaces afforded to inspection teams is strictly controlled, and "information barriers," which mediate the inspection of potentially sensitive items or activities in order to remove sensitive information but provide correct and authenticatable intended measurements.[4]

5.1.2 Applicability in the Cyber Realm

Many of the previously described characteristics of the SALT, START, and related treaty processes do not lend themselves for application in the cyber domain. Almost no aspects of hostile cyber actions are comparable to those that supported the adoption of the SALT. The specifics of those aspects and their imperfect application to the cyber domain are explained below.

Attributes that Merit Further Consideration

Bilateral format. The current international landscape of actors and capabilities in cyberspace is far more diverse than nuclear competition in the Cold War. This polycentric threat environment includes not only near-peer rivals Russia and China, but also rogue States (e.g., North Korea) and non-State actors (e.g., WikiLeaks) whose capabilities and assets may permit them to go toe-to-toe in the cyber domain with the great powers and punch far above their weight in the cyber domain. As such, a bilateral framework for cyber deterrence and stability between two superpowers as an international stabilizing

[1] Treaty Between the United States of America and the Russian Federation on Further Reduction and Limitation of Strategic Offensive Arms (START II) (signed January 3, 1993).
[2] The United States renounced the ABM Treaty in 2002; Russia renounced START II in response.
[3] The Treaty between the United States of America and the Russian Federation on Measures for the Further Reduction and Limitation of Strategic Offensive Arms (New START) (entered into force 5 Feb. 2011).
[4] See, e.g., U.S. Department of Energy, National Nuclear Security Administration & UK Ministry of Defence, Atomic Weapons Establishment, "Joint U.S.-U.K. Report on Technical Cooperation for Arms Control" (2015).

force would appear at first glance to be generally less effective than a multilateral framework open to universal membership. However, a multilateral format commensurately increases the difficulty in achieving consensus. A single bilateral agreement between States can be stabilizing,[1] and a network of bilateral agreements between multiple State parties can help generate international practice and norms that may otherwise be prohibitively difficult to codify through multilateral agreements.[2] **The bilateral format modeled by SALT and related agreements could thus be useful for promoting cyber stability, particularly if pursued individually with multiple partners, perhaps to lay the groundwork for a future multilateral agreement.**

Attributes with Limited to No Cyber Applicability

Emphasis on means of delivery. Any pathway for unauthorized access to a computer system or data in theory constitutes a means for "delivery" of a malicious "payload" or means of extracting or manipulating data. To illustrate, pathways for exploitation include any connection to a device—any network connection (internet, local area network or other), physical access, wireless (WiFi, Bluetooth, Cellular), or digital media (discs, flash drives) as well as applications used on the device, such as web browsers, email clients, document readers or media players. Additionally, manipulating the device's environment (e.g., using power supply or modifying the HVAC controls), exploiting peripheral devices (keyboards, mice, printers), and the supply chain of parts comprising a system can all have adverse impacts to the system. The universe of such "cyber delivery" methods is vast, in fact nearly infinite, and constantly evolving with technology.

Although analysis may reveal "chokepoints" at national boundaries or administrative/structural key nodes in the structure of the internet where some controls or restrictions might be imposed, the effectiveness and utility of such an approach would arguably be quite limited, while the costs could be high. Several countries engage in content and traffic filtration over their national segments of the internet,[3] which suggests some potential for border-focused capabilities to disrupt an in-progress attack that relied on the internet (such as "unplugging" or severing traffic flows believed to be part of an exploit). But traffic flows at borders are quite high, so it would be difficult to distinguish malicious from legitimate traffic. Moreover, the nature of internet protocols is such that discrete data communications between two points are separated into small data packets that can be routed by diverse pathways between the source and destination. Hence, **to reliably cut off malicious traffic would likely require severing all pathways between the attacker and target. Such restrictions to all traffic could cause significant collateral harm within a country**.

[1] "FACT SHEET: U.S.-Russian Cooperation on Information and Communications Technology Security," *Whitehouse.gov*, June 17, 2013, https://obamawhitehouse.archives.gov/the-press-office/2013/06/17/fact-sheet-us-russian-cooperation-information-and-communications-technol. (discussing how the United States and Russia collaboration in CBMs, including use of the Nuclear Risk Reduction Center to support information exchange about cybersecurity incidents of national concern).

[2] "Are BITs Representing the 'New' Customary International Law in International Investment Law? By Patrick Dumberry: SSRN," accessed August 31, 2017, https://papers.ssrn.com/sol3/papers.cfm?abstract_id=1666218. (Arguing that the extensive network of bilateral investment treaties is contributing to the consolidation and crystallization of rules of customary international law).

[3] Deibert, Robert J. "Black Code: Censorship, Surveillance, and the Militarisation of Cyberspace," *Millennium* 32, no. 3 (December 1, 2003): 501–30, doi:10.1177/03058298030320030801. (Noting observed national-level content and traffic controls in China, Saudi Arabia, Bahrain, Jordan, Syria, Tunisia, Pakistan, the United Arab Emirates, Yemen, Malaysia, Singapore, Vietnam, Myanmar and other States).

In theory, States could exercise controls to reduce the number of such pathways and impose additional layers of checks and authentication. Such "traffic restrictions" would generally be more effective and cause less collateral damage if they could be concentrated on the origin or intended target of an attack. However, such restrictions are likely to be situation-dependent, making it difficult to agree to limits on pathways between States in advance of an attack. **Unlike restrictions on strategic delivery vehicles for nuclear weaponry, which mostly impact military infrastructure, restrictions on internet traffic would impose burdens on all legitimate civilian traffic and infrastructure.** Additionally, internet traffic controls could also likely be spoofed or subverted. Skillful attackers generally can mask their purpose and point of origin, launch an attack from within a country, or attack without using the internet at all (such as by using infected USB sticks).

To summarize, this multidisciplinary team was not able to identify a finite set of cyber delivery "chokepoints" analogous to strategic delivery vehicles that are finite in number, physically visible, and tangible objects whose existence and location are amenable to verification by treaty parties. Some controls might be devised for malicious internet traffic, but these controls would have to be superimposed on a highly diverse and developed civilian information infrastructure environment, likely at great political and economic cost, and even still are unlikely to be effective.

Numeric limitations on "warheads." No physical "warheads" exist in the cyber realm, hence a more general and applicable description might be "**payload controls**." Understanding is limited of what could constitute an accountable or controllable cyber "warhead" or "payload" that might be meaningfully "countable." Malware and exploits rarely involve something analogous to a "weapon" that can reliably inflict physical damage and, as such, be displayed without disclosing information that could easily be used to render the asset inert or ineffectual. The utility of general malware or a tailored system exploit derives largely from an adversary's ignorance of a vulnerability or functionality of their system. **Since it is often possible to shore up vulnerabilities or take countermeasures at relatively little cost once a malicious approach is known, a cyber weapon-possessor has virtually no interest in disclosing any information about the existence or characteristics of its cyber weapons**. Even if countries did declare their cyber "weapons," given the lightning pace of technical advance in this field, such declarations would provide little stabilizing value, particularly for long-term stability.

States cannot be certain they are shown a complete accounting as code and data can easily be developed, hidden and altered, and are, for practical purposes, infinitely replicable. Whereas nuclear weapons manufacturing capacity and production can be tracked to some degree using tools like inspections and national technical means, malicious coding requires no comparable or trackable signature infrastructure. Information operations (disinformation, propaganda) that exploit computer systems are not analogous to physical weapons – they are designed to affect perceptions and, again, cannot be disclosed to an adversary without depriving them of their effect. For this reason, **a cyber treaty designed around the arms control concept of limited number of "weapons" or "payload" is unlikely to be successful**.

Limitations on destabilizing capabilities. Unlike the consensus that emerged in the 20th century regarding the destabilizing effects of MIRV technology or strategic missile defense, States do not have a common analogous understanding of a discrete cyber capability that could destabilize the cyber domain. Even if such capabilities existed, States would be reluctant to discuss them for fear of tipping their hand. The range of States' cyber weapon capabilities is not generally well-understood. In an environment of hidden capabilities, describing any "stability" is difficult; thus no particular capability can be readily characterized as "stabilizing" or "destabilizing." Such clarity is not likely to emerge anytime

soon. These circumstances generally do not appear conducive to entering into mutually agreed constraints on destabilizing capabilities.

Rather than focus on destabilizing aspects of specific cyber capabilities, a subset of concerned States have been more prepared to discuss norms relating to targeting.[1] In other words, **States are unlikely to agree to restricting inherently destabilizing cyber *capabilities*, but may be more amenable to agreements limiting States' destabilizing *actions***. Agreeing not to target critical infrastructure or nuclear forces would be examples of such a "code of conduct" approach. This is explored further in Sections 7, 8, and 9.

Mutual verification through inspections. As noted, there are unlikely to be meaningful, "inspectable" controls having to do with cyber weapons themselves. Nevertheless, the techniques employed for verification inspections in strategic arms control might be useful for other aspects of a cyber conflict control regime. As will be explored further in Section 5.2, international inspections to attribute banned attacks might be an element of such a regime. Using inspection modalities – such as managed access and information barriers – that would ease the inspection burden and risks for participating States would make such a regime more feasible.

5.2 Test Ban Treaties (Limited, Threshold, Comprehensive, Peaceful Nuclear Explosions Treaties)

In 1952 and 1953, respectively, the United States and Soviet Union tested their first hydrogen bombs. Over the succeeding years, the two countries conducted numerous additional tests of greater and greater magnitude. This increased the international alarm over the U.S.-USSR nuclear arms race and mounting environmental and public health concerns about atmospheric contamination due to the increasing frequency and yield of atmospheric testing. These concerns came to a head in 1954 when the United States tested what was supposed to be an 8-megaton bomb, but the resulting yield was nearly double that. It destroyed an atoll and contaminated a Japanese fishing boat and its crew.[2] This event prompted the United States, USSR, and the international community to consider limitations on nuclear weapon testing.

5.2.1 Principles of the Agreements

The discussions over limiting nuclear weapon tests led to several treaties on limiting the type and specific capabilities of tests. Summarized in brief, their general principles of limitations and verification include:

Multilateral format. The global impact of nuclear weapon tests meant that the entire international community had a stake in the outcome of the negotiations. Test ban negotiations began within a subcommittee of the UN Disarmament Commission in 1955.[3] While the main parties included the four

[1] "2015 UN GGE Report: Major Players Recommending Norms of Behaviour, Highlighting Aspects of International Law," *CCDCOE*, August 31, 2015, https://www.ccdcoe.org/2015-un-gge-report-major-players-recommending-norms-behaviour-highlighting-aspects-international-l-0.
[2] "Treaty Banning Nuclear Weapon Tests in the Atmosphere, in Outer Space and Under Water," *U.S. Department of State*, accessed August 31, 2017, http://www.state.gov/t/isn/4797.htm.
[3] Ibid.

nuclear weapon States at the time (United States, United Kingdom, USSR, and France), other Member States participated and helped maintain negotiating momentum.[1]

Limitations on types of tests. The first treaty to be negotiated was the Limited Test Ban Treaty (LTBT). It prohibited the testing of nuclear weapons in the atmosphere, in outer space, or under water, and prohibited testing of any type that would cause fallout or debris to extend into other countries. Significant discussion occurred over the course of negotiating the LTBT regarding verification. In the end, the LTBT did not include any verification mechanisms as the consensus was that verification could be confirmed through existing national technical means.

Limitations on testing thresholds. In 1974, the Threshold Test Ban Treaty (TTBT) was signed between the United States and the Soviet Union, limiting the size of underground nuclear weapon tests to at or below 150 kilotons. The treaty was strategically significant because it "remove[d] the possibility of testing new or existing nuclear weapons going beyond the fractional-megaton range….Of particular significance was the relationship between explosive power of reliable, tested warheads and first-strike capability."[2] The TTBT included a verification protocol to exchange detailed data regarding testing areas, coordinates of tests, and exchange of data on a number of actual tests to allow each party to calibrate their respective national technical means.[3] Given the technical difficulties of predicting yields at the time, the TTBT included statements noting that mistakes might occur where tests were larger than the threshold limits, but one or two minor breaches of the threshold yield would not immediately constitute a violation.[4] As of this writing, the Comprehensive Nuclear-Test-Ban Treaty (CTBT) (opened for signature in 1996) has not been ratified by the U.S. Senate nor entered into force. Once entered into force, the CTBT would go beyond previous test ban treaties to prohibit all forms of nuclear testing. To verify the CTBT, the international community developed and deployed the International Monitoring System (IMS) to continually monitor for characteristics of nuclear weapons tests. The IMS comprises hundreds of stations throughout the world to detect seismic, infrasound, hydro acoustic, and radionuclide indicators of a nuclear weapon test.[5]

Post-incident forensic analysis by a neutral international body. To verify compliance by State Parties, the CTBT establishes the Comprehensive Test Ban Treaty Organization and endows it with important capabilities and authorities. Above all, these are the IMS (described above) and the CTBTO's authority to perform certain on-site inspections.[6] Both of these are directed to create a neutral international technical capability to interrogate suspected violations of the treaty's ban on testing – first through detecting test nuclear explosions, and then providing for a means to dispatch expert teams to the suspected test site to collect measurements to assess whether a nuclear explosion took place. Under

---

[1] Ibid.
[2] "Treaty Between the United States of America and the Union of Soviet Socialist Republics on the Limitation of Underground Nuclear Weapon Tests," *U.S. Department of State*, accessed August 31, 2017, http://www.state.gov/t/isn/5204.htm.
[3] Ibid.
[4] "Treaty Between The United States of America and The Union of Soviet Socialist Republics on the Limitation of Underground Nuclear Weapon Tests (and Protocol Thereto) (TTBT)," *U.S. Department of State*, accessed August 31, 2017, http://www.state.gov/t/isn/5204.htm.
[5] "Comprehensive Nuclear Test-Ban Treaty (CTBT)," *U.S. Department of State*, accessed August 31, 2017, http://www.state.gov/t/avc/trty/16411.htm.
[6] Ibid., Art. IV (B.),(D.); Protocol, Parts I,II.

CTBT, State Parties may not refuse inspections, however an inspector team can only be dispatched upon the request of a State Party based on information from the IMS or national technical means, and only for the purpose of assessing compliance with the CTBT.[1] The CTBTO's authority to conduct on-site inspections will not vest until the CTBT enters into force, however it illustrates an important principle for verification of a multilateral arms control regime.

Limitations on destabilizing capabilities. While SALT I capped and the START treaties reduced the size of U.S. and Soviet deployed strategic nuclear arsenals, the TTBT focused on limiting destabilizing first-strike capabilities through detection and verification of testing events to enhance strategic stability. The TTBT ensured that neither party could continue to build and deploy higher yield and hence more destructive weapon systems than what existed when the TTBT entered into force. In the absence of testing, neither country would have high confidence in new weapons above 150 kilotons. This provided a strategically stabilizing outcome as the capability and destructive power of new nuclear weapons was limited—thus impacting, among other things, the first-strike capability of each side.[2] The CTBT would extend nuclear test limitations to all States party to the treaty. The CTBT would in essence freeze nuclear weapons capabilities among participating States at the time of entry into force and limit further development of destabilizing nuclear weapon systems.

Mutual declarations of infrastructure and capability. The TTBT included the exchange of a set of data that provided both parties with significant insight into the other's nuclear weapon testing program. The data exchange included geologic data regarding each country's test sites and the exchange of data for an agreed set of tests in order to calibrate seismic instrumentation to provide a better estimate of yield.[3]

Mutual verification of declarations through inspections. The TTBT was not ratified until 1990 due to technical verification challenges and delays associated with negotiation of the Peaceful Nuclear Explosions Treaty (PNET). Continued negotiations of the TTBT and PNET in 1987 led to an agreement to utilize both a better method to determine yield and the inclusion of on-site inspections and verification for tests exceeding 35 kilotons. To solve the verification issue, a set of experiments were held, one in each country, to confirm the accuracy of the proposed measurement techniques. This cooperation, and successful test result, led to the ratification of the TTBT. Once entered into force, parties to the CTBT can request on-site inspections to collect evidence within the border of States suspected of carrying out a prohibited nuclear weapon test detected by the IMS. These challenge inspections, if approved by the State subjected to the inspection, will attempt to collect evidence of a violation within six days of a request. The deterrence factor derived from challenge inspections leads to an increase in confidence of treaty compliance and its verifiability.[4]

5.2.2 Applicability in the Cyber Realm

As highlighted above, various characteristics of cyber weaponry do not lend themselves to being constrained by the principles of capability limitation central to the treaties that constrain nuclear testing. The concepts of a prohibition against cyber weapon testing in specific realms or limitations to the "yield"

[1] Ibid., Art. IV (B.)(34.-36.),(56.).
[2] Ibid.
[3] Ibid.
[4] "The Final Verification Measure: CTBTO Preparatory Commission," accessed August 31, 2017, https://www.ctbto.org/verification-regime/on-site-inspection/the-final-verification-measure/.

of a cyber weapon cannot be readily realized and are not generally analogous to the nuclear domain. **The on-site inspection verification mechanism of the PNET appears virtually irrelevant for cyber, but the exchange of data confidence-building paradigm of the TTBT may be a better analogue for what is relevant and possible in the cyber domain**. While it may often be possible to detect and measure the impact of a cyber event, attribution for such an event may be far more difficult than attributing responsibility for a high-yield nuclear test. Nonetheless, a few lessons learned from the test ban negotiation process may prove informative in a cyber context.

Attributes that Merit Further Consideration

Multilateral format. The landscape of actors and capabilities in cyberspace is increasingly global in nature, and the potential ramifications of a cyber incident can have a global impact. To increase effectiveness, negotiation of norms, prohibited activities, or limitations on tools or techniques should ultimately seek universal membership to engage all potentially affected parties. However, as the test ban negotiation process illustrates, pursuing negotiations in a multilateral format is typically time consuming and achieving consensus is difficult. Illustrative of the difficulty are the discussions surrounding cybersecurity tools and the Wassenaar Arrangement.

Limitations on destabilizing capabilities. Again, States do not have a common understanding of the scope of cyber weaponry, as the technology is continually evolving and potential applications are ever-expanding. At the time of writing, it is not clear that the limits and uses of cyber weaponry will become clearer or less ephemeral over the near term. This makes it more challenging for States to achieve consensus about what is "destabilizing" in terms of cyber arms races.

Given that the destructive potential of cyber weapons is at least as much a function of the target as the weapon, bilateral, regional, and multilateral discussions on norms and prohibitions should focus on the issue of targeting. Specifically, States could explore developing a norm to ban the targeting of specified critical infrastructure in both the civilian and military spheres. This driver could be analogous to the objective of the LTBT, which was driven by the need to reduce the health, safety, and environmental impacts of nuclear weapon tests. Similarly, an agreement limiting destabilizing targets or actions would be more meaningful and strategically significant in the cyber realm than an agreement that seeks to limit capabilities.

Attributes with Limited to No Cyber Applicability

Cyber weapon testing. No obvious parallel exists between nuclear and cyber weapon testing. Nuclear weapons are kinetic weapons developed and used by States that deploy their payload in a visible and measurable manner. **Nuclear testing is thought to serve as a deterrent since it demonstrates the nuclear capabilities of the testing State. In contrast, cyber weapons are not tested with the goal of deterrence and signaling in mind**. Also, cyber weapons are developed and used both by State and non-State actors (which, to date, have not acquired or used nuclear weapons, although they have sought them). While cyber weapons can result in kinetic effects, the payload may not always be visible and measurable. Additionally, testing of cyber weapons is typically conducted covertly since the attacker seeks to exploit unknown vulnerabilities which, if they were known, might be mitigated, thereby nullifying the effectiveness and deterrence value of the cyber weapon. Once deployed, the characteristics of a cyber weapon, the weapon's blueprint (code), and the vulnerability leveraged become known and defenses

against the weapon can be developed and deployed. Therefore, testing a cyber weapon bears little apparent deterrent value.

The LTBT did not include a verification regime because it was possible to attribute a test event to the country of origin using national technical means. Attribution of cyber events is not as simple. There are many ways to spoof, obfuscate, false-flag or repudiate attribution of a cyber-attack by the originating actor. Advanced tools and techniques that support attribution are often closely held as they can easily be nullified. The time and resources required for actionable attribution are significant—and States are generally unwilling to make that evidence public. Nonetheless, suspected actors have been called out by one State as the perpetrator of specific cyber-attacks or malicious activities. One example is the association of Russian activity to exploitation of networks and computer systems against a U.S. political party.[1] However, this tends to be the exception rather than the rule. **Therefore, treaty limits on cyber testing are not particularly realistic or useful in the cyber realm.**

Multilateral, Technical Verification. The CTBT includes a robust technical verification regime and the ability to request on-site inspections within a State Party suspected of conducting a nuclear weapon test. The verification regime is a distributed set of stations throughout the world to detect indicators of a nuclear test. Cyber weapon testing is likely to be performed and perfected in secret to ensure the payload can be successfully delivered when the attack occurs. Therefore, indicators with respect to a cyber weapon are likely only to be detected once the attack has begun. Additionally, utilizing the many methods available to obfuscate the source of the attack and the use of commodity malware (i.e., not targeted or customized malware) makes attribution difficult. **Secrecy and obfuscation reduce effectiveness of a cyber IMS to deter bad actors and challenge the efficacy of inspection regimes. Conversely, pseudo cyber IMSs exist with commercial antivirus and cyber threat intelligence companies that have global, though not universal, visibility of cyber-attacks. These significantly increase the cost required to remain stealthy**.

There are some further potential cyber applications of the CTBT verification paradigm. As will be noted later in the paper, a potential area for regulating hostile cyber activity includes agreeing not to engage in attacks against critical infrastructure or to engage in certain attacks with high risks of escalation. **A verification paradigm for such a cyber regime might look similar to the forensic combination of the CTBT's IMS and on-site inspection regime. Countries could declare that they have suffered a banned attack, and a neutral multinational technical body could be afforded access to assess the evidence of the attack to attribute it. The success of such a regime would likely hinge on the quality of and States' confidence in attribution.** To enhance effectiveness, the regime might include certain obligations on States to cooperate with inquiries where evidence suggests that their territories may have been exploited to carry out a forbidden hostile cyber action, for instance to preserve and facilitate access to evidence of such activity. This could result in a "globalized" attribution network similar to the IMS.

Cyber weapon thresholds. The effects of cyber weapon attacks can have varying destructive capability. The effect is more dependent on the target than the weapon. Therefore, there is little correlation between limiting the payload of a nuclear weapon, usually measured in kilotons of dynamite,

[1] National Cybersecurity and Communications Integration Center, "GRIZZLY STEPPE – Russian Malicious Cyber Activity", JAR-16-20296A, https://www.us-cert.gov/sites/default/files/publications/JAR_16-20296A_GRIZZLY%20STEPPE-2016-1229.pdf.

and limiting the payload of cyber weapon, which has no analogous measurement unit. Treaties that limit targets of attacks are discussed below.

Declarations on cyber infrastructure and capabilities. Participating States make declarations about their testing infrastructure and nuclear capabilities as required under the various test ban treaties to increase confidence in compliance. Such declarations assist the IMS in calibrating its explosion detection tools. However, such declarations about cyber infrastructure and capabilities are unlikely to make a meaningful contribution toward cyber stability. **Cyber weapon capabilities generally depend on the exploitation of unknown target vulnerabilities, so declarations of capabilities may identify those target vulnerabilities, thus rendering the capability useless in the future.** Therefore, States are unlikely to make meaningful declarations of their cyber testing infrastructure, intentions or history.

Nonetheless, data declarations in an arms control context traditionally increase confidence, predictability, stability, and transparency by giving all parties a greater understanding of the others' offensive and/or defensive capabilities and force structure. These data exchanges had a stabilizing effect between the United States and the USSR throughout the Cold War. Discussions on how to exchange information on certain aspects of cyber capabilities, doctrine and intent—in conjunction with other measures—likewise could raise the veil of extreme secrecy and distrust and enhance overall strategic stability between allies and adversaries. In this regard, **the development, presidential approval and release of an unclassified Cyber Posture Review (akin to the Nuclear Posture Review which declares U.S. nuclear doctrine) could be a useful supplementary measure.**

Verifiability. As described above, testing of cyber weapons does not occur publicly, and cyber arsenals are not bound to a specific geographic location. Therefore, such traditional verification mechanisms as on-site inspections or portal monitoring are not well-suited to the cyber domain. **States could potentially explore the technical feasibility of an agreement not to block the use of national technical means within agreed parameters, coupled with regular data exchanges, to increase cyber stability**.

The impetus for the negotiation and implementation of nuclear test ban treaties was the broad recognition of an escalating arms race, the destabilizing effect of the development and testing of ever-more powerful weapon systems, and the negative global environmental and health impacts of such testing. To a lesser extent, the same recognition is emerging in cyberspace. The increasing development and use of more destructive cyber weapons is destabilizing and the potential impact on the safety and security of the international community is also increasing. However, given the inherent differences between nuclear and cyber weapons, this paper does not recommend the solution of test ban treaties to enhance cyber stability. **Lessons learned from multilateral negotiations on the TTBT and LTBT, coupled with the utility of data exchanges, an unclassified Cyber Posture Review that declares U.S. policy on cyber-attack and defense, and other transparency measures hold potential to help limit or mitigate misperceptions and manage escalating tensions in times of crises**.

5.3 Biological and Chemical and Weapons Conventions

Both the Biological Weapons Convention (BWC)[1] and Chemical Weapons Convention (CWC)[2] prohibit classes of weapons generally understood as "weapons of mass destruction" (WMDs), owing to their highly indiscriminate and potentially lethal effects. However, the processes for these weapons' development, production, and delivery differ significantly both from nuclear weapons and from each other, as have their historic use for military purposes. As a result, the disparate regimes these treaties establish differ from treaties regulating nuclear weapons. **This report treats both conventions together in this report due to similarities in their principles of control, particularly in that their definitions of the weapons they regulate must be "open-ended" because new agents can be discovered as technology progresses and that the precursors for both biological and chemical weapons have widespread legitimate uses the make detection and prevention of prohibited activity challenging. Both of these features are highly applicable in the cyber realm** and are treated in more detail below.

Widespread international consensus against chemical weapons use followed World War I. The 1925 Geneva Protocol forbade the use of chemical and biological weapons. Nevertheless, the specific prohibition on chemical weapon use did not cover activities such as development, manufacture, possession, or transfer, and a number of countries engaged in these activities. Over the subsequent decades, both the BWC and CWC gradually took shape and share important similarities, but also retain some key differences.

BWC. The BWC, developed in the 1960s and opened for signature in 1972, currently has 178 signatories. The goal of this convention is to achieve "complete disarmament" of biological weapons through prohibiting the development, production, stockpiling, and delivery of biological materials "that have no justification for prophylactic, protective or other peaceful purposes…and [sic]…weapons, equipment, or means of delivery designed to use such agents or toxins for hostile purposes in armed conflict."[3] The BWC does not include a mechanism for verifying countries' compliance, relying instead upon voluntary confidence-building measures (CBMs) to help demonstrate compliance. Multiple efforts have developed measures for strengthening BWC verification, such as a proposed agreement to establish an Organization for the Prohibition of Biological Weapons, but no proposal has garnered sufficient support to be realized.[4]

Reliable verification of strictly peaceful uses of biological technology is quite difficult as biological research and development are inherently dual use (i.e., fermenters can be used for vaccine or biological

[1] Convention on the Prohibition of the Development, Production and Stockpiling of Bacteriological (Biological) and Toxin Weapons and on their Destruction (Biological Weapons Convention, or BWC), opened for signature 10 Apr. 1972, 1015 U.N.T.S. 163; 11 ILM 309.

[2] Convention on the Prohibition of Development, Production, Stockpiling and Use of Chemical Weapons and on their Destruction (Chemical Weapons Convention, or CWC), *opened for signature* Jan. 13, 1993, S. Treaty Doc. No. 103-21, 1974 U.N.T.S. 317.

[3] Ibid.

[4] A group of verification experts (VEREX) convened in 1991 to consider potential BWC verification measures, which led to the establishment of an Ad Hoc Group charged with developing a binding verification protocol. The Ad Hoc Group produced a draft text in March 2001 that proposed the establishment of an Organization for the Prohibition of Biological Weapons, a requirement to declare all relevant facilities and activities, and a system of inspections (both random and in response to alleged violations). However, the United States expressed opposition to the draft and the work of the Group was ultimately suspended and has not resumed. Ibid.

weapon production), comparatively small quantities of biological agents are easy to conceal yet still have potential lethal impacts, and the rise of "do it yourself" (or DIY) biological laboratories and the ease of obtaining equipment make verifying the intent of biological facilities difficult at best to assess when potential clandestine activities are taking place.[1] Moreover, such activities may also be carried out virtually anywhere, and evidence of the activities can be eliminated quickly; hence biological materials are quite difficult to account for and verify.[2]

To increase transparency among Member States Parties regarding biodefense-related activities, so-called CBMs were established in 1986 with six specific areas of reporting (Form A-G): facilities, defense programs, disease and toxin outbreaks, publications, past occurrence of non-defense/non-peaceful activities, and vaccine manufacturing plants.[3] These measures aim "to prevent or reduce the occurrence of ambiguities, doubts and suspicions."[4] To further improve transparency, prior years' information on CBM submissions is archived on the UN website, including an indication of which countries submit the annual forms.[5] In some cases, States may elect to have their complete report available as open source, while other countries opt to only make this information available to other States Parties; in 2016 approximately half the data was restricted from public view.[6] While State Parties have an obligation to report this information annually, only 66 of 178 countries submitted CBMs in 2016.[7]

When questions of compliance arise, the BWC has an established procedure for consultation among Member States, as well as a procedure for Member States to refer matters to the UN Security Council (UNSC) for investigating possible breaches and to take measures to bring States into compliance. To date, the BWC consultative process has been invoked only once—by Cuba against the United States in 1997—and no State has made a referral to the UNSC yet.[8] However, States have publicly noted alleged violations and compliance concerns with respect to themselves or other States on multiple occasions.[9] However, without a formal verification process, the ability to adjudicate such allegations is problematic. This has resulted in a lack of forward progress toward strengthening the BWC and establishing goals for the

---

[1] "Biological Weapons Convention (BWC) Compliance Protocol | NTI," accessed August 31, 2017, http://www.nti.org/analysis/articles/biological-weapons-convention-bwc/.
[2] "Biological Weapons Convention (BWC) Compliance Protocol | NTI," accessed August 31, 2017, http://www.nti.org/analysis/articles/biological-weapons-convention-bwc/.
[3] "Participating in the CBMs – UNODA," accessed August 31, 2017, https://www.un.org/disarmament/geneva/bwc/implementation/participating-in-the-cbms/.
[4] Second Review Conference of the Parties to the Convention on the Prohibition of the Development, Production and Stockpiling of Bacteriological (Biological) and Toxin Weapons and on their Destruction: Final Document, Part II, Final Declaration," BWC/CONF.II/13/II, 1986, p. 6.
[5] "BWC 2006: Building Transparency Through Confidence Building Measures | Arms Control Association," accessed August 31, 2017, https://www.armscontrol.org/act/2006_07-08/BWC2006#Note3.
[6] "Where Global Solutions Are Shaped for You | Disarmament | CBM Returns from 2000 Onwards," accessed August 31, 2017, https://www.unog.ch/80256EE600585943/(httpPages)/4FA4DA37A55C7966C12575780055D9E8?OpenDocument
[7] Ibid.
[8] Ibid.
[9] A few examples include the following: in 1986, the United States accused the USSR of violating the BWC; in 1992, the Russian Federation admitted that the USSR operated a large biological weapons program and committed to dismantle it; in 2001, the United States has accused Libya, Syria, Iraq, Iran, Sudan and North Korea of operating biological weapons programs. Ibid.

intersessional program, a five-year period between Review Conferences where, in years past, substantive efforts were made toward BWC-related measures.[1]

CWC. The CWC entered into force in 1997 and prohibits the use, development, production, stockpiling, and transfer of chemical weapons, thus also covering related activities not specifically forbidden by the Geneva Protocol. The CWC defines chemical weapons both in terms of specific chemicals and in terms of their characteristics and intended use for non-peaceful purposes. The CWC lays out a three-tiered system of declaration and verification requirements for specific chemicals, categorized roughly by those that have use for virtually no purpose except as weapons, that have some degree of commercial application and significant potential for use in weapons, or that are generally produced in large quantities for industrial purposes and have some potential for chemical weapons application. The CWC provides for a more stringent verification and compliance regime than the BWC through the Organisation for the Prohibition of Chemical Weapons (OPCW), with authority to conduct both "routine" prophylactic inspections of declared chemical production and storage sites to assure ongoing compliance, as well as challenge inspections in response to alleged violations. Owing to the widespread peaceful use of toxic chemicals, there are generally more sites than the OPCW can meaningfully inspect to achieve high confidence of strictly peaceful use.[2] Beyond its inspection and verification role, the OPCW serves as a dedicated forum for dialogue on and promotion of peaceful chemical practices, and as a repository of expertise.

5.3.1 Common Principles of the Treaties

Ban of class of weapon. Both the BWC and CWC ban the development, possession, production and use of entire classes of weapons. Both conventions also mandate the destruction of applicable weapon stockpiles and production facilities. Whether a particular biological or chemical agent is a prohibited "weapon" depends on whether its manufacture or use was for non-prohibited purposes—**a State's intent in engaging in activity involving the agent or its precursors determines whether the agent is a prohibited weapon**.

Open-ended definitions of prohibited classes of weapons. The BWC and CWC employ similar solutions to the common problem that no list of agents can be definitive as new weaponizable biological and chemical agents are rapidly discovered, synthesized and/or fabricated. Rather than relying solely on a proscriptive list of banned agents, the treaties establish general definitional criteria pertaining to State intent (for both the BWC and CWC) as well as, for the CWC, an open-ended list for which a specific procedure for review and amendment is provided.[3]

Exemptions for peaceful uses. Both conventions acknowledge the right of State Parties to retain listed agents where doing so is consistent with non-prohibited peaceful uses. Under the BWC, explicitly

---

[1] Jp Z and ers, "BTWC 8th RevCon Final Document," *The Trench*, November 25, 2016, http://www.the-trench.org/btwc-8th-revcon-final-document/.

[2] As of 2016, the OPCW had conducted over 3,000 on-site inspections, but had identified 4,772 facilities subject to inspection, hence has inspected each such facility an average of less than 0.63 times over the course of its existence. See OPCW, Report on the Implementation of the Convention on the Prohibition of the Development, Production, Stockpiling and Use of Chemical Weapons and on their Destruction in 2015, 7-8 (2 Dec. 2016).

[3] BWC, *supra* note at Art. I; CWC, *supra* note at Arts. II, XV.

named peaceful uses include prophylactic and protective purposes, which could include retaining samples for comparison purposes or for development of vaccines, antidotes, or other countermeasures to protect human health from known pathogens that may be encountered again, whatever their origin. The CWC goes further, explicitly allowing for use of agents against human populations for "law enforcement including domestic riot control purposes"[1]—that is, where used in a purely domestic context, not in an international armed conflict – though this is a matter of increasing controversy.[2] The regimes establish threshold criteria for presumed non-peaceful conduct, such as possessing certain enumerated quantities of agents for which there is presumed to be no peaceful use.

Declarations of accountable arms. All State Parties to the BWC and CWC are required upon accession to declare their applicable prohibited weapons stockpiles, equipment, and production facilities, with a view to securing their elimination in order to comply with the applicable convention.[3, 4]

Investigation of suspected violations. both the BWC and CWC contain procedures for investigating alleged instances of non-compliance. As noted above, the BWC provides that State Parties may lodge complaints regarding alleged breaches of the BWC with the UNSC, which may initiate an investigation.[5] The CWC's procedures are more specific and institutionalized. Under the CWC, the OPCW has the right to convene its own fact-finding missions to address potential violations, and State Parties have the right to request on-site challenge inspections of any facility or location in any other State to clarify and resolve questions of possible non-compliance.[6] Multiple violations have been alleged of the CWC and previous prohibitions on and norms against chemical weapons use.[7] The OPCW has played a very active role

---

[1] Chemicals are not considered chemical weapons when intended for "purposes not prohibited under [the] Convention." See CWC, supra note _ at Art. II (1) (a). Such purposes include "(a) Industrial, agricultural, research, medical, pharmaceutical or other peaceful purposes; (b) Protective purposes, namely those purposes directly related to protection against toxic chemicals and to protection against chemical weapons; (c) Military purposes not connected with the use of chemical weapons and not dependent on the use of the toxic properties of chemicals as a method of warfare; (d) Law enforcement including domestic riot control purposes." See ibid. at Art. 9. Accordingly, a tear gas grenade used by police for riot control would not be considered a chemical weapon under the convention, whereas it would be if used by military personnel against an enemy in an international armed conflict.

[2] "Aerosolisation of Central Nervous System-Acting Chemicals for Law Enforcement Purposes," accessed August 31, 2017, https://www.opcw.org/fileadmin/OPCW/CSP/C-21/national_statements/c21nat03_e_.pdf. (A number of States Parties to the CWC have undertaken efforts to strengthen the OPCW's ability to carry out its mandate to investigate allegations of the use of some highly toxic central nervous system-acting pharmaceutical chemicals, such as fentanyls, sometimes (arguably inappropriately) referred to as "incapacitating agents" in law enforcement scenarios. The British Medical Association first drew attention to these risks and expressed its opposition in 2007; more recently, Australia and 38 other State Parties have stressed that these chemicals pose a serious challenge to the Convention).

[3] "The Chemical Weapons Ban Facts and Figures," accessed August 31, 2017, https://www.opcw.org/news-publications/publications/facts-and-figures/#c1920. (With the CWC, for instance, as of 7 Jan. 2017, 90% of the State Parties' declared stockpile of banned chemical agents had been destroyed).

[4] Ibid.

[5] Ibid.

[6] CWC, *supra* note at Art. IX.

[7] The U.S. Department of State specifically cites numerous and systematic instances of use of chemical weapons by Syria every year since it acceded to the CWC in 2013, and notes its inability to certify that Russia, Iran and Iraq are in compliance with their CWC obligations. U.S. Department of State, Compliance with the Convention on the Prohibition of the Development, Production, Stockpiling and Use of Chemical Weapons and on Their Destruction Condition 10(c) Report (April 2016), https://www.state.gov/t/avc/rls/rpt/2016/255563.htm. In addition, despite the Geneva Protocol, there were various instances of application of chemical warfare agents in combat or against

investigating alleged violations of the CWC in Syria since 2013 under the authority of a Fact-Finding Mission convened by the OPCW Director General. To date, no State has requested a challenge inspection.[1]

Support for States Parties. The BWC and CWC contain provisions for economic and technological advancement in peaceful use of biology and chemistry. Support programs funded by States Parties hinder prohibited activities and foster trust and international cooperation. For example, Article X of the CWC outlines provisions for the coordination and delivery of protection against chemical weapons, including "detection equipment and alarm systems; protective equipment, decontamination equipment and decontaminants, medical antidotes and treatments, and advice on any of these protective measures." Further, States Parties are granted the authority to facilitate and participate in the "fullest possible exchange of equipment, material and scientific and technological information concerning means of protection against chemical weapons."[2] Similar directives are granted in the BWC. For example, Article V encourages States Parties to work together on BWC-relates challenges, while State Parties are directed to provide relief to countries that fall potential victims to a biological weapon attack.[3] Moreover, science and technical exchange (Article X) related to benign purposes is mandates between States Parties to foster additional understanding of the science and technology landscape.[4]

5.3.2 Applicability in the Cyber Realm

Certain aspects of biological and chemical weapons, as described below, pose challenges quite similar to challenges in the cyber realm, while other characteristics have less apparent relevance.

Attributes that Merit Further Consideration

Open-ended, purpose-oriented definitions of prohibited weaponry. Rather than monitoring and regulating weapons stockpiles, verification for any cyber control regime would likely be more effective if focused on State conduct. Given the technical complexity of cyber investigation and attribution, a neutral expert organization with previously agreed rights of access for investigation, like the OPCW, would likely make a regime more viable than reliance on an ad hoc investigation mechanism under the auspices of the UNSC per the BWC model, which would be subject to veto by any of its permanent members.

CBMs. Demonstrating good faith and cooperation through transparency measures could enhance trust among States and reduce the potential for hostilities. These good faith transparency measures could be modeled on the BWC CBMs, focused on specific research, technologies, and activities with the potential to be inherently dual use. This information could be compiled by State Parties and reported to an international body, similar to the Implementation Support Unit that is the administrative unit of the BWC.

---

civilian populations after 1925 but before either the drafting of the CWC, or accession to it by the States involved, such as by Nazi Germany, Japan, Iran, Iraq and Syria. See, e.g., OPCW, Brief History of Chemical Weapons Use, accessed April 4, 2017, https://www.opcw.org/about-chemical-weapons/history-of-cw-use/.

[1] The most intensive international response to alleged violations of the CWC to date – regarding multiple

[2] CWC, *supra* note at Art. X

[3] Convention on the Prohibition of the Development, Production and Stockpiling of Bacteriological (Biological) and Toxin Weapons and on their Destruction (Biological Weapons Convention; BWC). 26 March 1975. Accessed September 1, 2017, https://unoda-web.s3-accelerate.amazonaws.com/wp-content/uploads/assets/media/C4048678A93B6934C1257188004848D0/file/BWC-text-English.pdf

[4] Ibid.

Moreover, making this information publicly available would further boost transparency and help allay suspicion and provide additional measures of clarify regarding cyber activities.

Emphasis on State intent. The technical characteristics of cyber weaponry, or hostile cyber actions, could potentially be as varied as the code in the world—similar to the expanding variety of dangerous chemical and biological agents. To be meaningful, as with the BWC and CWC, any definition of controllable cyber weaponry or hostile cyber actions would have to be highly adaptable but clear enough to apply in specific cases. **Rather than attempt to capture specific (and rapidly evolving) technical parameters for controlled cyber weaponry, definition of terms and prohibitions on use based on State intent, as in the CWC and BWC, could likely provide a useful framework for monitoring State cyber conduct**.

Attributes with Limited to No Cyber Applicability

Declarations of accountable arms. States would likely be unwilling to disclose offensive cyber capabilities, and offensive cyber capability is not generally conducive to "accounting." Thus, **compliance with a cyber control regime would be better measured by the absence of offensive action, rather than some scheme of accountancy** for items or agents.

Weapons ban. Cyber weapons and cyber realm operations generally do not involve destructive health and physical effects comparable to those of biological, chemical, and nuclear weapons. Cyber weapons and operations could even be viewed as a means for States to achieve their desired ends without resorting to violence in many instances. In the aggregate, cyber weapons and operations tend to be comparatively low cost and high reward. These factors likely contribute to the perceived widespread and increasing State use of hostile cyber operations.[1] The likelihood of attaining widespread consensus on outright prohibition of cyber weaponry or hostile operations therefore seems low, nor would a full ban necessarily be advisable**. Widespread consensus among States against taking particularly destabilizing, escalatory, or harmful actions may nonetheless be possible.** The "Group of Government Experts" (GGE) has already endorsed a norm that States not engage in attacks against critical infrastructure, for instance.[2] States may be able to agree to similar norms or prohibitions against other, comparable high-consequence or escalatory attacks in peacetime.

5.4 United States-Soviet Union Incidents at Sea Agreement

In the late 1960s, the naval forces of the United States and Soviet Union had a number of dangerous close encounters, including ships bumping, high-risk fly-bys, and aggressive maneuvers that could have triggered reprisals or escalation toward nuclear war. The countries agreed to negotiate protocols for

[1] Eric Clapper, Dir. of National Intelligence, Statement for the Record, Worldwide Threat Assessment of the U.S. Intelligence Community (Feb. 26, 2015) (noting "[c]yber threats to US national and economic security are increasing in frequency, scale, sophistication, and severity of impact", and that "[t]he ranges of cyber threat actors, methods of attack, targeted systems, and victims are also expanding").

[2] General Assembly, Group of Governmental Experts on Developments in the Field of Information and Telecommunications in the Context of International Security, UN document A/70/174, 22 July 2015. Notably, there is not yet any international consensus on what constitutes "critical infrastructure."

conduct at sea to prevent such incidents from escalating, resulting in the Incidents at Sea Agreement of 1972.[1] Specifically, the agreement provides for the following:

- Steps to avoid collision
- Non-interference in the "formations" of the other party
- Avoiding maneuvers in areas of heavy sea traffic
- Requiring surveillance ships to maintain a safe distance from the object of investigation so as to avoid "embarrassing or endangering the ships under surveillance"
- Using accepted international signals when ships maneuver near one another
- Not simulating attacks at, launching objects toward, or illuminating the bridges of the other party's ships;
- Informing vessels when submarines are exercising near them
- Requiring aircraft commanders to use the greatest caution and prudence in approaching aircraft and ships of the other party and not permitting simulated attacks against aircraft or ships, performing aerobatics over ships, or dropping hazardous objects near them.[2]

The agreement also provides for (1) a notice three to five days in advance of any projected actions that might "represent a danger to navigation or to aircraft in flight," (2) information on incidents to be channeled through naval attachés assigned to the respective capitals, and (3) annual meetings to review the implementation of the Agreement. The sides also signed a protocol in 1973 in which each party pledged not to make simulated attacks against the nonmilitary ships of the other. The sides implemented the Agreement through adopting changes to their navies' standard operating procedures and coordinating other practical measures, such as establishing new supplementary signals unique to interactions between the U.S. and USSR (Russian) navies.[3] The agreement remains in force in 2017.

The agreement has no verification mechanism as such, as it does not regulate *capabilities* or force structures, but overt *conduct* that is readily apparent between the sides' militaries operating in international waters. The Agreement obligates the parties to exchange information concerning instances of collisions or incidents at sea between their ships and aircraft through their respective naval attachés.[4] The Agreement did not include specific provisions on violations, enforcement or resolution of disputes, hence these, by default, would be resolved in accordance with applicable international law.

5.4.1 Principles of the Agreement

The Incidents at Sea Agreement enhanced stability in part through the following principles:

---

[1] Agreement Between the Government of The United States of America and the Government of The Union of Soviet Socialist Republics on the Prevention of Incidents On and Over the High Seas [INCSEA Agreement] (entered into force May 25, 1972). Accessed April 13, 2017, https://www.state.gov/t/isn/4791.htm

[2] Ibid.

[3] See, e.g., U.S. Office of the Chief of Naval Operations, United States / Russian Federation Incidents at Sea and Dangerous Military Activities Agreements (10 Nov. 2008).

[4] INCSEA Agreement, *supra* note at Art. VII.

Refraining from specific, mutually recognized provocative actions. The sides committed to refrain from or notify certain actions that the sides agreed could be perceived as acts of aggression that might provoke a retaliatory response and lead to escalating conflict. The sides agreed and enumerated such specific actions based on their experience of incidents, particularly from the 1950s through the early 1970s.

Prior notification of certain potentially provocative acts. Prior notification of potentially dangerous maneuvers reduced the potential for unintended accidents with serious consequences that might otherwise have been interpreted as deliberate hostile acts.

Establishing channels for communication, dispute resolution, and deconfliction. Although rudimentary in this instance, the Incidents at Sea Agreement created clear points of contact for managing issues under the Agreement, which may have facilitated streamlined communications and reduced the potential for confusion during incidents.

5.4.2 Applicability in the Cyber Realm

Attributes that Merit Further Consideration

Limitations on Destabilizing Activities. Certain principles of avoidance of provocation might be cross-applied from the Incidents at Sea Agreement concept to the cyber realm. **Countries might agree to refrain from high-risk or particularly escalatory "maneuvers," such as probing or attacking sensitive national security assets or critical infrastructure in a manner that might be perceived as a prelude to or part of an armed attack**. Making more specific prohibitions would require that sides disclose the types of actions they might undertake, which, owing to secrecy considerations already described, may prove difficult. Short of mutual agreement on measures of restraint, **countries could unilaterally declare categories of assets they regard as sensitive where hostile cyber actions would prompt a response**.

Communication and information sharing. While the channels and protocols for communication, dispute resolution, and deconfliction are rudimentary in this Agreement, simply identifying clear points of contact may help ensure that criminal or non-State actors' cyber activities are not misconstrued and that States have the mechanisms in place to properly signal to one another.

Attributes with Limited to No Cyber Applicability

Prior notification of potentially provocative acts: The high seas and cyber realms are not close analogues in this aspect. Encounters between national navies on the high seas are characterized by the international character of open waters (where no State has territorial sovereignty or jurisdiction), the physical proximity of air- and seaborne vessels that are generally overt and visually apparent or apparent via remote sensing, and the destructive capacity of the vessels involved (typically equipped with weapons platforms). Except for segments that transit international spaces (outer space, the high seas), cyber "spaces" are almost always national in character (under the jurisdiction of a particular country, subject to

its laws and exclusive sovereignty).[1] Nation-State actions in other States' cyber spaces are therefore generally covert and likely more analogous to surveillance or espionage than to naval kinetic weapons platform maneuvers. Sides could realistically notify others of their "training exercises" as there is no international cyberspace in which such exercises could be conducted where forces would be in "proximity." One side "testing" a cyber tool or weapon on any country's systems or networks would presumptively be a hostile act regardless of any prior notice as such tests would infringe or violate the sovereignty of the country on whose servers or systems the test was conducted. Consequently, the value of communications and prior notification is relatively more limited for cyber operations.

5.5 Helsinki Process

The Helsinki Process sought to reduce tension between the Soviet Union and the Western Bloc during the 1970s and 1980s.[2] The first major milestone from the Process, the 1975 Helsinki Accords, involved three key elements significant for deterrence and stability:

- The first element focused on resolving the long-standing territorial dispute over recognition of the Soviet occupation of the Baltic States and reducing related tensions.[3]
- The second was a package of measures for transparency and dispute resolution in political-military matters, to be mediated through establishment of the joint Conference (later Organization) for Security and Cooperation in Europe (CSCE, OSCE).[4]
- The third was a joint commitment by the participating States to uphold the human, religious and political rights of their citizens.[5]

Whereas the USSR highly valued the first of these elements, the second two also proved to be tremendously valuable and consequential for East-West strategic stability and dispelling Soviet disinformation. The CSCE and OSCE became the forum for negotiating and implementing major deterrence and stability mechanisms—such as the Conventional Armed Forces in Europe and the Open Skies treaties—and remains one of the primary bodies for facilitating negotiations for mediating and monitoring of conflicts in the former East-West buffer zone, such as Transnistria, Nagorno-Karabakh and, more recently, Georgia and Eastern Ukraine.[6] The countries' human rights commitments under the Helsinki Process spawned several non-governmental human rights organizations dedicated to monitoring abuses in all the participating countries. These organizations provided credible information on the

[1] Tallinn Manual 2.0 on the International Law Applicable to Cyber Operations, Rule 9 (2017) ("A State may exercise territorial jurisdiction over: (a) cyber infrastructure and persons engaged in cyber activities on its territory; (b) cyber activities originating in, or completed on, its territory; or (c) cyber activities having a substantial effect in its territory.").
[2] U.S. Helsinki Commission, Commission on Cooperation and Security in Europe, "The Helsinki Process and the OSCE," Accessed April 4, 2017, https://www.csce.gov/about-csce/helsinki-process-and-osce.
[3] Ibid.
[4] Ibid.
[5] The process also included a "basket" devoted to economic provisions, which was less relevant for the purposes of this paper. Ibid.
[6] U.S. Helsinki Commission, Commission on Cooperation and Security in Europe, "The Security Dimension," Accessed June 1, 2017, https://www.csce.gov/about-csce/helsinki-process-and-osce/security-dimension.

situation within the Soviet Union which, some argue, constrained the USSR's ability to persist in peddling false information when it sought to undertake major reforms in the 1980s.[1]

5.5.1 Principles of the Agreement

Reduced to basic principles, the Helsinki Process involve the following:

Incorporating attenuated but valued regional interests of the parties within a single negotiating framework. Resolving or making progress on particularly intractable issues may require unusual creativity and flexibility, requiring the parties to broaden the scope of their negotiations until they can both see advantage in compromise.

Establishing a dedicated forum for dialogue and mediating disputes. The CSCE and OSCE served and continue to serve a stabilizing function, leading to and playing a prominent role in conflict monitoring and the negotiation of other key agreements.

Linking concessions on territorial disputes with affirmation of informational, civil, and political rights. The specific compromise reached under the Helsinki Accords—in which the West recognized the Soviet Baltic States in exchange for Soviet concessions on informational, civil, and political rights of Soviet citizens—is key to the agreement's success. It strengthened strategic stability and expanded the volume of information available to Soviet citizens and the world on events in the region, which may have helped to constrain subsequent excesses of the Soviet regime.

5.5.2 Applicability in the Cyber Realm

Attributes that Merit Further Consideration

Multilateral regional approaches. Negotiating strictly on cyber norms for their own sake may not be the most effective course as international players' interests in this area are not symmetrical. The United States has become more reliant on information technology as its market and democratic processes thrive on up-to-date, quality information; other societies have made a conscious choice to treat information technology much more cautiously and place much greater emphasis on State control of information. **A firmer foundation for agreement and progress on cyber issues may exist in adopting regional approaches that can incorporate other issues beyond the cyber domain to establish a greater symmetry of interests among the parties in question.** Russia and China both carry out much of what the United States views as their "bad" cyber behavior in a manner closely intertwined with regional issues. For Russia, engaging in regional hybrid warfare and disinformation may serve as a bulwark against encroaching European technocracy. For China, engaging in economic cyber espionage likely helps secure regional comparative advantage.

CBMs. Establishing a dedicated joint forum for confidence building through transparency, dialogue, and technical exchange can have a stabilizing effect and lead to the development of other stabilizing mechanisms. The OSCE has a dedicated Cyber/Information Communication Technology Security branch

[1] See, e.g., Gaddis, John Lewis. *The Cold War: A New History, xxvii* (2005); NY Times, "How Soviet Dissidents Ended 70 Years of Fake News" (10 Apr 2017).

that promotes stability and CBMs[1] and in 2016 it managed to reach agreement on initial CBMs in the cyber realm[2]. The role of the OSCE could be reinvigorated through political attention. **It may also be possible to develop an OSCE-like cyber and information security platform with a broader membership to promote cyber stability among all States, not just those in the OSCE**.

Communication and information sharing. Reaffirming norms of State conduct on human rights and information, even without strong formal or binding means of verification and enforcement, can be a strong counterweight to State disinformation and propaganda, a practice that has increased with alarming speed in the cyber realm. **The Helsinki Process experience suggests that formalizing the protections of independent fact-finders and media outlets—possibly in exchange for concessions on other regional issues—might help to reinforce independent voices, undermine State propaganda, and constrain State excesses**.

5.6 Regulating Piracy and Privateering

Some commentators compared malicious cyber activities to the phenomenon of piracy and suggest considering the evolution and regulation of piracy and privateering as a basis for grappling with cyber governance.[3]

Piracy. The evolution of the crime of piracy from one of national law to international law may be of particular interest for those wishing to clarify criminal jurisdiction questions for prosecuting cyber crime.

The accepted international law definition of piracy as a substantive crime is codified in the UN Convention on the Law of the Sea (UNCLOS) and its predecessor, the Geneva Convention on the High Seas. These documents define piracy as:

> (a) any illegal acts of violence or detention, or any act of depredation, committed for private ends by the crew or the passengers of a private ship or a private aircraft, and directed:
>
> (i) on the high seas, against another ship or aircraft, or against persons or property on board such ship or aircraft;
>
> (ii) against a ship, aircraft, persons or property in a place outside the jurisdiction of any State;
>
> (b) any act of voluntary participation in the operation of a ship or of an aircraft with knowledge of facts making it a pirate ship or aircraft;
>
> (c) any act of inciting or of intentionally facilitating an act described in subparagraph (a) or (b).[4]

Guilfoyle usefully summarizes the current state of international law with respect to piracy thus:

---

[1] See Org. for Security and Cooperation in Europe (OSCE), "Cyber/ICT Security" (12 Feb. 2014), accessed June 1, 2017, http://www.osce.org/secretariat/106324.

[2] See Org. for Security and Cooperation in Europe (OSCE), Permanent Council Decision No. 1202, accessed Sep 21, 2017, http://www.osce.org/pc/227281.

[3] See, e.g., Egloff, Florian. "Cybersecurity and the Age of Privateering: A Historical Analogy" in U Oxford Cyber Studies Programme Working Paper Series No. 1 (Mar. 2015).

[4] UN Convention on the Law of the Sea (UNCLOS), 1833 UNTS 3; 21 ILM 1261, Art. 101 (1982).

> "At the beginning of the twenty-first century piracy is best considered a national law crime for which international law provides a permissive rule of jurisdiction. […] There has been no serious attempt to establish an international piracy tribunal. There has been some judicial interest at the national level in enquiring into the meaning of piracy at international law if only to examine the compatibility of national laws and prosecutions with the provisions of UNCLOS. However, it appears generally accepted that the function of the international law of piracy is now to permit prosecutions by forum States lacking any conventional nexus to the crime rather than to directly criminalize conduct under international law in the manner of, for example, war crimes."[1]

Though the norm against piracy is now frequently cited as being a peremptory (*jus cogens*) norm of customary international law, Guilfoyle points out that tracing the origins of this norm can be a frustrating undertaking. However, it is useful to understand how a norm became customary international law to assess whether a similar undertaking might happen for cyber crimes. Guilfoyle writes, "[T]he meaning of piracy in the 17th, 18th and 19th centuries was bound up in very different ways with the laws of war. Broadly, the question was one of State sanction. **At a time before States generally had large standing navies, it was convenient for major powers to have a body of licensed privateers that they could incorporate into navies in times of war. Thus, a privateer was a State-licensed actor**,"[2] which was a valid criminal defense to charges of piracy. Guilfoyle then notes that the 20th century saw multiple efforts to codify laws against piracy—such as by the League of Nations Committee of Experts for the Progressive Codification of International Law and the Harvard Research in International Law—that lacked clear rigor. Nonetheless, those efforts influenced the current statements of customary law now codified in the Geneva Convention on the High Seas and UNCLOS.[3]

Generalizing the international law approach to piracy represents a rather haphazard process of defining general aspects of an individual crime, rather than a State crime, that would allow States to claim criminal jurisdiction to prosecute that crime even though traditional requirements of State jurisdiction were not met. Traditional elements of criminal jurisdiction include, for example, occurrence of the crime within the State or where the victim of the crimes are citizens of the State. Under UNCLOS, however, when outside the jurisdiction of any State, *every* State has jurisdiction to arrest and prosecute the pirates.[4]

Privateering. The State practice of licensing privateers may be a closer analogue to modern-day State-sponsored cyber actions mediated through private actors. State sponsorship of privateering was relatively commonplace starting in the 13th century, enjoyed its heyday in the 16th-18th centuries, and largely ceased with an international agreement concluded in 1856.[5] Ergloff relates the overall history of privateering.[6] He describes the initial deployment of the practice by the European powers as they were in ascendance and competing to establish themselves as naval powers during the 13th through 18th centuries, then the gradual abandonment of privateering following the Napoleonic Wars once the European powers'

---

[1] Guilfoyle, Douglas. "International Piracy," American Society of International Law, 6 (July 31, 2014), https://www.asil.org/sites/default/files/ERG_PIRACY.pdf.
[2] Ibid.
[3] Ibid. at 7.
[4] UNCLOS, *supra* note 4 at Art. 105.
[5] Ergloff, *supra* note at 3-6.
[6] Ibid.

status and commerce were more firmly established. While privateers were of some use during hostilities, the powers had to grapple with the problem that, after extended wars, privateers became professionalized and would continue preying on commerce even after formal hostilities had ceased. The English in particular, as the dominant sea power in the 18th and 19th centuries, sought to regulate privateering.[1] Ergloff notes that in the late 18th and early 19th century, less powerful players, like the United States and France, turned to the practice as a form of asymmetric warfare to weaken the British. The British then spearheaded the adoption of a declaration in the settlement to end the Crimean War in 1856 to end the practice, which ultimately won the support of the key European powers.[2] In general, Ergloff notes how the history of privateering reflects the rise of States capable of projecting force and control on the high seas. In particular, the practice fell out of favor with the development of both highly organized administrative States and modern weaponry against which private parties could not realistically compete.

5.6.1 Principles of the Agreement

A criminal law approach. The development of the international law norm against piracy did not rely on the development of a specific treaty instrument or the establishment and expansion of a particular regime. Rather, **States mutually came to recognize their collective authority to prosecute the crime of piracy—to respect the legitimacy of all States extending their national laws to cover a specific conduct that would normally have lacked a traditional basis for their exercise of jurisdiction**. In doing this, they made more credible the threat of criminal sanctions against private actors who might engage in piracy.

Projecting State authority in poorly governed spaces. Piracy and privateering both thrived because of States' limited capacity to project power or exercise significant control on the high seas. Both phenomena declined with the rationalization of power of modern administrative States, coupled with the rise of more modern military technologies that greatly outstripped the abilities of non-State actors to engage in violence.

Hegemonic domain dominance, overriding concern for orderly commerce. Per Ergloff's account, the particular agreement that banned privateering appears to have been brought about both because of the advocacy of the global naval hegemon at the time, and as part of a broader general settlement among the "great" European powers of the moment, having to do with the Crimean War.

5.6.2 Applicability in the Cyber Realm

Attributes that Merit Further Consideration

A "streamlined" criminal law approach. A pronounced tension exists in international relations between proponents of a criminal law (national law) approach and proponents of a State-focused (international law) approach to addressing cyber threats. The United States and Western Europe have generally promoted an approach of harmonizing national criminal laws and promoting mutual legal

[1] Ibid.
[2] Paris Declaration Respecting Maritime Law (30 Mar. 1856). Fifty-five States ultimately ratified the declaration, including the United Kingdom, Austria, France, Prussia, Russia and the Ottoman Empire.

assistance, most prominently manifested in the Council of Europe Convention on Cybercrime, generally arguing that international humanitarian laws, States' inherent right to self-defense, and international laws on State responsibility already apply in the cyber realm.[1] Russia and China have favored pursuing a State-focused, international law approach to setting norms of State conduct, less convinced that international humanitarian, self-defense, and State responsibility laws apply in the cyber realm.[2] More details on these two approaches are available in Appendix D.

From a deterrence standpoint, the significance of a national criminal law approach is to directly sanction individuals who may be involved in hostile cyber actions, rather than having to attribute responsibility to a State. International law on piracy serves to streamline the exercise of national criminal jurisdiction even where it might be lacking by traditional customary international law standards. To enhance deterrence with respect to States, States could consider seeking similar procedural "streamlining" to extend or apply national criminal jurisdiction to hostile cyber acts, or to relieve traditional sovereign immunity protections for such acts.

Where States may hide behind proxies to avoid responsibility for illegal hostile cyber actions, identification and prosecution of individual perpetrators can serve to deter participation in cyber-attacks. Where a State may not be subjected to criminal sanctions, individuals are highly motivated to avoid jail time, asset forfeiture, and travel restrictions. The more States can successfully prosecute cyber crime, the more difficult it will be for States to recruit proxies to obfuscate their role.

The prospect of criminal sanctions has been shown to impact both State behavior and individual behavior. For example, the United States under the George W. Bush Administration undertook pronounced efforts to subvert the jurisdiction of the International Criminal Court, which took on increased significance in connection with the U.S. and coalition forces' 2003 invasion of Iraq and subsequent conduct of military operations there.[3] The Russian Federation also appears to have gone to significant lengths to obfuscate its involvement in the conflict in Eastern Ukraine—such as sending troops without insignia on uniforms and claiming that its troops may be there without State sanction (on "vacation").[4] Such actions complicate the task of proving potential State responsibility for international crimes like aggression, war crimes, and crimes against humanity but suggest that State officials take those sanctions into account when making decisions. Thus, **international agreements that expedite the criminal prosecution of State proxies and individual criminals can deter on State behavior**.

Limited analogousness of governance, proxies, and conditions. Piracy and malicious cyber activities generally occur in spaces where States are challenged to meaningfully enforce laws or assert jurisdiction (the one on the high seas, the other often involving transnational criminal acts where perpetrators, instrumentalities, and victims can all be located in different States). State recruitment of

---

[1] Markoff, Michele. "Explanation of Position at the Conclusion of the 2016-2017 UN Group of Government Experts on Developments in the Field of Information and Telecommunications in the Context of International Security," United States Mission to the United Nations (June 27, 2017) available at: https://usun.state.gov/remarks/7880.
[2] NATO Cooperative Cyber Defence Centre of Excellence, "An Updated Draft of the Code of Conduct Distributed in the United Nations – What's New?" (10 Feb. 2015) available at: https://ccdcoe.org/updated-draft-code-conduct-distributed-united-nations-whats-new.html).
[3] Georgetown University Law Center, Georgetown Law Library, "International Criminal Court – Article 98 Agreements Research Guide" (Jan. 31, 2017), http://guides.ll.georgetown.edu/article_98.
[4] Voice of America, "NATO: 1,000 Russian Troops Remain in Ukraine" (Sept. 11, 2014); VOA, "Russian Troop Buildup Along Ukraine Border Raises War Fears," (Aug. 4, 2016).

witting and unwitting proxies to add layers in this attribution process may seem on a superficial level to recall privateering in that non-State actors are recruited as agents of States. This makes it unlikely that regulation of proxies in the cyber realm can follow the path of regulation of privateering. Privateering was a recognized defense to the crime of piracy—privateers' status was something they displayed openly, and States made no particular effort to hide their use of the practice. **States commissioned privateers to expand their own capabilities as their own navies had limited capacity. As privateering was overt, States' agreement to cease the conduct was meaningful. By contrast, modern States likely generally recruit cyber proxies precisely to obfuscate their role and capabilities and to avoid responsibility**. Acknowledging, regulating, and limiting use of cyber proxies would run counter to the whole premise of the practice. It tests the bounds of logic to imagine agreement to end or curtail a practice in which no one admits they are engaging.

Though irony would have it that world attention has re-focused on a dispute over the Crimea, the current overall outlook of factors relating to State cyber behavior do not resemble those of the high seas in 1856. Where States had a common but narrow overriding interest in preserving naval commerce, today there is considerable variety and disparity in States' reliance upon the internet (i.e., the greater "attack surface" of the United States), and the ubiquity of computing means that cyber-attacks can affect anything and everything; hence, the breadth and diffuseness of the issue makes it far less manageable and conducive to achieving consensus for action or regulation.

6.0 What is Cyber Deterrence?

As defined in Section 2, the authors propose that **the term "cyber deterrence" describes a set of conditions in which a State communicates, either explicitly or implicitly, a credible intention and capability to accurately attribute and impose substantial consequences for certain hostile cyber actions directed against it, and that communication dissuades other actors from undertaking such actions.** While "traditional" and cyber deterrence have in common a basic emphasis on cost-benefit calculus, effective deterrence in cyberspace is likely to assume different forms than nuclear or conventional deterrence in the Cold War. This section outlines a model of cyber deterrence, using relevant U.S. Government policies and official statements to explore the contours both of what cyber deterrence could look like generally, and of specific stated U.S. policy relating to cyber deterrence.

Whereas this paper focuses on deterrence of hostile cyber actions generally and does not dwell on analyzing the applicability of the law of armed conflict to hostile cyber actions, other experts have addressed that subject in detail. The threshold for armed conflict, both generally and with respect to hostile cyber actions specifically, remains a matter of controversy.[1] For the purposes of this paper, one of the prevailing views of experts is that any State-on-State cyber-attack may give rise to an international armed conflict.[2] Given the increasing prevalence of serious hostile cyber actions noted in Section 1, the

[1] See, e.g., Koh, Harold. "International Law in Cyberspace," USCYBERCOM Inter-Agency Legal Conference, Ft. Meade, MD, Sept. 18, 2012, available at http://cs.brown.edu/courses/cs180/static/files/lectures/readings/lecture19/koh_on_international_peace.pdf.

[2] Controversy exists as to the requisite level of violence for "hostilities," the existence of which is a precondition for international armed conflict. Hostilities may involve any combination of kinetic and cyber operations, or cyber operations alone. International Committee of the Red Cross commentary to the 1949 Geneva Convention notes that

fact that the law of armed conflict might apply does not appear to be significantly constraining or controlling State behavior.[1] Thus, **a law of armed conflict-based approach to date has proven inadequate in deterring States, whose actions suggest that they view the benefits of engaging in the attacks to outweigh whatever detriment they incur by potentially violating the international law of armed conflict.**

6.1 Cold War Deterrence *versus* Cyber Deterrence

As discuss previously, traditional deterrence theory can be difficult to usefully apply to cyberspace and cyber weapons, given how they differ from their conventional and nuclear counterparts. "Classic" Cold War deterrence was understood to describe extremely costly, high-consequence scenarios (i.e., nuclear or large-scale conventional war between the dominant superpowers) with relatively high general confidence of attribution and capabilities for retaliation that were relatively well-known among the competing adversaries. By contrast, Section 4 notes that countries usually do not display or publicly test their cyber weapons, as doing so could potentially compromise the weapons themselves. In addition, the cost of cyber conflict is typically significantly lower than the costs of establishing, maintaining or using kinetic forces.[2] The consequences of cyber-attacks can vary from negligible to potentially severe, but in the vast majority of cases do not result in physical injury, loss of life, or destruction of property (and do not inherently present that risk in any way comparable to nuclear weapons or large-scale conventional forces). Finally, States are frequently able to obscure their involvement in hostile cyber actions through various techniques, such as masking attacks' point of origin or generating false signature data. Thus **cyber conflict can be understood to differ from Cold War competition in part because it offers a low-cost, potentially low-consequence, and often non-attributable means for weakening an adversary**. These differing cost-benefit outlooks presumably help to explain the relative frequencies of the different types of attacks.

6.2 Cyber Deterrence Conceptual Model

The authors of this paper developed the following visual representation of cyber deterrence (see Figure 3). The components of that representation are discussed below, followed by an examination of these elements in light of U.S. cyber policy and related actions the United States has taken.

"*Any* difference arising between two States and leading to the intervention of armed forces is an armed conflict… It makes no difference how long the conflict lasts, or how much slaughter takes place." See Tallinn Manual 2.0 on the International Law Applicable to Cyber Operations, Rule 82, notes 11-12 (2017).

[1] Norms for conduct of war generally bar attacks directed against civilian infrastructure as such, however recent cyber-attacks on national civilian assets like Sony Pictures, Inc., and the Democratic National Committee suggest that, if these were perpetrated by States, those States are willfully disregarding such norms.

[2] See, e.g., Clapper, James R. Statement for the Record on Worldwide Threat Assessment of the US Intelligence Community, Senate Select Committee on Intelligence, 3 (Feb. 9, 2016), https://www.intelligence.senate.gov/sites/default/files/wwt2016.pdf.

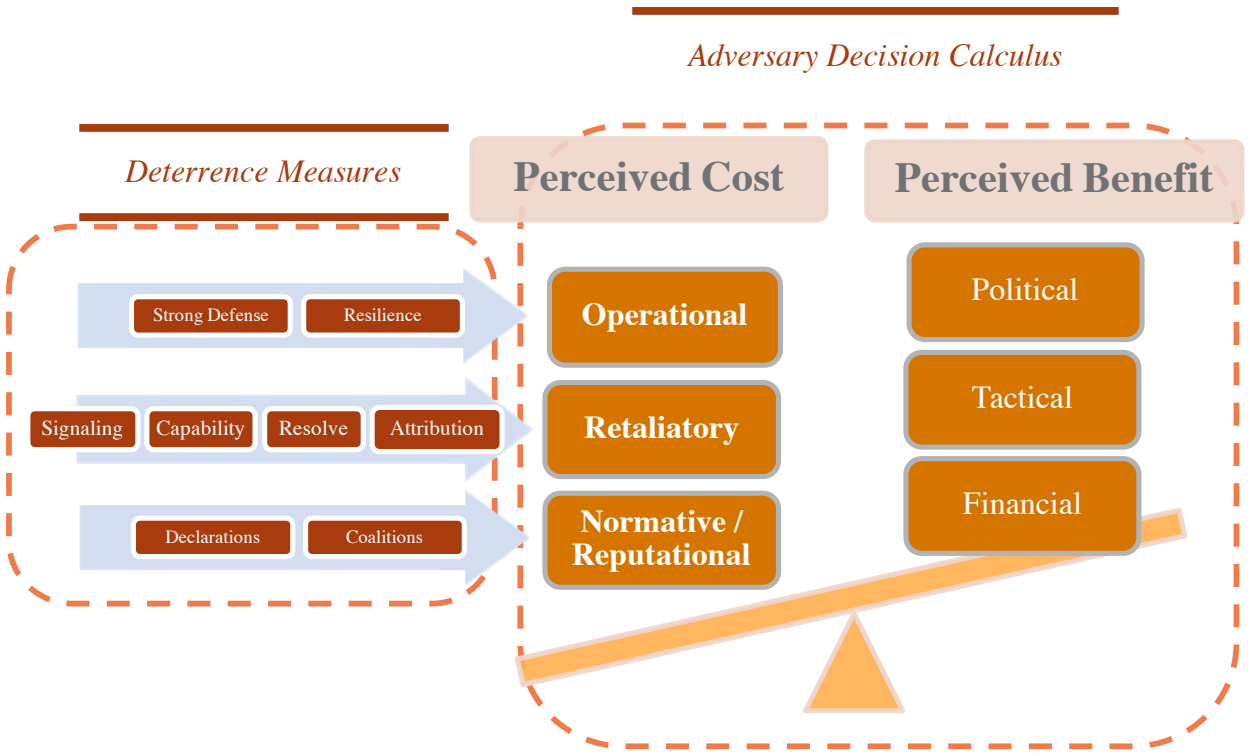

Figure 3. Successful Cyber Deterrence Visualized

To deter malicious actions, a State must ensure that any would-be attacker believes that the costs of the attack will likely outweigh the benefits. This can be achieved through demonstrating to an adversary that its attacks are unlikely to achieve its objectives, or that the consequences for an attack (even a failed one) will be unacceptably high. Without this ***strategic communication*** or ***signaling***, an adversary is much more likely to miscalculate and conduct harmful attacks. As Schelling notes, "[t]hreats are no good if they cannot be communicated to the person for whom they are intended; extortion requires a means of conveying the alternatives to the intended victim."[1] Further, since deterring any and all cyber-attacks is likely not feasible considering the huge number and variety of attacks, States will likely achieve more effective deterrence by articulating certain thresholds beyond which a retaliatory response is likely or assured. The United States has already articulated some such thresholds, discussed in detail below.

First, signaling can be implicit or explicit. T**he visible creation of Computer Emergency Response Teams and continued public efforts to reduce vulnerabilities, provision of assistance to industry in both resilience and reconstitution efforts, and open communication of development of attribution and offensive capabilities each offer implicit signals.** No specific adversary is mentioned, but the efforts increase the perceived cost and risk of executing attacks.

To demonstrate that attacks are unlikely to succeed in their objectives, States should deploy a ***strong defense***, which can consist of both "Active Defense" and "Passive Defense." Unlike active defense as

[1] Schelling, Thomas C. *The Strategy of Conflict: [With a New Preface]*, Nachdr. d. Ausg. 1980 (Cambridge, Mass.: Harvard Univ. Press, 20), 146.

found in traditional military operations, the Department of Defense's Dictionary of Military and Associated Terms defines active defense as "the process of analysts monitoring for, responding to, and learning from adversaries internal to the network."[1] Active defense has often, in public discourse, been misconstrued as being synonymous with "hacking back."[2]

In contrast, passive defense contains security measures that require no or minimal human intervention for their operation.[3] Passive defenses, in turn, are built on a strong cyber security **architecture**. *Cyber security architecture* is "the planning, establishing, and upkeep of systems with security in mind."[4] In addition to architecture, another recurrent theme that emerges in the aftermath of malicious cyber activity is the importance of cyber **education** for the populace. As one report notes, "some of the most notorious cybercrimes in recent history — such as the attacks on major banks, media companies and even security firms — started with just one person clicking on a spear-phishing email."[5]

A further means for a State to signal that cyber-attacks against it are unlikely to achieve their objectives is through establishing robust capabilities for cyber ***resilience*** and ***reconstitution***. Just as civil defense efforts of the early Cold War focused on building bunkers and shelters,[6] efforts can be made to ensure networks and systems are able to recover from attacks, and systems continue to operate in a cyber-degraded environment. Architecting cyber systems with security in mind will generally send a stronger signal and enhance the resilience of a target more than adopting security measures as ad hoc stop-gaps. Resilience as defined by Presidential Policy Directive 21 (PDD-21) is "…the ability to prepare for and adapt to changing conditions and withstand and recover rapidly from disruptions."[7] The PPD goes on to say, "Resilience includes the ability to withstand and recover from deliberate attacks, accidents, or naturally occurring threats or incidents,"[8] or to reconstitute themselves after a disruption.

The message potential adversaries receive from these signals is this: The more resilient the system, the lower the chances of success and the more costly the attack. Just as the nuclear triad complicates an adversary's targeting calculations—can *all* the weapons be destroyed in a first strike?—a resilient cyber adversary creates doubt in the attacker's mind about the chances for success. A power grid, for example, that can operate despite a loss of its network-enabled components is a much less an attractive target to adversaries. While the risk of getting caught remains the same, the chance of achieving the objective is much lower. Another example of improving resilience is the creation of, and investment in, the U.S.-CERT. The U.S.-CERT exists not just to identify, analyze, and issue alerts about malware and

---

[1] Ibid., 5.
[2] "The Sliding Scale of Cyber Security," 9, accessed January 23, 2017, https://www.sans.org/reading-room/whitepapers/analyst/sliding-scale-cyber-security-36240.
[3] Ibid., 8.
[4] Ibid., 5.
[5] "Spear Phishing Attacks - FireEye Whitepaper," accessed April 5, 2017, https://smlrgroup.com/wp-content/uploads/2016/11/wp-spear-phishing-attacks.pdf.
[6] Wohlstetter, Albert. 1959. "The Delicate Balance of Terror." *Foreign Affairs* 37 (2):25. Accessed January 24, 2017.
[7] 2013. Presidential Policy Directive 21 - Critical Infrastructure Security and Resilience. Edited by Executive Office of the President of the United States. Washington, DC. Accessed April 6, 2017. https://obamawhitehouse.archives.gov/the-press-office/2013/02/12/presidential-policy-directive-critical-infrastructure-security-and-resil
[8] Ibid.

other cyber threats, but also to provide a sort of emergency first response to industry. In other words, it functions as a public demonstration that cyber-attacks on these sectors may not be as impactful as an adversary hopes.[1]

This is further reinforced by other efforts by U.S. Government officials and Departments to promote resiliency and reconstitution in critical areas. The Department of Homeland Security (DHS) publicly lists which the U.S. Government considers critical infrastructure, including things like the defense industrial base, various electrical generation and transmission, emergency services, and, now, election infrastructure.[2] DHS need not reference any particular adversary. Through its public listing, this measure outlines what the United States places off limits.

Another element critical to cyber deterrence is ***detection.*** Detection is particularly important to the concept of cyber deterrence, as cyber weapons are frequently designed to be stealthy. Software or equipment failures may occur in normal operation; a determination of whether or not a cyber weapon is responsible requires the ability to detect the weapon. Furthermore, unlike with the launch of an ICBM, no robust system exists that can immediately detect the launch of any cyber-attack.[3] All too often cyber-attacks may go undetected for weeks, or even months. Knowing that an attack occurred is the predicate upon which attribution, reconstitution, and retaliation are all built. Without proper and timely detection, deterrence options are constrained.

Public efforts to improve detection capabilities—some undertaken by industry, some by government and other organizations—serve as yet another deterrent effort: an adversary cannot guarantee they will remain hidden. Though still alarmingly high, the decreasing length of time an adversary can typically remain in a network before discovery is a positive trend.[4] Continued improvement further reduces the value of an attack and raises the risk that the attacker will be uncovered.

Detection therefore is a close cousin to ***attribution***, as without the ability to reliably assign responsibility for attacks in short order, any threat of retaliation is far less credible than in other domains.[5] If an adversary can weaken a nation's resolve to retaliate by removing certainty about who or what to retaliate against, then retaliation is less likely to deter the adversary. As was explained in Section 3,

---

[1] 2017. "US-CERT: United States Computer Emergency Response Team." Department of Homeland Security, accessed June 13, 2017. https://www.us-cert.gov/about-us.

[2] The full DHS list of critical sectors includes: election infrastructure; chemical facilities; commercial facilities; communications; critical manufacturing; dams; defense industrial base; emergency services; energy; financial services; food and agriculture; government facilities; healthcare and public health; information technology; nuclear reactors, material, and waste; transportation systems; and water and wastewater systems. "Statement by Secretary Johnson on the Designation of Election Infrastructure as a Critical Infrastructure Subsector | Homeland Security," accessed May 4, 2017, https://www.dhs.gov/news/2017/01/06/statement-secretary-johnson-designation-election-infrastructure-critical.

[3] Nye, Joseph S. 2017. "Deterrence and Dissuasion in Cyberspace." *International Security* 41 (3):28. doi: 0.1162/ISEC_a_00266. Accessed February 6, 2017. https://www.belfercenter.org/sites/default/files/files/publication/isec_a_00266.pdf

[4] Mandiant Consulting M-Trends 2016 Report

[5] Lin, Herbet. 2016. Hoover Institution Aegis Paper Series on National Security, Technology, and Law. "Attribution of Malicious Cyber Incidents: From Soup to Nuts," Hoover Institution, Stanford University. Accessed October 20, 2016. http://www.hoover.org/sites/default/files/research/docs/lin_webready.pdf.

deterrence is the product of (capability) multiplied by (resolve) multiplied by (belief). Without belief, deterrence is weak. Attribution is the prerequisite for a credible threat of retaliation.[1]

As a result of the unacceptably destructive power of nuclear weapons, mounted on delivery systems that could not be reliably stopped, 20th century deterrence theory centered on the threat of unacceptable retaliation to any attack. If a State can reliably attribute cyber-attacks to a given adversary, said State must still have ***retaliation*** strategies, capabilities, and policies that serve to convince the adversary that the costs of acting outweigh the benefits. Having altered an adversary's cost-benefit calculus, the malicious action is deterred.[2]

Finally, while the cyber deterrence components outlined in this section interact with each other with varying degrees of feedback, ***international norms*** have a reinforcing effect on *all* the actions available to actors. Establishing a norm against destructive behaviors or actions, or conversely supporting constructive behaviors or actions, can serve as a useful tool in ensuring acceptable behavior. The Obama Administration's International Strategy for Cyberspace provides an example of this dynamic.

6.3 Certain U.S. Cyber Deterrence Implementation Measures

While the United States has already taken a number of actions to lay the foundation for this model for cyber deterrence, additional action is needed. As of the time of writing, President Trump had yet to fully develop a cyber security policy. His Executive Order (E.O.) 13800 on strengthening cyber security requests reports from executive agencies on a number of cyber security topics.[3] The requested reports are to cover the following: the best ways to support cyber security of critical infrastructure; the U.S. strategic options to deter adversaries from cyber threats; and international cyber cooperation priorities and others.[4] Because those reports are either not yet completed or have not been made publically available, and the Trump White House is still establishing its cyber priorities, this paper relies on policy statements from previous administrations.

E.O. 13694, issued by President Obama in April 2015 and revised in December 2016, provides clear language on what the United States considers unacceptable. The E.O., which targets the individuals engaging in cyber-attacks, lists a series of actions which would prompt a response:

(A) Harming, or otherwise significantly compromising the provision of services by, a computer or network of computers that support one or more entities in a critical infrastructure sector;

(B) Significantly compromising the provision of services by one or more entities in a critical infrastructure sector;

(C) Causing a significant disruption to the availability of a computer or network of computers; or

---

[1] Clark, David D. and Susan Landau. 2011. "Untangling Attribution." *Harvard Law School National Security Journal* 2. Accessed October 21, 2016. http://harvardnsj.org/2011/03/untangling-attribution-2/

[2] Whether or not certain adversaries can be deterred at all is beyond the scope of this work.

[3] 2017. E.O.13800 of May 11, 2017: On Strengthening the Cybersecurity of Federal Networks and Critical Infrastructure. Edited by Executive Office of the President of the United States of America. Washington, DC: Federal Register. Accessed September 1, 2017. https://www.whitehouse.gov/the-press-office/2017/05/11/presidential-executive-order-strengthening-cybersecurity-federal.

[4] Ibid.

(D) Causing a significant misappropriation of funds or economic resources, trade secrets, personal identifiers, or financial information for commercial or competitive advantage or private financial gain;[1]

(E) Tampering with, altering, or causing a misappropriation of information with the purpose or effect of interfering with or undermining election processes or institutions.[2]

E.O. 13964 aims to freeze the assets of individuals or entities outside the United States who are involved with malicious cyber actions directed toward the United States; its effectiveness is further improved by the well-known willingness of the United States to use economic measures in such a way. Other statements (less explicit in their consequences, perhaps, but no less explicit in their intent) can be found in the form of remarks and speeches given by agency heads and President Obama. One document produced by the White House in 2011, entitled "International Strategy for Cyberspace," states that:

> "The United States will, along with other nations, encourage responsible behavior, and oppose those who would seek to disrupt networks and systems, dissuading and deterring malicious actors, and reserving the right to defend these vital assets as necessary and appropriate. […] We reserve the right to use all necessary means -- diplomatic, informational, military, and economic -- as appropriate and consistent with applicable international law, in order to defend our Nation, our allies, our partners, and our interests."[3]

A stronger, more recent, response was given in response to alleged Russian interference in the 2016 Presidential Election. In the wake of the attacks, President Obama stated "I have issued an [E.O.] that provides additional authority for responding to certain cyber activity that seeks to interfere with or undermine our election process and institutions, or those of our allies or partners."[4,5] This single statement does two important things. First, it designated election infrastructure as critical infrastructure, thereby placing it under the aegis of the DHS. Secondly, it signaled to potential adversaries the value the United States places on election integrity and clearly indicated that interference in U.S. elections will provoke a response.

---

[1] 2015. E.O13694 of April 1, 2015: Blocking the Property of Certain Persons Engaging in Significant Malicious Cyber-Enabled Activities. Edited by Executive Office of the President of the United States of America. Washington, DC: Federal Register. Accessed April 10, 2017. https://www.treasury.gov/resource-center/sanctions/Programs/Documents/cyber_eo.pdf

[2] Ibid.

[3] 2011. International Strategy for Cyberspace: Prosperity, Security, and Openness in a Networked World. Edited by Executive Office of the President of the United States. Washington, DC. Accessed https://obamawhitehouse.archives.gov/sites/default/files/rss_viewer/internationalstrategy_cyberspace.pdf

[4] Secretary, Office of the Press. 2016. Statement by the President on Actions in Response to Russian Malicious Cyber Activity and Harassment. Washington, DC: Executive Office of the President of the United States. Accessed February 13, 2017. https://obamawhitehouse.archives.gov/the-press-office/2016/12/29/statement-president-actions-response-russian-malicious-cyber-activity

[5] 2015. E.O. 13694 of April 1, 2015: Blocking the Property of Certain Persons Engaging in Significant Malicious Cyber-Enabled Activities. Edited by Executive Office of the President of the United States of America. Washington, DC: Federal Register. Accessed April 10, 2017. https://www.treasury.gov/resource-center/sanctions/Programs/Documents/cyber_eo.pdf

The United States furthered signaled how seriously it values cyber security in President Trump's elevation U.S. Cyber Command to a Unified Combatant Command focused on cyberspace operations.[1] According to President Trump, doing so "demonstrates our increased resolve against cyber threats and will help reassure our allies and partners and deter our adversaries."[2]

With respect to international norms, the International Strategy for Cyberspace contains numerous statements that demonstrate the behaviors the U.S. finds acceptable, and the norms that evolve from them. It stresses the openness of the internet, particularly as a tool for the "free flow of information," trade and commerce,[3] along with the transformative impact it has had on people's daily lives, a salvo against the notion that cyberspace is one more domain dominated by militarization. It argued that cyberspace should "…remain a level playing field that rewards innovation, entrepreneurship, and industriousness"[4] and not the realm of national censorship and theft of intellectual property—chipping away at the "black box" model of States that certain authoritarian countries advocate in national internet schemes. It further opined that nations should cooperate in law enforcement, but in accordance with fundamental freedoms and privacy.

Though the document touches on behaviors that are to be avoided (e.g., saying that States should not "…arbitrarily disrupt the free flow of information to create unfair advantage"[5]), the Strategy does not explicitly identify behaviors or targets whose compromise is unacceptable to the United States; that is left to other statements and actions like E.O. 13964. **The International Strategy for Cyberspace demonstrates instead the idea that norms can have a reinforcing effect on the other elements of deterrence, thus increasing stability to the status quo.** Formal agreements to cooperate on cyber crime, such as the Council of Europe Convention on Cybercrime, can be seen as one part of this effort.

7.0 Fitting 20th Century Deterrence Concepts as Applicable to the Cyber Realm

This section outlines where traditional deterrence, and by extension arms control, concepts do or do not apply in cyberspace. For two reasons, this section focuses on comparing cyber to nuclear deterrence (as opposed to other types of weapons). First, as noted in Section 3, while deterrence as a concept far pre-dates the advent of nuclear weapons, deterrence theory and practice reached a zenith during the Cold War in connection with the grave potential consequences of nuclear attack. Second, as detailed in Section 6, some of the clearest examples of treaties and agreements impacting deterrence and imparting norms of behavior grew out of the need to curb the Cold War nuclear arms race.

---

[1] 2017. Statement by President Donald J. Trump on the Elevation of Cyber Command. Washington, DC: Executive Office of the President of the United States. Accessed September 1, 2017. https://www.whitehouse.gov/the-press-office/2017/08/18/statement-donald-j-trump-elevation-cyber-command.
[2] Ibid.
[3] 2011. International Strategy for Cyberspace: Prosperity, Security, and Openness in a Networked World. Edited by Executive Office of the President of the United States. Washington, DC. Accessed https://obamawhitehouse.archives.gov/sites/default/files/rss_viewer/internationalstrategy_cyberspace.pdf
[4] Ibid.
[5] Ibid.

To start, weaponization of cyberspace lacks many of the physical characteristics of weaponization of nuclear programs that make deterrence more effective for nuclear weapons than cyber. The **following are some crucial physical distinctions for deterrence purposes between cyber and nuclear weapons**.

- Cyber weapons and hostile cyber actions do not require large, costly industrial complexes to assemble. Thus, there is no extended break-out period within which diplomacy can be used to stop weapon development.
- Cyber weapons have more plausible deniability than highly-enriched uranium or separated plutonium; cyber activities have numerous beneficial and legitimate applications, while highly-enriched uranium and weapons-grade plutonium do not. When a country has vast quantities of fissionable materials, it is difficult for them to convince others that those materials are only for peaceful purposes.
- Cyber weapons are difficult to put on display in parades, even if a country decided it wanted to display its capabilities.
- Cyber weapons cannot be counted because they are easily destroyed and replicated.
- There is no cyber early warning system watching for the thermal trail of an incoming cyber missile.
- An army's cyber second-strike capabilities are always in question because on any given day a commercial vendor's patch may render a cyber weapon inert.

In addition to differing physical attributes, signaling one's capability and willingness to use a weapon in cyberspace differs from nuclear signaling. For example, during the Cold War, countries tested new nuclear weapon designs to demonstrate their strength and signal intent to use the weapons if provoked. But in the cyber realm, demonstrating a cyber weapon may in fact render it useless for the future. This is because cyber weapons are built to exploit vulnerabilities, often unknown vulnerabilities. But once a vulnerability is known, that vulnerability can often be eliminated with an inexpensive software patch or system reconfiguration. Once a highly prized "zero-day" attack is used (and detected), its value is greatly diminished as the target can often patch the vulnerability. Thus, cyber deterrence cannot be built upon the demonstration of a credible weapon.

Nuclear deterrence theory also is an imperfect concept for cyberspace given the challenge of attribution. For example, a nuclear ICBM launched from within a country's territorial borders is strong evidence that the attack can be attributed to that State (assuming that non-State actors have not gained access to nuclear weapons and ICBMs).[1] In cyberspace, however, even if it is clear that an attack was coming from within a State's borders, it can be more difficult to attribute that action to the State because use of a cyber weapon does not rely on State support and resources in the way that nuclear and missile programs traditionally do.[2] Moreover, a hallmark of cyber-attacks is their deliberate obfuscation of origin, taking circuitous routes from initiation to their target to reduce chances of detection and/or identification. Determining who is responsible for a cyber-attack and attributing that behavior to a State can be almost

---

[1] This is a simplification of the concept of attribution for a nuclear-launch. For greater details on how and when a wrongful act can legally be attributed to a State, see the International Law Commission's *Draft Articles on Responsibility of States for Internationally Wrongful Acts, with commentaries*[hereinafter *Draft Articles on Responsibility of States*], 2001, *available at* http://legal.un.org/ilc/texts/instruments/english/commentaries/9_6_2001.pdf.

[2] Again, this is still a simplification of attribution for nuclear-launch, assuming it is clear that a State would have to be involved in the development and launch of nuclear-tipped missiles. See *Draft Articles on Responsibility of States*, Article 4, Comment (3) ("That the State is responsible for the conduct of its own organs, acting in that capacity, has long been recognized in international judicial decisions.").

impossible, making it challenging to retaliate or threaten to retaliate because it is not clear against who or what to retaliate. Deterrence relies on the belief that aggression will be met with decisive retaliation, but that is simply not possible in the current cyber landscape.

The chart below uses the elements of deterrence identified in Figure 3 to compare key differences between nuclear deterrence and cyber deterrence.

Table 1. Elements of Deterrence - Comparison between Nuclear and Cyber Deterrence

Element	Nuclear	Cyber
Ability to develop a **Strong Defense**?	Low. Defenses against nuclear attack are limited and costly (such as missile defense).	Moderate. Defense against cyber-attack is dependent on architecture of target systems, and education level of population. It is costly and must constantly be maintained and refreshed as new vulnerabilities are published daily.
Ability to **Detect**?	High. Detection of launch and adversary nuclear programs relatively robust.	Low. No robust system for detection exists for cyber.
Ability to **Attribute**?	High. Attribution relatively straightforward given missile tracking and nuclear forensics.	Low. Attribution often highly difficult given ease of obfuscation by attackers, desire to not disclose sources and methods of cyber forensics.
Ability to **Retaliate**?	High. Massive retaliation for nuclear strike is well-established and credible; supported by relative ease of attribution.	Unclear. Retaliation for cyber strike is unclear in part because attack may not be detected and/or attributed to originator; proportionality for cyber-attacks remains unclear.
Ability to engage in **Strategic Communication** about intent, resolve and capabilities?	High. States have been involved in nuclear strategic communication for decades so communications are relatively well-understood; forward deploying bombers, visible demonstrations of capabilities, treaties and agreements, public statements, etc.	Low. Strategic communication of cyber capabilities is relatively nascent; States have incentives to keep their cyber capabilities opaque and cannot publicize capabilities due to the nature of weapons, targets, and technical sensitivities.
Ability to develop **Resiliency**?	High. Resiliency in nuclear deterrence is generally considered the ability to maintain second-strike capabilities.	High. Resiliency in cyber is dependent on system architecture and procedures.
Ability to **Reconstitute** after attack?	Low. Reconstitution after nuclear war is impossible for all practical purposes.	High. Reconstitution after cyber war is possible, but dependent on architecture and procedures; it may be costly.
Ability to rely on **International Norms** to deter malicious acts?	High. International norms caution against the use of nuclear weapons and proliferation	Limited. International norms related to cyberwar and aggression are still forming; and difficulty of attribution reduces effectiveness of norms.

The fact that many elements of "traditional" nuclear deterrence do not apply in cyberspace limits the ability of policy makers to apply similar deterrence concepts to address challenges posed by cyber aggression today. Importantly, the relative difficulty in detecting, attributing, and retaliating against an aggressor in cyberspace limits the effectiveness of classic arms control treaty models and related deterrence theories. Moreover, the ability to defend against and reconstitute after a cyber-attack lowers the threshold for use of a cyber weapon. However, ***this paper outlines certain features of various arms control treaties that may be useful to employ in thinking about a future treaty to promote cyber stability.*** Such a treaty may prove useful to policymakers, particularly if combined with a robust and diverse set of tools that can be used to better address cyber instability where traditional concepts of deterrence and arms controls fall short. The next section explores which arms control and deterrence elements may work in cyberspace and also outlines other tools or options policymakers may consider to promote stability.

8.0 How Might the United States Promote International Cyber Stability in Light of Cyber Aggression: Technical and Policy Options for Policy Makers

As outlined above, arms control and deterrence are tools used by the international community to promote international stability in the face of destabilizing weapons and aggression. However, given the differences between cyber conflict and kinetic conflict, arms control and deterrence alone are insufficient in bringing stability to the cyber realm. This section explores options to promote cyber stability, some relying on arms control and deterrence concepts and some that do not. These options are generally ordered from "easier" to "harder" based on estimated or perceived costs, difficulty in changing legislation/regulations, challenges in negotiating international agreements, and the time each option could take to complete.

8.1 Signaling

Signaling, as previously noted, is an important aspect of deterrence: it lets an adversary know where lines exist, which lines not to cross, and what to expect if they do. However, the concept is not well executed at present in the cyber domain as demonstrated by the countless cyber-attacks from State and non-State actors alike. Attempts at signaling have been made by the highest echelons of the U.S. government through public statements and written declarations, such as the "International Strategy for Cyberspace." However, the continuation of cyber-attacks demonstrates the message has not translated into a credible threat nor has it created fear among competitor States; both of which are necessary for deterrence to be successful. Below are a few options to ameliorate the problem.

8.1.1 Option 1. Clarifying Possible and Proportionate Retaliation Options for Malicious Cyber Acts

Retaliation is the driving force behind most deterrent paradigms of the past 70 years, and deterrence through the threat of punishment remains a key factor in cyberspace. In addition to the well-documented difficulties with attribution, questions remain as to when an act of cyber aggression rises to the level of a cyber-attack requiring a response. Without a clear consensus or articulated red-line in this area, the

credibility of retaliation is further undermined. The United States has previously made clear that acts of cyber aggression above a certain threshold will be met with kinetic response, and the United States will utilize a range of cyber and military options to respond to a given malicious cyber act. Such responses are often known as "reprisals," under international law and are only lawful when taken against a provoking State who violated international law after trying to resolve the matter without resorting to force, and with a proportionate response.[1] While there are legal limits on the lawful use of reprisals, for the purposes of this discussion the authors assume that all reprisals meet those legal standards. However, the views expressed in this paper are the views of the authors and do not constitute legal advice. After meeting the legal thresholds required as determined by legal counsel, the U.S. government could consider responding to malicious cyber acts with cyber reprisals, or possibly even physical reprisals.

To assist policy makers in deciding how to respond, the U.S. government could rate States according to their cyber dependency to determine if retaliation is proportionate, justified, and effective. This is the essence of a tailored deterrence strategy detailed below.

8.1.2 Option 2. Issuing Policy Statements, Clarifying Thresholds for Action/Reaction

One of the easier steps the U.S. government could take to promote stability with regards to cyberspace is to unilaterally declare the types of cyber-attacks that will trigger a response, and, in general terms, the spectrum of activities the U.S. might undertake as a proportionate response. While President Obama's E.O. 13694 clearly delineated what malicious cyber acts the United States perceives as threats to national security,[2] there is room for additional clarification. For example, the United States could issue statements , public and private, about how it will respond to cyber acts by specific countries, as a type of tailored deterrence. In July 2016, NATO recognized 'cyberspace as a domain of operations in which NATO must defend itself as effectively as it does in the air, on land, and at sea'[3] yet it remains unclear how the United States will respond to acts of cyber aggression against its allies. It has indicated that it is concerned with protecting its allies from cyber aggression as well.[4] As demonstrated by nuclear deterrence and explained in Section 7, the more clarity about a country's perception of and response to threats, the easier it is to reduce risks associated with miscommunication and miscalculation. While determining how to extend cyber deterrence to the U.S. allies could ultimately result in more formal agreements or treaties (a harder, more costly option for the U.S. government), the first step in that

---

[1] Mitchell, Andrew D. *Law of Reprisals: Does One Illegality Merit Another: Law of Belligerent Reprisals in International Law*, 170 Military L. R. 155, 156-57 (Dec. 2001), referencing Responsabilité de L'Allemagne a Raison des Dommages Causés dans les Colonies Portugaises du Sud de L'Afrique, 8 Trib. Arb. Mixtes 409 (1928) (Portugal v. Germany) (The Naulilaa Case), reprinted in 2 R. Int'l Arb. Awards 1011 (1949).

[2] See E.O. 13694, *supra* 1. 2015. E.O. 13694 of April 1, 2015: Blocking the Property of Certain Persons Engaging in Significant Malicious Cyber-Enabled Activities. Edited by Executive Office of the President of the United States of America. Washington, DC: Federal Register. Accessed April 10, 2017. https://www.treasury.gov/resource-center/sanctions/Programs/Documents/cyber_eo.pdf

[3] See Warsaw Summit Communique section 70, 09 Jul 2016, accessed Sep 21, 2017. https://ccdcoe.org/sites/default/files/documents/NATO-160709-WarsawSummitCommunique.pdf

[4] See E.O. 13694, *supra 152*. 2015. E.O. 13694 of April 1, 2015: Blocking the Property of Certain Persons Engaging in Significant Malicious Cyber-Enabled Activities. Edited by Executive Office of the President of the United States of America. Washington, DC: Federal Register. Accessed April 10, 2017. https://www.treasury.gov/resource-center/sanctions/Programs/Documents/cyber_eo.pdf

direction may be to simply issue more unilateral, public, clarifying statements about what the United States views as unacceptable and how it will respond to an adversary who violates U.S.-promoted norms.

Similarly, the United States could unilaterally clarify its position on cyber norms and possible responses using back channels or private communications with potential adversaries. Direct, bilateral communication of boundaries and intentions may allow a specific message to be tailored to the nation in question, allowing greater specificity than a general statement to the entire world.

8.2 Detection and Attribution

Attribution and detection are imperative for successful deterrence. Although capabilities in both processes have improved, more can be done to strengthen these elements.

8.2.1 Option 3. Improve Attribution and Detection Capabilities

Without detection of an attack, attribution and retaliation are impossible to conduct. Though progress has been made in this area—the average lifespan of a cyber compromise has decreased significantly in recent years[1]—the challenge remains that unlike kinetic attacks, malicious cyber-attacks may not be readily detected until well after they have been initiated. Improving detection at the national level (to parallel the radars and satellites that provide detection and early warning of a ballistic missile launch) may be more difficult but is an option worth considering. Increased investment into national-level detection systems and the policies to enable their effective use may yield fruitful results.

In addition and in relation to detection, attribution remains a problem. Attribution of an attack is perhaps the most oft-cited and important difference between the cyber and physical world; without knowing who is responsible for an action, that actor cannot be deterred with credible threats of retaliation.

Part of the difficulty with attribution is timely tracing of an attack definitively back to its originator, because the originating State has numerous tools at its disposal to obfuscate its involvement. Solving this problem will involve heavy investments in international law enforcement capabilities and coordination and even with such investments, there is no guarantee that all malicious action will be traceable.

Even when an attack can be traced back to its origin, attribution difficulties must still be overcome. Namely, neither U.S. citizens nor the international community, at large, are willing to accept an unsubstantiated assertion of attribution by a U.S. government agency. Transparency is needed, which means that substantial data must be provided to underpin the assertion. However, full transparency would require that sources and methods be revealed. Law enforcement is unlikely to choose to reveal all of their capabilities because doing so would enable adversaries to avoid detection in the future. Protocols need to be established that set forth norms for the amount of transparency required for cyber activities of varying levels of impact.

---

[1] According to Mandiant reporting the global attacker dwell time on a compromised system was 146 days in 2015 and dropped to 99 days in 2016. Mandiant RPT-M-Trends-2017 and M-Trends 2016-EMEA. Hau, Bill, Matt Penrose, Tom Hall, and Matias Bevilacqua. 2016. M-Trends. "M-Trends 2016: EMEA Edition," Mandiant, a FireEye Company. Accessed August 31, 2017. https://www2.fireeye.com/rs/848-DID-242/images/M-trends-2016-EMEA-NEW.pdf; 2017. "M-Trends 2017: A View From the Front Lines," Mandiant, a FireEye company. Accessed September 1, 2017, https://www2.fireeye.com/rs/848-DID-242/images/RPT-M-Trends-2017.pdf.

8.3 Retaliation

Instilling fear into an adversary's mind is at the very heart of any deterrence strategy. Sound retaliatory options and capabilities are a must for any successful deterrence strategy. Below are a few options to consider.

8.3.1 Option 4. Tailored Deterrence

The United States could develop tailored deterrence strategies specific for States considered to be adversaries. This option recognizes that not all adversaries are created equal, and therefore any retaliation could be most effective if adjusted accordingly. The pressure points of each State are different. For example, what motivates actors in a democracy differs from what motivates actors in a totalitarian State, and States reliant on cyberspace are different than those nations lacking widespread internet penetration and adoption. The United States, for example, has a large attack surface due to its dependence on cyberspace. This makes the U.S. both vulnerable to an attack and less effective at controlling upward movement through the escalatory ladder. As James Clapper points out, the reason the United States restrains itself from retaliating to cyber-attacks is the "lack of confidence in our ability to absorb a counter-retaliation."[1] Understanding these pressure points in advance, and updating these tailored deterrence strategies as each State evolves, could help policy makers communicate and signal appropriately to its different adversaries.

8.3.2 Option 5. Name and Shame

The United States could continue to publically expose and condemn perpetrators of cyber-attacks. By publically identifying States that are committing malicious cyber acts, the United States can build an international collation against that State. For example, the Obama Administration previously publicly accused North Korea for its alleged attack on Sony Corporation.[2] Such public approbation supported international efforts to tighten sanctions on North Korea. Naming and shaming tactic has been executed in several other cyber instances as well.[3] Moreover, by issuing public statements condemning certain actions, the United States can contribute to the development of customary international law against cyber-attacks.

8.3.3 Option 6. Enhanced Coordinated Use of National Criminal Laws

Given the aforementioned difficulty in attributing malicious action in cyberspace to a particular State entity, the United States may consider how to deter individual actors who work as proxies for States.

---

[1] "Hearing to Receive Testimony on Cyber Policy, Strategy, and Organization," § Committee on Armed Services, 10, accessed June 29, 2017, https://www.armed-services.senate.gov/imo/media/doc/17-45_05-11-17.pdf.
[2] "Remarks by the President in Year-End Press Conference," *Whitehouse.gov*, December 19, 2014, https://obamawhitehouse.archives.gov/the-press-office/2014/12/19/remarks-president-year-end-press-conference.
[3] "FACT SHEET: Actions in Response to Russian Malicious Cyber Activity and Harassment," *Whitehouse.gov*, December 29, 2016, https://obamawhitehouse.archives.gov/the-press-office/2016/12/29/fact-sheet-actions-response-russian-malicious-cyber-activity-and. See also "U.S. Charges Five Chinese Military Hackers for Cyber Espionage Against U.S. Corporations and a Labor Organization for Commercial Advantage," accessed June 29, 2017, https://www.justice.gov/opa/pr/us-charges-five-chinese-military-hackers-cyber-espionage-against-us-corporations-and-labor.

States may be co-opting criminal tactics in the cyber realm, using proxies to undermine cyber stability to accomplish the goals of the State. By implementing and enforcing strong criminal laws against cyber crimes in a coherent manner, coordinated across multiple U.S. agencies, the United States could deter the individuals who act on behalf of States in cyberspace. Though this paper has focused on almost exclusively on State-level actors for its analysis pf deterrence and arms control theory, it is nonetheless true that individuals make up the State apparatus and laws that make life difficult for State-sponsored cyber criminals could serve the U.S. national security interest and deter malicious cyber activity. Moreover, as cyber crime increases, it becomes easier for States to take malicious action in the cyber realm, hiding behind the general sense of lawlessness in the cyber realm. The more order that can be meted out in cyberspace, the more difficult it will be for States to hide behind criminal organizations, complicated server arrangements, and proxies.

The United States already has a number of criminal statutes with which to prosecute the cybercrimes that are the foundation of cyber aggression or foreign policy. The United States can use those criminal statutes to deter the individuals who would act as proxies for State adversaries. For example, U.S. criminal statutes prohibit "fraud and related activity in connection with computers,"[1] including criminalization of transmitting code with the intent to cause damage to computers,[2] and counterfeit access device fraud.[3] These statutes were used to indict Russian Federal Security Service officers and criminal conspirators in connection with the hacking of Yahoo email accounts, which allegedly included email accounts for U.S. and Russian government officials.[4] The U.S. Department of Justice previously filed criminal charges against individuals who were allegedly part of Chinese government efforts to engage in cyber espionage for commercial advantage.[5] These two examples demonstrate how prosecutions of cybercrimes can be used to deter would-be proxies.

However, simply indicting foreign nationals for cyber crimes is less effective where the United States does not have those actors in physical custody. To assist in apprehension of criminals, and in investigation and prosecution of cybercrime, the United States is party to a number of mutual legal assistance treaties (MLATs)[6] and the Budapest Convention on Cybercrime, which offers a framework for mutual legal assistance between States in the absence of bilateral or multilateral MLATs.[7] By leveraging MLATs and domestic criminal prosecutions, the United States may be able to deter would-be cybercriminals from operating as State proxies. Where physical jurisdiction is not possible, the United

---

[1] "[USC04] 18 USC 1030: Fraud and Related Activity in Connection with Computers," accessed June 29, 2017, http://uscode.house.gov/view.xhtml?req=%28title:18%20section:1030%20edition:prelim%29%20OR%20%28granuleid:USC-prelim-title18-section1030%29&f=treesort&edition=prelim&num=0&jumpTo=true.

[2] *Ibid*. § 1030 (a) & (c).

[3] *Ibid*. § 1029.

[4] "U.S. Charges Russian FSB Officers and Their Criminal Conspirators for Hacking Yahoo and Millions of Email Accounts," accessed June 29, 2017, https://www.justice.gov/opa/pr/us-charges-russian-fsb-officers-and-their-criminal-conspirators-hacking-yahoo-and-millions.

[5] Ibid.

[6] "COUNTRY PROFILE - MLAT," accessed June 29, 2017, https://mlat.info/country-profile/united-states.

[7] *See* Council of Europe, Convention on Cybercrime, *opened for signature* Nov. 23, 2001, 41 I.L.M. 282 (hereinafter "Convention on Cybercrime" or "the Convention"). As of writing, fifty-nine States have signed the Convention, including all of the members of the Council of Europe except for Russia (see Council of Europe, Chart of signatures and ratifications of Treaty 185, status as of June 26, 2017, *available at* http://www.coe.int/en/web/conventions/full-list/-/conventions/treaty/185/signatures.

States can nonetheless freeze the assets of proxy-criminals and restrict their freedom to travel. Restricting market access and reducing the freedom of State agents to travel freely with various countries for fear of being arrested and prosecuted, or extradited to the United States, can impact the motivation and ability of States to carry out acts of cyber aggression.

The United States could consider developing a credible and consistent cybercrime prosecution strategy that would serve national security interests. Rather than prosecuting cases in a one-off, haphazard manner, the United States could develop a policy for prosecution to support cyber deterrence that carefully considers both prosecutorial independence and complications related to foreign sovereign immunity. Prosecutorial independence is important for protecting the integrity and autonomy of the Department of Justice and is best preserved when prosecutors are given pre-prosecution direction, rather than being told mid- or post-prosecution that a prosecution should not move forward due to policy concerns.

Foreign sovereign immunity is the idea in customary international law that foreign States are not subject to jurisdiction of the courts of another State.[1] In the United States, foreign sovereign immunity generally denies U.S. courts jurisdiction over foreign governments, unless those governments are engaged in activity that could otherwise be carried out by private persons (such as commercial activity), the State explicitly or implicitly waives immunity, or in a number of other limited circumstances.[2] Foreign sovereign immunity generally would prevent U.S. courts from prosecuting a State agent for illegal acts in cyberspace. The United States would seek to prosecute a State proxy in U.S. courts in two possible outcomes. In the first outcome, the offending State could claim foreign sovereign immunity as a bar to prosecution, effectively acknowledging that the indicted individual is an actor of that State. In such a situation, domestic legal solutions would no longer be applicable, but then the United States would have solved the attribution puzzle and could pursue policy or military actions, as appropriate. In the second outcome, the offending State could deny any connection to the indicted individual, allowing its proxy to be prosecuted without State assistance. Such an outcome, while potentially protecting the State from retaliation, would undermine morale among those individuals acting on behalf of the State and might deter from them involvement in malicious cyber activities.

8.3.4 Option 7. International Sanctions

Among U.S. tools for coercion and retaliation, economic sanctions have increased in prominence significantly over the last three decades. The U.S. President has considerable powers, both inherent and legislatively authorized, to unilaterally impose sanctions at various levels against various subjects, from natural and juridical persons, to economic sectors, to countries.[3] While comprehensive overviews of sanctions powers have been produced by numerous authors,[4] this section will focus on how sanctions powers can be deployed in a tailored cyber deterrence framework.

---

[1] "Foreign Sovereign Immunities Act," accessed June 29, 2017, https://travel.state.gov/content/travel/en/legal-considerations/judicial/service-of-process/foreign-sovereign-immunities-act.html.
[2] Ibid.
[3] For instance, Congress has granted broad sanction authority under the International Emergency Economic Powers Act and the National Emergencies Act. 50 U.S.C. 1701 *et seq*.; 50 U.S.C. 1601 *et seq*.
[4] See, e.g., Corn, Geoffrey S., Jimmy Gurule, and Jeffrey D. Kahn. 2016. "13: Economic Powers and National Security." In *National Security Law and the Constitution* edited by Geoffrey S. Corn. Wolters Kluwer Law & Business.

The basic concept is the same as that already described for deterrence writ large: the United States would signal that it is willing to bring its sanctions powers to bear in response to certain serious hostile cyber actions in a manner that would target what potential adversaries value, and sketch elements of how a response would proceed in such a way as to make the threat credible.

During his tenure in office, President Obama took significant steps to align the President's existing authority to impose economic sanctions with a strategy to deter hostile cyber actions. In April 2015 and again in December 2016, President Obama issued E.O.s that plainly stated that the United States reserves the right to impose economic sanctions on designated persons who "engaged in cyber-enabled activities […] that are reasonably likely to result in, or have contributed to, a significant threat to the national security, foreign policy, or economic health or financial stability of the United States."[1] The 2016 order, E.O. 13757, further specified attempts to compromise elections processes or institutions as grounds for sanctions, and designated five entities and four individuals based in Russia believed to have been part of that country's assessed efforts to influence the 2016 U.S. presidential election.[2] President Trump has signed legislation imposing more sanctions on Russia, Iran, and the DPRK.[3] However, his Administration's sanctions policy with regards to cyber have yet to be articulated as the Administration awaits the outcomes of the reports requested in E.O. 13800.

Economic sanctions as a policy tool suffer from a number of well-documented limitations that would need to be accounted for in crafting a meaningful "tailored" response to deter hostile cyber actions. Here follows a summary of such limitations highlighted by Haas.[4] When targeted against specific individuals and entities, sanctions are especially easy to subvert. The persons in question are frequently able to register front companies to mask their beneficial ownership and carry on with ostensibly prohibited business, and their business counterparts generally have a strong economic interest in allowing transactions to go forward. Sanctions are also far less effective when implemented unilaterally by only one country; persons or countries subject to such sanctions can compensate for lost trade opportunities or acquire sensitive technology elsewhere. Sanctions tend to have greater effect, therefore, when they are more general in scope (on the sectoral or country level) and multilateral.

However, more general sanctions tend disproportionately to affect people other than those whom policy makers sought to influence through imposing the sanctions. Hostile foreign governments are often controlled by elites who are well-insulated from economic hardship, hence sanctions have greater impact on the broader population in a country than on decision makers. Elites in a country may feel an impact if discontent should boil over and the larger populace should seek major political change; however, this is rare. Sanctions' effects tend to be longer-term, such as economic malaise and general discontent, and may be unlikely to impose meaningful costs within the short time horizons that are often the focus of everyday

---

[1] E.O. 13694 of Apr. 1, 2015, Federal Register, Vol. 80, No. 63 (Apr. 2, 2015); E.O. 13757 of Dec. 28, 2016, Federal Register, Vol. 82, No. 1 (Jan. 3, 2017).
[2] U.S. Office of the Director of National Intelligence, "Background to 'Assessing Russian Activities and Intentions in Recent US Elections': The Analytic Process and Cyber Incident Attribution" (Jan. 6, 2017).
[3] 2017. Statement by President Donald J. Trump on Signing the "Countering America's Adversaries Through Sanctions Act". Washington, DC: Office of the Press Secretary. Accessed September 1, 2017. https://www.whitehouse.gov/the-press-office/2017/08/02/statement-president-donald-j-trump-signing-countering-americas.
[4] Haas, Richard N. 1998. Policy Brief. "Economic Sanctions: Too Much of a Bad Thing," Brookings Institution. Accessed August 31, 2017. https://www.brookings.edu/research/economic-sanctions-too-much-of-a-bad-thing/.

or crisis decision-making. To the extent sanctions do have effects, these may be different than correcting the behavior that prompted the sanctions. A more immediate means for authoritarian elites to address increasing public opposition may be to crack down on dissent, rather than alter the course that resulted in sanctions.

The history of U.S. economic sanctions bears out these issues. Through the 1990s and 2000s, onerous international sanctions are generally regarded to have had relatively little effect on the insular authoritarian regimes of Iraq and North Korea, despite imposing punishing economic costs on those societies.[1] The case of sanctions against Iran over its nuclear program stands in stark contrast, potentially a result of Iran's relatively more pluralistic government and oil-export-oriented economy, where the economy and policy over sanctions played a significant role in the country's presidential elections of 2013.[2] The success of sanctions against Russia for its occupation of Crimea and other recent aggressive conduct vis-à-vis Ukraine remains unclear; Russia's economy has contracted considerably since sanctions were imposed in 2014; however, the scope of the sanctions regime was narrower than the observed contraction, and much of Russia's economic difficulty can be explained by declines in world fossil fuel prices.[3]

To date record of the Obama Administration cyber sanctions policy implementation has been limited, hence its overall efficacy is difficult to gauge with precision. Following the high-confidence assessment of the U.S. intelligence community on the fact of Russia's attempt to interfere in the 2016 U.S. elections, the Obama Administration designated a small group of Russian persons and entities as being subject to retaliatory sanctions. As these sanctions were entity-specific and unilateral, they run the risk that the designated entities will evade them using front companies. The highly politicized nature of the underlying events and break in policy between administrations may further undermine the intended deterrent effect. Indeed, since the December 2016 U.S. sanctions, U.S. officials and the governments of various European allies of the United States continue reporting hostile cyber actions against those countries' election processes consistent with the 2016 U.S. experience, suggesting that the similar hostile Russian cyber actions may not have abated or been deterred.[4]

General lessons on how to apply sanctions to more effectively deter hostile cyber actions could include the following.

- Sanctions are likely to be more effective tools of deterrence when they have relatively broad scope (they cannot be defeated by simple re-registration for single entities) and have multilateral participation.
- Particularly as evidence mounts of a broader campaign of hostile cyber action against election infrastructure in multiple countries, a multilateral sanctions regime likely becomes more viable.

---

[1] See, e.g., Kaplan, Robert. 2009. "North Korea, the Next Iran?" *The Atlantic*, May 2009, 1. Accessed August 31, 2017. https://www.theatlantic.com/magazine/archive/2009/05/north-korea-the-next-iraq/307542/.

[2] See, e.g., Cassidy, John. 2013. "Iran Nuke Deal: Do Economic Sanctions Work After All?" *The New Yorker*, 25 November 2013, 1. Accessed September 1, 2017. http://www.newyorker.com/news/john-cassidy/iran-nuke-deal-do-economic-sanctions-work-after-all.

[3] See, e.g., World Bank. 2017. "Russia Economic Report 2017: From Recession to Recovery," World Bank Group. Accessed September 1, 2017. http://documents.worldbank.org/curated/en/782451497437509084/Russia-economic-report-2017-from-recession-to-recovery.

[4] See, e.g., "United States Senate Hearing to Receive Testimony on United States Cyber Command", May 9, 2017, Accessed September 6, 2017. https://www.armed-services.senate.gov/imo/media/doc/17-42_05-09-17.pdf.

- Sanctions have more impact against States with at least moderately effective channels for influence by the general population of a country over its government, and where the target country has a relatively vulnerable, export-oriented economy.
- Sanctions are more likely to be effective when accompanied by a comprehensive negotiation strategy coupled with other tools of pressure and influence, such as diplomatic engagement of all stakeholders and the targeted and complementary deployment of hard power.

8.3.5 Option 8. Loud Cyber Responses

In contrast to most cyber activity, in which stealth is paramount, "loud" cyber retaliations deliberately broadcast their presence and origin. A "Loud" cyber-attack leaves no question as to its perpetrator and is intended to be as much communication as attack. Such actions can be undertaken publicly, or tailored for notice by a specific agency or leader, in response to other cyber acts to demonstrate that the United States is aware of the adversary's cyber action and as a warning to cease and desist.

8.3.6 Option 9. Kinetic Response

A kinetic response to cyber aggression ought not be taken without utmost care. Misdirection and redirects are easier to execute in the cyber realm than in the physical world, thus for a kinetic response to be a feasible option in retaliation to a cyber event, attribution is essential. Additionally, the threshold for a kinetic response to cyber aggression is going to remain high—most likely considered in an event of loss of life. However, eventual kinetic response may be unavoidable if retaliations escalate into the physical domain.

8.4 Resilience

The United States, being heavily dependent on the cyber realm, must strengthen the resiliency of its critical infrastructure in order to counterbalance the large attack surface it represents on the international arena.

8.4.1 Option 10. Investing in Resilience and Reconstitution, including Architecture and Education

To deter would-be cyber adversaries, the United States could improve the cyber security posture of critical systems and infrastructure. Building stronger more defensible architectures, investing in education of the population to reduce risk of user errors, and building resilient systems would open other response options. The better protected a target, the costlier it is to attack and the less attractive it becomes.

Another way to improve defense could be to expand efforts within government and the technology industry to build security into systems and networks from the initial design phase. Defensible architectures present a smaller attack surface, limit the access of an attacker who does manage to infiltrate it, and can operate despite interference or damage. This simultaneously lessens the impact of a successful attack while increasing the cost to accomplish it.

The United States could also consider establishing or adopting standards for system architecture related to security, recovery, and reconstitution. Such standards could apply to the federal government and its contractors initially but might be expanded to include private sector actors as well, particularly those working in critical infrastructure. Properly designed and implemented architectures could enable a faster recovery in the event of a successful cyber-attack, giving decision makers both the extent of the damage and the tools to quickly reconstitute lost capabilities.

Innovative technological research and development to improve cyber defenses and critical infrastructure resilience will have a deterrent effect as well. Microgrid technology could be helpful in bolstering resiliency in cyberspace.[1] Increasing the resilience of critical infrastructure increases the cost to the adversary. Programs are already in place in both public and private sectors to increase the development and deployment of microgrid technology.[2]

Lastly, the United States could have two educational campaigns: teach technology users about cyber threats and groom the next generation of cyber security experts. Investing in education programs to encourage people to grow as future cyber security experts could help the United States stay ahead of those who would seek to exploit its cyber vulnerabilities. Such investments could be made to encourage students to pursue careers in cyber security, starting with grade school and continuing through post-graduate studies. The second campaign would target the human factor that is consistently exploited to compromise computer systems. This fact is underscored by the many successful spear-phishing attacks targeting high-profile individuals and organizations. Risk-based user education on likely cyber threats could be combined with refresher training on computer hygiene to raise to the adversarial cost of a successful attack. Greater understanding of user psychology may enhance these education efforts as well. Such training could be employed in U.S. critical infrastructure and government facilities nationwide.

8.4.2 Option 11. International Partnership and Collaboration

The U.S. can significantly enhance the effectiveness of its cyber deterrence posture by leveraging its long-standing alliances and international institutions. **One of the strongest elements of U.S.-backed deterrence in the Cold War was the establishment and maintenance of largely unified international blocs, underpinned by principles of free trade, democratic values, and mutual defense**.[3] These included, for instance, the mutual defense arrangements with Europe via NATO as well as with Japan and the Republic of Korea, which some argue would not have been possible without the high degree of complementary economic and political cooperation engendered in such institutions as the European Community (and then European Union), the Council of Europe, the Organization for Security and Cooperation in Europe (OSCE) and bilateral arrangements with Japan and Korea.[4] This web of strong

[1] "The Role of Microgrids in Helping to Advance the Nation's Energy System | Department of Energy," accessed June 12, 2017, https://energy.gov/oe/services/technology-development/smart-grid/role-microgrids-helping-advance-nation-s-energy-system.
[2] "Microgrid Portfolio of Activities | Department of Energy," accessed June 12, 2017, https://energy.gov/oe/services/technology-development/smart-grid/role-microgrids-helping-advance-nation-s-energy-syst-0.
[3] See, e.g., Holbrooke, Richard, "America, A European Power," Foreign Affairs (Mar 1995): 38; Martin Murphy, "The Importance of Alliances for U.S. Security," The Heritage Foundation 2017 Index of U.S. Military Strength (2016), http://index.heritage.org/military/2017/essays/importance-alliances-u-s-security/.
[4] Ibid.

alliances and interlinkages amplified adversaries' perceptions that there would be consequences for aggressive conduct and that these would be serious. It is likely no accident that a focus of recent Russian cyber disinformation campaigns appears to have been to stoke nationalist sentiment in U.S. allies, and to foment cynicism about and skepticism toward international institutions, particularly the European Union[1] **A logical countermeasure to present-day cyber aggression, therefore, is to reinvest in these institutions and to integrate into them robust counter-cyber arrangements and capabilities.** Wherever feasible, the U.S. should look to undertake its cyber deterrence actions together with allies, which will both isolate and increase the costs for would-be adversaries.

Serious consideration should also be given to the promotion of international partnership to assist U.S. allies and developing nations with cyber security and education. Cyber security is an international problem and will take international partnership and collaboration to solve. Christopher Painter, the former Coordinator for Cyber Issues at the State Department, stated, "International cooperation is critical to cybercrime investigations, which is why the United States has promoted international harmonization of substantive and procedural cybercrime laws through the Budapest Convention […] and promoted donor partnerships to assist developing nations."[2] Increased cooperation will assist law enforcement and diplomatic efforts to prosecute cyber crime and/or provide attribution to the national origin of the attackers. International partnership is essential to address obfuscation and use of proxies by States that engage in hostile cyber activities.

8.5 Regulation

Finally, to fulfill the balance of deterrence, as coined by Thomas Schelling, international arms control agreements must be put into place to minimize the impact and/or likelihood of conflict in the cyber realm.

8.5.1 Option 12. Treaties and Arms Control

While a set of credible power policies and capabilities likely serve as the foundation of cyber stability, mutually agreed norms and rules can solidify and reinforce a cyber stability framework. Numerous treaties, agreements, and international norms have enhanced international stability with respect to other types of forces in certain contexts. Given the differences between the cyber realm and the physical realm, no single past arms control treaty regime offers a single, strong analogue for regulating hostile cyber activities. However, as demonstrated in Section 6, numerous distinct principles from past treaty regimes may be applied to regulate State conduct in the cyber realm.

Though difficult and time consuming to conclude, multilateral or bilateral agreements could be an attractive option to promote stability in cyberspace. While the array of actors and capabilities in the cyber realm make multilateral formats desirable to address the global challenges of cyber security, bilateral agreements may be easier to conclude and could create a web of international norms piece-by-

[1] See, e.g., Daniel R. Coats, Statement for the Record, Worldwide Threat Assessment of the U.S. Intelligence Community, Senate Select Committee on Intelligence, May 11, 2017, p. 18, accessed Sept. 19, 2017, https://www.dni.gov/files/documents/Newsroom/Testimonies/SSCI%20Unclassified%20SFR%20-%20Final.pdf.
[2] Painter, Christopher. "Testimony of Christopher M. E. Painter, Coordinator for Cyber Issues U.S. Department of State," § Senate Foreign Relations Subcommittee on East Asia, the Pacific, and International Cybersecurity Policy, 8, accessed June 2, 2017, https://www.foreign.senate.gov/imo/media/doc/052516_Painter_Testimony.pdf.

piece. Bilateral agreements could happen in tandem with multilateral formats that could include regional fronts, alliance fonts, or even broader fronts potentially including adversarial States. While legal agreements are difficult to conclude, such agreements would likely be worth the effort.

These bilateral, regional, and multilateral agreements might limit who/what can legally be targeted with cyber measures, rather than limit capabilities or characteristics of cyber weapons. Cyber weapons verification is difficult, if not impossible, because cyber weapons cannot be demonstrated or counted, and are easy to conceal. Instead, certain targets such as civilian infrastructure (e.g., hospitals) might be placed entirely off-limits to attack. Such an agreement may reinforce the unilateral declarations by countries, discussed previously, to declare certain assets as sensitive, the targeting of which would elicit a response.

The Incidents at Sea Agreement offers a useful model. Countries could commit to refraining from high-risk or potentially escalatory "maneuvers," such as probing or attacking sensitive national security assets or critical infrastructure that could be perceived as a prelude to or part of an armed attack. Those escalatory maneuvers can be especially destabilizing given the limited time a decision maker has to respond to a cyber-attack. Agreements to refrain from these actions reduce the chances of miscalculation or accidental escalation.

The nuclear test ban and strategic arms reduction regimes provide further instructive principles for multilateral technical verification that might be cross-applied to the cyber realm. As with the CTBTO, it might be possible to vest a credible technical verification capacity in a neutral international body, for instance to investigate and attribute treaty-banned cyber attacks. The potential for such a regime will hinge on the confidence with which attribution might be effected. Tried and tested inspection techniques that balance required intrusiveness with protections of sensitive information – such as managed access and information barriers – may offer means to enhance both State participation in such a regime as well as the effectiveness of attribution that it could achieve. Technical and institutional options for international cooperation on attribution of hostile cyber actions could be a valuable topic for future research.

Another option is to pursue norms delegitimizing non-State actors in cyberspace. A starting point for this would be to confirm UNSC Resolution 1540's applicability to cyber terrorist organizations. Cooperation on one type of non-State cyber actor points to the cyber proxies certain nations used to pursue their foreign policies. Delegitimizing the actions of non-State cyber actors writ large would reaffirm States' monopoly on the use of force. Reaffirming this principle has parallels to the evolution of professional navies and decline of privateers that marked a major decline in piracy, and today, pirates are considered *hostis humani generis*, or the enemy of mankind. This option also reinforces efforts to harmonize and streamline criminal laws covering cyberspace with the caveat that piracy and terrorism are different and attempts to define terrorism in the UN run the risk of legitimizing authoritarian states' efforts to criminalize free speech.

Lastly, establishing a joint forum for transparency and CBMs and technical exchange may be a stabilizing option.

9.0 Conclusion

Throughout the 20th century, deterrence and arms control supported strategic stability between super powers. However, the weaponization of the cyber realm and its multipolar nature is undermining that stability. This paper has examined the role of deterrence and arms control in cyber stability. Many 20th century deterrence and arms control concepts are not applicable in the cyber realm. However, the United States can take lessons learned from deterrence and arms control to develop a strategy to support stabilization.

This paper highlights a spectrum of options for U.S. policy makers to consider to promote cyber stability based on elements of deterrence and arms control that are analogous to cyber. The options are not mutually exclusive and can be selected as necessary or expedient. Furthermore, the options can be used to develop tailored deterrence strategies for every would-be U.S. adversary, based on an adversary's motivations, tolerance for discomfort, and level of cyber infrastructure.

Sections 1 through 3 introduce the topics of cyber stability and deterrence and define key terms and outlines elements of 20th century deterrence theory, followed by a description of the cyber realm and its weaponization in Section 4.

Section 5 looks at trends in international cyber coordination and cooperation and Section 6 reviews a variety of 20th century arms control and related treaties to determine if similar treaties could promote cyber stability. The regimes analyzed and their applicability to the cyber realm are below.

- **Strategic Arms Limitation and Reduction Treaties** (SALT, ABM, and START) are of limited application as a model for cyber stability given their emphasis on numerical limitations of warheads and delivery vehicles, and verification through inspection. However, the bilateral framework may have application in the cyber realm, the Cold War's bipolarity is not found today in the cyber realm. Nevertheless, some verification modalities applied in these treaties might be incorporated in a cyber control regime, such as managed access and information barriers, which could facilitate international information sharing and cooperation on attack attribution.
- **Test Ban Treaties** lack a direct analogue to the cyber realm as there is no equivalent, verifiable way to limit testing of a cyber "weapon." However, these treaties demonstrate both a path for establishing multilateral norms against unacceptable State behavior as well as multilateral technical approaches to verification that may have potential to be cross-applied to the cyber realm. Additionally, information exchanges related to cyber capabilities, like those used in nuclear test ban treaties, could reduce secrecy that makes cyber weaponry so destabilizing.
- **BWC** and **CWC** both focus on intent related to the use of dual-use items. Chemical and biological weapons are similar to cyber weapons in that the precursors for weapons, such as a string of code is not inherently malicious but with malicious intent could become a weapon. The BWC and CWC's regulation of State intent, CWC's post-incident investigations, and both treaties' use of CBMs may be useful for promoting stability in the cyber realm.
- **United States-Soviet Union Incidents at Sea Agreement** has limited application in cyber stability, in part due to differences between the nature of international waters (where no State has jurisdiction) and cyber infrastructure (where different States have jurisdiction over various elements). However,

the Incidents at Sea defined States' red lines, helping other States know whether its actions might cause unintended escalation. A treaty regime with similar elements may be useful in the cyber realm.

- **The Helsinki Process** demonstrated how a dedicated joint forum for confidence building through transparency, dialogue, and technical exchange can stabilize otherwise tense situations. Such measures may be applicable to cyber security. Additionally, the OSCE, a product of the Helsinki Process, has a dedicated cyber security branch that could be expanded or used as a model for a multilateral cyber security forum.
- **The Regulation of Piracy** demonstrates that international agreements that expedite criminal prosecution of State proxies and individual criminals can be an effective deterrent to State behavior, which could be applied to the cyber realm.

Section 7 and 8 walk through a model of cyber deterrence, finding that traditional deterrence approaches are often not applicable in the cyber realm given States' reticence to "display" their cyber weaponry and the difficulty of attributing malicious cyber acts to a State.

Because traditional deterrence concepts are not perfect analogues to the cyber realm, Section 9 assesses how the United States might promote cyber stability with a variety of technical and policy options. The options described are not mutually exclusive and different options may be best suited for different adversaries, as explained below.

- **Signaling** – Clarify possible and proportionate retaliation options for malicious cyber acts and issuing policy statements establishing the thresholds for reaction or retaliation.
- **Improving detection and attribution capabilities** – Invest in new technologies to provide advance warning of cyber-attacks, if possible, and invest in international law enforcement capabilities to better identify perpetrators of cyber-attacks.
- **Defining tailored deterrence strategies** for specific cyber adversary States – Acknowledge that pressure points for different States are different.
- **Naming and shaming** – Publically identify State perpetrators of cyber-attacks to help build international coalitions against them and develop customary international law against such behavior.
- **Enhancing coordinated use of national criminal laws** – Prosecute State proxies and other criminal elements behind whom proxies hide.
- **Implementing sanctions** – Implement domestic sanctions. Garner international support for sanctions, as sanctions are more effective when multilateral.
- **Retaliating with "loud" cyber acts** – Demonstrate to adversaries that the United States is aware of the adversary's cyber action and warn them to stop.
- **Responding to cyber with kinetic attacks –** Use kinetic responses when attribution is certain as a last resort and only in a proportionate manner.
- **Investing in resilience and reconstitution** – Strong cyber architecture and education can deter attacks by reducing their effectiveness.

- **Collaborating and entering in agreements with other countrie**s – Work to enhance partner countries' counter-cyber capacity and leverage long-standing U.S. alliances in a coherent cyber-deterrence strategy to achieve the maximum deterrent effect vis-à-vis adversaries.

Weaponization of the cyber realm is already undermining international cyber stability. The United States has a number of options available to it to increase stability. Though neither 20th century deterrence nor traditional approaches to WMD arms control are perfect analogues for cyber stability, some elements from deterrence and arms control could be useful in the cyber realm. The United States should carefully consider its potential adversaries, its desired outcomes, and the options available in order to create a robust cyber stability strategy that deters malicious actors and encourages international cooperation in supporting a robust cyber infrastructure. Valuable topics for future research to support U.S. efforts in this area include, in particular, technical and institutional options for international cooperation on attribution of hostile cyber actions.

Appendix A
Glossary

Appendix A

Glossary

This appendix defines terms and concepts that are commonly used within this paper and related literature. **The key challenge in defining these terms in the cyber realm is that there is no international or national consensus on definitions.** Indeed, the terms are defined differently depending on who is writing. However, it is very important to have a consistent definition for the terms and concepts outlined in this paper. This paper relies on U.S. government sources where possible. Unless otherwise cited, definitions are drawn from the Department of Defense Dictionary of Military and Associated Terms.[1] Where terms were not defined in the Department of Defense Dictionary, definitions were drawn from other Department of Defense sources or government sources, if possible. Otherwise, other sources used are cited.

Active cyber defense: Synchronizing real-time detection, analysis, and mitigation of threats to critical networks and systems. It is active with in the networks it protects and not offensive in nature.[2]

Armed attack: No clear definition. At a certain level of damage, destruction, and/or casualties, an attack by cyber means becomes the equivalent of an armed attack, which under international law triggers the right of self-defense.

Attribution: Correctly identifying the attackers or assigning responsibility for an attack.[3]

Confidence-Building Measures (CBMs): Measures taken by States Parties to promote transparency and information sharing to build trust regarding defense-related activities.

Command and control: The ability of an intruder to communicate with a target's device or network. Once the command and control channel is established, intruders have "hands on the keyboard" access inside the target environment.[4]

Credibility: The ability to establish believability based on a demonstration of capability and resolve.

---

[1] Department of Defense Dictionary of Military and Associated Terms. Available online at http://www.dtic.mil/doctrine/new_pubs/dictionary.pdf.
[2] Summarized from "Enabling the Real-Time Defense of Critical Networks" accessed August 10, 2017 https://www.iad.gov/iad/programs/iad-initiatives/active-cyber-defense.cfm.
[3] Will Goodman, "Cyber Deterrence: Tougher in Theory than in Practice?" (DTIC Document, 2010), http://oai.dtic.mil/oai/oai?verb=getRecord&metadataPrefix=html&identifier=ADA528033.
[4] Hutchins, Erik M., Michael J. Clopper, and Rohan M. Amin. *Intelligence-Driven Computer Network Defense informed by Analysis of Adversary Campaigns and Intrusion Kill Chains* accessed July 12, 2017, http://www.lockheedmartin.com/content/dam/lockheed/data/corporate/documents/LM-White-Paper-Intel-Driven-Defense.pdf.

Cyber-attack: An attempt by actors to infiltrate a computer network or system with the intent to damage and/or destroy the system or affiliated auxiliary systems. For the purpose of this paper, the impact of a cyber-attack "goes beyond data collection to impose some form of harm on the United States…A large-scale cyber-attack on civilian critical infrastructure could cause chaos by disrupting the flow of electricity, money, communications, fuel, and water."[1]

Cyber realm: A global domain within the information environment consisting of the interdependent network of information technology infrastructures and resident data, including the internet, telecommunications networks, computer systems, and embedded processors and controllers.

Cyber deterrence: A set of conditions in which a State communicates, either explicitly or implicitly, a credible intention and capability to accurately attribute and impose substantial consequences for certain hostile cyber actions directed against it, and that communication dissuades other actors from undertaking such actions. The State's response may include but need not be limited to analogous cyber actions.[2]

Cyber espionage: Cyber intrusions with the intent to collect data.[3]

Cyber-intrusion: Unauthorized infiltration of a computer network or system.

Cyber security architecture: The planning, establishing, and upkeep of systems with security in mind.[4]

Cyber stability: A state of relations between States characterized by the absence of serious hostile cyber actions against one another, where the States have a sufficient common understanding of each other's capabilities and intentions so as to be inclined generally to avoid such actions, likely associated with a common belief that the costs of such conduct would outweigh the benefits.[5]

Cyber war: No clear definition. Commentators frequently use this term colloquially to describe phenomena such as "tit-for-tat" or simultaneous network cyber-attacks occurring over a prolonged duration and causing damage, distinct from kinetic operations. At present there is no generally recognized change in the legal rights and obligations between States engaged in such conduct, unless the conduct should cross recognized thresholds for armed attack or international armed conflict, which do entail such changes.[6]

Cyber weapons: Malicious code or exploitation of vulnerabilities in cyber networks or systems through various techniques intended to cause damage.

Damage: Physical, financial, or political harm caused to an asset in such a way as to impair its value, usefulness or normal function.

---

[1] Department of Defense Science Board Task Force on Cyber Deterrence, February 2017.
[2] Internal definition used by authors of this paper.
[3] "DSB CD Report 2017-02-27-17_v18_Final-Cleared Security Review.pdf," accessed April 14, 2017, https://www.armed-services.senate.gov/imo/media/doc/DSB%20CD%20Report%202017-02-27-17_v18_Final-Cleared%20Security%20Review.pdf.
[4] Ibid.
[5] Internal definition used by authors of this paper.
[6] Internal definition used by authors of this paper.

Defensive cyberspace operations: Passive and active cyberspace operations intended to preserve the ability to utilize friendly cyberspace capabilities and protect data, networks, net-centric capabilities, and other designated systems.

Delivery: Transmission of the weapon to the targeted environment.[1]

Deterrence: The prevention of action by the existence of a credible threat of unacceptable counteraction and/or belief that the cost of action outweighs the perceived benefits.

Extended deterrence: The ability of U.S. military forces to deter attack on U.S. allies and thereby reassure them.[2]

Exploitation: The successful activation of a cyber weapon. Most often, exploitation targets an application or operating system vulnerability, but it could also more simply exploit the users themselves or leverage an operating system feature that auto-executes code.[3]

Geopolitical symmetry: Countries with roughly equal retaliatory escalation power on the global stage.[4]

Hostile cyber action: Cyber-attack, use or deployment of cyber weaponry, cyber espionage, or cyber information warfare operations directed against a State, its nationals, or its interests, typically by another State or at another State's direction.[5]

Installation: Installation and propagation of malware on the victim system/network allowing the adversary to maintain persistence inside the environment.[6]

Kill chain: A systematic process to target and engage an adversary to create desired effects.[7]

Lethality: The capacity to cause death, serious harm or damage.[8]

---

[1] Hutchins, Erik M., Michael J. Clopper, and Rohan M. Amin. *Intelligence-Driven Computer Network Defense informed by Analysis of Adversary Campaigns and Intrusion Kill Chains* accessed July 12, 2017, http://www.lockheedmartin.com/content/dam/lockheed/data/corporate/documents/LM-White-Paper-Intel-Driven-Defense.pdf.

[2] "U.S. Nuclear and Extended Deterrence: Considerations and Challenges | Brookings Institution," *Brookings*, November 30, 2001, https://www.brookings.edu/research/u-s-nuclear-and-extended-deterrence-considerations-and-challenges/.

[3] Hutchins, Erik M., Michael J. Clopper, and Rohan M. Amin. *Intelligence-Driven Computer Network Defense informed by Analysis of Adversary Campaigns and Intrusion Kill Chains* accessed July 12, 2017, http://www.lockheedmartin.com/content/dam/lockheed/data/corporate/documents/LM-White-Paper-Intel-Driven-Defense.pdf.

[4] Ibid., 109.

[5] Internal definition used by authors of this paper.

[6] Hutchins, Erik M. Michael J. Clopper, and Rohan M. Amin. *Intelligence-Driven Computer Network Defense informed by Analysis of Adversary Campaigns and Intrusion Kill Chains* accessed July 12, 2017.

[7] Ibid.

[8] Internal definition used by authors of this paper.

Non-State Actor: Private persons or groups whose actions are not attributable to a State.[1]

Passive defense: Systems added to the architecture to provide reliable defense or insight against threats without consistent human interaction.[2]

Offensive cyberspace operations: Cyberspace operations intended to project power by application of force in or through cyberspace.

Proxy: A person or group of persons acting on the instructions of, or under the direction or control of, a State in carrying out the conduct in question.[3]

Resilience: The ability to provide acceptable operations despite disruption - natural or man-made, inadvertent or deliberate.[4]

Reconnaissance: Research, identification, and selection of targets, often represented as crawling internet websites such as conference proceedings and mailing lists for email addresses, social relationships, or information on specific technologies.[5]

Reconstitution: Those actions taken by one nation prior to, during, and following an attack by an enemy nation to minimize the effects of the attack, rehabilitate the national economy, provide for the welfare of the populace, and maximize the combat potential of remaining forces and supporting activities.

Signaling/Strategic communication: Focused efforts to understand and engage key audiences to create, strengthen, or preserve conditions favorable for the advancement of States' interests, policies, and objectives through the use of coordinated programs, plans, themes, messages, and products synchronized with the actions of all instruments of national power.

Stability: A state in which the international status quo is not easily upset, disturbed, or altered.[6]

State actor: Government body, person, or organization who is acting on behalf of a government body.[7]

Sunshine laws: Laws that support government transparency and accountability, for instance by requiring government agencies to make certain documents open to public disclosure.[8]

---

[1] Tallinn Manual 2.0 on the International Law Applicable to Cyber Operations, Rule 17, note 1 (2017). ..
[2] Internal definition used by authors of this paper.
[3] Tallinn Manual 2.0 on the International Law Applicable to Cyber Operations, Rule 17, note 1 (2017)
[4] Defense Science Board. 2013. "Task Force Report: Resilient Military Systems and the Advanced Cyber Threat," Department of Defense. Accessed October 13, 2016, http://www.acq.osd.mil/dsb/reports/ResilientMilitarySystems.CyberThreat.pdf.
[5] Hutchins, Erik K., Michael J. Clopper, and Rohan M. Amin. *Intelligence-Driven Computer Network Defense informed by Analysis of Adversary Campaigns and Intrusion Kill Chains* accessed July 12, 2017. http://www.lockheedmartin.com/content/dam/lockheed/data/corporate/documents/LM-White-Paper-Intel-Driven-Defense.pdf
[6] Internal definition used by authors of this paper.
[7] Internal definition used by authors of this paper.
[8] Internal definition used by authors of this paper.

Tailored deterrence: A deterrence strategy recognizing that each regime, each leadership, and each national situation is somewhat unique and therefore requires an approach to deterrence uniquely tailored to achieve maximum effect on that particular group of decision makers.[1]

Weaponization: Coupling a remote access trojan with an exploit into a deliverable payload, typically by means of an automated tool (weaponizer). Increasingly, client application data files such as Adobe Portable Document Format or Microsoft Office documents serve as the weaponized deliverable.[2]

Zero-day vulnerability: An unknown exploit that exposes a vulnerability in software or hardware and can create complicated problems well before anyone realizes something is wrong.[3]

---

[1] Schneider, Barry R., Patrick D. Ellis, and USAF Counterproliferation Center, *Tailored Deterrence: Influencing States and Groups of Concern*, 2012, 4.

[2] Hutchins, Erik M., Michael J. Clopper, and Rohan M. Amin. *Intelligence-Driven Computer Network Defense informed by Analysis of Adversary Campaigns and Intrusion Kill Chains* accessed July 12, 2017, http://www.lockheedmartin.com/content/dam/lockheed/data/corporate/documents/LM-White-Paper-Intel-Driven-Defense.pdf.

[3] "What Is a Zero-Day Exploit?," *FireEye*, accessed April 18, 2017, https://www.fireeye.com/current-threats/what-is-a-zero-day-exploit.html.

Appendix B
Other Treaty Regimes

Appendix B

Other Treaty Regimes

The authors considered certain other treaty regimes for arms control or for regulation of international or poorly governed spaces, but opted not to engage in a full-scope analysis in Section 6 due to perceived limited value of the analogy to hostile cyber actions or the cyber domain. Regimes considered and reasons for their exclusion include:

- **Intermediate-Range Nuclear Forces (INF) Treaty**.[1] INF banned a class of nuclear weapons considered to be particularly destabilizing, due in particular to the short times (potentially 4-6 minutes) in which the weapons could reach their targets.[2] These short delivery times hugely compressed the times in which hostile sides would have to make a determination on whether or not to massively retaliate, enhancing both the risk that an aggressor could carry out a pre-emptive strike to decapitate the leadership of its opponent before the opponent could order a retaliation, as well as the potential for mistakes or accidents in detection, command and control. Due to the lack of overt displays of cyber weaponry already described, it is unclear whether any type of cyber weaponry poses comparable risks for pre-emptive strike, escalation (intended or not), or decapitation. The authors view as unlikely that any class of cyber *weapon* as such (for instance, a distributed denial of service (DDoS) attack versus a tailored system exploit) would inherently pose similar risks; rather, the nature of a cyber *target* might present similar risks of escalation, such as a cyber-attack that appeared to place at risk command and control of nuclear forces.
- **The Outer Space Treaty (OST); the UNCLOS**. The OST and UNCLOS codify rules of conduct for States in physical spaces outside the physical spaces where States have traditionally exercised sovereignty according to customary international law.[3] Some of these rules are permissive (e.g., that a State may lawfully pursue, capture and prosecute criminals that flee the State's territory directly into international waters); others are prohibitive (e.g., that States may not lay territorial claims to celestial bodies).[4] The authors opted not to consider these regimes in detail due to the lack of analogousness between international physical spaces and the cyber realm. Jurisdiction (generally recognized authority of States to apply their laws) of at least one State almost always applies to actions in the cyber realm owing to several different factors, such as the nationalities of the actors involved and their physical location or the locations of actions, instrumentalities (computers, servers, cables), or

[1] Treaty Between The United States Of America And The Union Of Soviet Socialist Republics On The Elimination Of Their Intermediate-Range And Shorter-Range Missiles (INF Treaty) (signed Dec. 8, 1987).
[2] Fischer, Benjamin B. "A Cold War Conundrum: The 1983 Soviet War Scare," U.S. Central Intelligence Agency, Center for the Study of Intelligence (Mar. 19, 2007).
[3] Treaty on Principles Governing the Activities of States in the Exploration and Use of Outer Space (Outer Space Treaty; OST), Including the Moon and Other Celestial Bodies (signed Jan. 27, 1967; entered into force Oct. 10, 1967), https://www.state.gov/t/isn/5181.htm; United Nations Convention on the Law of the Sea (UNCLOS), 1833 UNTS 3; 21 ILM 1261 (1982).
[4] UNCLOS, supra FN 4 at Art. 111; OST, *supra* note 3 at Art. II._

effects within the territories of States.[1] Incidents often involve parties and/or infrastructure located in multiple States. Applying national laws and international cooperation in such cases can be complicated (reconciling competing jurisdictional claims and conflicts of laws, formally soliciting mutual legal assistance, arranging extradition, and so on); however, these are issues where there is extensive history of State practice and custom, and they are quite distinct from the case where *no* State can normally claim jurisdiction (i.e., categorically international spaces).

- **UNSCR 1540.** UNSC Resolution (UNSCR) binds all States to take various measures to prevent WMD proliferation and WMD terrorism by non-State actors.[2] The primary "deterrent" effect of UNSCR 1540 relates to non-State actors (e.g., groups and natural persons) in that, among other things, it increases the likelihood that such actors would apprehended and prosecuted for certain acts. UNSCR 1540 has little effect on deterrence among *States*, which is the focus of this paper.

[1] Am. Soc'y Int'l L., "Jurisdictional, Preliminary, and Procedural Concerns," in Benchbook on International Law § II.A (Diane Marie Amann ed., 2014), available at http://www.asil.org/benchbook/jurisdiction.pdf.
[2] UN Security Council, Security Council resolution 1540 (2004) [concerning weapons of massive destruction], 28 April 2004, S/RES/1540 (2004).

Appendix C
Agreement Attributes

Appendix C

Agreement Attributes

This appendix includes short definitions of the attributes used in Figure 2 to clarify what the authors meant by the short attribute names.

Ban of Weapon: A specific weapon or type of weapon is specifically banned in the agreement.

Bilateral: The agreement is between two States.

Communication and Information Sharing: The agreement defines a communication protocol and information that must be shared to reduce miscommunication and enhance efficacy of signaling.

Confidence Building Measures: The agreement defines specific activities that enable the signatories to build upon successes and de-escalate conflict.

Declarations of Accountable Arms: The agreement defines weapons, delivery systems, production systems, or arms that must be declared by signatories of the agreement.

Facilitates Criminal Prosecution: The agreement establishes global norms on what is considered criminal behavior vs State-sanctioned activities.

Limitation on Testing: The agreement outlines limitations on testing specific weapons or systems.

Limitations on # of Delivery Vehicles: The number of delivery vehicles for a given weapon is specifically defined in the agreement.

Limitations on # of Warheads: The number of warheads or individual payloads in a weapon are specifically defined in the agreement.

Limitations on Destabilizing Activities: The agreement defines regulations and restrictions on specified activities that are considered destabilizing.

Multilateral: The agreement is between multiple States.

Regulate by State Intent: The agreement seeks to regulate intent versus specific weapons or agents that likely have dual use. In these cases, indications of intent may be easier to assess than banning or verifying the existence or lack of specific weapons or agents.

Verification of Suspected Violation: The agreement outlines verification mechanisms that are enacted when an incident has occurred or suspected to have occurred.

Verification through Inspections: The agreement outlines verification mechanisms, including inspections, that are regularly conducted under the agreement.

Appendix D
Trends in International Cyber Coordination and Cooperation

Appendix D

Trends in International Cyber Coordination and Cooperation

International cooperation to date on curbing malicious cyber conduct can be summarized as reflecting three primary trends. First, certain States have advocated for and developed a variety of measures in support of treating international malicious cyber actions under the rubric of **national criminal law**, opting to harmonize substantive law and streamline procedures for international investigative and prosecutorial cooperation. The most prominent expression of this approach is the Council of Europe Convention on Cybercrime (the Budapest Convention).[1] A second camp of States has advocated for a regulation of cyber issues under international law using an **arms control treaty-analogous framework**, which would serve to assert the primacy of States and their right to exercise a high degree of control over their respective cyber environments. The third trend is marked by international efforts to find common ground for these two camps, primarily under the auspices of **UN-convened GGEs,** which engaged in several rounds of consultations to articulate and advance international consensus.

D.1 The National Law Approach

A number of States, particularly those in the Council of Europe (including the United States), have taken extensive measures to harmonize their substantive criminal laws on cybercrime and to streamline international cooperation on investigation and prosecution. The prime example of this is the Budapest Convention (hereinafter in this section, simply "the Convention"). From a substantive standpoint, the Convention sets forth general definitions of crimes of "illegal (unauthorized) access," "illegal interception," "data interference," "system interference," "misuse of devices," computer-related "fraud" and "forgery," offenses related to child pornography, infringements on copyright, and aiding, abetting and attempt of these crimes.[2] From a procedural standpoint, the Convention requires that States establish round-the-clock central points of contact, mechanisms for expedited preservation of computer and traffic data, authorities for engaging in search and seizure of computer data, real-time collection of traffic and content data, and then a relatively standard set of mutual legal assistance-type features.[3] The approach remains silent on matters of the global governance structure of the internet, on the nature of State sovereignty with respect to the internet, and on questions of cyber conflict among and between States. The implication is that the status quo where international law applies in the cyber realm continues to apply.

---

[1] Council of Europe, Convention on Cybercrime, *opened for signature* Nov. 23, 2001, 41 I.L.M. 282 (hereinafter "Convention on Cybercrime" or "the Convention"). As of writing, fifty-five States have signed the Convention, including all of the members of the Council of Europe except for Russia and San Marino (see Council of Europe, Chart of signatures and ratifications of Treaty 185, status as of May 9, 2016, *available at* http://www.coe.int/en/web/conventions/full-list/-/conventions/treaty/185/signatures).

[2] See ibid., Arts. 2-11.

[3] Ibid., Arts. 14-21; 23-34. Article 32 is a potential exception to the blanket characterization of "standard" mutual legal assistance; Art. 32 controversially allows certain trans-border searches without prior consent of the State on whose territory the computers or servers are to be searched.

There is considerable overlap between the State-sanctioned "hostile cyber actions" that are the focus of this paper and criminal acts articulated under the Budapest Convention. Thus, it may be possible to prosecute individuals and non-State entities who perpetrate, assist in, or attempt to commit a hostile cyber action even where a State denies its involvement.

D.2 The International Law Approach

Other States, prominently including Russia and China, advocate for a different approach to international cyber governance that emphasizes the regulation of the conduct of States under international law. Russia and China have persistently refused to accede to the Budapest Convention: Russia, ostensibly, based on objections to certain trans-border searches that it authorizes, and China based on objections over their non-inclusion during the treaty's initial development.[1] Under the auspices of the Shanghai Cooperation Organization, Russia and China have put forward multiple drafts of an "International Code of Conduct for Information Security,"[2] key features of which have included that States commit to

- comply with the Charter of the UN by highlighting the respect for sovereignty and territorial integrity;
- not use information and communication technologies (ICT) for hostile activities and aggression and not to proliferate information weapons or related technologies [revised in 2015 to omit the phrase "information weapons" in favor of the more neutral but vague "*carry out activities which run counter to the task of maintaining international peace and security"]*;
- cooperate in combating criminal and terrorist activities that use ICT;
- promote the establishment of a democratic and multilateral internet management system;
- promote the "important role of the United Nations in formulating international norms"; and
- *endeavor to ensure the supply chain security of ICT products and services,* **especially not to take advantage of its dominant position in the sphere of information technology**.[3]

Certain ideological and political features are plain. The emphasis on State sovereignty suggests that States should be the prime agents in setting their respective national policies for cyberspace, viewed by some specialists as a way to legitimize censorship and State control over internet.[4] The emphasis on equal, democratic, and multilateral internet management suggests a deliberate move away from the multi-stakeholder, "Western-dominated" architecture of the internet, illustrated by U.S.-based Internet Corporation for Assigned Names and Numbers' prominence in international domain registry, for instance, again in favor of a State-centric governance structure.[5] The Code has also been criticized for neglecting

---

[1] See, e.g., Alex Grigsby, "Coming Soon: Another Country to Ratify the Budapest Convention," Council on Foreign Relations (Dec. 11, 2014), https://www.cfr.org/blog/coming-soon-another-country-ratify-budapest-convention.
[2] See, e.g., UN General Assembly, Letter Dated 9 January 2015 from the Permanent Representatives of China, Kazakhstan, Kyrgyzstan, the Russian Federation, Tajikistan and Uzbekistan to the United Nations Addressed to the Secretary General, A/69/723 (13 Jan. 2015).
[3] NATO Cooperative Cyber Defence Center of Excellence, "An Updated Draft of the Code of Conduct Distributed in the United Nations – What's New?" (10 Feb. 2015).
[4] Ibid.
[5] Ibid.

cooperation on cross-border law enforcement by putting a strong emphasis on combating terrorism, secessionism or extremism.[1]

The Code does not commit to details on substantive criminal law or obligate States to take specific procedural measures, and in any case, owing to the political issues noted, has generally not won the support of the U.S. and other Western States. As a result, **cooperation on cyber criminal matters has not significantly progressed between the two groups of States. This has had the practical effect of allowing attacks based in countries like Russia and China to proceed against the U.S. with no mechanism for their investigation, prosecution or other resolution, and likely exacerbating the cyber conflict outlook among these countries**.

D.3 The Group of Government Experts: Room for Compromise?

A considerable, perhaps fundamental, philosophic divide has occurred between the above approaches. For the former, criminalizing strictly malicious or unauthorized access leaves as a tacit presumption that other online conduct is permitted, or at least that there not sufficient grounds for States to give prior consent to sacrifice their discretion to cooperate on investigation and prosecution. The national law approach de-emphasizes the role of the State in engaging in regulation of individuals' substantive conduct on the internet. The international law approach flips this emphasis on its head. By opting to regulate the conduct of States and suggesting that broad swaths of online conduct could be described as "terrorism," the approach presumptively leaves to States broad discretion to regulate their respective national cyber spaces, including on the level of speech. **It may be that this gulf can be bridged, however, as the two approaches largely regulate different things and may not be mutually exclusive. For example, States might agree both to criminalize malicious or unauthorized access by non-State actors and to refrain from certain attacks against one another.**

The GGE has de facto served as a forum for States to explore common ground. The UN General Assembly began considering information security issues in the 1990s, and the UN First Committee, through Office of Disarmament Affairs, has now convened five successive GGE meetings starting in 2003 to articulate areas of consensus among participating States.[2] The work of the GGEs has generally been characterized by iterative and methodical progress, with occasional setbacks. The 2013 and 2015 GGEs established and reaffirmed a normative framework for international cybersecurity by stating that the UN Charter, international law, and the principles of State sovereignty applied to cyberspace.[3] The 2015 GGE further recommended that States should cooperate to prevent harmful ICT practices and should not conduct or support ICT activity that damages or impairs critical infrastructure.[4] However, the 2017 GGE was unable to issue a consensus report, with Russia and China reportedly refusing to either reaffirm or specify in greater detail the statement that international law applies in cyberspace.[5]

---

[1] Ibid.

[2] 2016. "Report of the International Cyber Security Issues Workshop Series," United Nations Institute for Disarmament Research. Accessed 31 August 2017. http://www.unidir.org/files/publications/pdfs/report-of-the-international-security-cyber-issues-workshop-series-en-656.pdf (noting that the GGEs issue reports and recommendations according to a strict rule of consensus).

[3] Ibid. at 6.

[4] Ibid. at 7.

[5] See, e.g., ibid.; American Society of International Law, "International Law and the Trump Administration: Strengthening Cybersecurity" (19 July 2017).

D.4 Past Efforts at International Cooperation and Coordination on Cyber Conflict

While this report offers only a brief summary of the work of the GGEs, it should help to establish that international cooperation in this area seems only to be just beginning in earnest. Also illustrative of this point, **to date relatively few studies have analyzed the potential for treaties or other formal legal international mechanisms to regulate States' potentially hostile conduct in cyberspace, and these have not typically surveyed or analyzed in any depth specific mechanisms for constraints on State cyber conduct or capabilities**.[1]

[1] See, e.g., Arimatsu, Louise. "A Treaty for Governing Cyber-Weapons: Potential Benefits and Practical Limitations" Chatham House (2012). Hughes, Rex. "A Treaty for Cyberspace," *International Affairs* 86: 2 (2010) 523–541.

Pacific Northwest
NATIONAL LABORATORY

Proudly Operated by **Battelle** *Since 1965*

U.S. DEPARTMENT OF
ENERGY

902 Battelle Boulevard
P.O. Box 999
Richland, WA 99352
1-888-375-PNNL (7665)

www.pnnl.gov