

THE CYBER DEFENSE REVIEW

International Conference on Cyber Conflict (CYCON U.S.)

November 14-15, 2018

Cyber Conflict During Competition



Combining Recurrence Quantification
Analysis and Adaptive Clustering to
Detect DDoS Attacks

*Marcelo Antonio Righi
Raul Ceretta Nunes*

Predicting Enterprise Cyber Incidents
Using Social Network Analysis on the
Darkweb Hacker Forums

*Soumajyoti Sarkar
Mohammad Almukaynizi
Jana Shakarian
Paulo Shakarian*

Feed the Bears, Starve the Trolls:
Demystifying Russia's Cybered
Information Confrontation Strategy

*Dr. Nina A. Kollars
Dr. Michael B. Petersen*

Strategic Cyber: Responding to Russian
Online Information Warfare

Dr. Matthew J. Flynn

Defense Support to the Private Sector:
New Concepts for DoD's National
Cyber Defense Mission

*Jason Healey
Major Erik B. Korn*

INTRODUCTION

CYCON U.S. 2018:

Cyber Conflict During Competition

Colonel Andrew O. Hall

THE CYBER DEFENSE REVIEW

◆ SPECIAL EDITION ◆

THE CYBER DEFENSE REVIEW

A DYNAMIC MULTIDISCIPLINARY DIALOGUE

EDITOR IN CHIEF

Dr. Corvin J. Connolly

MANAGING EDITOR

Dr. Jan Kallberg

DIGITAL EDITOR

Mr. Tony Rosa

AREA EDITORS

Dr. Harold J. Arata III
(Cybersecurity Strategy)

Prof. Robert Barnsby, J.D.
(Cyber & International Humanitarian Law)

Maj. Nathaniel D. Bastian, Ph.D.
(Cybersecurity Optimization/Operationalization)

Dr. Aaron F. Brantly
(Policy Analysis/International Relations)

Dr. Chris Bronk
(National Security)

Dr. Dawn Dunkerley Goss
(Cybersecurity Optimization/Operationalization)

Dr. David Gioe
(History/Intelligence Community)

Col. Paul Goethals, Ph.D.
(Operations Research/Military Strategy)

Dr. Michael Grimaila
(Systems Engineering/Information Assurance)

Dr. Steve Henderson
(Data Mining/Machine Learning)

Ms. Elsa Kania
(Indo-Pacific Security/Emerging Technologies)

Maj. Charlie Lewis
(Military Operations/Training/Doctrine)

Dr. Fernando Maymi
(Cyber Curricula/Autonomous Platforms)

Lt. Col. William Clay Moody, Ph.D.
(Software Development)

Sgt. Maj. Jeffrey Morris, Ph.D.
(Quantum Information/Talent Management)

Ms. Elizabeth Oren
(Cultural Studies)

Dr. David Raymond
(Network Security)

Dr. Paulo Shakarian
(Social Threat Intelligence/Cyber Modeling)

Dr. David Thomson
(Cryptographic Processes/Information Theory)

Dr. Robert Thomson
(Learning Algorithms/Computational Modeling)

Lt. Col. Mark Visger, J.D.
(Cyber Law)

EDITORIAL BOARD

Col. Andrew O. Hall, Ph.D. (Chair.)
U.S. Military Academy

Dr. Amy Apon
Clemson University

Dr. Chris Arney
U.S. Military Academy

Dr. David Brumley
Carnegie Mellon University

Dr. Martin Libicki
U.S. Naval Academy

Ms. Merle Maigre
CybExer Technologies

Dr. Michele L. Malvesti
Financial Integrity Network

Dr. Milton Mueller
Georgia Tech School of Public Policy

Dr. Hy S. Rothstein
Naval Postgraduate School

Dr. Bhavani Thuraisingham
The University of Texas at Dallas

Ms. Liis Vihul
Cyber Law International

Prof. Tim Watson
University of Warwick, UK

CREATIVE DIRECTORS

Sergio Analco
Gina Daschbach

LEGAL REVIEW

Courtney Gordon-Tennant, Esq.

PUBLIC AFFAIRS OFFICER

Capt. Lisa Beum

KEY CONTRIBUTORS

Clare Blackmon
Nataliya Brantly

Kate Brown
Erik Dean

Martha Espinoza
Shane Fonyi

Col. John Giordano
Lance Latimer

Eric Luke
Alfred Pacenza

Diane Peluso
Michelle Marie Wallace

CONTACT

Army Cyber Institute
Spellman Hall
2101 New South Post Road
West Point, New York 10996

SUBMISSIONS

The Cyber Defense Review
welcomes submissions at
mc04.manuscriptcentral.com/cyberdr

WEBSITE

cyberdefensereview.army.mil

The Cyber Defense Review (ISSN 2474-2120) is published quarterly by the Army Cyber Institute at West Point. The views expressed in the journal are those of the authors and not the United States Military Academy, the Department of the Army, or any other agency of the U.S. Government. The mention of companies and/or products is for demonstrative purposes only and does not constitute endorsement by United States Military Academy, the Department of the Army, or any other agency of the U.S. Government.

© U.S. copyright protection is not available for works of the United States Government. However, the authors of specific content published in The Cyber Defense Review retain copyright to their individual works, so long as those works were not written by United States Government personnel (military or civilian) as part of their official duties. Publication in a government journal does not authorize the use or appropriation of copyright-protected material without the owner's consent.

This publication of the CDR was designed and produced by Gina Daschbach Marketing, LLC, under the management of FedWriters. The CDR is printed by McDonald & Eudy Printers, Inc. ∞ Printed on Acid Free paper.

CONFERENCE PROFILE

2018 International Conference on Cyber Conflict U.S. (CyCon U.S.)

November 14th, 2018 - November 15th, 2018
 Ronald Reagan Building and International Trade Center
 Washington D.C.

INTRODUCTION

COLONEL ANDREW O. HALL	9	CYCON U.S. 2018: Cyber Conflict During Competition
-------------------------------	---	---

SESSION 1

MARCELO ANTONIO RIGHI RAUL CERETTA NUNES	15	Combining Recurrence Quantification Analysis and Adaptive Clustering to Detect DDoS Attacks
---	----	---

VAUGHN H. STANDLEY ROXANNE B. EVERETTS	31	The Calculus of Protecting Interstate Competition from Cyberattack
---	----	---

COLIN BROOKS	45	Critical Infrastructure Protection at the Local Level: Water and Wastewater Treatment Facilities
---------------------	----	--

SESSION 2

DAVID M. PERLMAN, PH.D.	67	Applied Computational Social Choice Theory as a Framework for New Cyber Threats
--------------------------------	----	---

SOUMAJYOTI SARKAR MOHAMMAD ALMUKAYNIZI JANA SHAKARIAN PAULO SHAKARIAN	87	Predicting Enterprise Cyber Incidents Using Social Network Analysis on the Darkweb Hacker Forums
--	----	--

THOMAS KLEMAS REBECCA K. LIVELY NAZLI CHOUCRI	103	Cyber Acquisition: Policy Changes to Drive Innovation in Response to Accelerating Threats in Cyberspace
--	-----	---

SESSION 3

JOBEL KYLE P. VECINO	123	United by Necessity: Conditions for Institutional Cooperation Against Cybercrime
NINA A. KOLLARS, PH.D. MICHAEL B. PETERSEN, PH.D.	145	Feed the Bears, Starve the Trolls: Demystifying Russia's Cybered Information Confrontation Strategy
MAJOR GENERAL (RET.) JOHN A. DAVIS MAJOR CHARLIE LEWIS, U.S. ARMY RESERVES	161	Beyond the United Nations Group of Governmental Experts: Norms of Responsible Nation-State Behavior in Cyberspace

SESSION 4

DR. CHAR SAMPLE DR. CONNIE JUSTICE DR. EMILY DARRAJ	171	A Model for Evaluating Fake News
MATTHEW J. FLYNN, PH.D.	193	Strategic Cyber: Responding to Russian Online Information Warfare
ANNACHIARA ROTONDO PIERLUIGI SALVATI	209	Fake News, (Dis)Information, And The Principle Of Nonintervention. Scope, Limits, And Possible Responses To Cyber Election Interference In Times Of Competition

SESSION 5

JASON HEALEY ERIK B. KORN	227	Defense Support to the Private Sector: New Concepts for the DoD's National Cyber Defense Mission
MICHAEL WARNER	245	Borders in Cyberspace: Strategic Information Conflict since 9/11
MICHAEL P. FISCHERKELLER RICHARD J. HARKNETT	267	Persistent Engagement, <i>Agreed Competition</i> , and Cyberspace Interaction Dynamics and Escalation

THE CYBER DEFENSE REVIEW

◆ INTRODUCTION ◆

SPECIAL EDITION

CYCON U.S. 2018: Cyber Conflict During Competition

Colonel Andrew O. Hall
Director, Army Cyber Institute
United States Military Academy



INTRODUCTION

The Army Cyber Institute at West Point, in partnership with the NATO Cooperative Cyber Defence Centre of Excellence, proudly presents the Proceedings from the 2018 International Conference on Cyber Conflict U.S. (CyCon U.S.).

CyCon U.S. 2018 took place on 14-15 November 2018 at the Ronald Reagan Building in Washington, D.C. The conference theme was *Cyber Conflict during Competition*. As the U.S. 2018 National Defense Strategy states, “Inter-state strategic competition, not terrorism, is now the primary concern in U.S. national security.” During competition, U.S. and Allied forces actively campaign to advance and defend national interests in an environment that is short of armed conflict. Adversaries are increasing activities to separate alliances, devalue partnerships, and challenge traditional methods of deterrence by conducting operations that make unclear the distinctions between peace and war. These actions include political, economic, informational, and military efforts that exceed steady-state diplomacy, yet are short of violence. Conflict during competition combines cyber, electronic, and information operations to infiltrate systems and infrastructure, influence the sentiments of the populace and national decision-makers, destabilize partners and allies, and set conditions for a 'fait accompli' campaign with conventional forces. How democratic states and non-state actors anticipate, adapt, and innovate during competition will dictate the success of the democratic world and its citizens.

To explore *Cyber Conflict during Competition*, the 2018 conference was composed of a magnificent international spectrum of keynotes and panels that presented fresh ideas, relevant and actionable content, insight into future trends, and the perspectives of industry, government, and military leaders, cyber innovators, and pioneers in the discipline.

This is a work of the U.S. Government and is not subject to copyright protection in the United States. Foreign copyrights may apply.



Colonel Andrew O. Hall is the Director of the Army Cyber Institute at the United States Military Academy (USMA) located at West Point, New York. In his position as Director, Colonel Hall leads a 70-person multi-disciplinary research institute and serves as the Chairman of the Editorial Board for *The Cyber Defense Review* (CDR) journal; and Conference Co-Chair for the International Conference on Cyber Conflict U.S. (CyCon U.S.). He has a B.S. in Computer Science from the USMA, an M.S. in Applied Mathematics from the Naval Postgraduate School, and a Ph.D. in Management Science from the University of Maryland. Colonel Hall additionally teaches in the Department of Mathematical Sciences and the Department of Electrical Engineering and Computer Science at the USMA. Since 1997, Colonel Hall's military career has been focused on operations research and solving the Army's most challenging problems using advanced analytic methods. Colonel Hall also serves as the President of the Military Applications Society of the Institute for Operations Research and the Management Sciences. His research interests include Military Operations Research, Cyber Education, Manpower Planning, and Mathematical Finance.

The CyCon U.S. Paper Review Committee selected these 15 papers for publication at the conference from the 38 full papers submitted for consideration (a 40% selection rate). These papers were presented under the following four session topics: Technical, Analytics and Acquisition, Strategy and Policy, and Information Warfare.

We thank all those who submitted papers, attended the conference, and worked so hard behind the scenes to make this a spectacular event. Enjoy these papers, and may their insights guide our continued journey together. 🛡️

The views expressed in these conference papers are those of the authors and do not reflect the official policy or position of the Department of the Army, the Department of Defense the United States Government or the North Atlantic Treaty Organization.

SESSION

◆ 1 ◆

Combining Recurrence Quantification Analysis and Adaptive Clustering to Detect DDoS Attacks

Marcelo Antonio Righi

Cyber Defense

CommandQGEEx / SMU - Brazilian Army

Brasilia, DF, Brazil

Raul Ceretta Nunes

Applied Computing Department

Federal University of Santa Maria

Santa Maria, RS, Brazil

ABSTRACT

The high number of Distributed Denial of Service (DDoS) attacks executed against a lot of nations has demanded innovative solutions to guarantee reliability and availability of internet services in cyberspace. In this sense, different methods have been used to analyze network traffic for denial of service attacks, such as statistical analysis, data mining, machine learning, and others. However, few of them explore hidden recurrence patterns in nonlinear network traffic and none of them explore it together with Adaptive Clustering. This work proposes a new method, called *DDoSbyRQA*, which uses the Recurrence Quantification Analysis (RQA) based on the extraction of network traffic dynamic features and combination with an Adaptive Clustering algorithm (A-Kmeans) to detect DDoS attacks. The experiments, which were performed using the Center for Applied Internet Data Analysis (CAIDA) and University of California, Los Angeles (UCLA), databases, have demonstrated the ability of the method to increase the accuracy of DDoS detection and apply the method in real-time.

Keywords— DDoS, RQA, Adaptive Clustering, A-Kmeans.

I. INTRODUCTION

Maintaining internet services is critical during conflicts and crises when nations need to be able to send information and be increasingly resilient to the challenges that cyber conflict can provide. The ability to use and deploy Intrusion Detection Systems to networks can be crucial for enabling communication, especially in a hostile environment. DDoS can disrupt the ability to maintain communication between interested actors. Military or civilian areas can be impaired and lose the freedom to continue fighting in the cyberspace, putting at risk the security of a region or country.

© 2019 Marcelo Antonio Righi, Raul Ceretta Nunes

Detection of DDoS can be an excellent cyberspace security solution. Detection of DDoS has mechanisms that can indicate in real-time a possible attack and enable actions to be taken in a timely manner because only in this way may the success of the mitigation process be satisfactory.

To detect DDoS attacks, different techniques are used. From^[1], the detection techniques could be aggregated in at least four relevant methods: statistical-based, data mining-based, knowledge-based, and machine learning-based. However, as noted in^[2], many of them still have limitations, and their quality of service can be affected due to an excessive number of false alerts. The existence of traffic with nonlinear dynamic behavior instead of just stationary behavior can be one of these limiting factors^[3]. Network traffic contains the properties of self-similarities^[4], long-range dependence^[5], and recurrence^[3].

RQA^[6] is a mathematic technique that allows the analysis of the behavior of a nonlinear signal that repeats itself over a specific period. In the network security field, RQA already has been applied in other works^[3,7,8]. However, in the current paper, we have changed the way that it is applied. To provide better results on DDoS attack detection, this paper explores RQA to extract knowledge from dynamic features of network traffic in combination with an Adaptive Clustering method. The Adaptive Clustering method (A-Kmeans)^[9], which automatically calculates the number of clusters rather than using a fixed amount of these, is combined with RQA, which extracts dynamic features of a set of network flow attributes selected to effectively express DDoS behavior^[10].

Using RQA it is possible to extract various dynamic features of specific behaviors for each system – this is called Recurrence Quantification Measures (RQMs). Examples of RQMs are Recurrence Ratio (RR), Determinism (DET), Entropy (ENTR), TREND, and Laminarity (LAM), among others. Developing RQA over these RQMs allows us to obtain an analysis focused on the dynamic features of the traffic rather than an analysis focused on, for example, traffic statistical variability.

This work proposes the *DDoSbyRQA*, a new method for DDoS attack detection that combines RQA based on extracting dynamic features (RQMs) with an Adaptive Clustering to classify DDoS network traffic (Transmission Control Protocol (TCP) Flood, User Datagram Protocol (UDP) Flood, and Internet Control Message Protocol (ICMP) Flood). Applied on the CAIDA and UCLA databases, *DDoSbyRQA* demonstrates the power of this combination. This is a more accurate method when compared to similar methods. The main contributions of this work are: (1) to demonstrate that the use of RQA can be applied on DDoS detection, not only to analyze adopted network flow attributes, but, also, their dynamic features; (2) to demonstrate that an Adaptive Clustering method (A-Kmeans), which automatically calculates the number of clusters, can be a good partner of RQA to increase the efficiency of DDoS detection; and (3) to demonstrate that the method can be used in real time to take effective action during DDoS attacks.

The remainder of this paper is organized as follows: section II presents related works and section III presents a theoretical review of RQA. Section IV presents details of the implementation of the proposed *DDoSbyRQA* method and section V presents our experiments and results.

Finally, section VI presents the work conclusions.

II. RELATED WORKS

The traditional method for characterizing and detecting DDoS attacks is to attribute extraction based on network traffic behavior and construct an analysis of the behavior. For example, in^[11], the authors propose a method for detecting DDoS attacks using a classifier based on a decision tree algorithm (C 4.5). The authors use 16 attributes to describe a normal network traffic pattern behavior. However, the rate of false positives (FPs) is incremented when network traffic increases^[11], denoting a less effective method in situations in which there is increased flow on normal network traffic. Also, the choice of network traffic attributes did not consider important features for DDoS since the chosen attributes do not contemplate the variance of packets size and variance of time arrival packets (time among received packets). These variances tend toward zero during a DDoS attack^[10].

In^[12], the authors present a method for the detection of DDoS attacks that explores different classifiers – the Apriori algorithm, FCM, and K-Means clustering – demonstrating that the combination of multiple classifiers can improve the accuracy of detection. From these works, it is easy to comprehend that the performance of a detector depends on extracted attributes and the chosen classifier. In contrast, our work explores these factors when applying RQA combined with a self-adaptor classifier (A-Kmeans) on a set of attributes of network traffic that could effectively characterize a DDoS attack.

RQA was used in other works^[3,7,8]. In^[3], the authors demonstrate that RQA can be applied to offer qualitative and quantitative observations on detecting anomalous events in complex traffic (nonlinear). They suggest that network traffic exposes itself to the omnipresent properties of self-similarity and long-range dependence, which are correlations in a wide range of time scales. In^[7], the authors focus on demonstrating the visual analysis of RQMs in Recurrence Plots (RPs) and their power in regard to detecting anomalies. Visual tools like web RP (www.recurrence-plot.tk/glance.php) and graphical application programming interface of the Weka data mining tool were used to determine whether changes visually indicate a DDoS attack. In [8], the authors extend the work performed in^[7] to demonstrate that RQA can be applied to detect DDoS on Voice over Internet Protocol networks, but the authors maintain the empirical analysis based on visual tools of RPs. The authors do not consider the need for alert generation automatically and in real-time. In contrast to the above works, our approach looks at attributes and a method that automatically analyzes the dynamic features (RQMs) over RPs. In addition, we also explore the use of Adaptive Clustering (A-Kmeans) in combination with RQA.

In^[10], a method is presented that characterizes DDoS attacks from seven attributes extracted directly from network traffic. According to the authors, from these attributes, a classifier can effectively distinguish this kind of attack. The authors use the K-NN algorithm^[13], which is a similarity-based, supervised learning algorithm that makes classifications based on the nearest neighbor rule. The choice of k neighbor is fixed and determined by the researcher. However,

the use of a classifier to operate directly on the attribute time series (TS) can significantly limit attempts to achieve efficiency of DDoS detection. In addition, manually setting the algorithm number of neighbors is a challenge and a limitation. In^[14], the authors perform a combination of Wavelet Transform and RQA and clustering algorithm to classify the traffic. The authors used K-Means clustering, which has a predefined, fixed number of clusters; in addition, wavelet preprocessing is time-consuming. In contrast, adopting the set of attributes proposed in^[10], our work explores the combination of RQA and Adaptive Clustering (A-Kmeans^[9]), showing that the method does not require a fixed number of clusters and achieving better results than nonadaptive methods.

III. RECURRENCE QUANTIFICATION ANALYSIS (RQA)

RQA corresponds to the construction of RPs, a visual graph of recurrence quantification of a given TS, and its interpretation. The RP (see example in fig. 1), which was proposed in^[15] as a technique of nonlinear dynamic analysis systems, provides behavior visualization of the space trajectory of multidimensional phases^[8]. In practice, RP is a two-dimensional square array that represents the evolution of dynamic system states and is populated by black and white dots. The black dots indicate recurrence – namely, the states of the dynamical system for these orbiting points in regions near each other in the trajectory of the phase space. Such regions are called the Recurrence Areas. A black dot marked at the coordinate (i, j) of the RP represents the recurrence of system states at time i and j ^[16,6]. In other words, considering the RPs of fig. 1, generated in the testing phase of this work, each state of the Average Packet Size (AVG_PAC_SIZE) in each moment (i) is compared with all other states in each moment $(j, j + 1, \dots, n)$. In case of recurrence, a black dot will be marked from each result of each comparison; otherwise, it will be a white dot. Now its state $(i + 1)$ will again be compared with all other states $(j, j + 1, \dots, n)$ and so on until the end of the TS for each used attribute. The result is a square matrix of black and white dots that indicates the recurrence of the interesting attribute.

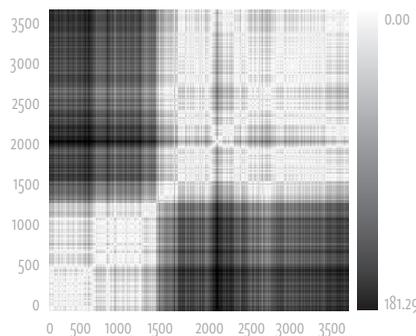


Fig. 1. RP of the Average Packet Size TS in an instance of normal traffic. The axes correspond to the number of traffic system states considered in RQA (i.e., the RP demonstrates the recurrence over N states of the TS).

Given a network traffic TS $X\{x_i\}$, where $i = 1, 2, \dots, n$ ^[16,17], the traffic system states can be ex-

pressed by X_j (see equation (1)). In (1), m is the embedded dimension (represents how many delays are used in relation to the initial TS) and τ is the duration of the delay (time to wait between states). Note that n is the total number of samples in X and N is the number of states.

$$X_j = [x_j, x_{j+\tau}, \dots, x_{j+(m+1)\tau}], j = 1, 2, \dots, N \quad (1)$$

After collecting the traffic from pre-defined attributes, the RP is built to each one according to (2).

$$R_{ij} = \theta(\varepsilon - \|x_i - x_j\|) \quad (2)$$

R_{ij} corresponds to an element of the recurrence matrix (RP), where ε is the adopted threshold called Recurrence Radius, x_i and x_j are the states of the system in the m -dimensional phase space under analysis, N is the number of states considered, and θ is the decision function defined in (3). According to (3), if the distance between the states x_i and x_j is smaller than the threshold ε , then the value of $\theta(\varepsilon)$ is 1 and there is a black dot in position (i, j) of RP; otherwise, the value of $\theta(\varepsilon)$ is 0 and there is a white dot (i, j) in RP.

$$\theta(f(\varepsilon)) = \begin{cases} 0 & (\varepsilon - \|x_i - x_j\| \leq 0) \\ 1 & (\varepsilon - \|x_i - x_j\| > 0) \end{cases} \quad (3)$$

This highlights that the ε is an important parameter in the RQA. This radius corresponds to a threshold that defines the recurrent points on the RP and depends on each type of system that is being analyzed and their objectives^[16]. The literature does not provide a specific method for establishing the ideal Recurrence Radius to define recurrence points, taking it to be adjusted according to the type of application.

Despite RP allowing the visual analysis of recurrence, this type of analysis is human-based and can lead to different interpretations. Thus, to obtain more precision to the analysis, RQMs^[16] can quantify the behavior structures in the RP. RQMs can be computed and analyzed by algorithms. From^[16], the main RQMs are RR, DET, Average Length of the Diagonal Lines, Maximum Length of the Diagonal Lines, Shannon ENTR, TREND, LAM, Average Length of Vertical Structures, and Maximum Length of Vertical Structures.

The RQA can be applied in the analysis of short, nonstationary series. However, compared to other techniques of nonlinear dynamic analysis, one of the main advantages offered by RQA is that it enables the analysis of anomalies in nonstationary systems, minimizing the bias in the analysis when overloads occur in parameters of the sampling system.

IV. THE DDoSbyRQA METHOD

This section presents the *DDoSbyRQA* anomaly detection method. It can distinguish between network traffic due to DDoS attacks versus benign traffic. Fig. 2 shows the architecture of the detection solution, where the Attack Detection Module highlights the main functionalities of the proposed method.

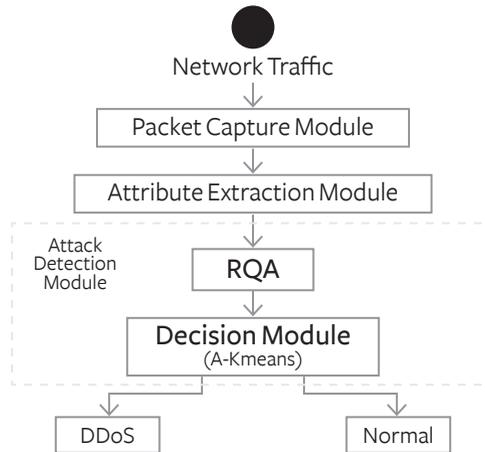


Fig. 2. The architecture of the attack detection by *DDoSbyRQA* method.

In general, *DDoSbyRQA* is supported by a Packet Capture Module, which collects data on the network, and by an Attribute Extraction Module, which selects desired attributes for RQA. The Attack Detection Module encapsulates the method that combines RQA and Adaptive Clustering (A-Kmeans) to detect DDoS attacks. The subsections A, B, and C detail each module of the architecture and subsection D presents the algorithm that implements the *DDoSbyRQA* method.

A. Packet Capture Module

The Packet Capture Module is a module that corresponds to a network sniffer. It selects the inbound traffic to a network under analysis by *DDoSbyRQA*. After captured, the data is sent to the Attribute Extraction Module.

B. Attribute Extraction Module

The extraction of attributes corresponds to the phase of selection network attributes that potentially provide relevant information to the problem of interest (DDoS detection).

For detection of DDoS attacks, RQA application requires attributes that characterize the anomalies of interest in a TS. From^[10], it is known that seven attributes are enough to identify DDoS attacks. These attributes are illustrated in table I.

The function of the Attribute Extraction Module is to extract these seven attributes from network traffic and send them to the Attack Detection Module. The result value of each attribute corresponds to statistical values extracted from network traffic flow at each second, as described in table I. Every 60 seconds, a new TS is formed and sent to the detection module. Thus, the output of this module is seven TSs, one for each attribute described in table I, at each minute.

TABLE I. ATTRIBUTES USED BY RQA. ADAPTED FROM [10]

Attributes used by RQA	
Attributes	Description
N_PAC	Number of packets
N_BYTES	Number of bytes
AVG_PAC_SIZE	The average packets size
VAR_T_PAC	The variance of the time arrival packets
VAR_S_PAC	The variance of the packets size
R_PAC	Total packets rate
R_BYTES	Total bytes rate

C. Attack Detection Module

The Attack Detection Module is the central module of the proposed solution (see fig. 2). It is composed of (i) the RQA Module and (ii) the Decision Module centered in an Adaptive Clustering classifier.

1) RQA Module: It is important to highlight that in the RQA Module, the method also extracts dynamic features (RQMs) of the network traffic (for example, ENTR) which aim to enable recurrence analysis through RPs.

In this module, the RQA and RQMs compute and analyze the RPs. Each attribute received through the Attribute Extraction Module is represented in the RQA Module by a TS (60 seconds) modeled by samples held in equidistant periods. Every TS, one for each attribute that expresses DDoS attacks or normal traffic (table I), results in RPs with their RQMs extracted mathematically. From each TS, one RP is built, as defined in section III. After the formation of the RP, three dynamic features are extracted: RR, ENTR, and DET. These features correspond to RQMs used in *DDoSbyRQA* for DDoS detection. Our goal is to analyze anomaly occurrences over these RQMs and not over network traffic statistical attributes.

To extract the dynamic features from each network attribute, the quantification calculations (calculation of RR, DET, and ENTR) applied to the RP in *DDoSbyRQA* are performed as follows.

a) Recurrence Ratio (RR): Measures the density of recurrence points on the RP. See (4) for RR computation.

$$RR = \frac{1}{N^2} \sum_{i,j=1}^N R_{i,j} \quad (4)$$

b) Determinism (DET): The ratio between the number of recurrence points that makes the diagonal structures and all points of recurrence.

$$DET = \frac{\sum_{l=l_{\min}}^N IP(l)}{\sum_{i,j=1}^N R_{i,j}} \quad (5)$$

In (5), $P(l)$ is the number of recurrence points for each diagonal formed and l is the RP diagonal length.

c) **Shannon Entropy (ENTR):** Represents the frequency distribution of the lengths of the diagonal lines.

$$ENT = \sum_{l=l_{\min}}^N p(l) \log_2 p(l) \quad (6)$$

$$p(l) = \frac{P(l)}{\sum_{l=l_{\min}}^N P(l)} \quad (7)$$

Through these 21 dynamic features (3 for each of the 7 attributes), the RQA Module forms a set of data to express through the recurrence properties the network behavior. This set is then forwarded to the Decision Module to be clustered and classified.

2) Decision Module: The Decision Module has the function of classifying the set of dynamic features received from RQA Module. The data is first partitioned by similarity (clusters) and then classified as a DDoS attack (anomalous) or not (no anomalous).

In order to avoid the difficulty of defining the optimal number of clusters, the *DDoSbyRQA* method applies the A-Kmeans algorithm^[9]. This algorithm works on a set of 21 RQMs derived from the values of ENTR, DET, and RR of 7 network attributes (see table I). The A-Kmeans automatically calculates the number of clusters (value of “k” is automatic) and compares each of them with preset thresholds during the training phase with the normal traces databases.

The decision of the module is then centered on the calculation of centroids (central points of each cluster) of the set of dynamic features (RQMs) received from the RQA Module. If the majority of the formed clusters are classified as anomalous, then the traffic will be classified as a DDoS attack.

In short, the Decision Module is also enhanced with an Adaptive Clustering method to provide more flexibility in the calculation of the number of clusters used to classify the traffic. A-Kmeans does it automatically. The automatic calculation improves the minimization of accuracy errors of the classifier. For example, in the K-means^[14] method, the researcher determines the number of clusters.

D. DDoSbyRQA Algorithm

The following steps detail the algorithm that implements the *DDoSbyRQA* method.

Entry: TS traffic (seven attributes).

Output: An indication of DDoS attack or normal traffic.

Step 1: For each traffic series X (one for each of the seven attributes), calculate the dynamic features (RR, ENTR, and DET) as described in subsection IV.C. This process is illustrated in (8), (9) and (10).

$$F_1 = f(X_{N_PAC}) F_2 = f(X_{N_BYTES}) F_3 = f(X_{AVG_PAC}) F_4 = f(X_{VAR_T_PAC}) \quad (8)$$

$$F_5 = f(X_{VAR_S_PAC}) F_6 = f(X_{R_PAC}) F_7 = f(X_{R_BYTES}) \quad (9)$$

$$F_n = \{RR_n, ENT_n, DET_n\}, \quad n = 1,2,3,4,5,6,7 \quad (10)$$

Step 2: Group the 21 dynamic features (from step 1) to describe the dynamic patterns of network traffic behavior synthesized in F in (11).

$$F = \bullet \bullet \{[RR_n, ENT_n, DET_n]\} \quad (11)$$

Step 3: From the A-Kmeans algorithm, groups of dynamic characteristics in F are built within different clusters and the traffic behavior is classified as a DDoS attack (if the majority of clusters are anomalous) or “Normal.”

V. EXPERIMENTS AND RESULTS

This section presents the experiment setup (subsection A); the tests and results of *DDoSbyRQA* (subsection B), including an FP rate comparison with other methods (subsection C); and a demonstration of the performance tests (subsection D).

A. Experiment Setup

First, the dataset is chosen. Second, the *DDoSbyRQA* operational parameters are set.

Performing real experiments on DDoS attacks is a challenge and requires good databases. Some authors^[17,18] have used databases like CAIDA 2008^[19] and CAIDA 2007^[20] to characterize normal traffic and DDoS attack traffic. In addition, the UCLA Cambridge Structural Database (CSD)^[21] is well known and contains interesting datasets with and without attacks. The CAIDA 2007 database contains one hour of DDoS attacks (ICMP Flood and TCP Flood) divided into files of type “*pcap*” sanitized with five minutes each. The CAIDA 2008 database contains 16 hours of traffic without attack divided into files of type “*pcap*” sanitized with 1 hour each. The data was collected for 16 days on the network in Chicago and San Jose in the United States. UCLA CSD contains traces of 1 hour of DDoS attacks (UDP Flood) and traffic traces without attacks collected on 10 different days. Assuming that these databases contained workloads to test *DDoSbyRQA*, the experiments in this paper used these three databases.

From these datasets, seven attributes were extracted, as described in table I, resulting in a TS X for each attribute of interest. Thus, the experiments were arranged in two phases, one for training and another for tests. Normal traffic (without DDoS attacks) was used in the training phase and anomalous traffic (with DDoS attacks) was used in the testing phase. In the training phase, the goal of the experiment was to calibrate the threshold values of the *DDoSbyRQA* method. In order for the operation to be correct, it was important to identify the behavior of each dynamic feature (RQM) in traces with and without attacks. To characterize the normal traffic, the experiments in this phase used 62 minutes of traces from the CAIDA 2008 database and 152 minutes of traces from UCLA CSD. All of these traces were without attacks. To characterize anomalous traffic, only datasets with traces containing DDoS attacks, one with 66 minutes from the CAIDA 2007 database and another with 56 minutes from UCLA CSD, were used.

The *DDoSbyRQA* method was set up to work with a TS corresponding to a sample of 60 seconds and containing a network traffic attribute for each one. Thus, without loss of generality, we chose to set the duration of delay (τ) to 1 second and the embedded dimension (m) to 60. Based on the experiments performed in^[14,15], in this work the RPs were generated with the Recurrence Radius (ϵ) set to a rate of 10 percent. Of course, these parameters of RQA could differ, but to demonstrate the power of the method, we decided to fix the threshold ϵ (the most influential parameter) on a value already used in similar works. The parameters τ and m have less influence on RQA^[14] and, thus, our choice followed the chosen TS structure.

B. Testing and Results

The first test step was to evaluate the significance of the adopted MQRs. We highlighted the chosen MQRs derived from^[14], a previous work on network anomaly detection with RQA. To be significant, an MQR must present different behavior to normal (training) and abnormal (testing) traces. Fig. 3 illustrates the results of the training phase for the dynamic features RR of the AVG_PAC_SIZE (one of the seven selected attributes). The analysis of dynamic features of other attributes follows the same methodology and, as a result, its demonstration was removed to eliminate redundancy. In fig. 3, the RR for the training dataset is shown to be stationary, with an RR level of around 25 percent (line 2). For the testing dataset, which contained only traces with attacks, the stationary behavior remained observable, but the level of RR was increased (line 1) to almost twice the observed value in the series without attacks. These results demonstrate the feasibility of threshold adoption for distinguishing between normal traffic and DDoS attacks using dynamic features (RQMs).

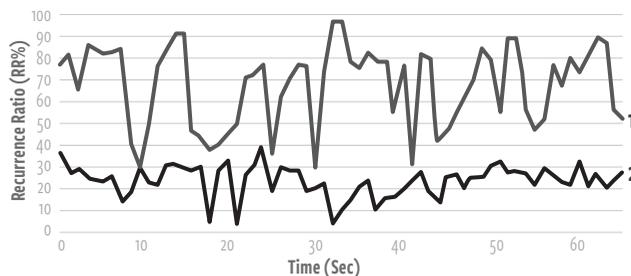


Fig. 3. RR for Average Packet Size (AVG_PAC_SIZE).

The second test step was to evaluate the accuracy of *DDoSbyRQA*. Table II (TCP Flood / ICMP Flood) and table III (UDP Flood) present the results of the testing phase. The experiment evaluated the proportion of True Positives (TPs), FPs, and the resulting Accuracy (AC), with AC defined as follows.

For purposes of comparison, tests were conducted with (i) K-Means algorithms, (ii) RQA + K-Means, (iii) A-Kmeans, and (iv) RQA + A-Kmeans. The latter corresponds to the *DDoSbyRQA* method. The goal of these tests was to allow the evaluation of the impact of the RQA and Adaptive Clustering inclusion. These tests also considered two datasets: a dataset merging of databases CAIDA 2007 and CAIDA 2008 (see results in table II) and other merging datasets from UCLA CSD Normal and UCLA CSD with DDoS (see results in table III).

When comparing the results in both cases (with attacks and without attacks), shown in tables II and III, it was possible to observe an improvement in the efficiency of classifiers when applied in conjunction with RQA. The TP when the RQA is associated with a K-Means classifier improved more than 13 percent (13.88 percent for the CAIDA dataset and 18.15 percent for the UCLA CSD dataset) and more than 19 percent when associated with the A-Kmeans (20.03 percent for the CAIDA dataset and 19.69 percent for the UCLA CSD dataset). According to values in tables III and IV, the reduction of the FPs was also significant. As a result, with both datasets, there was an increase in the accuracy of classifiers when in conjunction with RQA, reaching an improvement of 10.54 percent for A-Kmeans on the CAIDA dataset. The tests also demonstrated that the association of RQA and A-Kmeans provided a more effective result when compared with RQA + K-Means. This result demonstrates the effectiveness of Adaptive Clustering proposed by the *DDoSbyRQA* method. With the CAIDA dataset, AC was improved by 12.42 percent, and with the UCLA CSD dataset, AC was improved by 8.62 percent, demonstrating better results in DDoS detection.

TABLE II. RESULTS FOR CAIDA 2007/2008 (TCP/ICMP FLOOD)

ALGORITHM	AC (%)	TP (%)	FP (%)
K-Means	70,96	69,23	28,33
RQA+K-Means	85,99	83,08	13,54
A-Kmeans	85,96	75,35	12,31
RQA+A-Kmeans	98,41	95,38	1,54

TABLE III. RESULTS FOR UCLA CSD NORMAL / DDOS (UDP FLOOD)

ALGORITHM	AC (%)	TP (%)	FP (%)
K-Means	84,34	60,63	11,25
RQA+K-Means	88,26	78,78	10,48
A-Kmeans	94,23	74,24	4,54
RQA+A-Kmeans	96,88	93,93	3,03

C. Comparison with other methods

Table IV demonstrates that the FP rates of other similar DDoS detection methods are higher than with the *DDoSbyRQA* method. It can be seen that the *DDoSbyRQA* method has an excellent performance when compared with others. Our method results in 1.54 percent FPs (see table II) and the most effective other method results in 2.40 percent (see table IV).

TABLE IV. FP RATES TO DDOS DETECTION METHOD CITED

REFERENCES	METHOD	FP (%)
[11]	C 4.5 (Decision Tree)	2,40
[12]	Apriori+ FCM + K-Means	2,45
[13]	KNN	8,11
[14]	RQA+TW+ K-Means	8,91
[17]	Centroid-Based Rules	3,23

D. Performance test

The performance test of *DDoSbyRQA* was executed on an Intel® Core™ i7 4510U CPU 2.60GHz with eight cores and eight gigabytes of memory. The operating system was the Debian GNU / Linux 7 with kernel 3.2.0-4-amd64. The compiler used was the GNU C Compiler, version 4.7.2-5. Each execution time represents the average of 20 execution times.

The experiments measure three algorithm times: (i) the time spent to extract network traffic statistical attributes from data collected during a 60-second traffic window; (ii) the time spent to compute the RP graph and its RQMs; and (iii) the time spent to make a decision with the adaptive classifier. Table V shows the results of the performance test. The results demonstrate that *DDoSbyRQA* can decide in less than one second. This performance result enables the proposed method to be applied in real-time applications that operate over network traffic statistics collected with time windows higher than one second.

TABLE V. PERFORMANCE TEST RESULTS OF *DDOSBYRQA*.

<i>DDoSbyRQA</i> Step	Average Execution Time (ms)
Extraction of network traffic attributes	285
Computation of RP and its RQMs	324
Adaptive Clustering and decision	325
Total	934

VI. FINAL CONSIDERATIONS

In this paper, we discussed how DDoS detection on a computer network could overcome many of the limitations and security challenges posed to cyberspace during conflicts and crises that are exploited by adversary nations. To avoid damage to the communication system of any country, this paper presented an effective way to detect DDoS attacks in order to react accurately and quickly.

The effectiveness of anomaly-based DDoS detection methods has been a challenge for designers of detection algorithms. Thus, the use of the RQA combined with A-Kmeans technique is a new option for improving the quality of service of these algorithms. Until now, in the context of detecting anomalies in network traffic, RQA has been explored with limitations. This work has contributed evaluations of RQA in conjunction with a small and known group of network traffic attributes and an Adaptive Clustering algorithm (A-Kmeans).

This work showed that from only seven network traffic attributes, which characterize DDoS, it is possible to extract relevant dynamic features (RQMs) that allow increases in the accuracy of DDoS detection. This method also aimed to enable anomaly detection with RQMs, making it possible to overcome the negative influence of variability in traffic attributes, which could lead to erroneous detection. We highlight that this is possible because RQA looks for a recurrence domain instead of a traffic domain.

The experiments have shown that the use of RQA increases accuracy in identifying DDoS attacks mainly by for two reasons. First, the method classifies dynamic features of recurrence instead of traffic attributes (the tests evaluated classifiers with and without RQA). The benefit, in this case, was an increment of up to 10.54 percent in accuracy of detection. It is important to note that this result is associated with a significant increase in TPs and decrease in FPs. Second, without sudden variations in traffic, the method allows the observation of changes in behavioral patterns of recurrence that help the classifiers correctly generate clusters. With normal abrupt changes (not caused by DDoS attacks), the method allows observation of the regularity of recurrence behavior.

The work also demonstrated that the use of the A-Kmeans algorithm, an Adaptive Clustering algorithm that automatically calculates the number of clusters, fits well with DDoS detection based on RQA and improves accuracy when combined with RQA. The improvement in detection accuracy was by 8.62 percent when compared with a nonadaptive cluster algorithm (K-Means). The worst performance of K-Means clustering reflects the difficulty of calibrating a nonadaptive cluster, which can be observed by the variability of accuracy when explored with two databases of different characteristics.

Not only effective for DDoS detection, the proposed *DDoSbyRQA* method can also be explored in other contexts of network behavioral analysis and other types of cybernetic attacks, mainly by its characteristic of enabling analysis in the domain of recurrence while minimizing the negative influence of variability that causes deviations in the analysis of traditional traffic statistics. ♥

NOTES

1. M. Gyanchandani, J. L. Rana, and R. N. Yadav, "Taxonomy of Anomaly Based Intrusion Detection System: A Review," In: *International Journal of Scientific and Research Publications*, v.2, n.12, 2012.
2. A. S. Raut, and K. R. Singh, "Anomaly Based Intrusion Detection-A Review," *Int. J. on Network Security*, vol. 5, 2014.
3. F. Palmieri, and U. Fiore, "Network anomaly detection through nonlinear analysis," *Computers & Security*, 29(7), pp. 737–755, 2010.
4. W. Willinger, V. Paxson, and M. S. Taqqu, "Self-similarity and heavy tail: structural modeling of network traffic," *A Practical Guide to Heavy Tails: Statistical Techniques and Applications*, ISBN: 0-8176-3951-9, pp. 27-53, BirkhRäuser, Boston, USA, 1998.
5. M. Grossglauser, and J. C. Bolot, "On the relevance of long-range dependence in network traffic," *IEEE/M Transactions on Networking*, 7(5): pp. 629-640, 1999.
6. C. L. Webber, and N. Marwan, "Recurrence Quantification Analysis: Theory and Best Practices," *Springer series: Understanding Complex Systems*. Springer International Publishing, Cham Switzerland, 2015.
7. N. Jeyanthi, J. Vinithra, S. Sneha, R. Thandeewaran, and N.C.S.N. Iyengar, "A Recurrence Quantification Analytical Approach to Detect DDoS Attacks," In: *Computational Intelligence and Communication Networks (CICN)*, Washington, DC, USA, pp. 58-62, 2011.
8. N. Jeyanthi, R. Thandeewaran, and J. Vinithra, "RQA based approach to detect and prevent DDoS attacks in VoIP networks," In: *Cybernetics and Information Technologies*. v.14, n.1, pp. 11-24, 2014.
9. S. K. Bhatia, "Adaptive K-Means Clustering. American Association for Artificial Intelligence," Copyright. Palo Alto, California 94303 USA. Copyright, 2004.
10. T. T. Oo, and T. Phyu, "A Statistical Approach to Classify and Identify DDoS Attacks using UCLA Dataset," *International Journal of Advanced Research in Computer Engineering & Technology (IJARCET)*, vol. 2, Issue 5, 2013.
11. Y. C. Wu, H. R. Tseng, W. Yang, and R. H. Jan, "DDoS detection and traceback with decision tree and grey relational analysis," *International Journal of Ad Hoc and Ubiquitous Computing*, 7, pp. 121–136, 2011.
12. R. Zhong, and G. Yue, "DDoS detection system based on data mining," *Proceedings of the 2nd International Symposium on Networking and Network Security*, Jingtangshan, China, 2-4 April, pp. 062–065. Academy Publisher, 2010.
13. H. Nguyen, and Y. Choi, "Proactive Detection of DDoS Attacks Utilizing k-NN Classifier in an Anti-DDoS Framework," *International Journal of Electrical and Electronics Engineering*, vol. 4, n° 4, 2010.
14. J. Yuan, R. Yuan, and X. Chen, "Network Anomaly Detection based on Multi-scale Dynamic Characteristics of Traffic," *INT J COMPUT COMMUN*, ISSN 1841-9836, 9(1), pp. 101-112, 2014.
15. J. P. Eckmann, S. O. Kamphorst and D. Ruelle, "Recurrence plots of dynamical systems. *Europhys.*" *Lett*, 56 (5), pp. 973-977, 1987.
16. N. Marwan, and C.L. Webber Jr, "Mathematical and computational foundations of recurrence quantifications," In: *Recurrence Quantification Analysis: Theory and Best Practices*. Springer Series: Understanding Complex Systems. Springer International Publishing, Cham, Switzerland, pp. 1-41, 2015.
17. W. Bhaya, and M.E. Manaa, "The Proactive DDoS Attack Detection Approach Using Data Mining Cluster Analysis," *Journal of Next Generation Information Technology (JNIT)*, vol. 5, no. 4, 2014.
18. M. Suresh, and R. Anitha, "Evaluating Machine Learning Algorithms for Detecting DDoS Attacks," In *4th international Conference on Advances in Network Security and Applications (CNSA)*, pp. 441-452, 2011.
19. "The CAIDA UCSD Anonymized Internet Traces 2008," Access in 05 may 2015 11:12h, <https://data.caida.org/datasets/passive-2008/>.
20. "The CAIDA "DDoS Attack 2007" Dataset," Access in 15 may 2015 11:12h, <https://data.caida.org/datasets/security/ddos-20070804/>.
21. "UCLA CSD packet traces," <http://www.lasr.cs.ucla.edu/ddos/traces/public/usc>.

The Calculus of Protecting Interstate Competition from Cyberattack

Vaughn H. Standley

*College of Information and Cyberspace
National Defense University
Washington, DC, U.S.A.*

Roxanne B. Everetts

*College of Information and Cyberspace
National Defense University
Washington, DC, U.S.A.*

ABSTRACT

Lethal conflict may be approximated using power law statistics which, on a log-log plot of exceedance probability (EP) versus severity, is characterized by constant slope $-q$. Values of $q < 1$ violate probability axioms and describe high-risk systems. Consistent with reports that q for war improves after 1950 from 0.41 to 0.75 due to increases in military alliances, q is argued to be a sensitive function of network variables. Low-risk interstate competition is achieved when $q > 1$ and allows for the use of Bayesian hypothesis tests based on q to serve as a decision criterion about when to react to threats, leading to a set of parameters that determine whether conflict will escalate and to the conclusion that redundant networks, deterrence, and attack detection stabilize competition against cyber conflict. Examples of the importance of the Bayesian parameters in creating and adapting networks to stabilize competition are provided.

Keywords - network; likelihood ratio; power law; resiliency; Bayesian

I. INTRODUCTION

This paper describes a first-order analytical model that associates a set of network parameters with the potential for cyberattacks to escalate a state of peaceful interstate competition to violent conflict.

The *Stability-Instability Paradox*^[1], developed by Snyder in 1965^[2], describes an interrelationship between all-out war and lesser forms of conflict where strategic-level peace, achieved through nuclear deterrence, leads to regional armed conflict. Competition is a form of nonviolent conflict that includes political, economic, informational, and military efforts that exceed normal peaceful relations. “During competition, U.S. and Allied forces actively campaign to advance and defend national interests in an environment that is short of armed conflict”^[3]. Thus, competition arises in the absence of regional armed conflict in the same way that regional conflict arises in the absence of all-out war.

This is a work of the U.S. Government and is not subject to copyright protection in the United States. Foreign copyrights may apply.

Deterrence is about “decisively influencing an adversary’s decision calculus to prevent attack or the escalation of a conflict”^[4]. Thus, deterrence is first and foremost an information and decision process and not simply derived from an assessment of risk from military capabilities. There is no a priori reason to think that interstate competition is different because people decide, and nation-states do not.

Because deterrence is within the domains of information and decision theory, “Shannon’s Maxim” comes into play. The “father of information theory,” Claude Shannon, argued that one ought to design cyber systems under the assumption that the enemy will immediately gain full familiarity with them^[5]. While Shannon’s Maxim is more widely known as being fundamental to public key infrastructure (PKI), on which all cybersecurity is based today, its parallel in deterrence is that a nation’s military capabilities should be assumed to be known by its adversaries. Emerging technologies like Blockchain, extend the level of openness described by Shannon’s Maxim and, for this reason, appear to be stronger than PKI. While it seems certain that there will always be a role for encryption, “policy and political push for more transparency could prove to be the deciding factor” in selecting open technology over traditional cybersecurity methods^[6]. In the same way that encryption methods evolved from “security through obscurity” to those based on a presumption of being known, information assurance seems set to evolve to be more open and networked, rather than more closed. The extreme importance of information methods and systems to deterrence suggests they too are better served by more open and more networked systems. Similarly, if we applied the *Stability-Instability Paradox*, then this assumption would apply to regional armed conflict and nonviolent interstate competition as well.

II. THE POWER LAW, WAR, AND STRATEGIC COMPETITION

The power law of statistics is used to describe the relationship between two values when a change in one results in a change of proportional size in the other. Power laws have been defined across numerous disciplines, including science, statistics, physics, engineering, etc.^[7]. A common application of power laws is the use in defining a probability distribution. Unpredictable and catastrophic failures in networked systems are often observed to follow a power law relationship, meaning that the probability of a severity S that exceeds a severity level s is equal to s raised to a negative constant q and multiplied by a constant C . Intuitively, the power law means that smaller consequence events happen exponentially more often than larger events. Formally, it is described by the following formula:

$$P(S>s) = Cs^{-q} \quad (1)$$

In 1960, Lewis Fry Richardson was the first to fit armed conflict with power law constants in his famous *Statistics of Deadly Quarrels*. This work was later confirmed by Cederman^[8]. On a plot of $\log P(S>s)$ versus $\log(s)$, Cederman’s fit to the data describes a straight line with slope equal to -0.41 and a y -intercept equal to 1.27. See figure 1.

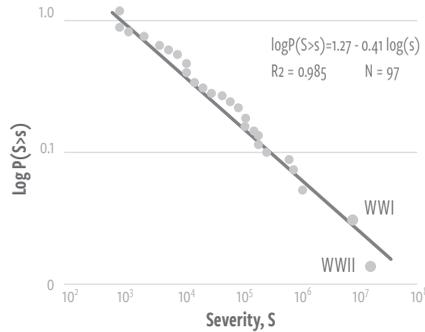


Fig 1. Dots indicate wars. The solid line is a power law fit to the data. Data Source: Correlates of War Project [8].

When not written in log-log form, the power law fit to this data may be written as eq. (2):

$$P(S>s) = 18.6s^{-0.41} \tag{2}$$

Power law phenomena are identified as “low-risk” if $q > 1$ because the rate of increasing severity is outpaced by decreasing likelihood. Conversely, phenomena are “high-risk” if $q < 1$ because severity increases faster than the likelihood decreases. Accordingly, q for war is a high-risk phenomenon. However, the power law violates the axioms of probability if $q < 1$ and is, therefore, untrustworthy beyond what is indicated directly by data.

Rational decisions are based on risk, not just likelihood. Risk is equal to likelihood multiplied by severity. In the case of figure 1, severity is the number of persons killed. Figure 2 illustrates how extremely severe wars cannot be discounted, even though their likelihood is small. For example, the war line (red) in figure 2 increases as severity s increases. Believing that the fit to the data in figure 1 holds until $s = 10^8$, a hypothetical nuclear war killing 100 people (8 on the x-axis) is riskier than a war killing 1 thousand (3 on the x-axis). This agrees with our observation that nations build and maintain military defenses for extremely unlikely wars. For systems having a q value greater than one, risk decreases with increasing consequences. The decreasing gray line that we call “Competition” in figure 2 is derived from $q = 1.10$. The consequences of increasingly improbable conflicts may be ignored. That is, they are low-risk.

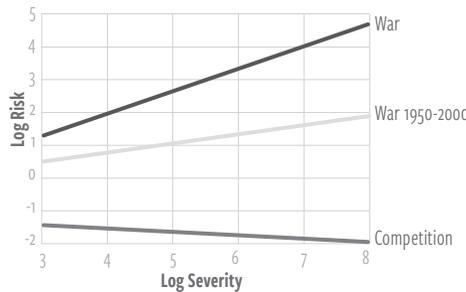


Fig 2. Log Risk versus Log Severity s for three q values.

Reflecting on the *Stability-Instability Paradox*, it seems that it is simply evidence of the power law nature of war. The power law requires that smaller conflicts be likely in the absence of larger ones, and vice versa, even if the underlying mechanisms are not clear.

III. WAR AND COMPETITION NETWORKS

The underlying mechanisms for war have been proposed and debated for centuries. One idea is that war is a network phenomenon^[9]. In this theory, nations seeking to expand their network eventually seek to absorb nodes of other nations’ networks, analogous to the natural process that often leads larger companies to acquire smaller ones. The first 40 years of Durant’s acquisition of various automobile companies (Oldsmobile, Cadillac, etc.), that eventually lead to the formation of General Motors, is a notable example. This process is often referred to as “preferential attachment.” “Species distribution and many other phenomena are observed empirically to follow power laws where preferential attachment process is a leading candidate mechanism to explain this behavior”^[10]. It is called “Competitive Exclusion” in different contexts.

A parametric model of this network process is obtained through the EP used in traditional quantitative risk management. Failures start and propagate in a network according to the vulnerability of nodes in terms of probability, γ , and the network spectral radius, ρ . Spectral radius embodies the main characteristics of a bidirectional network, which are the density of links and size of heavily connected hubs. The measure of network resilience, z , is proportional to both the inherent fractal dimension of the network, q , and to $\gamma\rho$, where $z < 1$ indicates low-risk, $z > 1$ is high-risk, and $z \gg 1$ indicates the potential for catastrophe. The spectral radius can be seen as a measure of “reachability” from any one node to any other node along a chain of network hops. As reachability increases, vulnerability to cascading failure increases. The product $\gamma\rho$ will determine the degree to which failures propagate. Survivability of a network can be achieved by hardening or isolating nodes from the network as soon as the nodes have been compromised. For example, in a network model for the communicability of a human disease, nodes in the network represent humans who may receive preventive treatment to reduce infectiousness, decreasing γ . Or, links in the infection network are cut by enforcing a quarantine to reduce ρ . Lewis^[11] reports that networked critical infrastructure sectors (e.g., communication, transportation, electricity, etc.) obey a *Fundamental Resilience Equation*, the log-linear relationship in eq. (3), where b and k are constants:

$$\log(q) = b + k\gamma\rho \tag{3}$$

This equation, where k is negative, indicates that vulnerable and/or large networks subject to cascading failure lessen q . Raising severity s to the value on both sides of eq. (3) reveals that q changes exponentially with linear changes to $b + k\gamma\rho$, as in eq. (4):

$$q = s^{b + k\gamma\rho} \tag{4}$$

Based on extensive simulation work, Lewis reports that the average b , k , and ρ are 0.5, -0.42,

and 8, respectively. What this means is that there is only a narrow band, between 0 and 0.14, where the product $\gamma\rho$ results in $q>1$. Greater values result in $q<1$, leading to estimates of EP that cannot be trusted. Within the narrow range, however, q will be extremely sensitive to the product $\gamma\rho$ if war is a network phenomenon.

Jackson and Nei^[12] report that political, military, and economic alliances increase resistance to cascading failure. In other words, war is a network phenomenon. Their findings allow us to estimate a larger value of q for war between the years of 1950 and 2000 that is apparently due to increased network redundancy and hardness that decreases the exponent:

“The number of wars per pair of countries per year from 1950 to 2000 was roughly a 10th as high as it was from 1820 to 1949. This significant decrease in the frequency of wars correlates with a substantial increase in the number of military alliances per country and the stability of those alliances.”^[12]

Though there has been no employment of nuclear weapons since 1945, their presence and proliferation bears some comment, since the possibility of nuclear warfare has, in some way, affected all subsequent wars involving nuclear states and their surrogates. The consequences of war changed when the U.S. and USSR gained nuclear capabilities. The effects of nuclear arsenals – particularly those delivered by intercontinental ballistic missiles – on whether or not states go to war may not be an easy thing to measure, but surely exist. In addition, the number of countries over the last two centuries varied considerably due to the rise and fall of European colonialism. In short, it may not be an “apples-to-apples comparison.” For the moment, however, we simply accept the assertion by Jackson and Nei^[12].

The q in eq. (2) was adjusted until $P(10^3>s)$ became one-tenth the value, as for for $q=0.41$. The value $q=0.75$ was obtained. Since the year 2000, the internet and global positioning system (GPS) enabled smart phone technologies, to name a few, have further extended interstate networks, decreasing the negativity of the network term and increasing the overall exponent (i.e., increasing q). Note that the values -0.41 and -0.75 are typical according to Lewis^[11] but outside the band where q is a valid parameter of a probability distribution. Figure 3 illustrates decreasing $P(S>s)$ as a function of s for war ($q=0.41$), war from 1950 to 2000 ($q=0.75$), and Competition ($q=1.1$).

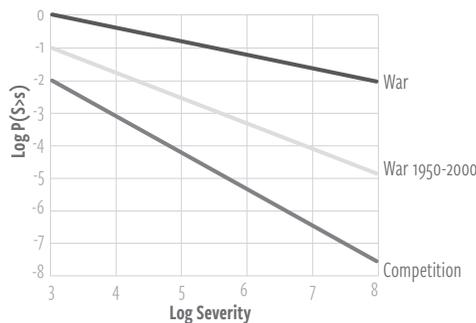


Fig 3. Log EP $P(S>s)$ versus Log severity s for three q values.

IV. ATTACK CALCULUS DURING WAR AND COMPETITION

To have operational significance, the power law data must be incorporated into a decision-making formula. A Bayes' hypothesis test provides a means for making decisions based on probabilities and evidence and serves as a simple quantitative model of how friendly and hostile nations would react to threats rationally under different circumstances.

A Bayesian hypothesis test may be derived from a dichotomous form of Bayes' theorem. Let A be an event, such as a military attack or cyber operations that have comparable consequences; (i.e., deaths). Participants do not know if A will happen, only that it may happen. The probability of A happening is $P(A)$ and the probability of A not happening is $P(\bar{A})$. The consequences, or cost, of inaction are C if A is true. The cost of unnecessary action when A is not true is \bar{C} . Let d be some data that is a probabilistic indication of A . Knowledge of d helps decide whether an action should be taken to avoid A . The probability of observing d given A is $P(d|A)$ and the probability of observing d given \bar{A} is $P(d|\bar{A})$. These definitions of cost may be thought of as the consequences of deciding contrary to the truth. One chooses to act on the belief that A is true if the following is true:

$$P(d|A)CP(S>s|A)P(A) > P(d|\bar{A})\bar{C}P(\bar{A}) \tag{5}$$

This particular dichotomous formulation of Bayes' theorem includes the costs for mistakes because rational decisions cannot be made without taking into account risk; (i.e., cost multiplied by probability). Note: For the purposes of calculating risk, we rewrite $P(S>s)$ in eq. (1) in the conditional form as $P(S>s|A)$ to indicate that the severity of an attack depends on whether an attack occurs. Thus, the risk of A is $CP(S>s|A)P(A)$ and the risk of \bar{A} is $\bar{C}P(\bar{A})$. Intuitively, when the consequences of inaction (the left-hand side of eq. (5) exceed the cost of incorrect action (the right-hand side of eq. (5)), then one chooses to act.

Eq. (5) may be written with both conditionals on the left-hand side as the ratio of the true-positive over the false-positive, and, on the right side, the ratio of the risks of both choices:

$$\frac{P(d|A)}{P(d|\bar{A})} > \frac{\bar{C}P(\bar{A})}{CP(S > s|A)P(A)} \tag{6}$$

The left-hand side of eq. (6) is referred to as the likelihood ratio, L (also called a Bayes factor), while the right-hand side is referred to as the critical likelihood ratio, L^* , weighted with consequences. To make a rational decision favoring a belief in A , L must be greater than L^* . If L is not greater than L^* , then the decision should be to do nothing:

$$L > L^* \tag{7}$$

Note that this is a simplified construct. Probabilities are more accurately defined by density functions. And, and other considerations, such as morality, ethics, economy, etc., would need to be weighed separately. However, this formulation will serve to demonstrate some of the characteristics of conflict during competition and war.

The math should be indifferent to whether a conflict is conventional or nuclear, so consider a hypothetical example of the critical likelihood ratio (L^*) where two nations, X and Y , have equivalent nuclear weapon capabilities. Both arsenals have the ability to destroy the other’s population unless the nation under attack sends its population to hardened bunkers within 30 minutes, the amount of time it takes for the first missiles to arrive. However, every time such an emergency is declared, many people will die in the panicked rush to get to safety. Event A is where nation Y intends a surprise nuclear first strike against X , using all of its nuclear forces. Nation X has a launch warning system that provides data (d) indicating that Y ’s attack is underway. C is the consequence of deciding \bar{A} when A is true and \bar{C} is the consequence of deciding A when \bar{A} is true.

The probability of an attack $P(A)$ against the U.S. can be estimated based on historical data. Using the Correlates of War (COW) Project data, we do that now. Between years 1816 and 2007, inclusive, of the years covered by the COW project, there were 239 intrastate and interstate wars. Thus, on average, there is approximately 1.25 wars per year. Assuming a Poisson distribution (“a statistical distribution showing the likely number of times that an event will occur within a specified period of time”), this yearly average of wars leads to a probability of 0.71 that there will be at least one war in the world in any given year. Of the 239 wars, the U.S. was involved in 13. Therefore, $P(A)=0.71 \times 0.054=0.038$. Conversely, $P(\bar{A})=0.96$.

Following the 2018 missile attack false -alarm in Hawaii, Fisher ^[13] argued that the Soviet downing of Korean Airlines Flight (KAL) 007 in 1983 could be considered as an example of a nuclear war false alarm. Reportedly, the Soviet Union mistook KAL 007 for an American spy plane conducting pre-nuclear war operations. All 269 passengers and crew were killed. For illustration, this number is rounded to two significant digits and used as the value of \bar{C} , such that $\bar{C}/C=270/s$, where s is severity in deaths. Therefore, the critical likelihood ratio for war is given by eq. (8). It is shown as the lower line of figure 4:

$$L^*(War) = \left(\frac{270}{s \cdot 18.6s^{-0.41}} \right) \left(\frac{0.96}{0.038} \right) \tag{8}$$

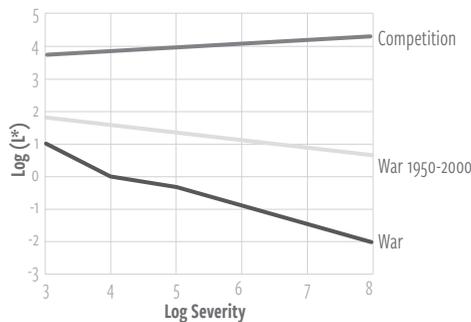


Fig 4. Log critical likelihood ratio L^* versus log severity s for three q values and where $\bar{C}/C=270/(s \times 18.6s^{-q})$.

On the right -side of eq. (8), the left bracketed ratio is the ratio of the false-alarm consequence in deaths over the inaction consequences in deaths. We call this “the deterrence ratio.” The right-most bracketed ratio is the ratio of the probability that the U.S. will not be attacked in any given year over the probability that it will be attacked in the same year. Another way to interpret eq. (8) is to regard the numerator as the risk of incorrect action and the denominator as the risk of inaction. Even though this form of the equation can be simplified, we choose not to do so to retain some clarity.

To decide that an attack is underway, a likelihood ratio $P(d|A)/P(d|\bar{A})$ should exceed the critical likelihood ratio, L^* , in figure 4. L^* should be greater than one, zero on the log scale. But this is not the case for war above 10 thousand deaths, or 4 on the log scale, meaning the hypothesis test is useless because no indication of an attack is enough to send citizens to bunkers. If this result seems non-intuitive, consider that L^* is dominated by the severity s in the denominator of eq. (8), which is consistent with our understanding of a high-risk phenomenon. The only way for L^* to become greater is to increase \bar{C}/C , which serves as a deterrent for attack. A different result is obtained for war for the years 1950 to 2000. The detection ratio stays above 1, starting near 1 hundred for 1 thousand deaths and lowering to about 3.7 for 1 hundred million deaths. See the middle line in figure 4. The significance is that there is a high threshold for deciding that an attack is underway, even though the exponent q has changed only slightly. The middle line is given by eq. (9) below:

$$L^*(War\ 1950 - 2000) = \left(\frac{270}{s^{18.6s^{-0.75}}}\right) \left(\frac{0.96}{0.038}\right) \quad (9)$$

Low-risk interstate competition is achieved when $q > 1$. The smallest increase above 1 involving two significant digits finds the exponent equal to 1.1. Although arbitrarily chosen to be greater than 1 one, our confidence in this assumption is bolstered by the works of Overill and Jantje, who report that cybercrime may be fitted to a power law with a q of 1.6^[14], suggesting that an overall q of between 1.0 and 2.0 is realistic. The equation for L^* in this case is eq. (10) below:

$$L^*(Competition) = \left(\frac{270}{s^{18.6s^{-1.1}}}\right) \left(\frac{0.96}{0.038}\right) \quad (10)$$

To decide that an attack is underway during competition, a likelihood ratio $P(d|A)/P(d|\bar{A})$ should exceed the critical likelihood ratio, L^* , following eq. (10). It is shown as a blue line in figure 4. It indicates that the positive likelihood ratio must be greater than 10 thousand, 4 on log scale, to choose in favor of A . The significance is that the evidence for attack during competition must be very high, whereas, for war, it is small.

V. CONFLICT ESCALATION PARAMETERS

The forgoing analysis suggests that cyberattacks, or any kind of conflict, could escalate if there are significant impacts to any one of the following sets of parameters: the fractal dimension q , the likelihood ratio $P(d|A)/P(d|\bar{A})$, or the deterrence ratio, \bar{C}/C . These are discussed in turn.

A. The Fractal Dimension

The network term $k\gamma\rho$ which is related to the fractal dimension q by double exponential (i.e., an exponent raised to an exponent), impacts the Bayesian hypothesis tests in an extreme way. It is comprised of the spectral radius, ρ , which may increase without bound with the number of nodes. Nodes may represent nations. Segments connecting nodes may represent alliances between nations. But nodes could just as easily represent internet server farms in different nations and the segments be fiber-optic transmission lines between them. For an n -by- n connection matrix, ρ ranges from $\sqrt{n-1}$ to n . The vulnerability, γ , is associated with forces that diminish or nullify the network specified by ρ . For example, the network term may be the target of information and cyber operations, reducing its contribution to the negative exponent and making the system more riskier. The fractal dimension will normally be measured or obtained through simulation where the underlying mechanism need not be immediately evident. Such is the case for the q -value for war. However, in terms of network parameters k , γ and ρ the effect of a cyberattack on the exponent of eq. (4) should be calculable.

The theory that war is a network phenomenon posits that networks help prevent or mitigate war. While this seems counter to the *Fundamental Resiliency Equation*, eq. (3), which attributes cascading failure to the network itself, the two ideas are not incompatible. The addition of network components, such as the alliances described by Jackson and Nei^[12], can overlay an existing network. The cascading failure of one such network thus will be mitigated by the redundant network leading to a larger q .

B. Likelihood Ratio

The parameter representing attack detection is the likelihood ratio (L). Obtaining $P(d|A)$ or $P(d|\bar{A})$ separately or together as a ratio depends on the performance of real information systems or processes. Ideally, the performance of a detection system is assembled into what is called a Receiver Operator Characteristics (ROC) graph. Detection systems face many technical challenges and are a natural target of hostile information or cyber operations. Manipulating or interfering with a target nation's detection network could facilitate a surprise attack by decreasing L or decreasing L^* , suppressing the target's reaction time.

C. The False-Negative/False-Positive "Deterrence Ratio"

The "false-alarm" consequence will depend on what specific action is taken, whether it is sending people to shelters, the launching of a counterstrike, or some other action that intends to mitigate the impact of an attack. The value of \bar{C} is a vexing question, more so than the value for C . The cost of a false-alarm is not easy to justify without real data. Furthermore, unnecessary action in response to a false-alarm can lead to a series of counter-actions that may escalate out of control. That is, a false-alarm could result in the same consequences as a true-positive: war. The deterrence ratio must be increased in order to prevent war. Figure 5 shows the critical likelihood ratios for war and competition if \bar{C}/C is equal to 1.0. This is the case where the cost of unnecessary action is war itself and might be considered an extreme case. All three lines

in figure 5 indicate that $P(d|A)/P(d|\bar{A})$ must be dramatically higher to choose in favor of an attack, A . For the case of strategic war, the red line, L^* exceeds 100. Similarly, L^* for competition starts high and increases even more dramatically.

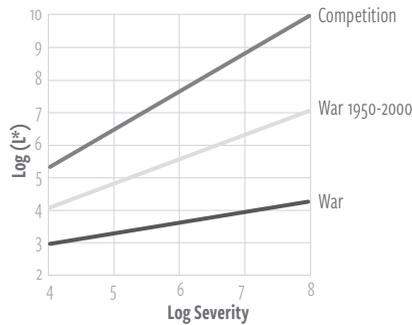


Fig 5. Log L^* versus log severity s for three q values and $\bar{C}/C=1$.

VI. EXAMPLE ADAPTATIONS THAT PROTECT COMPETITION

In this section we discuss specific examples of adaptations involving redundant networks, deterrence, and attack detection to illustrate how they protect peaceful competition from escalating as a result of cyber conflict.

A. Trade and Diplomatic Alliances Incorporating Detection

Larger L^* means a greater threshold for conflict escalation. Figure 6 shows how L^* increases exponentially with linear changes in q , indicating increased robustness against an attack false alarm, provided that γp decreases. Redundant networks compensate for effects that would otherwise cause a cascading affair failure. Extending or strengthening the alliances of the type cited by Jackson and Nei^[12] should help stabilize competition, beyond that for war between 1950 and 2000.

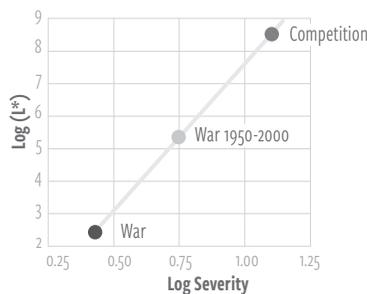


Fig 6. Log L^* versus q .

Adding to an alliance a method for detecting attacks is an adaptation that amplifies an existing or redundant network. Bilateral and multi-lateral treaties may include verification provisions. The Nuclear Nonproliferation Treaty (NPT), for example, is a long-lived diplomatic network started as part of the Atoms for Peace program in 1953 that has been maintained by

the International Atomic Energy Agency since 1970. It has provisions for continuous verification of nuclear materials and technologies by a cadre of nuclear safeguards inspectors. Under the NPT, the sharing of nuclear technology is encouraged for the use of civilian power, but not allowed if the technology is used for military applications. Thus, not only does this network increase stability of competition through the network parameters, it increases the true-detect/false-alarm ratio to improve rational choices about potential escalation of conflict.

B. Open Technologic Networks

Returning to Shannon's Maxim, there is an increasing abundance of evidence that open electronic networks and software contribute to the stability of interstate dynamics. Open systems are computer systems that provide a combination of interoperability, portability, and open software standards. They allow for increased communication, negotiation, and vetting of cybersecurity processes. Examples include the internet, the Unix operating system, and the Firefox web browser. The top reasons individuals or organizations choose open source software are: lower cost, security, no vendor lock-in, and better quality^[15]. For these reasons, open systems comprise the bulk of computer technology in the world. These technologies increase the spectral radius, ρ , and decrease node vulnerability, γ .

Some open technologic networks were originally closed national security systems. The internet is the most famous example of a national security technology that was made open to the public. Despite the difficulties it creates, its net effect is to help keep peace within and between nations. There are other examples. Many people don't know that GPS navigation signals were originally classified. After the Soviet downing of KAL 007, it was decided that it was decided the benefits of declassifying these signals outweighed the disadvantages^[16]. Once GPS was made publically available, the private sector miniaturized the electronics which enabled the receiver to be added to cell phones. Despite navigation telemetry being broadcast only (i.e., unidirectional broadcast), GPS public availability increases network infrastructure by making information globally available via a small spectral radius while being simultaneously less vulnerable to cyberattacks due to its unidirectionality^[17]. An adaptation that decreases vulnerability and spectral radius while maintaining global availability diminishes the loss of network resilience resulting from cascading failure.

C. Deterrence of Cybercrimes Crimes

By way of the Bayesian likelihood ratio weighted with consequences, we have shown that a rational enemy will attack unless there is a deterrent. One reason that cybercrime is so rampant is that there is little deterrent. There's a low probability of the offending individual or nation facing punishment or sanctions, so attacks are likely to continue. An adaptation of cyberspace that clarifies what constitutes national and international offenses and ensures commensurate responses with a high probability would help prevent attacks and help stabilize interstate competition. Goldman and McCoy argue that, "imposition of financial sanctions, public/private partnerships to disrupt tools of cybercrime, and activities to disrupt payment networks run by criminals who sell fraudulent goods over the Internet"^[18] decrease cybercrime.

Their recommendations emphasize that it is as important to punish criminals if convicted as it is to lessen the chances that they will benefit from their crime.

Schwartz contends that deterrence of cybercrime is a myth due to the unique nature of the medium and attackers^[19]. Thus, detection of cybercrime and cyberattacks appears as important as dispensing punishment and denying benefit. The Bayesian attack formula shows how detection is intertwined with prior probability, conditional probability, and consequences, suggesting that another adaptation be the automation of detection based on techniques such as the Bayesian hypothesis test.

VII. CONCLUSION

The power law of statistics, Shannon's Maxim, network failure theory, and Bayes' theorem were brought together in this study to create a parametric model of war and nonviolent interstate competition to enable a study of the tendency of cyberattack and other forms of attack to escalate conflict from competition to war and, conversely, how to lessen this tendency by modifying network parameters.

We manipulated the *Fundamental Resiliency Equation* to show that conflict, as represented by q , is in theory related to network variables b , k , γ and ρ by double exponential. The significance of this is that the EP will be extremely sensitive to network variables over the domain of their validity; (i.e., $q > 1$). Our investigation relies on the researched conclusion that war has a q value reported by Cederman of less than one. Other q values have been reported^[20].

Examples of creating and adapting networks to stabilize and protect competition were provided. Open networks, standards, and software continue to create technologic interstate alliances that further stabilize competition. Further stability can be achieved by making military information systems publically available, as was done with GPS in the 1980s. Trade and diplomatic networks that build -in systems for detecting conflict should be expanded. Finally, deterrence and detection of attack seem to be inseparable for the case of cybercrime.

The method described in this paper for estimating an attack detection threshold is built on a rigorous mathematical framework on which to conduct further research. Preliminary results reported by Standley, Nuño, and Sharpe indicate that the severity of war follows log-normal statistics with a mean of 7,900 deaths, standard deviation of 10, and validity between 1one and 15 million deaths; obeys probability axioms in all cases; and is equally applicable to the Bayesian hypothesis test method^[21]. These findings suggest that the power law is an approximation that is valid only for a narrow range of deaths and is not indicative of the underlying phenomena; (e.g., preferential attachment).♥

The opinions, conclusions, and recommendations expressed or implied are the authors' and do not necessarily reflect the views of the Department of Defense or any other agency of the Federal Government, or any other organization.

NOTES

1. Rauchhaus, R., Evaluating the Nuclear Peace Hypothesis: A Quantitative Approach, *Journal of Conflict Resolution*, Vol. 53, Issue 2, January 27, 2009.
2. Snyder, G., 1965. The balance of power and the balance of terror. In *Balance of power*, ed. Paul Seabury. San Francisco: Chandler.
3. 2018 Nuclear Posture Review, U.S. Department of Defense, <https://media.defense.gov/2018/Feb/02/2001872886/-1/-1/1/2018-Nuclear-Posture-Review-Final-Report.PDF>.
4. *Ibid.*
5. Shannon, C. E., "Communication Theory of Secrecy Systems", *Bell systems Technical Journal*, Vol 28-4, page 656 – 715, Oct. 1949.
6. Gabison, G., "Policy Considerations for the Blockchain Technology Public and private Applications," *Science and Technology Law Review*, Vol 19, 327 – 350, 2016.
7. Andriani, P.; McKelvey, B. (2007). "Beyond Gaussian averages: redirecting international business and management research toward extreme events and power laws". *Journal of International Business Studies*. 38 (7): 1212–1230. doi:10.1057/palgrave.jibs.8400324.
8. Cederman, L.E., "Modeling the Size of Wars: From Billiard Balls to Sandpiles." *The American Political Science Review* 97.1 (2003): 135-50. JSTOR. Web. 27 Apr. 2015.
9. Duffield, M., War as a Network Enterprise, *Cultural Values: The Journal for Cultural Research*, Jan-April 2002, 6 (1&2):
10. Simon, H. A. (1955). "On a class of skew distribution functions". *Biometrika*. 42 (3–4): 425–440. doi:10.1093/biomet/42.3-4.425.
11. T.G. Lewis, *Critical Infrastructure Protection in Homeland Security: Defending a Networked Nation*, 2e. Wiley, 2015. Gause, Georgii Frantsevich (1934). *The Struggle For Existence* (1st ed.). Baltimore: Williams & Wilkins.
12. Networks of military alliances, wars, and international trade, Matthew O. Jackson and Stephen Nei, *PNAS* December 15, 2015. 112 (50) 15277-15284.
13. Fisher, M., Hawaii False Alarm Hints at Thin Line Between Mishap and Nuclear War, *The New York Times*, Jan. 14, 2018, <https://www.nytimes.com/2018/01/14/world/asia/hawaii-false-alarm-north-korea-nuclear.html>.
14. Overill, R.E. and J.A.M Silomon, Single and Double Power Laws for Cyber-Crimes, Department of Informatics, King's College London, Strand, London WC2R 2LS, UK, <http://citeseerx.ist.psu.edu/viewdoc/download?doi=10.1.1.225.6194&rep=rep1&type=pdf>.
15. Guseva, I., "Bad Economy Is Good for Open Source". *Cmswire.com*. 2009-03-26.
16. Pace, S. and G.P. Frost, Irving Lachow, David R. Frelinger, Donna Fossum, Don Wassem, and Monica M. Pinto, *The Global Positioning System: Assessing National Policies*. Santa Monica, CA: RAND Corporation, 1995.
17. Standley, V. and E. Boucheron, Space-based Unidirectional Networks and Resiliency, (Accepted), *QRS* 2018.
18. Goldman, Z.K. and D. McCoy, "Economic Espionage: Detering Financially Motivated Cybercrime," *8 J. Nat'l Sec. L. & Pol'y* 595 (2015- 2016).
19. Schwartz, M.J., "The Myth of Cybercrime Deterrence," *Bank Info Security*, June 1, 2015. Online. Available at: <https://www.bankinfosecurity.com/blogs/myth-cybercrime-deterrence-p-1867>.
20. Clauset, A., Trends and fluctuations in the severity of interstate wars, *Science Advances* 21 Feb 2018, Vol. 4, no. 2, <http://advances.sciencemag.org/content/4/2/eaao3580.full>.
21. Standley, V.H., F.G. Nuno, J.W. Sharpe, "The Validity of Log-Normal Statistics for the Severity for High Magnitude War," 2018, abstract submitted to NATO Science & Technology Organization, Systems Analysis and Studies Panel.

Critical Infrastructure Protection at the Local Level

Water and Wastewater Treatment Facilities

Colin Brooks

*National Defense University
College of Information and Cyberspace
Washington, DC*

ABSTRACT

The increasing number of Industrial Control System (ICS) vulnerabilities, coupled with continuing revelations about ICS compromises, emphasizes the importance of securing critical infrastructure (CI) against cyber threats^{[1],[2]}. The ability to adversely affect the operation of an ICS through cyberspace is exacerbated by the increasing use of automation and implementation of common routing protocols to communicate with control devices^[3]. Local water treatment facilities are particularly vulnerable to this attack vector due to the need to manage key functions with minimal staff. Reacting to specific cyber risks without developing a holistic method for managing risk provides only a modicum of protection. This monograph demonstrates how focusing on risk management as a mitigation strategy – not individual risks – maximizes the security efforts at the local level.

Some basic information technology (IT) security practices such as access control, physical security, and operations security can be applied to ICS security. However, determining which security controls to select and evaluating their effectiveness requires a process or framework that holistically considers risk across the enterprise. A risk management framework (RMF) allows an organization to assess risk in terms of impact to overall business operation, instead of assessing risks isolated to particular divisions within the organization. The National Institute of Standards and Technology (NIST) RMF, National Infrastructure Protection Plan RMF (NIPP-RMF), and NIST Cybersecurity Framework for CI are three complementary frameworks water facilities can employ to facilitate risk mitigation in a cost-effective way^{[4],[5],[6],[7],[8]}.

Keywords - industrial control system; cyber; critical infrastructure; water treatment facilities; wastewater.

This is a work of the U.S. Government and is not subject to copyright protection in the United States. Foreign copyrights may apply.

I. INTRODUCTION

Over the last century the position of the United States as a world leader depended on a strong economy, strong democracy, and exceptional military capability. As technological improvements increased the capability and capacity of the United States to maintain its position in the world, these improvements simultaneously created greater dependencies on CI.

According to Presidential Decision Directive (PDD) 63, CI is composed of physical and cyber assets essential to the minimal operation of the economy and the government. Homeland Security Presidential Directive (HSPD) 7 provided further details on what types of acts would compromise CI^[9]. President Obama’s Executive Order 13636, in concert with Presidential Policy Directive (PPD) 21 (which replaced HSPD-7), expounds on the work of earlier administrations by specifically defining 16 different CI sectors and reiterates which government agencies support each sector. “Water and wastewater treatment” is identified in all four executive directives and orders as a CI sector, the Environmental Protection Agency (EPA) is assigned as the government proponent for water sector protection in HSPD-7, and this is reiterated in PPD-2^{[10], [11], [12]}.

Water and wastewater treatment is essential for ensuring clean drinking water, preventing disease, and protecting the environment^[13]. Efforts at the beginning of the 20th century were primarily aimed at ensuring the purity of drinking water. In the late 1990s and early 21st century, protecting water sector resources from malicious actors was recognized as a security priority as awareness of vulnerabilities grew^[14].

Particular concern about vulnerabilities in ICSs – the systems responsible for controlling CI operation (figure 1) – increased as experts identified the possibility of exploiting vulnerabilities remotely through the internet^{[1], [2]}. ICSs are composed of multiple devices, including Supervisory Control and Data Acquisition (SCADA), Human Machine Interface (HMI) devices, Radio Terminal Units, Main Terminal Units, and Programmable Logic Controllers (PLCs), each of which have vulnerabilities. Increased use of common routing protocols to communicate with these devices exacerbates the issue of ICS cybersecurity^[3].

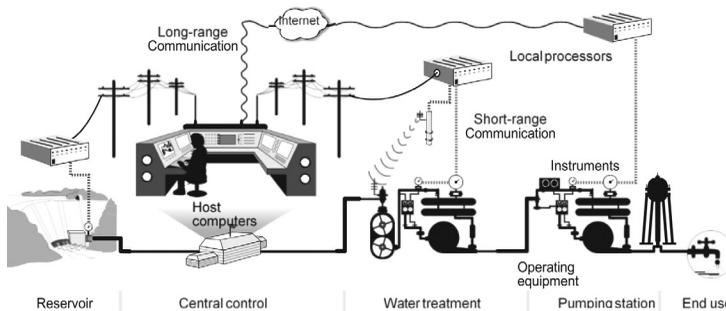


Fig. 1 Components of a control system in a water treatment and distribution facility (p.3)^[3].

II. THREATS AND VULNERABILITIES IN ICSS

Many different CI sectors have been adversely affected through cyberspace. Disruption to air traffic control systems in Worcester, Massachusetts, in 1997 was caused by a teenager disabling part of the phone network. In 2000, a disgruntled contractor at the Maroochy Shire Water Treatment facility in Australia caused hundreds of thousands of gallons of sewage to flow into streams by controlling facility equipment from a laptop computer. In 2003, the Structured Query Language worm Slammer disabled safety monitoring systems at the Oak Harbor, Ohio, nuclear power plant for nearly five hours^[15].

Recent findings by members of both the public and private sectors exacerbate the concern over the vulnerability of ICSs to attack. In 2016, the Industrial Control System Cyber Emergency Response Team of the Department of Homeland Security (DHS) found 700 security vulnerabilities in the 300 systems it analyzed^[16]. Positive Technologies, Inc., a network security company, identified 197 vulnerabilities in ICS components of major manufacturers in 2017^[17].

In late 2017, Schneider Electric, a major manufacturer of ICS components, revealed its components had been compromised by hackers. The malware, labeled Triton, was a zero-day (previously unknown) vulnerability in Triconex Tricon safety system firmware. The malware escalated privileges and then dropped a remote access tool (RAT) in the system to await further instructions. The RAT was intended to manipulate emergency shutdown processes to keep the system operational, allowing further invasive action. Triton continued system analysis and reconnaissance as it worked, exfiltrating information back to the source. The attacker, who was never identified, demonstrated an elevated level of sophistication^[18].

In 2010, the malicious code known as Stuxnet was revealed as the cause of the degraded capability of the Iranian nuclear refinement facility at Natanz. Specifically, it attacked Siemens PLCs that controlled the centrifuges, causing them to spin at erratic rates^[19]. This attack, which is widely considered to be the first confirmed act of cyber war, is believed to be an effort of the U.S. and Israel to thwart the Iranian nuclear weapon development program^[20]. This initially generated a great deal of excitement in the IT community, but many members of the ICS sector believed the attack was not important to their operations, as it targeted centrifuges belonging to Iran, not U.S. infrastructure^[1].

While cyber threats to CI in general have been more prevalent in the last two decades, there is a long history of attacks on the water sector. During World War II, the Japanese poisoned Soviet water sources with typhoid bacteria; Soviets flooded the area south of the Istra Reservoir near Moscow to slow the German advance in 1944; Israeli water infrastructure was attacked by Yasar Arafat's Fatah in 1965; neo-Nazis attempted to poison urban water supplies in the U.S. in 1972; and two Al-Qaeda operatives were arrested in 2002 with plans describing how to poison U.S. water systems^{[21],[22]}.

Fear of terrorist attacks, especially on water facilities and water supplies, increased in the 1990s and early 2000s, leading to formalized efforts to protect CI. In 1998, PDD-63 aligned federal agencies with particular infrastructure sectors to better coordinate protection efforts. PDD-63 established Information Sharing and Analysis Centers (ISACs) for public-private security

cooperation to facilitate threat data sharing between the government and the private sector^[10]. In response to the 2001 terrorist attacks, the Bush Administration passed the Public Health Security and Bioterrorism Preparedness and Response Act of 2002. It directed that vulnerability assessments of CI be conducted in each sector, allocated funding for the protection of water sector facilities, and increased penalties for attacks on water^{[23],[24],[10]}.

Water is a particularly vulnerable resource. Approximately 17 percent of the drinking water treatment facilities in the U.S. provide service to 92 percent of the populace^[13]. This means a terrorist or other malicious actor targeting one of approximately 2,700 facilities could have an inversely proportional impact on public health and may be able to delay the detection of a compromise. One way to execute an attack is to introduce toxic substances through a service point (a fire hydrant, for example) via backflow. Backflow occurs when the pressure gradient of the water in the distribution system is overcome by a source with higher water pressure (figure 2). This can accidentally occur when backflow prevention devices, like check valves, fail due to wear or nonmalicious acts^[25].

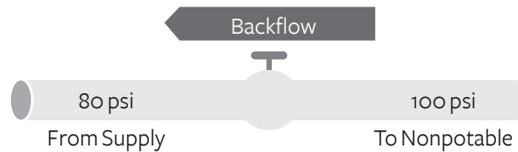


Fig. 2. Backflow due to Backpressure ^[25].

There are numerous examples of such accidental incidents, including: a glycol contamination of a West Virginia county health department due to a faulty check valve; the failure of a backflow preventer on an elementary school boiler feedline, causing drinking water contamination; and, ironically, an incident at a Boston hotel in 1974 where an American Water Works Association conference was being held (chromium entered the drinking water through a submerged inlet cross-connection to the building air conditioning system)^[25].

Backflow devices are designed to prevent accidental contamination but can be defeated by a determined attacker and are not a reliable safeguard against malicious actors. Attacking through backflow only requires the actor to overcome the ambient water pressure with a pump capable of creating a higher pressure and injecting a contaminant. If injected correctly, a contaminant can be carried throughout the rest of the system from a strategic point. Using a highly toxic contaminant only requires a few gallons to be introduced to have widespread impact. Devices that detect contamination are not ubiquitous and could be modified to present a false negative to personnel monitoring them^[22].

As shown in figure 3, a marked increase in attacks on water sector ICSs occurred from 1999-2012. Although some of the upward trends can be attributed to late disclosure or better detection of vulnerabilities, the increasing number of ICS equipment able to be accessed remotely makes it more vulnerable to attack. In the U.S., the connection of ICS components to the internet increased by 10 percent from 2017 to 2018^[17].

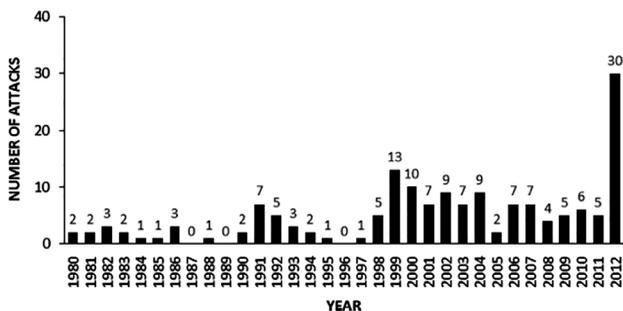


Fig. 3. Recorded trends on water CI (p.4) [21].

Further compounding the issue is the recent development of system control applications for mobile platforms. This improves the productivity and efficiency of local water facilities but exposes ICSs to cyber threats not previously encountered^[26]. For example, Bolshev and Yushkevich found 147 vulnerabilities in 34 vendor applications used for managing ICS components^[3]. Another research team, Rios and McCorkle, set out to find 100 security flaws in ICS software in 100 days but found 665 flaws in the same amount of time; 75 of the flaws were easily exploitable. The latter team’s research was based on open source information from the internet^{[27],[1]}.

Terrorists are not the only ones who could exploit such ICS vulnerabilities. Cybercriminals may target the systems because they are less secure and serve as a means to another end. In 2006, a computer used for controlling water system devices in Harrisburg, Pennsylvania, was compromised and used for spam email distribution^[28].

Feasible attacks on water sector assets through cyberspace are only one facet of a complex security problem. Interdependency between the water sector and other CI sectors amplifies the potential for catastrophic damage (see figures 4 and 5). The water sector depends on CI such as electricity to operate pumps, petroleum for backup generators, and the chemical sector for the disinfection of water. Conversely, other CI sectors need water for manufacturing, cooling equipment, and agricultural production.

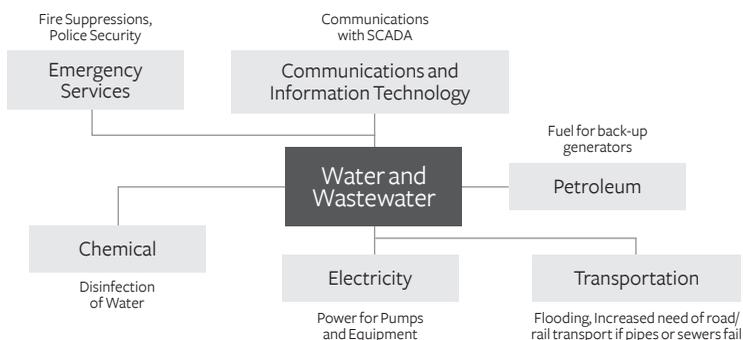


Fig. 4. Dependence of the water sector on other CI (adapted from [21]).

CRITICAL INFRASTRUCTURE PROTECTION AT THE LOCAL LEVEL

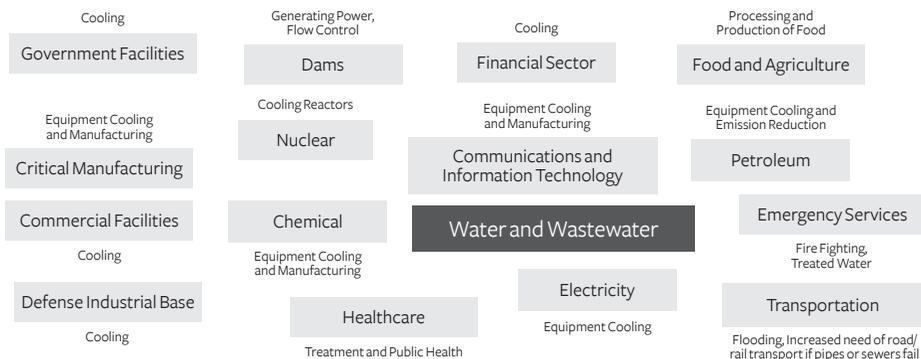


Fig. 5 Dependencies of other CI on the water sector (adapted from [21], [38], [1], [33]).

Denial or disruption of water service can have cascading effects. For example, an uncontrolled release of a large volume of wastewater, as happened in Australia in 2000, could have catastrophic effects on public health, environmental well-being, and commercial facilities^[29]. Attacks on transport systems used to pipe water from sources to agricultural production sites could cause significant financial harm^[24]. Catastrophic damage to water mainline pipes inflicted by opening and closing main gates too rapidly, causing a hammering effect, could collapse sections of pipe, immobilizing traffic and delaying emergency service response time. In addition, it could cause backsiphonage (figure 6).

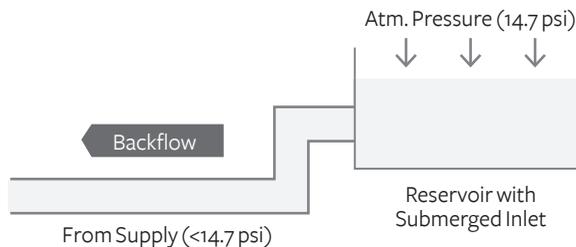


Fig. 6 Backsiphonage [25].

Backsiphonage is a type of backflow caused by a zone of negative pressure in a water system – if a cross-connection exists, atmospheric pressure pushing against a contaminant will force it into the water supply that contains zero negative pressure^[25]. These types of attacks on distribution systems and other CI are a concern expressed by many in the sector^{[30], [31]}.

III. CHALLENGES TO SECURING THE WATER SECTOR

Securing facilities from cyber threats is challenging for many reasons. These include funding, the age of equipment, and education^{[1],[8],[31]}. One of the main challenges water sector decision-makers face in securing their facilities is obtaining enough funding. Organizations’ funding can vary depending on the size of their facilities and the number of people they service. Organizations with larger facilities have better opportunities to account for security in planning their budgets because they are better resourced than organizations with smaller facilities^{[2],[32]}.

Though it serves fewer people than a large urban facility, denial of service to a rural facility could have an equivalent impact by degrading public confidence in water supplies and causing other second and third-order effects. These could include pressure on local and state government to provide potable water for extended periods of time, decreased revenue from business and tourism, and disruption to agricultural and manufacturing operations^{[33],[27],[2],[21],[26]}.

Most of a local water facility budget is earmarked for operations and maintenance. The Congressional Budget Office noted that 67 percent of funding for water infrastructure is spent on operations and maintenance by state and local governments^[8]. Such a limited budget for efforts other than infrastructure maintenance requires conscious decisions to invest in security by facility and sector leadership. Therefore, efforts by local water facilities to implement monitoring software or hardware security appliances may be limited or impractical.

Another factor in securing ICSs is the age of their equipment. Securing SCADA, PLCs, and HMIs is challenging because much of it is 20 to 30-years-old and designed with reliability and safety in mind, not security^{[1],[8],[31]}. Systems initially used obscure, proprietary protocols for communication and were isolated from other early computer systems. “Security through obscurity” was a common approach^[14]. The growing interconnections between previously isolated systems and the internet, along with the use of common protocols like Transmission Control Protocol / Internet Protocol, expose ICSs to previously unidentified threats^[3]. Like the use of mobile computing platforms, using newer technologies to manage equipment designed before the advent of the internet poses risks.

Some gaps in ICS security exist due to a lack of awareness of cyber threats and their impact to operations. An example is the focus on cybersecurity of IT (corporate network) versus operations technology (OT) security. Engineers understand the process flow and operation of ICS components, but are often not aware of the vulnerabilities in their connected systems. Conversely, IT personnel often do not understand the unique nature of SCADA systems and how patching vulnerabilities might interfere with system processes^[1]. Reviews by the National Cybersecurity and Communications Integration Center identified common network issues, such as the improper use of virtual machines; poor configuration of Virtual Local Area Networks; improper management of Bring Your Own Device implementations; and, where IT and OT efforts were combined, a lack of OT monitoring^[34].

Staff at a local water facility in New England interviewed by this author corroborated many of the challenges noted in other reports and studies. They stated that their operation was largely dependent on revenue from the businesses and households they service. Much of their revenue has been reinvested in maintaining the infrastructure, while the majority of the budget allotted for wastewater treatment was spent on the removal and incineration of sludge. Most of the pump stations dated to the 1980s and remote connectivity to the system were limited but possible through the telephone system. While the operators and supervisors were highly skilled at their jobs, their understanding of how cyber threats associated with an IT network could affect their OT network was less developed.

IV. MANAGING RISK

In light of these vulnerabilities and challenges, steps can be taken to advance the security of the water sector. Some basic IT security practices, such as access control, physical security, and operations security, can be applied to ICS security. However, determining which security controls to select and evaluating their effectiveness requires a process or framework that holistically considers risk across the enterprise. An RMF allows an organization to assess risk in terms of impact to overall business operations, instead of assessing risks isolated to particular divisions within the organization. The NIST RMF, NIPP-RMF, and NIST Cybersecurity Framework for CI are three complementary frameworks a water facility can employ to facilitate risk mitigation in a cost-effective way^{[13], [4], [29], [35],[36],[37]}.

A. NIST RMF

The NIST RMF was developed to improve information security, strengthen risk management processes, and encourage reciprocity between federal agencies. It is a holistic approach to risk that incorporates IT security into enterprise risk management, emphasizing continuous monitoring and linking of risks to organizational and executive-level operational decisions. It is the successor to the Department of Defense Information Assurance Certification and Accreditation Process (DIACAP). DIACAP emphasized compliance with the patching of system vulnerabilities, whereas the RMF broadly considers many facets of information system security^[6].

The NIST RMF consists of six steps (figure 7). Step one categorizes the system and information processed based on an impact analysis. The second step identifies a set of basic security controls based on categorization and tailored to the organization’s assessment of risk. Step three implements the selected security controls, documenting how they were deployed. The fourth step assesses the security controls to determine effectiveness in meeting security requirements. Step five authorizes system operation based on determination of acceptable risk to operations, assets, individuals, and other organizations. The last step is continuous monitoring of controls for effectiveness, documentation of changes to the system or environment, and reporting of the security state to organization officials^[38].

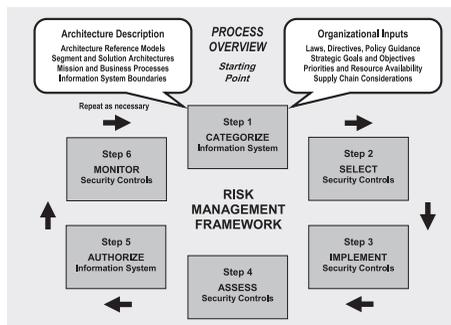


Fig. 7 NIST RMF^[4].

The NIST RMF is a baseline framework that can be applied to both governmental and non-governmental organizations^[38]. The process can be applied to any type of IT system. It does not consider specific types of systems.

B. NIPP-RMF

The NIPP-RMF is specifically designed with CI in mind. Presented in the 2013 NIPP, it recognizes the importance of a public-private partnership and the differing constraints on private versus government organizations^[5]. NIPP-RMF is broad in its application, accounting for dissimilar operating environments and both natural and man-made threats. It emphasizes the importance of information sharing to build resilience and improve threat reduction. Figure 8 provides an outline of its main components^[5].

The NIPP-RMF complements other efforts, such as the Threat and Hazard Identification and Risk Assessment process conducted by regional and urban jurisdictions to establish capability priorities^[5]. The CI community shares information and builds upon best practices and lessons learned to fill gaps in security and resilience through the RMF.

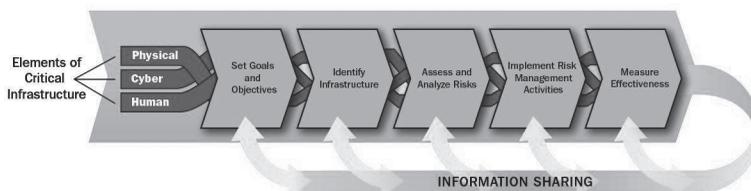


Fig. 8 NIPP-RMF^[3].

The first step is set at the national level, with input from each CI sector. The second step includes identification of all assets, systems, and networks for continued operation, considering dependencies and interdependencies. Step three, assess and analyze risks, rely on the analysis of threats, vulnerabilities, and consequences. Information sharing is essential in this step. Step four, implementing risk management strategies, involves the prioritization of activities to manage risk based on costs and potential to reduce risk. The final step in the process measures the effectiveness of controls. Continuous monitoring is essential to the risk management process, as is informing leadership whether the controls in place are effectively mitigating risk^[5].

C. NIST Cybersecurity Framework for CI

The Cybersecurity Framework for CI is a risk management construct developed specifically for CI cybersecurity by NIST and numerous stakeholders in the private sector. It is composed of three distinct sections, including the Framework Core, Framework Implementation Tiers, and Framework Profile^[6]. The framework uses holistic business risks as drivers for cybersecurity activity instead of the compliance-related endeavors previously associated with cybersecurity^[39]. Integrating cybersecurity with the overall business operations process informs decision-makers where they can best apply resources to enable operations.

The functions of identify, protect, detect, respond, and recover are part of the Framework Core. They provide a strategic view of the life cycle management of cybersecurity risk. The core provides a method for communicating industry standards, guidelines, and practices across the organization, from the strategic level to the operational and tactical levels. It identifies key categories and subcategories for each function and correlates them with existing guidelines and best practices for desired outcomes. The five primary core categories are shown in figure 9^[6].

Function identifiers	
Categories	Functions
ID	Identify
PR	Protect
DE	Detect
RS	Respond
RC	Recover

Fig. 9 Function identifiers^[6].

Framework Implementation Tiers define how an organization views cybersecurity risk and how it manages risk. They describe the level of management, from reactive to adaptive and agile. This permits an organization to “see” itself and determine how risks are managed. For instance, intrusion detection and response may have a well-developed process, while a natural disaster contingency may have little planned response action, providing the organization an assessment of agile in the first instance and reactive assessment in the second. Identifying differences between response levels informs the Framework Profile^[6].

The Framework Profile represents the outcomes based on the business needs selected from the framework categories and subcategories. Profiles can be used by an organization to identify areas for cybersecurity improvement. Profiles can inform the current state of security and present the desired end state. Based on the gaps between current and end state profiles, the organization can assess risk and allocate resources based on what is most important for business operations^[6].

Implementation of the framework is not without challenges. The Government Accountability Office (GAO) found that many CI sectors have not implemented the cybersecurity framework due to a lack of resources, lack of knowledge and skills to implement it, and regulatory and industry requirements preventing implementation. Some CI sectors had concerns over the disclosure of vulnerabilities or other priorities, such as physical security and direct support to customers. Some sectors perceived no cyber threat at all and believed that there was no need to use the framework^[32].

While some of these arguments are relevant, they indicate a lack of knowledge of the framework’s purpose and intent. The Cybersecurity Framework for CI clearly states^[6]:

The Framework complements, and does not replace, an organization's risk management process and cybersecurity program. The organization can use its current processes and leverage the Framework to identify opportunities to strengthen and communicate its management of cybersecurity risk while aligning with industry practices. Alternatively, an organization without an existing cybersecurity program can use the Framework as a reference to establish one (p.4).

Addressing cybersecurity concerns within a limited budget with personnel who are primarily involved in operating facilities or performing IT functions is difficult at best. The framework maps to industry standards without dictating which ones a facility must use. How leadership applies the resources they have depends on the risks they identify and their perceived threat to business operations.

V. PRACTICAL TOOLS FOR ASSESSING RISK

Risk assessments are critical in determining where the greatest vulnerability and return on investment are for a facility. All three frameworks call for assessing risk. Several tools are available to water facilities at no cost to help organizations practically identify and mitigate risks. Some of these tools are automated programs that map the network to help operators understand the flow of data, while others are computer-driven queries that populate a spreadsheet with recommended best practices. Several of these tools are discussed below^{[40],[41]}.

The Cybersecurity Evaluation Tool (CSET) is free, downloadable desktop software that guides operators and system owners through a step-by-step guide to assess cybersecurity practices^[40]. It correlates answers obtained through queries with accepted industry practices for securing networks. Data entered into the system is protected by the Protected Critical Infrastructure Information program; this enables private sector entities to pass information to DHS without fear of litigation or public disclosure^[40].

The Vulnerability Self-Assessment Tool (VSAT) is a water sector-specific tool developed by the EPA to help water facilities identify the most vulnerable areas and find the most cost-effective measures to reduce those risks^[40]. Like CSET, it is freely downloadable, but can be run from a web browser. Data is not retained by the EPA, protecting sensitive information about individual facilities.

A third tool is the Design Architecture Review (DAR) assessment, which reviews network architecture and security controls, looking at data flow, communication sharing, and proper communication channels^[42]. The Network Architecture Verification and Validation (NAVV) assessment, another type of review, passively monitors data traffic to determine whether there are leaks across boundaries and identifies anomalous behavior^[40]. Neither of these assessments requires connection to the OT or IT network at a facility.

National Cybersecurity Assessment and Technical Services is a team that can conduct penetration testing to test the security measures implemented by a facility. This is a valuable

resource to determine whether measures put in place after a security review are effective, achieving step 5 of the NIPP-RMF^[40].

The Cyber Resilience Review (CRR) is the sixth type of assessment freely available through DHS. It can be done as a self-assessment program or facilitated by DHS experts. It is designed to help organizations use the cybersecurity framework. The CRR addresses efficiency by balancing risks and costs, provides a roadmap by determining the best standard for an organization to use, and addresses the internal and external challenges of an organization^[43].

The risk assessment tools outlined above are free of charge. As an example, VSAT can be used to assess risk and increase the security posture of a facility. Beginning with the choice of quantitative or qualitative method for assessing risk, it leads a user through specific questions about the water utility, including questions about assets, countermeasures, and threats. The current risk to the facility based on the threats/assets input and existing countermeasures is provided as an output. Improvement recommendations are presented after completing the baseline assessment and a cost/risk analysis is used to develop new packages of countermeasures that conform to existing budgets or can be executed over a period of time. Finally, the VSAT can generate analysis result reports developed using the inventories of assets, threats, and countermeasures.

The tool has a demonstration mode with prefilled data to enable new users to understand the relationship between different values and the impact on operations if a component fails or is attacked. Key parameters and areas where data are entered are outlined below.

The Asset Selection screen is where facility-specific assets can be selected for analysis. The screen is prepopulated with common assets, such as generators, pumps, wells, instrumentation, and valves. Customization can be done by editing existing assets for system-specific items.

The countermeasures section of the VSAT allows user-defined countermeasures to threats to be entered. Similar to the asset selection, it is populated with common countermeasures. The countermeasure inputs, along with the asset inputs, form the baseline risk assessment for the facility. Unique inputs can be added to the countermeasure screen to tailor it to the water facility.

The Baseline Analysis performs analysis on one asset/threat combination at a time. It indicates the relative financial cost of a compromise. It queries the ability to reduce the consequence levels of an incident, given the ability to detect, delay, or respond. The system asks for the likelihood of occurrence and, combined with the previous responses, provides baseline risk and resiliency metrics.

Subsequent queries request potential improvements to existing countermeasures and the likelihood of damage if a vulnerability is successfully exploited. These queries provide results of cost savings and reduced likelihood of damage, expressed as percentages. These queries allow a facility to compare its existing security posture to its future posture if countermeasures are improved and displays this as a monetized amount of risk reduction.

Finally, the Results and Reports section summarizes the vulnerability assessment. The section can represent the data in a narrative format or as a chart. The section can also display the monetized risk metrics and resilience metrics of the assessment. The Results section may be used to drill down on the specific risks related to an asset/threat combination. Figure 10 shows the monetized risk output associated with the threats and vulnerabilities and other data input in the earlier portions of the query.

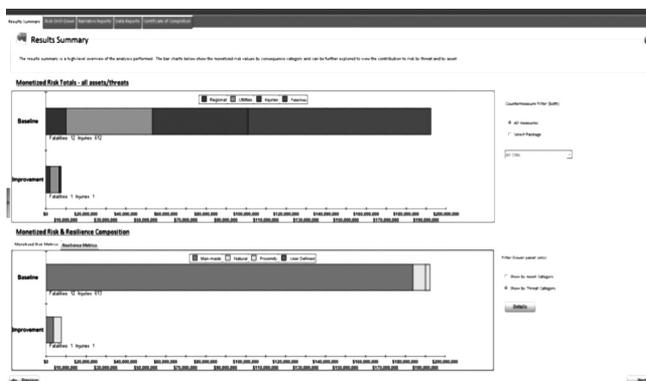


Fig. 10 Results Summary^[44].

On the whole, the water sector has completed more assessments to identify vulnerabilities than any other sector^[42]. While this places water and wastewater facilities ahead of peer CI, the challenge of securing decades-old SCADA equipment remains.

VI. PRACTICAL WAYS TO IMPLEMENT AN ACTION PLAN

Based on the assessment results, decisions can be made regarding which areas are most important to address. In reality, a local facility will still have a small budget for security and may not be able to apply resources to some areas highlighted as a risk, nor have the operational capacity to maintain them over the long term. However, some security improvements can be made at a low cost.

Information sharing and coordination is an area where risk management gains can be made with minimal effort. Free information updates from organizations such as the Water ISAC (WaterISAC) are available for water facility managers to stay abreast of trends in cyber threats^[44]. Coordinating with local emergency services, critical partners (such as electric service providers), and public health agencies prior to an incident can improve response and recovery operations^[45].

Training, education, and coordination are first steps, but the implementation of software, hardware, and physical security requires finesse. OT and IT networks have similarities, but the specialized nature of ICS equipment sometimes prevents patching or other standard IT security measures from being implemented^[7]. Updating ICSs by replacing old equipment in

wholesale fashion is not feasible for most facilities^[14]. Costs associated with expansive security software and hardware implementation are often prohibitive for local facilities^[8].

Using technology such as preprocessors can be an inexpensive and effective way to reduce some common risks to water sector ICSs (figures 11 and 12). Researchers at the University of Louisville demonstrated this concept in 2012. A preprocessor is a security module built on a small circuit board that is placed before a field SCADA device with either a software interface at the HMI point or another board in the same location to allow control of the field unit. This does not require replacement of equipment being added in-line to existing architecture. A Gumstix[®] circuit board was used in this experiment at the cost of only a few hundred dollars^{[7],[48]}.

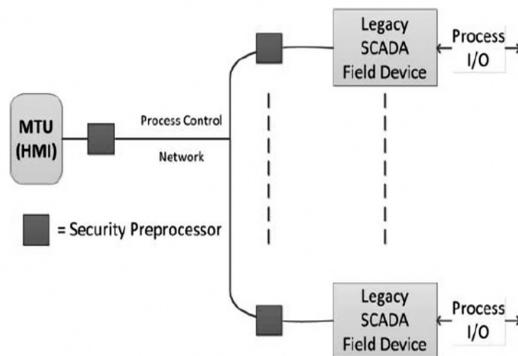


Fig. 11 Preprocessor integrated with ICS architecture ^[7].

The device provides authentication and authorization on behalf of the SCADA device. By configuring the Modbus protocol – a common protocol used in ICSs – to incorporate a connection request, challenge, and challenge-response, and incorporating Role Based Access Control (RBAC), users are only able to perform functions for which they have authorization (see figures 11 and 12). The device uses a simple operating system known as “OKL4” to reduce overhead. Further research by Schreiber indicates that a Bloom filter is a viable option for enforcing RBAC that limits the amount of bandwidth required to operate^{[7],[48]}.

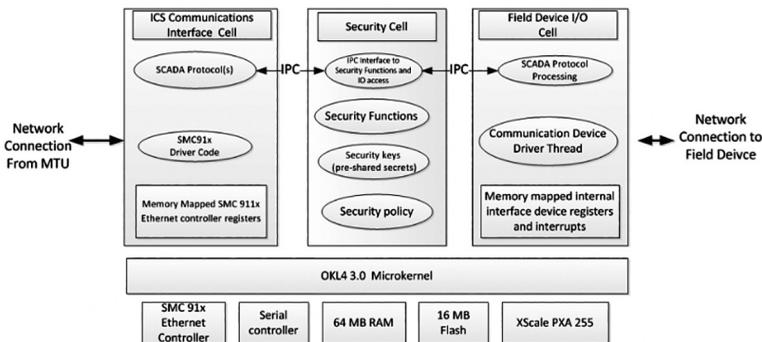


Fig. 12 Preprocessor architecture ^[7].

VII. FIELD IMPLEMENTATION

The implementation of the cybersecurity framework and the tools previously highlighted by water and wastewater treatment facilities has varied. In its February 2018 report on framework implementation by the GAO, the EPA reported it does not have the statutory ability to collect information on implementation of the framework by the water sector, and that it had no plans to implement a methodology to do so^[32].

This perspective was not unique to the water sector. A dearth of information on framework implementation was ubiquitous across all 16 CI sectors^[32]. The water sector is the most proactive of the CI sectors in leveraging assessment resources, however. From 2009 to 2014, 128 on-site assessments were conducted by the sector, which was double the number conducted by the next closest sector in the same amount of time^[43].

Reasons for not leveraging security assessment tools at the local level included lack of awareness of tool availability, limited understanding of cyber threats to the facility or sector, lack of personnel to dedicate to conducting security risk assessments, reluctance to share sensitive information, and an absence of directives from higher echelons to implement risk assessments^[32], [2]. The primary focus of the facilities is to provide the service which they are mandated to provide. While security was not entirely ignored, water reclamation and purification was prioritized over other activities. Time to dedicate to security considerations was limited^{[32],[52]}.

One local facility manager who was interviewed depended on the state to manage security concerns. The manager was unaware of WaterISAC or the tools available. While the importance of security was not misunderstood, daily operations had primacy.

In 2015, the EPA published the results of a pilot test of a contamination warning system (CWS) conducted jointly with five water utilities across the U.S. Its purpose was to examine detection of and response to drinking water contamination. Cybersecurity was an important component of the program, with an emphasis on the detection of contamination (with a minimum of false positives), operational reliability, and early detection to improve response time^{[32],[52]}.

The report highlighted the importance of communicating the value of the program to personnel, the impact to daily operations, and how it enhanced core job functions. Support from senior management, education of key leaders, and inclusive engagement across the staff were particular lessons learned. In the latter instance, it was discovered one pilot site did not engage its IT personnel and found the design of the information system to be infeasible because it conflicted with IT requirements. While the report focused on a CWS, the challenges of incorporating the multiple facets of a new process are applicable to instituting and assessing cybersecurity at the local level of the water sector^[52].

VIII. CONCLUSION

The increasing number of ICS vulnerabilities identified by researchers and industry experts, coupled with continuing revelations about ICS compromises, emphasizes the importance of securing CI. The security of water sector ICSs is undeniably important in its own right, but is also important for other CI sectors. Water sector ICS security is necessary for safe drinking water, environmental safety, growing food, cooling equipment for businesses and hospitals, and manufacturing.

As the water sector ICSs increasingly leverage routing protocols and automation equipment to reduce manning requirements and increase productivity, the potential for system vulnerability exploitation will increase. Evolving threats to water CI through cyberspace place an increased burden on local water facilities to protect their resources. They are especially challenged as they often do not have the training or equipment to identify and mitigate the risks to their systems. They may be able to apply only limited risk reduction measures by allocating personnel, funding, and materiel against specific threats.

Defending water sector ICSs from attack cannot be viewed as a separate function relegated to IT personnel or system operators; rather, it must be viewed as part of a whole-of-business approach to risk. Leveraging the NIST RMF, NIPP-RMF, and Cybersecurity Framework for CI as methodologies for categorizing cyber risk will aid organizations in holistically viewing risk across the enterprise. These RMFs aid organizations in allocating resources to achieve the greatest returns on their investments.

Several assessment tools exist to help executives and operations personnel apply the principles of the NIST RMF, NIPP-RMF, and Cybersecurity Framework for CI. Some, like CSET, CRR, and VSAT, can be performed at a local level without external support. Others, like NAVV and DAR, are facilitated by DHS at no cost to the local facility; these tools help identify vulnerabilities on the network and areas for improving network security. Some cost-effective measures, such as installing preprocessors at legacy water sector facilities to prevent unauthorized system access, can be implemented.

Using the NIST RMF, NIPP-RMF, and Cybersecurity Framework for CI with best network security practices, local water sector leaders can advance the security of their facilities while preserving the operational purpose of their facilities.♥

ACKNOWLEDGMENT

I would like to thank Dr. Roxanne Everetts for her mentorship and guidance, Dr. Chen for reading my paper and providing input, and my wife Erin and les trois soeurs for their patience and support.

NOTES

1. B. Ireland, "Security Risks," EC&M Electrical Construction and Maintenance, pp. 10-16, January 2012.
2. C. Copeland, "Terrorism and Security Issues Facing the Water Infrastructure Sector," Congressional Research Service, Washington D.C., 2010.
3. A. Bolshvov and I. Yushkevich, "SCADA and Mobile Security in the Internet of Things Era," 6 January 2017. [Online]. Available: [https://ioactive.com/pdfs/SCADA-and-Mobile-Security-in-the-IoT-Era-Embedi-FINALab%20\(1\).pdf](https://ioactive.com/pdfs/SCADA-and-Mobile-Security-in-the-IoT-Era-Embedi-FINALab%20(1).pdf). [Accessed 28 January 2018].
4. National Institute of Standards and Technology, SP 800-37r1 Information Security: Guide for Applying the Risk Management Framework to Federal Systems, Gaithersburg: NIST, 2010.
5. Department of Homeland Security, "National Infrastructure Protection Plan 2013," Department of Homeland Security, Washington D.C., 2013.
6. National Institute of Standards and Technology, "Framework for Improving Critical Infrastructure Cybersecurity," February 2014. [Online]. Available: <https://www.nist.gov/sites/default/files/documents/cyberframework/cybersecurity-framework-021214.pdf>.
7. J. Hieb, J. Graham, J. Schreiber and K. Moss, "Security Preprocessor for Industrial Control Networks," in Proceedings of the International Conference on Information Warfare and Security, 2012.
8. Congressional Budget Office, "Public Spending on Transportation and Water Infrastructure, 1956 to 2014," Congress, Washington D.C., 2015.
9. G. W. Bush, "Homeland Security Presidential Directive 7," 17 December 2003. [Online]. Available: <https://www.dhs.gov/homeland-security-presidential-directive-7>.
10. W. Clinton, "Presidential Decision Directive NSC-63," 16 October 1998. [Online]. Available: <https://fas.org/irp/offdocs/pdd/pdd-63.htm>. [Accessed 18 February 2018].
11. B. Obama, "Executive Order 13636 Improving Critical Infrastructure Cybersecurity," White House, Washington D.C., 2013.
12. B. Obama, "PPD 21," White House, Washington D.C., 2013.
13. Department of Homeland Security and Environmental Protection Agency, Water and Wastewater Sector-Specific Plan 2015, Washington D.C.: EPA, 2015.
14. L. Van Leuven, "Water/Wastewater Infrastructure Security: Threats and Vulnerabilities," in Handbook of Water and Wastewater Systems Protection, Springer, 2011, pp. 27-46.
15. Government Accountability Office, "Critical Infrastructure Protection: Multiple Efforts to Secure Systems are Underway but Challenges Remain," Government Accountability Office, Washington D.C., 2007.
16. Department of Homeland Security, "ICS-CERT Annual Assessment Report FY2016," Department of Homeland Security, Washington D.C., 2017.
17. Positive Technologies, "ICS Security: 2017 in Review," Positive Technologies, 2018.
18. L. Newman, "Menacing Malware Shows the Dangers of Industrial System Sabotage," 18 January 2018. [Online]. Available: <https://www.wired.com/story/triton-malware-dangers-industrial-system-sabotage/>. [Accessed 20 January 2018]
19. K. Zetter, "An Unprecedented Look at Stuxnet, the World's First Digital Weapon," November 2014. [Online]. Available: <https://www.wired.com/2014/11/countdown-to-zero-day-stuxnet/>. [Accessed 20 February 2018].
20. K. Zetter, "Iran: Computer Malware Sabotaged Uranium Centrifuges," 29 November 2010. [Online]. Available: <https://www.wired.com/2010/11/stuxnet-sabotage-centrifuges/>. [Accessed 20 February 2018].
21. D. M. Birkett, "Water Critical Infrastructure and Its Dependencies," Journal of Terrorism Research, vol. 8, no. 2, pp. 1-21, May 2017.
22. D. Kroll, K. King, T. Engelhardt, M. Gibson, K. Craig and Hach Homeland Security Technologies, "Terrorism Vulnerabilities to the Water Supply and the Role of the Consumer: Water Security White Paper," March 2010. [Online]. Available: <http://www.waterworld.com/articles/2010/03/terrorism-vulnerabilities-to-the-water-supply-and-the-role-of-the-consumer.html>. [Accessed 20 February 2018].
23. 107th Congress, "The Public Health Security and Bioterrorism and Response Act of 2002," 12 June 2002. [Online]. Available: <https://www.gpo.gov/fdsys/pkg/PLAW-107publ188/pdf/PLAW-107publ188.pdf>. [Accessed 18 February 2018].

NOTES

24. T. G. Lewis, "Critical Infrastructure Protection in Homeland Security : Defending a Networked Nation," in *Critical Infrastructure Protection in Homeland Security : Defending a Networked Nation*, Palo Alto, John Wiley & Sons, 2015, pp. 185-203.
25. M. Lewis, "Cross-connection and Backflow Devices," January 2011. [Online]. Available: https://www.wvdhhr.org/oehs/eed/swap/training&certification/cross-connection&backflow/documents/Cross_Connection_Backflow_Prevention.pdf. [Accessed 20 March 2018].
26. S. Jerome, "Federal Officials Warn Rural Water Systems Of Cyber Threats," 17 March 2017. [Online]. Available: <https://www.wateronline.com/doc/federal-officials-warn-rural-water-systems-of-cyber-threats-0001>. [Accessed 3 March 2018].
27. B. Rios and T. McCorkle, "McCorkle and Rios: 100 bugs in 100 days," 7 October 2011. [Online]. Available: https://www.youtube.com/watch?v=29S_Beg71dA. [Accessed 17 February 2018].
28. R. McMillan, "Hackers break into water system network," 1 November 2006. [Online]. Available: <https://www.infoworld.com/article/2659670/security/hackers-break-into-water-system-network.html>. [Accessed 3 March 2018].
29. M. Abrams and J. Weiss, "Malicious Control System Cyber Security Attack Case Study-Maroochy Water Services Case Study," 28 July 2008. [Online]. Available: https://www.mitre.org/sites/default/files/pdf/08_1145.pdf. [Accessed 30 January 2018].
30. R. A. Clark and R. A. Deninger, "Protecting the Nation's Critical Infrastructure:," *Journal of Contingencies and Crisis Management*, vol. 8, no. 2, pp. 73-80, 2000.
31. Government Accountability Office, "Securing Wastewater Facilities: Costs of Risk Assessments, Risk Management Plans, and Alternative Disinfection Methods Vary Widely," March 2007. [Online]. Available: <https://www.gao.gov/assets/260/258480.pdf>. [Accessed 2 February 2018].
32. Government Accountability Office, "Critical Infrastructure Protection: Additional Actions Are Essential for Assessing Cyber-security Framework Adoption," Government Accountability Office, Washington D.C., 2018.
33. P. Pederson, D. Dudenhoffer, S. Hartley and M. Permann, "Critical Infrastructure Interdependency Modeling: A Survey of U.S. and International Research," Idaho National Laboratory, Idaho Falls, 2006.
34. P. Valentia, *Water Sector Cyber Threat Briefing*, 2018.
35. American Water Works Association, "Process Control System Security Guidance for the Water Sector," AWWA, 2017.
36. E. G. Bachman, "Pre-planning for Emergencies at Water Treatment Facilities," *Fire Engineering*, vol. 156, no. 8, pp. 120-130, August 2003.
37. WaterISAC, "10 Basic Cybersecurity Measures," June 2015. [Online]. Available: https://ics-cert.us-cert.gov/sites/default/files/documents/10_Basic_Cybersecurity_Measures-WaterISAC_June2015_S508C.pdf. [Accessed 24 February 2018].
38. D. Verner, F. Petit and K. Kim, "Prioritization in Critical Infrastructure," *Homeland Security Affairs*, vol. 13, October 2017.
39. National Institute of Standards and Technology, "SP 800-39 Managing Information Security Risk: Organization, Mission, and Information System View," March 2011. [Online]. Available: <https://csrc.nist.gov/publications/detail/sp/800-39/final>.
40. DOD, "DODI 8510.01 DIACAP," November 2007. [Online]. Available: <http://www.dtic.mil/dtic/tr/fulltext/u2/a551538.pdf>.
41. Department of Homeland Security, "NCCIC ICS-CERT Assessments FAQ," May 2016. [Online]. Available: https://ics-cert.us-cert.gov/sites/default/files/documents/NCCIC%20ICS-CERT%20Assessment%20FAQ_S508C.pdf. [Accessed 23 February 2018].
42. Environmental Protection Agency, "Conduct a Water or Wastewater Utility Risk Assessment," EPA.gov, 12 October 2017. [Online]. Available: <https://www.epa.gov/waterriskassessment/conduct-drinking-water-or-wastewater-utility-risk-assessment>. [Accessed 19 February 2018].
43. M. McWhirt, "August 20, 2014 Cybersecurity Assessments and Tools DHS," 20 August 2014. [Online]. Available: <https://www.waterisac.org/portal/august-20-2014-cybersecurity-assessments-and-tools-dhs>. [Accessed 23 February 2018].
44. Environmental Protection Agency. (2017, October 26). VSAT 6.0. Retrieved July 2, 2018, from EPA.gov: <https://vsat.epa.gov/vsat/>.
45. K. Dillon, "Cybersecurity Assessments and Tools by DHS," 20 August 2014. [Online]. Available: <https://www.waterisac.org/portal/august-20-2014-cybersecurity-assessments-and-tools-dhs>. [Accessed 23 February 2018].
46. WaterISAC, "About Us," 2017. [Online]. Available: <https://www.waterisac.org/about-us>. [Accessed 5 March 2018].

NOTES

47. Environmental Protection Agency, "Learning from State Water Emergency Response Exercises," May 2012. [Online]. Available: <https://www.epa.gov/waterresiliencetraining/learn-state-water-emergency-response-exercises>. [Accessed 2 February 2018].
48. Government Accountability Office, "Wastewater Facilities: Experts' Views on How Federal Funds Should Be Spent to Improve Security," GAO, Washington D.C., 2005.
49. Environmental Protection Agency, "National Primary Drinking Water Regulations," 21 August 2016. [Online]. Available: https://www.epa.gov/sites/production/files/2016-06/documents/npwdr_complete_table.pdf. [Accessed 18 February 2018].
50. J. Schriever, Role based access control and authentication for SCADA field devices using a dual Bloom filter and challenge response, 1281 ed., Kentucky: University of Louisville, 2012.
51. National Institute of Standards, "Guide to Industrial Control System (ICS) Security," May 2015. [Online]. Available: <https://csrc.nist.gov/publications/detail/sp/800-82/rev-2/final>. [Accessed 24 February 2018].
52. Environmental Protection Agency, "WSI Pilot Summary Report," October 2015. [Online]. Available: https://www.epa.gov/sites/production/files/2015-12/documents/wsi_pilot_summary_report_102715.pdf. [Accessed 4 September 2018].

SESSION

◆ 2 ◆

Applied computational social choice theory as a framework for new cyber threats

David M. Perlman, Ph.D.

CogSec Technologies Inc

San Francisco, California, USA

ABSTRACT

Social media and “big data” have combined to create a new era of marketing, political campaigning, and hostile propaganda. The *tactics*, such as microtargeting of ads, have recently received intense public scrutiny. However, little has been publicly said about the tools and techniques of *strategy*. In this context, Applied Computational Choice (ACSC) refers to a framework for analyzing data, modeling tactics, and planning strategy. Here we describe an ACSC framework derived from the work being done by some of the main actors and apply it to show how a few simple scenarios can be modeled and realistic behaviors predicted, as well as illuminate possible motivations for certain patterns observed in the real world. We introduce the concept of vulnerability assessment applied to voting systems by analyzing the cost of influence operations on simple model voting systems. We believe this framework reflects those being used by a number of different actors with goals and hope that this article helps to provide an overview and introduction to the field.

Index Terms—data science, political science, propaganda, influence, information warfare, narrative warfare, weaponized demographic

I. INTRODUCTION

Although adversarial propaganda is as old as war itself, recently new techniques have been implemented with unprecedented power, speed, and effectiveness in a number of political contests around the world. Technological advance in the application of computational social choice theory^[1], mass profiling^[2], and microtargeting^{[3], [4]} have been developed by a number of sources, including tech media companies, private marketing and campaign data businesses, as well as hostile state and/or non-state actors. In addition to the societal ethical concerns, it is now apparent that hostile actors have developed extensive art and proficiency in using these technologies offensively in ways that are critically relevant to national security and the military^{[5]-[8]}. Broad awareness of this threat is newly dawn-

© 2019 Dr. David M. Perlman

ing and terminology is not yet standardized; various terms are used, including Cyber-enabled Information Operations (CyIO)^{[6], [9]}, Information Warfare and Influence Operations (IWIO)^[10], and Information/Influence Warfare and Manipulation^[5]. The battlefield of the new information warfare is the information environment^{[5], [7]}, especially “social media”^[10], the globally pervasive sphere of media-rich, personal and social communication that has evolved out of the original handful of social networks and which is perfectly suited to communicating emotional information that bypasses rational filtering. The weapons and tactics are narratives^[11]; the delivery vehicles are “memes” (in the original sense, as well as the contemporary meaning), which are units of information that are tuned for, first, rapid propagation by the humans in the social network, and, ultimately, assimilation into the mass mentality for the promotion of disruptive and harmful politicians or agendas and other political and social goals^[12]. With respect to these cutting-edge social media techniques, a large portion of the publicity has recently been focused on Cambridge Analytica, Facebook (FB), and the Internet Research Agency in St. Petersburg, but it is virtually certain that actors all over the world are engaged in research into and deployment of these techniques.

In part, ACSC is a natural extension of advertising and marketing science developed in the context of this social media sphere and a highly competitive, largely unregulated marketplace. The presence of a significant fraction of the world’s population on social networks turns mass psychology and behavior manipulation into computational big data problems. Social choice theory and the models described here originate in the economics and political science literature dating from as far back as 1957^[13], with extensive theoretical work being performed on the topic in the 1960s and 1970s^{[14]-[17]}. Computational approaches to social choice theory appeared more recently^[1] and continue in earnest. Practical, technological applications of computational social choice theory became possible only very recently in the age of big data and social media, and serious academic research has only become prominent since around 2016. There is, obviously, an enormous market for understanding and influencing population psychology and behavior; the most famous companies of our era, such as Google, Facebook, etc., spend much of their effort studying this domain and developing techniques, tools, algorithms, and other kinds of expertise. In addition, it seems clear by now that some state actors have devoted tremendous time and attention to understanding the role of social media and the internet in population influence. However, both the private corporations and the governments doing this work have powerful incentives to keep their innovations secret, so little has been published regarding the formalism and techniques useful in this domain. When private research has been published, it has generated strident criticism^{[18], [19]}. Thus it is not surprising that research into and implementation of applications of these technologies are largely hidden behind a veil of secrecy. Here we share a general framework or formalism for political data science which we believe is representative of how some of these actors may be operating and then apply the formalism to suggest qualitative outlines of how several plausible scenarios could be conducted.

The information in this article has been inferred from extensive conversations with a number of individuals across various related fields, and synthesized with general principles of data

science and linear algebra to form this framework. Some of the contributing individuals have reviewed the framework and, without divulging trade secrets, have agreed that it is compatible with many of the important considerations they would raise. It is my hope that this article will serve as a useful introduction to what might be possible and provide a common terminology and framework for those wishing to study ACSC more openly. The goal of this article is to synthesize a framework that is optimized for practical applications in industry and defense, specifically the application of identifying, understanding, and countering large-scale influence operations in the big data context of online digital platforms. To serve this goal, the primary considerations have been utilitarian rather than theoretical: little of the material is theoretically novel; it has been selected from other sources to be useful for this endeavor, and no effort has been made to comprehensively cover mathematical voting theory or any other existing field. The specific examples given are not intended to prove that any specific activities actually happened in the real world (although we have our suspicions!), but instead are intended to stimulate intuition with respect to what is possible in this domain and encourage and support more researchers wishing to enter this field.

We have intentionally avoided two crucial topics, not because they are unimportant, but because they are already receiving extensive attention: individual psychology and agent and network-based simulations of organic social behavior. There is currently a cottage industry around fake news, misinformation, manipulation, and bias, and a thoughtful awareness of much of the most important work in the field is readily available to the public and professionals alike. Likewise, many people are investigating the natural patterns of propagation of information through social networks and the formation of cliques and cults. The goals that we hope to serve with this paper ultimately will rest on a foundation of knowledge of individual cognition and emotion and emergent, aggregate phenomena. Here we address only the edge of the field that we believe to be most critically underserved.

II. THE FRAMEWORK

A. Ideological Space

We will follow the traditional structure of computational social choice theory^{[14], [16], [17], [20], [21]}. We construct a preference space over political ideologies which we will call “ideological space” or “policy space”. Consider a population BN of N individual agents in a particular society who share some set of K issues of political or ideological interest. Without loss of generality, we can represent each political/ideological issue as a real number in $[-1, 1]$ corresponding to the agent’s response to an issue question on a continuous, Likert-type agree-disagree scale. Each individual’s preferences are represented by a K -dimensional vector \mathbf{b} ; we will write \mathbf{b}_i for the preference vector for individual i . We will assume a K -dimensional Euclidean coordinate space defined by taking each of the K issues as a dimension. (In some unusual applications, the assumption of a Euclidean space may be limiting, but is a tremendously useful starting point because it facilitates conceptual intuition as well as computation.) Within this Euclidean space, the “ideological space” of all possible configurations of beliefs is P^K , the K -dimensional cube

centered on the origin with edges of length 2. Thus the complete population preference set B_N^k is a set of N points inside the K -dimensional cube P^K , $B_N^k = \{\mathbf{b}_i \in P^K : i \in [1, N]\}$.

Of the K issues/dimensions, it is likely that some issues are highly correlated with each other. Assume that it is possible to apply an appropriate dimension-reduction process to determine the actual number of underlying, latent, ideological dimensions k . This defines a reduced ideological space P^k , a k -dimensional cube in k -dimensional Euclidean space, and a reduced population preference set B_N^k :

$$B_N^k = \{\mathbf{b}_i \in P^k : i \in [1, N]\} \quad (1)$$

The k dimensions might be thought of as the fundamental philosophical, moral, emotional, etc., beliefs that form the foundation for individuals' preferences on policy issues. It would require empirical investigation to assign meaning to these k reduced ideological dimensions and, in practice, in large-scale applications of this framework it may not be possible to extract dimensions that appear meaningful in human terms. In fact, in big data applications generally, dimension reduction is a non-obvious problem. It is likely that much work is happening behind closed doors to develop these basic techniques.

B. New Technology for Surveying the Ideological Space

The traditional method of estimating B_N^k is, essentially, opinion polling. Political polls that attempt to model "likely voters" are attempting to estimate a reduced set $B_{n_{vt}}^k$ and a turnout function VT (see equation 2 below). Recently, there has been a great deal of attention given to the possibility of measuring psychological profiles, political preferences, etc., from online social media behavior, search history, and other internet sources^{[2]-[4], [22]-[24]}. Most of the discussion of this topic has focused on the individual level: the privacy implications of collecting and modeling personal information without an individual's knowledge and the implications for personal autonomy of precisely microtargeted advertising or propaganda. These new techniques may also allow greatly improved speed and accuracy of estimation of B_N^k : increased speed due to the use of massive online databases that contain daily or even faster updates for many people, and increased accuracy due to the freedom from response biases with surreptitious modeling as well as the very large sample sizes available. This increased power is what makes possible rapid and powerful operations such as short-term manipulation of political preferences before an election, effective on the order of days or weeks^[25], leaving no time for any effective tactical response. This is especially true for Western liberal democracies that do not currently have national defense capabilities in this domain at all.

C. The Curse of Ideological Dimensionality

The first insights from the framework can be gleaned by considering the question of the actual dimensionality k of the ideological space P^k for real-world societies. A plausible guess is that it would be comparable to the number of distinct political issues identifiable in the news and other media of the society at any given time. In general, the specification of the analytical

framework will be different depending on the goals of the given application, especially the population in question and the time frame in question. The necessary value of k for a useful application of the model will be greater if we wish to study how a population's beliefs evolve over time because some issues will be forgotten as new ones arise; however, the model must contain dimensions for all of them in order to represent this drift. If we wish to study a population over a long stretch of time, then the necessary k may be very large. The necessary k will also be greater if we wish to study a larger and/or more diverse population because, although any individual may only be aware of a small number of issues, across a large population, a larger number of issues will be represented. Importantly, in many real-world societies, it is also likely that k has recently increased as the proliferation of digital news media sources has increased the total number of different issues of which the citizens are collectively aware.

As the number of ideological dimensions k increases, a number of practically relevant phenomena can be expected based on the “curse of dimensionality”^{[21], [26], [27]}:

- 1) ***Individuals disagree with each other more:*** The expected distance between any two randomly chosen points increases.
- 2) ***The population overall becomes more dissatisfied with any platform that specifies a complete set of policies (e.g., the actual government policy at any moment):*** The average distance between all of the points in B_N^k and the single point G^k that represents the platform increases.
- 3) ***The potential for insurmountable disagreements increases:*** The maximum possible distance between any two points increases even faster than the expected distance.
- 4) ***A smaller proportion of the ideological space is taken up by moderates and a greater proportion by the fringe, even for a highly inclusive definition of moderate:*** We can choose a reasonable definition of “moderate” in the policy space as being “near the center” and represent this with $S_{r_{mod}}^k$, the centered k -ball of radius r_{mod} (i.e., the region of policy space that is within the distance r_{mod} of the center). Then the ratio of the volume of $S_{r_{mod}}^k$ to the volume of the policy space P^k goes to zero as k increases, even if we choose $r_{mod} = 1$.
- 5) ***Extending that, depending on the population's distribution of beliefs, an increasingly greater share of the population will find that its values and beliefs fall outside of any of the available political parties:*** If we represent p political parties as p non overlapping k -balls of radius r_i , S_{i,r_i}^k , the ratio of the total volume of all the parties' territories $\sum_{i=1}^p (S_{i,r_i}^k)$ to the total volume of P_k also goes to zero.
- 6) ***The two previous points, taken together, imply that if a political party uses this kind of data analysis in its electoral or marketing strategic planning:*** it will be motivated to expand its ideological-space territory to include more and more of the fringe in an attempt to capture more of the electorate.

See the appendix for mathematical derivations of these effects.

III. DEMOCRACY: POLICY FROM POPULATION

A. Voting Models

Democracy, in a very general sense, refers to a system in which the policies and actions implemented by the government are intended to be consistent with the will of the population. Intuitively, this means that the point G^k that represents actual government policy in ideological policy space ought to be “in the middle” of the density of the point cloud B_N^k . We define a policy-choice function g that takes B_N^k as input and yields G^k as output. We further define g to be the composite $g = h(\tau)$, where τ is a turnout function and h is a complete turnout voting function. The turnout function τ determines a subset of $B_N^k, B_{n_\tau}^k$:

$$B_{n_\tau}^k = \tau(B_N^k) \quad (2)$$

The voting function h takes $B_{n_\tau}^k$ as input and yields a single point for G^k . Thus:

$$G^k = g(B_N^k) = h(\tau(B_N^k)) = h(B_{n_\tau}^k) \quad (3)$$

B. Turnout

A great deal of complexity and uncertainty is hidden in τ . Polling services devote large resources to modeling voter turnout, with limited success. Historically there have been a number of famously embarrassing and disruptive prediction errors based on errors in turnout modeling, such as the classic “Dewey Defeats Truman” headline^[28]. The outcomes of elections can be dramatically altered by changing turnout, and there is already reason to believe that hostile actors have engaged in microtargeted social media campaigns primarily oriented around voter suppression. Future expansion of this manuscript will include examples that consider the implications of changing τ in this framework.

C. Simple Examples

a) Technocratic Direct Democracy: The intuitive criterion that a democracy should yield actual government policies G^k that are “somewhere in the middle” carries over into h . As the (mathematically, not practically) simplest possible example, one can imagine a hypothetical “technocratic direct democracy” (TDD) where the full population has its policy preferences measured and then G^k is set at the average or centroid of B_N^k :

$$G_{\text{TDD}}^k = \bar{\mathbf{b}} = \text{mean}(B_N^k) \quad (4)$$

b) Two-Party Direct Democracy: Now we expand that reductionist model to include one additional element of complexity. Consider now the simplest possible example of a two-party voting system, which we might call “2-party direct democracy” (2PDD). B_N^k is divided into $B_{n_1}^k$ for the n_1 voters of Party 1 and $B_{n_2}^k$ for the n_2 voters of Party 2. Party 1 evaluates the preferences of their constituency and defines a platform G_1^k as

the centroid¹ of $B_{n_1}^k$ and Party 2 likewise defines G_2^k as the centroid of $B_{n_2}^k$. The implementation of the voting function then yields

$$G_{2\text{PDD}}^k = h(B_{n_1}^k \cup B_{n_2}^k) = G_j^k$$

$$j = \arg \max_{i \in \{1,2\}} n_i \quad (5)$$

c) Dictatorship: For comparison purposes, we can also describe G_{dictator}^k as a G^k that is dictated without regard to B_N^k . This may be thought of as a voting function g that is constant. Or, if other factors are known and available to be modeled, g may be a function of those other factors.

d) Other examples: We can also describe a number of other simplified example scenarios. Future expansion of this manuscript will describe:

- ◆ Indirect democracy: Voting districts and an “electoral college”
- ◆ Analyzing the effects of different voting systems, such as First Past The Post and Ranked Choice Voting
- ◆ Turnout defined over districts or other clusters
- ◆ Turnout as a selector function versus probability field T' over P_k
- ◆ Iterative feedback between political parties’ selection of issue-space territories and voters’ party alignment

IV. POPULATION INFLUENCE

A. Influence Cost Function

Among the three examples of G_{dictator}^k , G_{TDD}^k , and $G_{2\text{PDD}}^k$, we can consider what would be necessary for an influence operation to change policy by looking at how changes in B_N^k affect G^k . To allow comparisons, we can define a metric of “influence cost” for a change from B_N^k to $B_N'^k$. The simplest metric is based on the unweighted sum of Euclidean distances moved by each individual:

$$C_{\text{unweighted}}(B, B') = \sum_{i=1}^N |(\|B_i'^k - B_i^k\|_2)| \quad (6)$$

where $|\cdot|$ is the numerical absolute value and $\|\cdot\|_2$ is the (k -dimensional) Euclidean norm applied row-wise to the differences of the i th rows of $B_i'^k - B_i^k$. This can also be written

$$C_{\text{unweighted}}(B, B') = \|B_i'^k - B_i^k\|_{1,2} \quad (7)$$

where $\|\cdot\|_{1,2}$ is the $L_{1,2}$ matrix norm for row-wise data points in a matrix.

We can abbreviate:

$$C_{\text{unweighted}}(\Delta B) = \|\Delta B\|_{1,2} \quad (8)$$

¹ The centroid or mean of the constituency in the Euclidean space is neither plausibly realistic nor strategically optimal as an actual real-world choice of platform for a party. Any number of other considerations would come into play in the real world, especially turnout, loyalty, and other non-policy effects. In addition, there are also evolutionary and iterative effects in the emergence of parties; see, for example, [21] and [29] as a tiny, arbitrary selection of (not at all centroidal) examples of greater complexity. Our use of the centroid here is purely motivated by the choice of the computationally simplest starting point for this exposition.

B. Weighted Cost Functions

Different individuals will have different susceptibility to influence. To take this into account we can add weights w_i for each individual, represented in an $N \times N$ diagonal weight matrix W :

$$C_W(\Delta B) = \|W \Delta B\|_{1,2} \tag{9}$$

Different preference dimensions of the ideological space may have differing degrees of “stickiness”, as well; some may be easier to change people’s minds about than others. To account for this we can add another $k \times k$ diagonal weight matrix V with the weights v_j for each of the k preference dimensions:

$$C_{WV}(\Delta B) = \|W \Delta BV\|_{1,2} \tag{10}$$

C. Example scenarios

We will look at some simple “back-of-the-envelope” calculations of the cost of influence operations to explore the possibilities within the framework.

a) Dictatorship: In G^k_{dictator} , g is not a function of B_n^k . The most direct way to change G^k dictator would be to influence the dictator individually. Influence on B_n^k leads to changes in G^k dictator only to the extent that the dictator notices, cares, and reacts to the population change or, under a coup.

Call the region of ideological space that represents willingness to act on a grassroots coup Q , and call the minimum number of individuals necessary for a coup n_Q . Then the cost metric for influencing a coup is

$$C_{\text{dictator}}(Q) = \sum_{n_Q} \min(\|B_i^k - Q\|) \tag{11}$$

Here $\min(\|B_i^k - Q\|)$ refers to the distance from the point B_i^k to the nearest point in Q .

Qualitatively, with a few straightforward assumptions, we can interpret:

- 1) The cost of influencing a coup is proportional to the number of people who must be induced to participate, which is determined by the strength of the regime.
- 2) The cost of influencing a coup depends on how far the relevant slice of the population is from the “boiling point” Q^k . In other words, it’s easier to induce a coup in a population that is already dissatisfied.

This suggests that it may be possible for an actor with access only to data such as search and social media to remotely estimate the likelihood of regime change with little direct interaction.

b) Technocratic Direct Democracy: Consider the goal of moving G^k_{TDD} to a target Q . According to (4)

$$\begin{aligned}
 G_{\text{TDD}}^k &= \text{mean}(B_N^k) \\
 Q &= \text{mean}(B_N^k) \\
 \Delta B = Q - G_{\text{TDD}}^k &= \text{mean}(B_N^k) - \text{mean}(B_N^k) \\
 &= \text{mean}(B_N^k - B_N^k) \\
 &= \text{mean}(\Delta B_N^k)
 \end{aligned}
 \tag{12}$$

The specification of a target does not uniquely determine the influence cost because we do not know the trajectories of all of the individuals; however, the minimum possible influence cost arises in the situation where each individual moves in parallel to the overall movement, in the “forwards” direction:

$$\min C_{\text{TDD}} = N \times \|\Delta B_N^k\|
 \tag{13}$$

Under this hypothetical, minimal system (and ignoring for the moment the per-person and per-issue weights), there are no influence shortcuts: the cost of influence is proportional to the total population and the magnitude of the targeted change. In future work, we will present calculations suggesting that adding further mechanisms of complexity to the system will lead to more complex influence cost functions, which will have some variables that lead to greater costs and others that lead to less. An actor that performs a detailed analysis of the complete set of mechanisms of a voting system will be able to identify weak points that constitute the influence version of attack surfaces and design influence campaigns fine-tuned for the maximum sociopolitical impact with minimum cost. In turn, this tells us that a democracy must perform this same detailed vulnerability assessment of its own voting systems in order to defend effectively against influence attacks which could have devastating, paralyzing consequences.

V. THE OVERTON HULL

A. *The Original Overton Window*

The Overton window is a concept first put forth by Joe Overton of the Mackinac Center^[30] to refer to the range of public political discourse that is tolerated within a given society’s media ecosystem. The original concept referred to the segment on a unidimensional, left-right, political spectrum that represents the positions that, say, a politician can publicly profess and still expect to be taken seriously. It is important to clarify that the Overton window is a population-level concept: while it may be reasonable to talk about a “window” that an individual is willing to tolerate, we are interested in studying a society as a whole, so the concept in question relates to the emergent “window” across the society’s whole media ecosystem.

B. *Extending to k dimensions*

The idea of a unidimensional, left-right spectrum is certainly used for simple rhetoric, but in order to make the Overton window practically useful, we extend it here to a k-dimensional “blob,” the region within the ideological space P^k that represents those views that are acceptable within the media ecosystem of the society in question. Although it is rarely stated

explicitly, discussions of the Overton window universally assume that the window is a single, connected line segment. We can generalize that to define the “Overton hull” H_o as a convex region of P^k that represents the range of political, ideological views that are acceptable within the media ecosystem of population B_N .

C. Estimating the Overton Hull

Discussions of the Overton window usually assume that it is approximately centered around the bulk of the distribution of the population along the political spectrum; in other words, the majority’s political beliefs are within the window. We can make a first pass at a simple working definition of a measured H_o with P^k and B_N^k as our starting point.

First we postulate that B_N^k is a sample drawn from a distribution with probability density function f_B defined over the sample space P^k . Then H_o can be defined as the convex hull of the region of P^k in which f_B is over a threshold value d_{thr} :

$$H_o = \text{Conv} \{ \mathbf{a} \in P^k \mid f_B(\mathbf{a}) \geq d_{thr} \} \tag{14}$$

VI. WEAPONIZED DEMOGRAPHICS

“Useful idiots” is a term widely used since the Cold War to refer to individuals who are easily manipulated into serving hostile propaganda purposes, even though they may not actually support or even understand the issues at stake. In order to study the role of useful idiots in influence operations at the population, rather than individual, level, we can describe a useful idiot demographic B_{UI} as a subpopulation whose ideological preferences are particularly easy to manipulate. This ease of manipulation can be represented as low values of the cost function weights below a UI threshold $w_i < w_{UIthr}$ for these individuals:

$$B_{UI} = \{ \mathbf{b}_i \in B_N^k \mid w_i < w_{UIthr} \} \tag{15}$$

With these definitions, we can describe the Weaponized Useful Idiot Demographic (WUID), a mass-influence technique derived from the “door-in-the-face” (DITF) frequently discussed in the literature on the Overton window^{[31]-[33]}. We will also describe influencing H_o using traditional mass propaganda to provide a baseline for comparison.

A. Mass Propaganda, or The Bulk Move

Consider a target point Q which is outside the Overton hull H_o ; using traditional methods of nontargeted mass propaganda operating on the population at large, the attacker wants to move H_o to include Q . Let q_{surf} be the closest point to Q on the hull of H_o and q_{sq} be the vector from q_{surf} to Q , so that $\|q_{sq}\|$ is the minimum distance from H_o to Q . In order to “bulk move” the whole population’s average preferences over until Q is just inside H_o , we know from the consideration of C_{TDD} above that the cost will be approximately $C_{BM} \approx N \|q_{sq}\|$. We can now use this a baseline for comparing WUID.

B. The Weaponized Useful Idiot Demographic (WUID)

The WUID is a formalized variant of the “Door in the Face” technique frequently discussed in the literature on the Overton window. Let the attacker choose a “dummy target” point Q' near $lq_{sq} + q_{surf}$, which represents a “more extreme” version of the real target Q in the sense that it is farther away from H_0 , with the factor l determining how much more extreme it is. Because H_0 is a convex hull, if the density function f_B can be raised above d_{thr} in even a tiny region around Q' , then Q will immediately be included well within the Overton hull.

To accomplish this, choose a small subpopulation of M individuals from the useful idiots demographic, $B_{WUID} \subset B_{UI}$, where $M = m \times N$, $m \ll 1$. Although we can reasonably anticipate that the individuals in B_{WUID} will have markedly different individual preferences than the population at large, in the absence of any reason to believe they have a specific direction of political bias, assume that to begin with the average preferences of B_{WUID} are approximately the same as the average for the population at large, $\bar{b}_{WUID} \approx \bar{b}$. Then

$$\begin{aligned} \|q_{WUID}\| &\approx l \times \|q_{sq}\| + \|q_{surf} - \bar{b}\| \\ &\approx l \times \|q_{sq}\| \text{ if } l \gg 1 \end{aligned} \tag{16}$$

If the extremeness factor l is great enough, then the target movement distance $\|q_{WUID}\| \approx l \times \|q_{sq}\|$. This means that we expect the influence distance to be much greater in this case. However, consider the influence cost. Based on the derivation for C_{TDD} , we can see that

$$\begin{aligned} C_{WUID} &\approx \bar{w}_{UI} \times M \times l \times \|q_{sq}\| \\ &\approx \bar{w}_{UI} \times m \times N \times l \times \|q_{sq}\| \\ &\approx \bar{w}_{UI} \times m \times l \times C_{BM} \end{aligned} \tag{17}$$

By definition, we know that $l \gg 1$, but $m \ll 1$ and $\bar{w}_{UI} \ll 1$. This means that there is ample opportunity for a well-planned operation to have influence cost much lower than that of traditional mass propaganda, $C_{WUID} < C_{BM}$.

In qualitative terms, we can describe this operation as follows. First, the attacker identifies a particularly gullible demographic of useful idiots who are likely to be scattered around the fringes of society in their various beliefs. The attacker uses social media, search history, etc., to profile them and prepare targeted, narrative weaponry. The narratives might extensively incorporate the language of conspiracy theories to appeal to the fringe psychology. Next, the attacker uses microtargeted, viral, and mass-media delivery vehicles for the narrative weaponry to “lasso the fringe” into a WUID over which the attacker now has some degree of control. The WUID is induced to create a media-noticeable prevalence of dummy target ideology Q' , which immediately opens up the Overton hull to include Q , thus accomplishing the attacker’s goals faster and with much less cost than would be possible with traditional mass propaganda.² In fact, the attacker receives even more benefit from the WUID: this is, in essence, a reusable weapon; once the WUID has become accustomed to taking its cues from certain sources, it is likely to remain open to those sources for some time.

2 If the WUID were to be used to influence an election and install a puppet government, perhaps that government could then be referred to as a useful idiocracy.

C. Defense and Counter-offense

Having considered the potential power of the WUID attack, naturally questions of defense and counter-offense arise. In the long run, the best defense against this attack or any other techniques of influence and propaganda is a well-educated population with a strong sense of national identity founded on principles of tolerance, generosity, openness to diversity, and service to others; especially important are critical thinking skills, the ability to weigh evidence and reject implausible fringe theories, and a realistic respect for the value of established authorities and institutions. The only way to prevent narrative warfare from spreading out from individual victims to mass societal effect is to reduce the systemic vulnerabilities and attack surfaces, the “cracks in our society” that come from ignorance and divisive factionalism. However, in the short run, there is a pressing need for rapidly deployable tactics. We believe that here, as in other forms of narrative warfare, playing defense is a losing strategy. While prevention is the best strategy, once an attack has taken place and the WUID has become entrenched, we believe the most effective tactic is a counteroffensive. The key observation of the WUID is that, in order to be effective, it must remain coordinated. In order for the density spike created in f_B by B_{WUID} to remain high enough to exceed the threshold d_{thr} , the individuals must be clustered close together in the preference space P^k ; if they drift apart, then they are no longer an effective weapon. This exposes a weakness in the attacker’s weapon that could be exploited by instigating counteroffensive, targeted narratives designed to disrupt the unity of B_{WUID} as well as disrupt the narratives the attackers use to direct the WUID. In future work, we will consider possible counteroffensive techniques in greater detail.

VII. CONCLUDING REMARKS

Narrative warfare and propaganda are as old as warfare itself. Self-propagating units of information are a newer concept in the digital age, and are core to the established field of cybersecurity. Viruses, worms, Trojan horses, etc., are well understood and thoroughly monitored. However, the recent surge of innovation in sociotechnical systems, including social networks, weaponized memes, and big data, has opened up a new era of conflict in which an adversary can, for example, rapidly manipulate popular sentiment to swing an election with only days or hours of lead time, faster than any currently possible response. This paper does not attempt to solve these problems immediately. Rather, our goal has been to describe a framework and a way of thinking about ACSC and CyIO that experienced actors already use to analyze populations and plan operations. By making this introduction widely available to friendly actors, we hope to support defensive innovation and lead to improvements in the current situation, in which the United States and its allies have been severely outpaced in this domain.

Information warfare takes place on a high-dimensional, abstract battlefield, which makes monitoring and planning extremely difficult. One promise of this framework is the development of technology for situational awareness and battlefield visualization in near-real-time.

There is already a rich field of research and practical application of tools for meaning extraction and visualization of high-dimensional data sets, which could be adapted to create tactical battlefield displays for real-time awareness, planning, and defense against CyIO operations as they unfold.

With moves such as the U.S. Department of Homeland Security designation of elections systems as critical infrastructure, the world is acknowledging the need for physical security and cybersecurity in election systems. However, the emergence of powerful CyIO capabilities is a qualitatively new development, and it is currently debatable whether it even falls in the wheelhouse of cybersecurity and cyberwarfare. The concept of national security risk from population-level influence weaknesses, attack surfaces, and vulnerabilities not in voting machines³ but elections and social choice systems themselves may not be on anyone's radar screen at all. And yet, these weaknesses may have already been recognized, analyzed, and exploited by hostile actors against democratic systems around the world. We believe that this type of approach can help analyze the vulnerabilities of our own voting systems and recommend improvements. We also hope that a more formal framework for such analysis could generate greater clarity and objectivity about risks and recommendations, and that this objectivity in turn might politicize election security.

As an example of applying this framework to understand and describe narrative warfare attacks, we explored the idea of a WUID. This technique creates a population-level "hammer" that could be wielded with great effectiveness in CyIO operations. The framework of ACSC allows the comparison and evaluation of different operations and the tentative measurement of their potential cost and effectiveness. We suspect that WUID-type operations may have already been used with great economy and effectiveness against the United States, and so it is particularly important to study this type of tactic and develop defenses and counter-offenses.

It is important to note that the assumptions of a linear space and Euclidean metrics are limiting; a more realistic model would be a manifold embedding as commonly found in large machine-learning applications. In fact there is empirical evidence of subpopulations moving through "wormholes" in the sense that they abruptly shift from one region of P^k to another without seeming to traverse the intervening territory, as well as other strange effects. We introduce only the most elementary election theory here and acknowledge that there is huge body of literature that we are glossing over. In particular, recent work critiquing Median Voter Theory is relevant and will certainly inform refinement and changes in this type of framework in the future [34], [35]. Furthermore, the relegation of all of psychology into the weight matrices W and V is a radical oversimplification, to say the least. Nevertheless, the first step toward making any new field theoretically and computationally tractable is to define a mathematical framework that fits reasonably well. This allows the deviations from simple behavior to be quantified, which in turn allows the model to be expanded with appropriately defined weight matrices, locally Euclidean manifold techniques, etc. Our goal with this paper is to help open the doors to this field; we have no illusions of completeness of anything presented here.

³ We, of course, are strong supporters of conventional cybersecurity and especially election security and protection from hacking of databases and voter disenfranchisement. However, these are not covered in this paper.

Much has been written elsewhere about the greater susceptibility of Western liberal democracies to propaganda and narrative warfare. Many nations censor or block the flow of information, even entertainment, from the United States and other Western nations, while we freely allow input from anywhere and enjoy media from every nation of the world. Dictatorships are insulated from the vulnerability of democracy to population-level influence because dictatorial power does not even nominally derive from the will of the people. However, the United States and its allies do enjoy significant advantages that can potentially be applied in CyIO, especially for defense and/or counter-offense. For one, the people living under dictatorships fueled by lies and corruption usually become inured to any messaging at all from their own government, which may be leveraged by careful introduction of narratives from outside sources. The United States leveraged cultural “soft power” in the defeat of the Soviet Union in the previous Cold War^[36] and, at our best, our values of openness, service, and tolerance can inspire and bring much of the world over to our side— a capability which we must regain. At a more technical level, the battlefield of CyIO itself is literally owned by U.S.-based corporations^[8] who may, in some cases, be willing to assist in defending against these attacks. In the current climate, increased regulation of social media companies is already all but inevitable; without a solid foundation in the principles of CyIO, these regulations are likely to have no beneficial effects on these risks, or may even make them worse. We hope an improved understanding of the role of social media in CyIO will guide regulatory efforts to be useful and effective towards global peace and security.♥

APPENDIX

Here we provide derivations of the various facets of the “Curse of Dimensionality” listed earlier.

1) *Individuals disagree with each other more:* The expected distance between any two randomly chosen points increases.

The formula for the expected distance between two points chosen on independent and identically distributed (IID) uniform distributions in a k -dimensional cube is extensively explored in^[37]. The distance has a lower bound of $\frac{1}{3}\sqrt{k}$.

2) *The population overall becomes more dissatisfied with any platform that specifies a complete set of policies (e.g., the actual government policy at any moment):* The average distance between all of the points in B_N^k and the single point G^k that represents the platform increases.

For simplicity we consider IID normal distributions and place G^k at O . Clearly, the average distance we refer to here is simply $E(B^k)$, which approaches $\sigma\sqrt{k}$ for large k ^[38].

3) *The potential for insurmountable disagreements increases:* The maximum possible distance between any two points increases even faster than the expected distance.

This is simply the distance between opposite corners of a k -dimensional hypercube, which can easily be seen to be $2\sqrt{k}$.

4) A smaller proportion of the ideological space is taken up by moderates and a greater proportion by the fringe, even for a highly inclusive definition of moderate: We can choose a reasonable definition of “moderate” in the policy space as being “near the center” and represent this with $S_{r_{mod}}^k$, the centered k -ball of radius r_{mod} (i.e., the region of policy space that is within the distance r_{mod} of the center). Then the ratio of the volume of $S_{r_{mod}}^k$ to the volume of the policy space P_k goes to zero as k increases, even if we choose $r_{mod} = 1$.

The volume of the unit k -ball goes to zero rapidly for large k , as shown in^[39], and so necessarily the volume ratio to the unit cube goes to zero even more rapidly. Scaling to radius $r_{mod} < 1$ simply adds a factor of $r_{mod}^k < 1$, which makes the convergence even more rapid.

5) Extending that, depending on the population’s distribution of beliefs, an increasingly greater share of the population will find that its values and beliefs fall outside of any of the available political parties: If we represent p political parties as p nonoverlapping k -balls of radius r_i , $S_{r_i}^k$, the ratio of the total volume of all the parties’ territories $\sum_{i=1}^p (S_{r_i}^k)$ to the total volume of P_k , also goes to zero.

This can be seen to follow obviously from the previous item: the sum of several volumes, all of which converge to zero, also converges to zero.

6) The two previous points, taken together, imply that if a political party uses this kind of data analysis in its electoral or marketing strategic planning, it will be motivated to expand its ideological-space territory to include more and more of the fringe in an attempt to capture more of the electorate.

If the effective value of k is large, then expansions of a party’s “ball of appeal” by a radius ratio $1 + \epsilon$ increase the volume by $(1 + \epsilon)^k$. More realistically, consider the expansion of the “ball of appeal” along only one dimension. This expansion is equivalent to adding a k -dimensional, cylindrical chunk of volume to the original ball. For simplicity’s sake, consider a party that expands its appeal in two specific dimensions by a distance equal to the diameter of the ball; this is qualitatively what you might consider becoming “twice as fringe” in two topics alone while leaving all others the same. We choose the diameter rather than the radius because this adds a cylinder that perfectly circumscribes the original ball, and we choose two dimensions rather than one because the derivation of the formula for the volume of a k -ball^[40] has a simple mathematical ratio:

$$V_k^B = \frac{2\pi}{k} V_{k-2}^B \tag{18}$$

The new volume added into the territory is equal to the volume of a cylinder that circumscribes the original sphere, where the cylinder has 2 “straight” dimensions and k 2-dimensional balls as the “ends.” The volume ratio of the k 2-dimensional ball to the k -dimensional cylinder generated by translating the ball twice is, by simple geometry, $V_k^C = 4V_{k-2}^B$.

Finally we can see that the ratio of the volume of the original k -ball to the volume of the added cylinder is:

$$\frac{V_k^B}{V_k^C} = \frac{\pi}{2k} \quad (19)$$

If $k = 10$, for example, the ratio is about 0.16. If we proceed boldly with the formalism, this implies that if a party expands its reach by just 16 percent on only 2 out of 10 salient issues, it can double the volume of issue space that its territory encompasses; or, conversely, likewise with $k=10$, if a party doubles its “fringeyness” on only 2 of the 10 issues, it now encompasses more than 7 *times* as much volume in its territory!⁴ It is somewhat frightening to consider the possible superadditive effects of these incentives for fringe expansion, together with the power of fringe manipulation available through the WUID paradigm.

ACKNOWLEDGMENTS

I would like to thank MAJ Mike Klipstein, Ph.D., of the Army Cyber Institute; Lt Col Jennifer J. Snow, USAF, USSOCOM Donovan Group and SOFWERX; and CDR Pablo C. Breuer, USN, USSOCOM Donovan Group and SOFWERX. All of them provided invaluable insights, perspective, and feedback, as well as much-needed support and encouragement. I would also like to thank C.P. Frost, Ph.D., for his expertise and feedback on voting and policy. A few individuals who wish to remain anonymous provided critical insights into the principles and mathematical terrain of online profiling, targeting, and population analysis; these made this work possible. Finally, I would like to thank my father, Dr. Michael D. Perlman, Ph.D., for teaching me math and statistics as a child so that I was inclined to view and question the world through that lens.

⁴ This is a rule of thumb of the curse of dimensionality: spheres are weak; almost anything other than inflating a sphere will do a better job of capturing more territory.

NOTES

1. F. Brandt, V. Conitzer, U. Endriss, J. Lang, and A. D. Procaccia, Eds., *Handbook of Computational Social Choice*. New York: Cambridge University Press, 2016.
2. M. Kosinski, D. Stillwell, and T. Graepel, "Private traits and attributes are predictable from digital records of human behavior," *PNAS*, pp. 5802–5805, Apr. 2013.
3. Z. Tufekci. (2014, Jul.) Engineering the public: Big data, surveillance and computational politics. [Online]. Available: <http://firstmonday.org/ojs/index.php/fm/article/view/4901/4097>.
4. D. G. Wilson, "The ethics of automated behavioral microtargeting," *AI Matters*, vol. 3, no. 3, pp. 56–64, Oct. 2017.
5. H. Lin and J. Kerr, "On Cyber-Enabled Information/Influence Warfare and Manipulation," *SSRN*, Aug. 2017.
6. C. Watts, "Cyber-Enabled Information Operations," Apr. 2017.
7. R. Waltzman, "The weaponization of information: the need for cognitive security," Apr. 2017.
8. J. C. Inglis, "Statement of Chris Inglis before the Senate Armed Services Committee," Apr. 2017.
9. D. Reynolds. (2018) Cyber-enabled information operations: The battlefield threat without a face. *Jane's Defence Weekly*. [Online]. Available: <https://www.janes.com/images/assets/438/77438/Cyber-enabled-information-operations-The-battlefield-threat-without-a-face.pdf>.
10. H. Lin. Developing Responses to Cyber-Enabled Information Warfare and Influence Operations. *Lawfare Blog*. [Online]. Available: <https://www.lawfareblog.com/developing-responses-cyberenabled-information-warfare-and-influence-operations>.
11. A. K. Maan and P. L. Cobaugh, *Introduction to Narrative Warfare: A Primer and Study Guide*. Narrative Strategies, LLC, Jun. 2018.
12. M. B. Prosser, "Memetics-a growth industry in US military operations," Master's thesis, United States Marine Corps School of Advanced Warfighting, 2006.
13. A. Downs, *An economic theory of democracy*, New York, 1957.
14. O. A. Davis, M. J. Hinich, and P. C. Ordeshook, "An Expository Development of a Mathematical Model of the Electoral Process," *The American Political Science Review*, vol. 64, no. 2, pp. 426–448, Jun. 1970.
15. O. A. Davis and M. J. Hinich, "A mathematical model of policy formation in a democratic society," in *Mathematical Applications in Political Science*, J. L. Bernd, Ed. Dallas: Southern Methodist Press, 1967.
16. R. Axelrod, "The structure of public opinion on policy issues," *Public Opinion Quarterly*, vol. 31, pp. 363–371, Spring 1967.
17. R. D. McKelvey and R. E. Wendell, "Voting Equilibria in Multidimensional Choice Spaces," *Mathematics of Operations Research*, vol. 1, no. 2, pp. 144–158, May 1976.
18. A. D. I. Kramer, J. E. Guillory, and J. T. Hancock, "Experimental evidence of massive-scale emotional contagion through social networks," *Proceedings of the National Academy of Sciences of the United States of America*, vol. 111, no. 24, pp. 8788–8790, Jun. 2014.
19. Goel, Vinu, "Facebook Tinkers With Users' Emotions in News Feed Experiment, Stirring Outcry," *New York Times*, Jun. 2014.
20. F. Brandt, V. Conitzer, U. Endriss, J. Lang, and A. D. Procaccia, "Introduction to Computational Social Choice," in *Handbook of Computational Social Choice*, F. Brandt, V. Conitzer, U. Endriss, J. Lang, and A. D. Procaccia, Eds. New York: Cambridge University Press, 2016, pp. 1–29.
21. D. Xefteris, "Multidimensional electoral competition between differentiated candidates," *Games and Economic Behavior*, vol. 105, pp. 112–121, Sep. 2017.
22. M. Kosinski, S. C. Matz, S. D. Gosling, V. Popov, and D. Stillwell, "Facebook as a research tool for the social sciences," *American Psychologist*, vol. 70, no. 6, pp. 543–556, Sep. 2015.
23. R. J. Gonzalez, "Hacking the citizenry?" *Anthropology Today*, vol. 33, no. 3, pp. 9–12, Jun. 2017.
24. K. K. Roberts, "Privacy and Perceptions," *The Elon Journal of Undergraduate Research in Communications*, vol. 1, no. 1, pp. 24–34, Mar. 2010.
25. The Economist Data Team. Support for Britain's exit from the EU is waning. *The Economist*. [Online]. Available: <https://www.economist.com/graphic-detail/2018/06/22/supportfor-britains-exit-from-the-eu-is-waning>.
26. B. D. Bernheim and S. N. Slavov, "A Solution Concept for Majority Rule in Dynamic Settings," *The Review of Economic Studies*, pp. 33–62, Dec. 2008.
27. R. E. Bellman, *Dynamic Programming*, Princeton University Press, 1957.

NOTES

28. Dewey defeats Truman. Wikipedia. [Online]. Available: https://en.wikipedia.org/wiki/Dewey_Defeats_Truman.
29. Door-in-the-face technique. Wikipedia. [Online]. Available: https://en.wikipedia.org/wiki/Door-in-the-face_technique.
30. Russell, Nathan J. An Introduction to the Overton Window of Political Possibilities. Mackinac Center for Public Policy. [Online]. Available: <https://www.mackinac.org/7504>.
31. J. Lehman. A brief explanation of the overton window. Mackinac Center. [Online]. Available: <https://www.mackinac.org/overtonwindow>.
32. T. B. Edsall, “Nothing in Moderation,” *New York Times*, Oct. 2014. [Online]. Available: <https://www.nytimes.com/2014/10/29/opinion/nothing-in-moderation.html>.
33. E. Levitz, “Democrats Can Abandon the Center — Because the Center Doesn’t Exist,” *New York Magazine*, Jul. 2017. [Online]. Available: <http://nymag.com/daily/intelligencer/2017/07/demscan-abandon-the-center-because-the-center-doesnt-exist.html>.
34. J. S. Nye, “Squandering the U.S. ‘Soft Power’ Edge,” *International Educator*, pp. 4–6, Jan. 2007.
35. E. W. Weisstein. Hypercube Line Picking. Mathworld—A Wolfram Web Resource. [Online]. Available: <http://mathworld.wolfram.com/HypercubeLinePicking.html>.
36. V. Chandrasekaran, B. Recht, P. Parrilo, and A. S. Willsky, “The convex geometry of linear inverse problems,” in *eprint arXiv:1012.0621v3*, Apr. 2012.
37. E. W. Weisstein. Ball. Mathworld—A Wolfram Web Resource. [Online]. Available: <http://mathworld.wolfram.com/Ball.html>.
38. P. A. Mariano. The volume of the unit ball in n dimensions. University of Connecticut Department of Mathematics. [Online]. Available: <http://www2.math.uconn.edu/84mariano/research/MathClubspl4%20.pdf>.

Predicting enterprise cyber incidents using social network analysis on dark web hacker forums

Soumajyoti Sarkar

Arizona State University
Tempe, Arizona, U.S.A.

Mohammad Almukaynizi

Arizona State University
Tempe, Arizona, U.S.A.

Jana Shakarian

Cyber Reconnaissance, Inc.
Tempe, Arizona, U.S.A.

Paulo Shakarian

Arizona State University
Tempe, Arizona, U.S.A.

ABSTRACT

With the rise in security breaches over the past few years, there has been an increasing need to mine insights from social media platforms to raise alerts of possible attacks in an attempt to defend conflict during competition. We use information from dark web forums by leveraging the reply network structure of user interactions with the goal of predicting enterprise cyberattacks. We use a suite of social network features on top of supervised learning models and validate them using a binary classification problem that attempts to predict whether there would be an attack on any given day for an organization. We conclude from our experiments, which gathered information from 53 forums on the dark web over a span of 12 months and attempted to predict real-world cyberattacks across 2 security incidents, that analyzing the path structure between groups of users is better than merely studying centralities like Pagerank or relying on user-posting statistics in forums.

INTRODUCTION

With recent data breaches at organizations such as Yahoo, Uber, and Equifax¹ emphasizing the increasing financial and social impacts of cyberattacks, there has been an enormous requirement for technologies that could alert such organizations to possible data breaches. These breaches are a direct or indirect result of cyber, electronic, and information operations to infiltrate systems and infrastructure as well as gain unauthorized access to information, thus setting an example of conflict during competition. On the vulnerability

© 2019 Soumajyoti Sarkar, Mohammad Almukaynizi, Jana Shakarian, Paulo Shakarian

Some of the authors are supported through the AFOSR Young Investigator Program (YIP) grant FA9550-15-1-0159, ARO grant 11NF-15-1-0282, and the DoD Minerva program grant N00014-16-1-2015.

¹ <https://www.consumer.ftc.gov/blog/2017/09/equifax-data-breach-what-do> <https://www.consumer.ftc.gov/blog/2016/09/yahoo-breach-watch>

front, Risk Based Security's VulnDB database² published a total of 4,837 vulnerabilities in a quarter of 2017, which was around 30 percent higher than the previous year. This motivates the need for extensive systems that can utilize vulnerability-associated information from external sources to alert organizations to such cyberattacks. The dark web is one such place on the internet where users can share information on software vulnerabilities and ways to exploit them^{[1],[15]}. Surprisingly, it might be difficult to track the actual intentions of those users, thus making it necessary to use data mining and learn to identify the discussions among the noise that could potentially raise alerts on attacks on external enterprises. In this paper, we leverage the information obtained from analyzing the reply network structure of discussions in dark web forums to understand the extent to which dark web information can be useful for predicting real-world cyberattacks.

Most of the work on vulnerability discussions on trading exploits in underground forums^{[9],[13],[14]} and related social media platforms like Twitter^{[2],[8],[15]} have focused on two aspects: (1) analyzing vulnerabilities discussed or traded in the forums and the markets, thereby giving rise to the belief that the "life cycle of vulnerabilities" in these forums and marketplaces and their exploitation have a significant impact on real-world cyberattacks^{[13],[14]}; and (2) prioritizing or scoring vulnerabilities using these social media platforms or binary file appearance logs of machines to predict the risk state of machines or systems^{[7],[11]}. These two components have been used in silos; however, and in this paper, we ignore the steps between vulnerability exploit analysis and the final task of real-world cyberattack prediction by removing the preconceived notions used in earlier studies where vulnerability exploitation was considered a precursor towards attack prediction. We instead hypothesize about user interaction dynamics conceived through posts surrounding these vulnerabilities on these underground platforms to generate warnings for future attacks. We note that we *do not* consider whether vulnerabilities have been exploited in these discussions since a lot of zero-day attacks^[11] might occur before such vulnerabilities are even indexed and their gravity might lie hidden in discussions related to other associated vulnerabilities or some discussion on exploits. We based our research on the dynamics of all kinds of discussions on dark web forums; however, we attempted to filter out the noise to mine important patterns by examining whether pieces of information gained traction within important communities.

To this end, the major contributions of this research investigation are as follows:

- ◆ We create a network mining technique using the directed reply network of users who participate in dark web forums to extract a set of specialized users we term *experts* whose posts with *popular vulnerability mentions* gain attention from other users in a specific time frame.
- ◆ Following this, we generate several time series of features that capture the dynamics of interactions centered around these *experts* across individual forums as well as general feature time series based on social network and forum posting statistics.

² <https://www.riskbasedsecurity.com/2017/05/29-increase-in-vulnerabilitiesalready-disclosed-in-2017/>

- ◆ We use these time series features and train a supervised learning model based on logistic regression with attack labels for two different incidents from an organization to predict daily attacks. We obtain the best results with an F1 score of 0.53 on a feature that explores the path structure between *experts* and other users compared to the random (without prior probabilities) F1 score of 0.37. Additionally, we identify instances of superior feature performance based on discussions involving vulnerability information rather than network centralities and forum posting statistics.

The rest of the paper is organized as follows: We introduce several terms and the dataset related to the vulnerabilities and dark web in section II; the general framework for attack prediction, including feature curation and learning models, in section III; and, finally, our experimental evaluations in section IV.

II. BACKGROUND AND DATASET

In this section, we describe the dataset that we used in our research to analyze the interaction of patterns of dark web users and the real-world security incident³ data that we used as ground truth (GT) for the evaluation of our prediction models.

A. Enterprise-Relevant External Threats (Ground Truth (GT))

Our GT was based on data on cyberattacks on Armstrong Corporation systems occurring between April 2016 and September 2017; we obtained this data from the Intelligence Advanced Research Projects Activity Cyberattack Automated Unconventional Sensor Environment program⁴. Some of the relevant attributes in this data are “event type” and “event occurred date.” Event type is the type of attack and event occurred date is the date on which a particular attack occurred. The event types examined in this study are “malicious email” and “endpoint malware.” Malicious email refers to an incident in which an individual in the organization received an email that contained a malicious attachment or link. Endpoint malware refers to malware discovered on an endpoint device. This includes, but is not limited to, ransomware, spyware, and adware. As shown in figure 1, the distribution of attacks over time is different for the events. The total number of incidents reported for the events are as follows: 119 tagged as *endpoint-malware* and 135 as malicious-email events, resulting in a total of 280 incidents over a span of 17 months that were considered in our study.

B. Dark web data

The dark web forms a small part of the deep web, the part of the web not indexed by web search engines, although sometimes the term “deep web” is mistakenly used to refer to the dark web. We obtained all dark web data used in this study through an application programming interface provided by a commercial platform⁵.

³ We often use the terms “attacks,” “incidents,” and “events” interchangeably.

⁴ <https://www.iarpa.gov/index.php/research-programs/cause>

⁵ Data is provided by Cyber Reconnaissance, Inc., www.cyr3con.ai

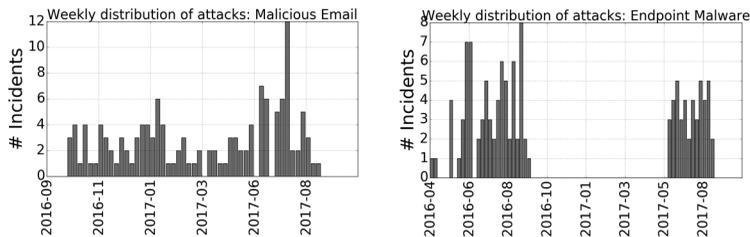


Fig. 1: Weekly occurrence of security breach incidents of different types (a) malicious email and (b) endpoint malware

The structure of a dark web forum is hierarchical: each forum consists of several independent threads, a thread caters to a particular discussion on a topic, and each thread spans several posts initiated by multiple users over time. We note that one user can appear multiple times in the sequence of posts depending on when and how many times the user posted in that thread. However, the dataset we obtained does not contain the hierarchical information of reposting - it does not provide us with which user a particular user replied to, while posting or replying in a thread. We filter out forums based on a threshold number of posts that were created in the time frame of January 2016 to September 2017. We gathered data from 179 forums in which the total number of unique posts was 557,689, irrespective of the thread to which they belonged. The number of forums with less than 100 posts was large and, therefore, we only considered forums that had greater than 5,000 posts during that time period, which gave us a total of 53 forums. We denote the set of 53 forums used in this dataset with the symbol F .

Common Vulnerabilities and Exposures (CVEs): The database of CVEs maintained on a platform operated by the MITRE Corporation⁶ provides identity mapping for publicly known information security vulnerabilities and exposures. We collect all of the information regarding the vulnerability mentions in dark web forums during the period between January 2016 and October 2017. The total number of CVEs mentioned in the posts across all forums during this period was 3,553.

CVE - Common Platform Enumeration (CPE) Mapping: A CPE is a structured naming scheme for identifying and grouping clusters of information technology systems, software, and packages maintained on the National Vulnerability Database (NVD) platform operated by the National Institute of Standards and Technology (NIST)⁷. Each CVE can be assigned to different CPE groups based on the naming system of CPE families, as described in [9]. Similarly, each CPE family can have several CVEs that conform to its vendors and products to which the CPE caters. In order to cluster the set of CVEs in our study into a set of CPE groups, we use the set of CPE tags for each CVE from the NVD database maintained by NIST. For the CPE tags, we only consider the operating system platform and the application environment tags for each unique CPE. Examples of CPEs include Microsoft Windows_95, Canonical ubuntu_linux, and, Hp elitebook_725_g3. The first component in each of these CPEs denotes the operating system platform and the second component denotes the application environment and its versions.

⁶ <http://cve.mitre.org>

⁷ <https://nvd.nist.gov/cpe.cfm>

III. FRAMEWORK FOR ATTACK PREDICTION

The mechanism for attack predictions can be described in three steps: (1) Given a time point t for which we need to predict an enterprise attack of a particular event type, (2) we use features from the dark web forums prior to t and (3) we use these features as input for a learned model that predicts attacks on t . So one of the main tasks involves learning the attack prediction model for each event type. Below we describe steps (2) and (3) - curating features and building supervised learning models.

A. Curating Features

We first describe the mechanism in which we build temporal networks, following which we describe the features used for the prediction problem. We build three groups of features across forums: (1) Expert-centric; (2) User/forum statistics; and (3) Network centralities.

Dark Web Reply Network: We assume the absence of global user identification (IDs) across forums⁸ and therefore analyze the social interactions using networks induced on specific forums instead of considering the global network across all forums. We denote the directed reply graph of a forum $f \in F$ by $G^f = (V^f, E^f)$ where V^f denotes the set of users who posted or replied in some thread in forum f at some time in our considered time frame of data; and E^f denotes the set of three-tuple (u_1, u_2, rt) directed edges where $u_1, u_2 \in V^f$ and rt denotes the time at which u_1 replied to a post of u_2 in some thread in f , with $u_1 \rightarrow u_2$ denoting the edge direction. We denote by $G_\tau^f = (V_\tau^f; E_\tau^f)$, a temporal subgraph of G^f , τ being a time window such that V_τ^f denotes the set of individuals who posted in f in that window; in addition, E_τ^f denotes the set of tuples $(v_1; v_2; rt)$ such that $rt \in \tau, v_1; v_2 \in V_\tau^f$. We use two operations to create temporal networks: "Create," which takes a set of forum posts in f within a time window τ as input and creates a temporal subgraph G_τ^f and "Merge," which takes two temporal graphs as input and merges them to form an auxiliary graph. To keep the notations simple, we drop the symbol f when we describe the operations for a specific forum in F as context but which does apply for any forum $f \in F$. We describe these two operations in brief; however, a detailed algorithm relating the network construction is given in algorithm 1 of appendix A1.⁹ We adopt an incremental analysis approach by splitting the entire set of time points in our frame of study into a sequence of time windows $\Gamma = \{\tau_1, \tau_2, \dots, \tau_Q\}$, where each subsequence $\tau_i, i \in [1, Q]$ is equal in time span and non-overlapping and the subsequences are ordered by their starting time points for their respective span.

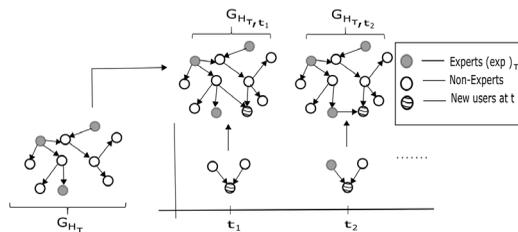


Fig. 2: An illustration to show the Merge operation: $G_{H, \tau}$ denotes the historical network which is used to compute the experts shown in gray. $\{G_{t_1}, G_{t_2}, \dots\}$ denote the networks at time $t_1, t_2, \dots \in \tau, t \in \Gamma$.

⁸ Note that even in the presence of global user IDs across forums, a lot of anonymous or malicious users would create multiple profiles across forums and create multiple posts with profiles; identifying and merging such profiles is an active area of research.

⁹ Online appendix: <http://www.public.asu.edu/~ssarka18/appendix.pdf>

CREATE: *Creating the reply graph* - Let h be a particular thread or topic within a forum f containing posts by users $V_h^f = \{u_1, \dots, u_k\}$ posted at corresponding times $T_h^f = \{t_1, \dots, t_k\}$, where k denotes the number of posts in that thread and $t_i < t_j$ for any $i > j$; that is, the posts are chronologically ordered. To create the set of edges E_f , we connect two users $(u_i, u_j) \in V_h^f$ such that $i > j$; that is, user u_i has potentially replied to u_j , and is subject to a set of spatial and temporal constraints (appendix A1). These constraints make up for the absence of exact information about the reply as to whom u replied to in a particular post in h .

MERGE: *Merging network* - In order to create a time series feature $\mathcal{T}_{x,f}$ for feature x from threads in forum f that maps each time point $t \in \tau$, $\tau \in \Gamma$ to a real number, we use two networks: (1) the historical network G_{H_τ} , which spans over time H_τ such that $\forall t \in H_\tau$, and $t \in \tau$, so that we have $t' < t$; and (2) the network G_t^f induced by user interactions between users in E_t , which varies temporally for each $t \in \tau$. We note that the historical network G_{H_τ} is different for each subsequence and same for all $t \in \tau$, so that as the subsequences $\tau \in \Gamma$ progress with time, the historical network G_{H_τ} also changes; in addition, we discuss the choice of spans $\tau \in \Gamma$ and H_τ in section IV. Finally, for computing feature values for each time point $t \in \tau$, we merge the two networks G_{H_τ} and G_t^f to form the auxiliary network $G_{H_\tau,t} = (V_{H_\tau,t}, E_{H_\tau,t})$, where $V_{H_\tau,t} = V_{H_\tau} \cup V_t$ and $E_{H_\tau,t} = E_{H_\tau} \cup E_\tau$. A visual illustration of this method is shown in figure 2. Now we describe the several features we used that would be fed to a learning model for attack prediction. We compute time series of several features x , $T_{x,f}[t]$ for every time point t in our frame of study and for every forum f separately.

1. Expert-Centric Features

We extract a set of users we term *experts* who have a history of CVE mentions in their posts and whose posts have gained attention in terms of replies. Following that, we mine several features that explain how attention is broadcast by these experts to other posts. All of these features are computed using the auxiliary networks $G_{H_\tau,t}$ for each time t . Our hypothesis is based on the premise that any unusual activity must spur attention from users who have knowledge about vulnerabilities.

We focus on users whose posts in a forum contain the most-discussed CVEs belonging to important CPEs during the time frame of analysis, where the importance will shortly be formalized. For each forum f , we use the historical network $G_{H_\tau}^f$ to extract the set of *experts* relevant to time frame τ ; that is, $exp_t^f \in V_{H_\tau}^f$. First, we extract the top CPE groups CPTop in the time frame H based on the number of historical mentions of CVEs. We sort the CPE groups based on the sum of the CVE mentions that belong to the respective CPE groups and take the top five CPE groups by the sum in each H_τ . Using these notations, the experts exp_t^f from history H_τ considered for time span τ are defined as users in f with the following three constraints:

- (1) Users who have mentioned a CVE in their posts in H_τ . This ensures that the user engages in the forums with content that is relevant to vulnerabilities.

- (2) Let $\theta(u)$ denote the set of CPE tags of the CVEs mentioned by user u in his/her posts in H_τ , such that it follows either $\theta(u) \in CP_\tau^{top}$, where the user’s CVEs are grouped in less than five CPEs, or, $CP_\tau^{top} \in \theta(u)$ in cases where a user has posts with CVEs in the span H_τ grouped into more than five CPEs. This constraint filters out users who discuss vulnerabilities that are not among the top CPE groups in H_τ .
- (3) The in-degree of the user u in GH_τ should cross a threshold. This constraint ensures that there are a significant number of users who potentially responded to this user, thus establishing u ’s central position in the reply network. Essentially, the set of experts exp from H_τ would be used for all of the time points in τ .

We curate path- and community-based features based on the experts listed in table I. These expert-centric features try to quantify the distance between an expert and a daily user (non-expert) in terms of how fast a post from that user receives attention from the expert. In that sense, the community features also measure the like-mindedness of non-experts and experts.

Group	Features	Description
Expert centric	Graph Conductance	$\tau_x[t] = \frac{\sum_{e \in experts} \sum_{u \in V_t \setminus experts} \pi(expert) P_{xy}}{\pi(expert)}$ where $\pi(\cdot)$ is the stationary distribution of the network $G_{H_\tau, t}$. P_{xy} denotes the probability of random walk from vertices x to y . The conductance represents the probability of taking a random walk from any of the <i>experts</i> to one of the users in $V_t \setminus experts$, normalized by the probability weight of being on an expert.
	Shortest Path	$\tau_x[t] = \frac{1}{ experts } \sum_{e \in experts} \min_{u \in V_t \setminus experts} s_{e,u}$ where $s_{e,u}$ denotes the shortest path from an expert e to user u following the direction of edges.
	Expert replies	$\tau_x[t] = \frac{1}{ experts } \sum_{e \in experts} OutNeighbors(e) $ where $OutNeighbors(\cdot)$ denotes the out neighbors of user in the network $G_{H_\tau, t}$.
	Common Communities	$\tau_x[t] = \mathcal{N}(c(u)) \cap c_{experts} \cap u \in V_t \setminus experts $ where $c(u)$ denotes the community index of user u , $c_{experts}$ that of the experts and $\mathcal{N}(\cdot)$ denotes a counting function. It counts the number of users who share communities with experts.
Forum/User Statistics	Number of threads	$\tau_x[t] = \{h \mid \text{thread } h \text{ was posted on } t\} $
	Number of users	$\tau_x[t] = \{u \mid \text{user } u \text{ posted on } t\} $
	Number of expert threads	$\tau_x[t] = \{h \mid \text{thread } h \text{ was posted on } t \text{ by users } u \in \text{experts}\} $
	Number of CVE mentions	$\tau_x[t] = \{CVE \mid \text{CVE was mentioned in some post on } t\} $
Network Centralities	<i>Outdegree</i> _k	$\tau_x[t] = \text{Average value of top } k \text{ users, by outdegree on } t$
	<i>Outdegree</i> _k CVE	$\tau_x[t] = \text{Average value of top } k \text{ users with more than } 1 \text{ CVE mention in their posts, by outdegree on } t$
	<i>Pagerank</i> _k	$\tau_x[t] = \text{Average value of top } k \text{ users, by Pagerank on } t$
	<i>Pagerank</i> _k CVE	$\tau_x[t] = \text{Average value of top } k \text{ users with more than } 1 \text{ CVE mention in their posts, by pagerank on } t$
	<i>Betweenness</i> _k	$\tau_x[t] = \text{Average value of top } k \text{ users, by Betweenness on } t$
<i>Betweenness</i> _k CVE	$\tau_x[t] = \text{Average value of top } k \text{ users with more than } 1 \text{ CVE mention in their posts, by betweenness on } t$	

Fig. 1: Weekly occurrence of security breach incidents of different types (a) malicious email and (b) endpoint malware

Why focus on experts? To show the significance of these properties in comparison those of other users, we examine the time periods of 3 widely known security events: the Wannacry ransomware attack that happened on May 12, 2017, and the vulnerability MS-17-010; the Petya cyberattack on June 27, 2017, with the associated vulnerabilities CVE-2017-0144, CVE-2017-0145, and MS-17-010; and the Equifax breach attack that occurred primarily on March 9, 2017, and vulnerability CVE-2017-5638. We consider two sets of users across all forums - exp_τ , where G_{H_τ} denotes the corresponding historical network prior to τ in which these three events occurred and the second set of users being all U_{alt} who are not experts and who fail either one of two constraints: they have mentioned CVEs in their posts which do not belong to CP^{top} or their in-degree in G_{H_τ} lies below the threshold. We consider G_{H_τ} being induced by users in the last 3 weeks prior to the occurrence week of each event for both cases and we consider the

total number of interactions, ignoring the direction of reply of these users with other users. Let deg_{exp} denote the vector of counts of interactions in which the experts were involved and deg_{alt} denote the vector of counts of interactions in which the users in U_{alt} were involved. We randomly pick a number of users from U_{alt} equal to the number of experts and sort the vectors by count. We conduct a two-sample t -test on the vectors deg_{exp} and deg_{alt} . The null hypothesis H_0 and the alternate hypothesis H_1 are defined as follows:

- ◆ $H_0 : \text{deg}_{\text{exp}} \leq \text{deg}_{\text{alt}}$
- ◆ $H_1 : \text{deg}_{\text{exp}} > \text{deg}_{\text{alt}}$

The null hypothesis is rejected at significance level $\alpha = 0.01$ with a p-value of 0.0007. This suggests that with high probability, experts tend to interact more prior to important, real-world cybersecurity breaches than other users who randomly post CVEs.

Now, we conduct a second t -test where we randomly pick 4 weeks not in the weeks considered for the data breaches to pick users U_{alt} with the same constraints. We use the same hypotheses as above and, when we perform statistical tests for significance, we find that the null hypothesis is not rejected at $\alpha=0.01$ with a p -value close to 0.05. This empirical evidence from the t -test also suggests that the interactions with exp are more correlated with an important cybersecurity incident than those of users who post CVEs not in top CPE groups and that, therefore, it is better to focus on users exhibiting our desired properties as experts for cyberattack prediction. Note that the t -test evidence also incorporates a special temporal association since we collected events from three interleaved time frames corresponding to the event dates.

2. User/Forum Statistics Features

We try to see whether the forum or user-posting statistics are themselves indicators of future cyberattacks; for this, we compute Forum/User Statistics, as described in table I.

3. Network Centrality Features

In addition, we also tested several network centrality features mentioned in table I. The purpose is to check whether the emergence of central users in the reply network $G_t, t \in \tau$, is a good predictor of a cyberattack. We note that, in this case, we only use the daily reply networks to compute the features, unlike in the case of the expert-centric network features, where we use $G_{H\tau, t}$.

B. Building Supervised Learning Models

In this section we explain how we use the time series data $T_{x,f}$ to predict an attack at any given time point t . We consider a supervised learning model in which the time series T_x is formed by averaging $T_{x,f}$ across all forums in $f \in F$ at each time point t and then used for the prediction task. We treat the attack prediction problem in this paper as a binary classification problem in which the objective is to predict whether there would be an attack at a given time point t . Since the incident data in this paper contains the number of incidents that

occurred at time point t , we assign a label of 1 for t if there was at least one attack at t and 0 otherwise.

In [4], the authors studied the effect of longitudinal sparsity in high-dimensional time series data. They propose assigning weights to the same features at different time spans to capture the temporal redundancy. We use two parameters: β , which denotes the start time prior to t from where we consider the features for prediction, and n , the time span for the features to be considered. In figure 3, to predict an attack occurrence at time t , we use the features for each time $t_h \in [t_{-n-\beta}, t_{-\beta}]$. Here we use logistic regression with longitudinal ridge sparsity that models the probability of an attack as follows with X being the set of features and β being the vector of coefficients:

$$P(attack(t) = 1 | \mathbf{X}) = \frac{1}{1 + e^{-(\beta_0 + \sum_{k=\eta+\delta}^{\delta} \beta_k x_{t-k})}} \tag{1}$$

The final objective function to minimize over N instances where N is the number of time points spanning the attack time frame: $l(\beta) = -\sum_{i=1}^N (y_i(\beta_0 + \mathbf{x}_i^T \beta) - \log(1 + \exp^{\beta_0 + \mathbf{x}_i^T \beta}) + \lambda \beta^T \beta$, with y being the instance label.

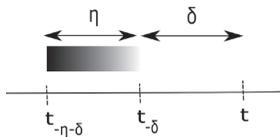


Fig. 3: Temporal feature selection window for predicting an attack at time t

One of the major problems of the dataset is the imbalance in the training and test dataset, as will be described in section IV; thus, in order to use all features in each group together for prediction, we use three additional regularization terms: the L1 penalty, the L2 penalty, and the group lasso regularization [5]. The final objective function can be written as:

$$l(\beta) = -\sum_{i=1}^N \log(1 + e^{-y_i(\beta^T \mathbf{x}_i)}) + \frac{m}{2} \|\beta\|_2^2 + l \|\beta\|_1 + g \cdot GL(\beta) \tag{2}$$

where m , l , and g are the hyper-parameters for the regularization terms and the $GL(\beta)$ term is $\sum_{g=1}^G \|\beta_{I_g}\|_2$, where I_g is the index set belonging to the g^{th} group of variables, $g = 1 \dots G$. Here each g is the time index $t_h \in [t_{-n-\beta}, t_{-\beta}]$, so this group variable selection selects all features of one time in history while reducing some other time points to 0. It has the attractive property that it performs variable selection at the temporal group level and is invariant under (group-wise) orthogonal transformations, like ridge regression. We note that while there are several other models that could be used for prediction that incorporate the temporal and sequential nature of the data, like hidden markov models and recurrent neural networks, the logit model allows us to transparently adjust to the sparsity of data, especially in the absence of a large dataset.

IV. EXPERIMENTAL EVALUATIONS

In our work, the granularity for each time index in the T function is 1 day; that is, we compute feature values over all days in the time frame of our study. For incrementally computing the values of the time series, we consider the time span of each subsequence $\tau \in T$ as 1 month, and for each τ , we consider $H_\tau = 3$ months immediately preceding τ . That is, for every additional month of training or test data that is provided to the model, we use the preceding 3 months to create the historical network and compute the corresponding features on all days in τ . For choosing the experts with an in-degree threshold, we select a threshold of 10 to filter out users having an in-degree of less than 10 in G_{H_τ} from exp_τ . For the centrality features, we set k to be 50; that is, we choose the top 50 users sorted by the corresponding metric in table I. We build different learning models using the GT available from separate *event – types*.

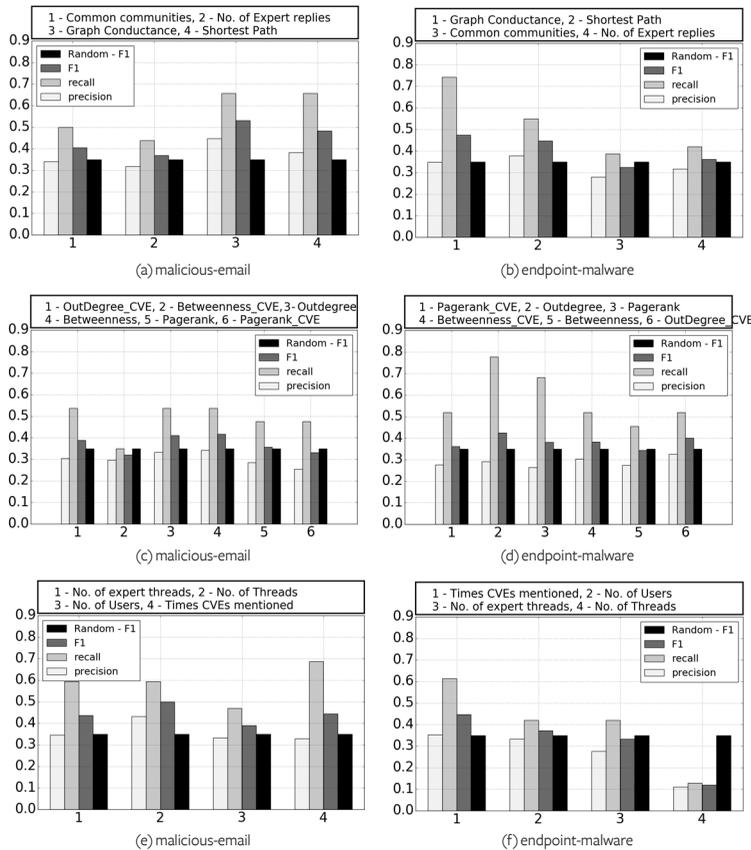


Fig. 4: Classification results for the features considering the logistic regression model: = 7 days, = 8 days.

As mentioned in section III-B, we consider a binary prediction problem in this paper. We assign an attack flag of 1 for at least 1 attack on each day and 0 otherwise. For malicious email, out of the 335 days considered in the dataset, there were reported attacks on 97 days;

this constitutes a positive class ratio of around 29 percent. For endpoint malware, the total number of attack days was 31 out of the 306 days considered in the dataset, which constitutes a positive class ratio of around 26 percent. For evaluating the performance of the models on the dataset, we split the time frame of each event into 70 percent - 30 percent, averaged to the nearest month separately for each event type; that is, we take the first 80 percent of time (in months) as the training dataset and the remaining 20 percent in sequence for the test dataset. We avoid shuffle split as generally being done in cross-validation techniques in order to consider consistency in using sequential information when computing the features. As shown in figure 1, since the period of the attack information provided varies in time for each of the events, we use different time frames for the training model and the test sets. For the event type malicious email, which remains our primary test bed evaluation event, we consider the time period from October 2016 to June 2017 (9 months) in dark web forums for our training data and the period from July 2017 to August 2017 (3 months) as our test dataset. For endpoint malware, we use the time period from April 2016 to September 2016 (6 months) as our training time period and June 2017 to August 2017 (3 months) as our test data for evaluation.

We consider a 1-week time window while keeping $w = 8$ days. From among the set of statistics features that were used for predicting malicious email attacks shown in figure 4(e), we observe the best results using the number of threads as the signal, for which we observe a precision of 0.43, recall of 0.59, and an F1 score of 0.5 against the random F1 of 0.44 for this type of attack. From among the set of expert-centric features in figure 4(a), we obtain the best results from graph conductance with a precision of 0.44, recall of 0.65, and an F1 score of 0.53, which shows an increase in recall over the number of threads measure. Additionally, we observe that the best features in terms of F1 score are graph conductance and shortest paths, whereas the number of threads and vulnerability mentions turns out to be the best among the statistics. For the attacks belonging to the type endpoint malware, we observe similar characteristics for the expert-centric features in figure 4(b), where we obtain a best precision of 0.34, recall of 0.74, and an F1 score of 0.47 against a random F1 of 0.35, followed by the shortest paths measure. However, for the statistical measures, we obtain a precision of 0.35, recall of 0.61, and an F1 score of 0.45 for the vulnerability mentions, followed by the number of threads, which gives us an F1 score of 0.43. Although the common community features do not help much in the overall prediction results, in the following section, we describe a special case that demonstrates the predictive power of the community structure in networks. On the other hand, when we investigate the centrality features with respect to the prediction performance in figure 4(c), we find that just looking at network centralities does not help. The best values we obtain for malicious-email event predictions are from the out-degree and betweenness metrics, both of which give us an F1 score of 0.41. Surprisingly, we find that when the metrics are used for only the users with CVE mentions, the results worse, with the best F1 score for out-degree CVE having an F1 score of 0.38. This calls for a more

complex understanding of path structures between users, rather than just focusing on user significance solely. The challenging nature of the supervised prediction problem is not just due to the issue of class imbalance, but also to the lack of large samples in the dataset which, if they had been present, could have been used for sampling purposes. As an experiment, we also used random forests as the classification model, but we did not observe any significant improvements in the results over the random case.

For the model with the group lasso regularization in equation 2, we set the parameters m , l , and g , and 0.3, 0.3, and 0.1, respectively. We obtained better results for each group of features together for the malicious email event type, with an F1 score of 0.55 for expert-centric, 0.51 for forum/user statistics, and 0.49 for network centrality-based features..

Prediction in High-Activity Weeks

One of the main challenges in predicting external threats without any method for correlating them with external data sources like the dark web or any other database is that it is difficult to validate which kinds of attacks are most correlated with these data sources. To this end, we examine a controlled experiment setup for the malicious email attacks in which we only consider the weeks which exhibited a high frequency of attacks compared to the overall time frame. In our case, we consider weeks that had more than five attacks in the test time frame. These high numbers may be due to multiple attacks in 1 or a few specific days or a few attacks on all days. We run the same supervised prediction method but evaluate them only on these specific weeks.

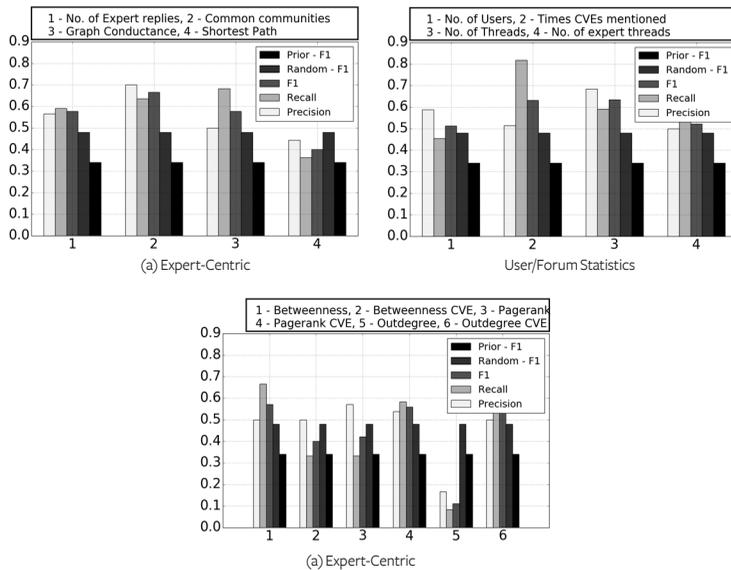


Fig. 5: Classification results for *malicious-email* attacks in high-frequency weeks, $\partial = 7$ days and $n = 8$ days.

From the results shown in figure 5, we find that the best results were shown by the common communities feature, which had a precision of 0.7, a recall of 0.63, and an F1 score of 0.67, compared to the random (no priors) F1 score of 0.48 and a random (with priors) F1 score of 0.34 for the same time parameters. Among the statistics measures, the highest F1 score was 0.63 for the vulnerability mentions feature. From among the set of centrality features, we find that the betweenness measure had the best F1 score (0.58), with a precision of 0.5 and a recall of 0.78. This also suggests the fact that analyzing the path structure between nodes is useful since betweenness relies on the paths passing through a node. Additionally we find that, unlike the results over all of the days, for these specific weeks, the model achieves high precision while maintaining comparable recall, emphasizing the fact that the number of false positive was also reduced during these periods. This correlation between the weeks that exhibit huge attacks and the prediction results imply that network structure analytics can definitely help generate alerts for cyberattacks.

V. RELATED WORK AND CONCLUSION

Using network analysis to understand the topology of dark web forums was studied at breadth in ^[6], where the authors used social network analysis techniques on the reply networks of forums. There have been several attempts to use external social media data sources to predict real-world cyberattacks^{[2],[7],[8]}. The use of machine learning models to predict security threats ^[2] presents many research opportunities, including predicting whether a vulnerability would be exploited based on dark web sources^{[3],[9]}. The availability of large external data sources makes the use of machine learning methods to predict cyberattacks more promising. Previous studies also included the use of time series models to forecast the number of cyber incidents^[16], which increases the demand for the use of such models in cyberattack prediction. The authors in^[17] look at text-mining techniques to understand the content of the posts which provide threat intelligence on various social media platforms. In this study, we argue that the dark web can be a reliable source of information for predicting external enterprise threats. We leverage the network and interaction patterns in the forums to understand the extent to which they can be used as useful indicators. Our study also opens further research possibilities surrounding sentiment analysis on these discussions, which could help track malicious discussions and hence defend against cyber conflict during competition.🛡️

NOTES

1. Samtani, Sagar, Ryan Chinn, and Hsinchun Chen. "Exploring hacker assets in underground forums." IEEE (ISI), 2015.
2. Liu, Yang, et al. "Cloudy with a Chance of Breach: Forecasting Cyber Security Incidents." USENIX Security Symposium. 2015.
3. Nunes, Eric, et al. "Darknet and deepnet mining for proactive cybersecurity threat intelligence." IEEE ISI (2016).
4. Xu, Tingyang, Jiangwen Sun, and Jinbo Bi. "Longitudinal lasso: Jointly learning features and temporal contingency for outcome prediction." ACM, KDD 2015.
5. Meier, Lukas, Sara Van De Geer, and Peter Bühlmann. "The group lasso for logistic regression." Journal of the Royal Statistical Society: Series B (Statistical Methodology) 70.1 (2008): 53-71.
6. Almukaynizi, Mohammed, et al. "Predicting cyber threats through the dynamics of user connectivity in darkweb and deepweb forums." ACM Computational Social Science. (2017).
7. Liu, Yang, et al. "Predicting cyber security incidents using featurebased characterization of network-level malicious activities." 2015 ACM International Workshop Security and Privacy Analytics.
8. Khandpur, Rupinder Paul, et al. "Crowdsourcing cybersecurity: Cyber attack detection using social media." ACM CIKM 2017.
9. Almukaynizi, Mohammed, et al. "Proactive identification of exploits in the wild through vulnerability mentions online." IEEE CyCON, 2017.
10. Thonnard, Olivier, et al. "Are you at risk? Profiling organizations and individuals subject to targeted attacks." International Conference on Financial Cryptography and Data Security. Springer 2015.
11. Bilge, Leyla, and Tudor Dumitras. "Before we knew it: an empirical study of zero-day attacks in the real world." Proceedings of the 2012 ACM conference on Computer and communications security.
12. Sabottke, Carl, Octavian Suci, and Tudor Dumitras. "Vulnerability Disclosure in the Age of Social Media: Exploiting Twitter for Predicting Real-World Exploits." USENIX Security Symposium. 2015.
13. Herley, Cormac, and Dinei Florêncio. "Nobody sells gold for the price of silver: Dishonesty, uncertainty and the underground economy." Economics of information security and privacy. Springer, Boston, MA, 2010. 33-53.
14. Allodi, Luca, Marco Corradin, and Fabio Massacci. "Then and now: On the maturity of the cybercrime markets the lesson that blackhat marketeers learned." IEEE Transactions on Emerging Topics in Computing 4.1 (2016): 35-46.
15. Chen, Hsinchun. "Sentiment and affect analysis of dark web forums: Measuring radicalization on the internet." IEEE ISI 2008.
16. Okutan, Ahmet, et al. "POSTER: Cyber Attack Prediction of Threats from Unconventional Resources (CAPTURE)." Proceedings of the 2017 ACM SIGSAC.
17. Sapienza, Anna, et al. "Early warnings of cyber threats in online discussions." Data Mining Workshops (ICDMW), 2017.

Thomas Klemas
U.S. Air Force

Rebecca K. Lively
U.S. Air Force

Nazli Choucri
*Professor of Political Science
Massachusetts Institute of Technology (MIT)
Cambridge, Massachusetts, USA*

ABSTRACT

The United States of America faces great risk in the cyber domain because our adversaries are growing bolder, increasing in number, improving their capabilities, and doing so rapidly. Meanwhile, the associated technologies are evolving so quickly that progress toward hardening and securing this domain is ephemeral, as systems reach obsolescence in just a few years and revolutionary paradigm shifts, such as cloud computing and ubiquitous mobile devices, can pull the rug out from the best-laid defensive planning by introducing entirely new regimes of operations. Contemplating these facts in the context of Department of Defense (DoD) acquisitions is particularly sobering because many cyber capabilities bought within the traditional acquisition framework may be of limited usefulness by the time that they are delivered to the warfighter. Thus, it is a strategic imperative to improve DoD acquisitions pertaining to cyber capabilities. This paper proposes novel ideas and a framework for addressing these challenges.

Keywords—DDoS, RQA, Adaptive Clustering, A-Kmeans.

I. INTRODUCTION

Almost everyone agrees that growing threats to cybersecurity are undermining the Nation's safety. Not a day goes by without reports on new breaches and exploitations. Indeed, an entire industry has developed around evaluating the impacts of cybersecurity incidents, reporting on trends, and assessing impacts. Far more compelling is the evidence that the United States is facing escalating cyber hostilities with increasing frequency from a growing number of diverse adversaries^{[1],[2],[3],[4],[5],[6]}. The challenges posed by the near-instantaneity of cyber action have no precedent. Given the fluidity, complexity, and ambiguity of the cyber domain, framing an adaptive, dynamic, and reliable policy response amounts to a critical imperative. It is a necessity, not a choice.

NOTE: The views expressed in this paper are the authors' and do not necessarily represent the views of the U.S. Air Force (USAF), the DoD or the United States

The contributions of Thomas Klemas and Rebecca K. Lively are the work of the U.S. Government and is not subject to copyright protection in the United States. Foreign copyrights may apply.

© 2019 Nazli Choucri

Shaping and retaining advantage in the cyber domain requires a comprehensive approach that leverages all aspects of national power, including diplomatic, economic, informational, technological, and military elements. This paper focuses on the military dimension of national power and concentrates on one major factor—namely, equipping the force with innovative and necessary cyber tools through the acquisition process. Our purpose is to motivate cyber-specific enhancements to existing policy. More specifically, we seek to reduce, if not eliminate, powerful obstacles that prevent the rapid development and delivery of cyber capabilities that are crucial to defending U.S. systems and infrastructure.

This paper presents a logical basis for necessary changes to existing policy and empirical data which compel essential, cyber-specific changes to current acquisition processes. In addition, this paper proposes a specific approach to enhancing the process so that cyber acquisition can be responsive to the rapidly changing threat landscape. Considering the current cyber domain and the overall environment, we demonstrate that the current acquisition process is too slow to: (1) meet current and likely future cyber warfighter needs; (2) too slow to respond to cyber adversaries who are frequently moving faster than the United States; and (3) keep pace with the rapidly changing threat environment. These factors, among others, highlight the fundamental differences between cyber requirements and traditional acquisition.

We proceed as follows: Section II highlights the new strategic imperatives that create the context for both cyber and traditional acquisition and the general imperative driving the urgency of cyber acquisition reform. Section III explores the expanding roster of hostile states and criminal organizations and growing adversary progress and cyber strength, as reported in publicly available materials. Section IV describes cyberspace dynamics, including the impacts of dramatic information technology (IT) change, and then points to how these factors will continue to impact the defense posture of the United States. Finally, section V presents an acquisition policy framework which can address these compelling issues and contribute to U.S. cyber superiority.

II. NEW STRATEGIC PARAMETERS

There is a growing awareness that acquisition reform is crucial to the national defense and that traditional acquisition approaches are measured in completely different timescales than the pace required by cyber realm approaches. In fact, timelines for operational needs are quite short in the cyber domain. Some capabilities are needed within only a few weeks and are often used only one time by cyber warfighters. But traditional acquisition processes take many years, sometimes even more than a decade to complete.

Recent attempts to streamline the acquisition process^[7] targeted improvements that would result in a 5- to 7-year process. In 2016, the DoD disclosed that the estimated median duration for Major Defense Acquisition Programs was more than 6.9 years^[8]. However, Major Automated Information System life cycles had an estimated median of 3.2 years for programs after 2009^[8].

It is noteworthy that both of these figures exceed most cyber need timelines, potentially by orders of magnitude. Based on these considerations, it is evident that traditional processes, even if improved to achieve the goals in^[7], are not sufficiently rapid to keep pace with technological evolution or to acquire cutting-edge cyberspace capabilities. The mismatch of traditional acquisition timelines with cyber needs and useful lifespans virtually guarantees that the military will be equipped with aging cyber capabilities that may have limited usefulness or rapidly become obsolete^[50]. So long as acquisition does not have the mechanisms to keep pace with needs, the military will be forced to utilize increasingly inferior capabilities^[9]. All of this is embedded in the very reality of a process shaped by criteria other than time. More to the point, it sheds a dim view of a situation seen through the lens of timelines for warfighter needs.

“America’s military has no preordained right to victory on the battlefield”^[30]. This is especially true in the face of “rapid technological changes” and an environment where “inter-state strategic competition, rather than terrorism, is the primary concern in US National Security”^[30]. Thus, “[t]his is truly a period in history in which we are falling behind if they are merely holding our position in the overall movement to forge new capabilities”^[10]. However, existing acquisition processes were designed to develop warfighting systems that sometimes last for decades. They were not designed for any features of the cyber domain, nor for the extremely rapid cyber decision and action process. A number of U.S. airplanes have been operating for more than 40 years, an extreme example being the Boeing B-52, which may survive past 100 years^[11]. For the most part, cyber power rests on speed and agility, not on stability and longevity. Cyber capabilities have a lifespan of weeks; months; or, at most, a few years, often only persisting that long through frequent upgrades.

III. ACCELERATING THREATS

The current intensity of cyber incidents and sophistication of advanced cyber threats is a defining feature of the 21st century, and barriers to effective defense are high^{[1],[2],[3],[4],[5]}. As a direct result, demands are mounting on U.S. cyber forces. Additionally, new malicious activities cause features of the cyber domain to change and sometimes create a need for new tools, new skills, and new training. In this section, we will substantiate that the cyber adversaries challenging the United States today are well resourced, increasing in number, constantly striving to improve and diversify their capabilities, growing bolder, displaying a high degree of freedom of action, and perhaps outpacing the United States in some regards.

A brief overview of cyber threat history, including recent malicious activities, intrusions, and responses, is necessary to provide context, justify the principal motivational elements, and distill key insights that will guide discussion and substantiate the proposed approach. Especially relevant is the fact that many of our adversaries are not hampered by an acquisition process anchored in institutional and historical experience and resistant to rapid adaptation to changing circumstances. Two of the countries that represent the greatest overall threat to U.S. interests – Russia and China – seem to display a remarkable level of hostile cyber intent.

The progression of Microsoft Cloud Azure Service reports^{[1],[2],[4]} from 2016 to 2018 suggests a notable escalation in malicious activities on Microsoft Cloud virtual machines that seem to originate from Russian internet protocol (IP) addresses. The 2018 data reported an almost 16-percent rate of total incoming attacks which seem to originate from Russia, up from previous levels below 10 percent.

We have learned the surprising extent of Russian moves to interfere with U.S. elections, signaling an elevated degree of the Russian intelligence intent to penetrate and influence civil society. The Office of the Director of National Intelligence released^[6], which describes some of the national intelligence analytical assessments regarding Russian interference in the 2016 elections. The analysis indicates that the campaign was well coordinated and financed, consisting of operations organized by the General Staff Main Intelligence Directorate which included exfiltration of a significant quantity of data from the U.S. Democratic National Committee and the leveraging of internet trolls from the Saint Petersburg-based Internet Research Agency, a close Putin ally with ties to Russian intelligence. These activities highlight the growing “grey zone” behaviors of state actors who take actions below the international law threshold which would permit a kinetic military response^[12]. All was done without the use of one single bullet or the loss of one single life. An adversary has unilaterally changed the “rules of the game” and made civil society its operational target.

Beyond election interference, an alarming set of other significant cyber activities have occurred during the past several years that appear to have originated from the Russian Federation. Here we summarize just a few of the more prominent incidents, referenced from the Center for Foreign Relations data set^[3]. In March 2015, Ukrainian officials were targeted by cyber espionage attempts. In September 2016, the World Anti-Doping Agency (WADA) computer systems were compromised and data was leaked regarding athletes in the 2016 Rio de Janeiro Olympics, presumably in response to the previous WADA report that outlined systematic Russian use of performance-enhancing substances during the 2014 Sochi Olympic Games. Shortly thereafter, several U.S. think tanks that focused on international relations and national security were targeted by compromise attempts. In July 2017, the NotPetya malware encrypted data in numerous European, Australian, and U.S. organizations, to disrupt financial operations (tax filings). During early 2018, numerous actions targeted Winter Olympics sports entities following the ban on Russian Winter Olympic athletes. Also during this period, several spear phishing attempts appeared to target a European defense agency and several foreign ministries.

Despite the prominence and targeting of malicious Russian activities, China's actions have also been prolific during the past several years. The same Microsoft Cloud Azure Service reports^{[1],[2],[4]} referenced above found that almost 33 percent of all malicious activities on its virtual machines came from IP addresses in China in 2018, a dramatic upswing in activity from 2016 and 2017 and an indication of targeted aggression.

Considering only virtual machines that were penetrated, 54 percent communicated with IP addresses in China. While IP address attribution is not definitive, these statistics do suggest actors in Russia and China are principal cyber adversaries. China's state exploits have concentrated on business and industry and gained considerable notoriety. China has been rapidly growing its cyber operational capabilities. Especially important is the rapid rate of cyber skill development in a government-controlled labor force. A new social credit system introduced in China—whereby citizens are observed and rewarded for good behavior—all but assures China's almost total knowledge of and potential control over its citizens and facilitates the possibility of government-controlled, crowd-sourced activities^{[13],[14]}.

The Council on Foreign Relations incident data set^[3] contains at least 85 major cyber incidents attributed to China since 2006. The incidents described in this section are just a few of the more recent activities linked to China and the Chinese Government. In April 2017, an operation called “Cloud Hopper” tried to penetrate internet service providers to access customer data in 15 countries, including the United States^[16]. The global scope of this activity suggests the deployment of a significant level of resources. Notable for the use of multiple types of malware, including Remote Access Trojans and Microsoft file signatures, this campaign employed targeted phishing utilizing Microsoft Office documents that contained modifications to exploit system vulnerabilities and leveraged hundreds of variations of malware and customized, open-source tools to exfiltrate data, even compressing and encrypting the data to avoid detection.

The variety, customization, and diversity of techniques employed by China establish it as a very advanced threat actor. In October 2017, another group referred to as “Bronze Butler” staged numerous hacks targeting industry, manufacturing, and infrastructure in Japan, South Korea, Russia, and even entities within China, apparently for espionage purposes^[17]. This group demonstrated advanced techniques, including the development of custom malware, elimination of traces of infiltration, and encryption of command and control communications. In June 2016, government systems and critical infrastructure were targeted within Myanmar, the United States, Canada, South Korea, Singapore, Germany, and India^[18]. After that, in October 2017, entities associated with the maritime industry were targeted within Asia, the United States, the Philippines, and Hong Kong. Then, in November 2017, hackers from a Chinese internet security company attempted to steal trade secrets from Trimble, Siemens, and Moody's Analytics^[19]. The internet security company associated with the hacking has been linked closely to the Chinese People's Liberation Army and is believed to receive state sponsorship for its activities. The intent in all but one of these cases appeared to be espionage and theft of intellectual property, signaling key differences between the Russian and Chinese actions during this period.

The news has been so saturated with discussion of Russian election interference and Chinese cyber technology espionage activities that it is easy to overlook other incidents. However, recent history is replete with mounting reports of North Korean and Iranian intrusions, as well

as those of other nation-states. The Council on Foreign Relations incident data set^[3] listed more than 20 incidents that gained news attention that was attributed to Iran between 2010 and 2018, 7 of which were between 2017 and 2018 alone. Additionally, about 20 incidents were attributed to North Korea between 2009 and 2018.

Perhaps slightly below the radar, Iran has been quite active. In March 2018, it was discovered that almost 150 U.S. universities, and a similar number in over 20 other countries, had been compromised as part of malicious activity by the Mabna Institute, an entity believed to have ties to the Iranian National Guard^[20]. In June 2017, Iran-linked hackers attempted to infiltrate and compromise email accounts of British Parliament members^[21]. Investigations revealed that hackers gained access to 30 accounts out of the more than 9,000 targeted. This event was noteworthy more for its boldness than its sophistication. In July 2017, Iran targeted universities; the defense industry; and IT companies in Germany, Saudi Arabia, Israel, Jordan, and the United States^[22]. This intrusion was notable for the diversity of techniques employed to achieve its objectives and the introduction of custom tools, although the hackers were noisier than normal for advanced threat actors, which accelerated detection and response.

A few months later, in November 2017, another event, labeled “Muddy Water”^[23], promulgated by a group known as “Unit 42,” targeted numerous Middle Eastern nations with the apparent goal of espionage. The techniques employed, which did not seem to display tremendous diversity, leveraged open-source tools but evolved over time. However, these intrusions featured documents that were delivered to the targets and designed to entice users with customizations related to their geographic region or relevant organizations. Even more nefarious, in many cases, actual documents were stolen from compromised accounts, modified to introduce malware, and sent onwards to additional targets that were already expecting the original documents.

Significant activity during the past few years also appears to have originated from North Korea. The Center for Foreign Studies data set cites several such actors as having perpetrated cybercrimes in February 2018. One actor, known as “Group 123,” targeted South Korea^[24]. This actor initiated numerous campaigns that received publicity: “Golden Time,” “Evil New Year,” “Are you Happy?,” “Free Milk,” “North Korean Human Rights,” and “Evil New Year 2018.” Prominently featured in this campaign was spear phishing with maliciously modified documents. Another well-known example, “WannaCry,” was ransomware that struck hundreds of companies around the world in May 2017, causing about \$4 billion in losses^{[25],[26]}. This activity exploited a known and patched vulnerability for Windows, but over 200,000 unpatched systems were still affected. Additionally, in September 2017, hackers targeted U.S. electrical companies with an apparent objective of early-stage surveillance^[27]. Many of the actions attributed to North Korea seem designed for disruption (warning) or to show national determination, build wealth by theft or fraud, or conduct espionage. Clearly, the activities demonstrate a boldness that usually accompanies impunity.

Overall, the cyber aggression attributed to Russia, China, Iran, and North Korea exhibits a

pronounced freedom of action buttressed by advancing capabilities, enabling the increasingly complex scenarios demonstrated by these countries. On balance, the cyber domain appears to be a great leveler, emboldening states^{[1],[2],[3],[4],[5]} and freeing them from limitations in kinetic capability. To all of this we must add the rapid growth of cybercrime and potential asymmetries inherent to cyber that suggest how many non-state actors can pose significant threats to national security. In these situations, the clear advantage of the aggressor and the significant stresses placed on the defense cannot be denied.

The record of threat actors and cyber intrusions constitutes powerful evidence of growing cyber needs that reinforce the disparity between such cyber needs and the timeliness of the acquisition process. This disparity amounts to a massive opportunity cost in the form of an institutional handicap imposed on warfighters and corroborates the notion that the current acquisitions process is not providing U.S. cyber warriors the resources they need to maintain superiority over adversaries. More to the point, this disparity is creating powerful constraints, potentially crippling the effectiveness of the cyber force. But there are added factors that reinforce this corroboration. .

IV. UNRELENTING CYBER TRANSFORMATION

In cyberspace, as in most competitive spaces, having a faster pace of advancement is an advantage. But in the cyber domain, the speed of innovation coupled with rapid procurement is far more than an advantage—it is a matter of basic survival. The United States has long been a leader in advanced technology. If other countries develop new, advanced capabilities more quickly or implement them more efficiently, we will find ourselves in dire circumstances. It goes without saying: in order to succeed in a sword fight, when your opponent strikes a blow, you must be at least fast enough to dodge or parry the blow in *real* time and have the requisite speed to respond or counterattack. At a minimum, you should not be equipped with a heavy, cumbersome, and blunt sword, or no sword at all.

To serve as a suitable analog for the cyber battlespace, the sword fight example must be extended so that both the swords and the fight environment are also continually changing to account for the constant and rapid evolution of cyber tools, networks, and computer technologies. Risks are amplified dramatically by the speed at which the cyber environment evolves, the frequency of security vulnerabilities, and the degree of asymmetry that is possible in this realm. In fulfilling its cyber missions, the DoD must not only protect against malicious activity, but also account for the rapid technological changes and equip cyber warriors with powerful capabilities that will provide critical leverage in battle.

Numerous technology-based technology shifts are occurring at this time. Cloud computing serves as an example of the speed at which the cyber environment is changing; it represents a dramatic paradigm shift with impacts on cybersecurity. Prior to the 2000s, the term “cloud computing” was not even used, but more than \$33 billion was spent on cloud services in

the year 2015, making it the most expensive category in IT infrastructure^[28]. Mobile device computing has also exploded^[29]. Almost 95 percent of Americans own a cell phone, and smart-phone ownership increased from 35 percent in 2011 to 77 percent in 2018, according to a Pew Research Center study. Correspondingly, mobile device vulnerabilities have also risen as malicious actors attempt to exploit the mobile devices, connections to the internet, connections to peripherals, and organizational infrastructure.

Clearly, many, if not most, of the activities noted in section III and the technological transformations described early in section IV bear directly on national security. And more change is on the horizon with advances in artificial intelligence and quantum computing. Thus, it is incumbent on the DoD to remain at the edge, if not transcend, the current frontier of cyber capabilities to defend against and even respond to cyber-enabled aggression. To address the cyber domain, section V will explore alternative acquisition constructs that have demonstrated success and other approaches. .

V. ENHANCING CYBER ACQUISITION

This paper demonstrates that many factors, including warfighter needs, adversary progress, and rapid environmental change, demand a faster cyber acquisition process. General George S. Patton is often quoted as saying, “A good plan violently executed now is better than a perfect plan executed next week.” General Patton’s demand for strong and immediate progress is particularly apropos for cybersecurity. For the United States to simply keep up with cyber change is insufficient. We must lead, developing cutting-edge technology and approaches, despite the breakneck speed of cyber environmental dynamics, because this is the only way to ensure that the United States maintains superiority over our adversaries. The only way to achieve the required advances is to address the acquisition shortcomings. Thus, it is imperative that the United States adopt an approach suitable for rapid cyber acquisition that addresses operational needs.

The previous sections substantiate that cyber needs, posed by the existing environment and threats, mandate a much shorter life cycle than other capabilities. This section will present the recommended policy changes intended to enable cyber acquisition to meet cyber warrior needs. While cyber is not the only acquisition category in which the warfighter needs to outpace the existing acquisition constructs, cyber is at the shortest extreme of the acquisition needs timescale. Accordingly, cyber acquisition is a useful case study for acquisition approaches designed to meet cyber needs.

There is no dispute that the current federal acquisition system is too slow, especially for cyberspace capabilities. DoD leadership has mandated change, Congress wants to see change, and it seems that the DoD is taking steps to enact change. Reference^[30] makes this imperative clear—we must “[d]eliver performance at the speed of relevance.” However, despite the clear impetus for change, it is difficult to determine how best to change. With a system as complex

as the federal acquisition system, it is challenging to identify the root cause (or root causes) of the problems. Indeed, over 300 studies have been completed in the last 3 decades^[9], resulting in hundreds of findings of inefficiency and recommendations for reform.

This section first discusses some of the recognized problems with the current acquisition system, especially with regard to cyberspace; next, discusses some of the promising DoD acquisition pilot programs for delivering innovation more quickly; and, ultimately, makes three broad recommendations for reforming policy to better meet the DoD objective of delivering performance at the speed of relevance, especially in cyberspace. The three recommendations are as follows: (1) Manage rather than avoid risk—especially time-based risks; (2) Delegate authority to the lowest reasonable level; and (3) Treat different problems differently.

A. The Existing System is Flawed

“Current [DoD] processes are not responsive to need; the Department is over-optimized for exceptional performance at the expense of providing timely decisions, policies, and capabilities to the warfighter^[30].”

As the above quote demonstrates, DoD leadership has identified a link between acquisition reform and national security—recognizing that our current processes put the warfighter at risk. However, while the DoD clearly recognizes that there is a problem, determining the necessary reforms to solve the problem is not as straightforward. That’s not to say that the DoD and Congress are not trying to identify the problem and implement fixes. Since 1986, over 300 formal studies into the DoD acquisition system have been directed, both by the DoD and by Congress. Some of the findings of these studies are discussed below and represent some of the common complaints about what is wrong with the acquisition system.

For example, in^[31], Congress directed the DoD to establish an advisory panel composed of recognized experts in acquisition and procurement policy from the public and private sectors. The Section 809 Panel is charged with reviewing acquisition regulations applicable to the DoD “with a view toward streamlining and improving the efficiency and effectiveness of the defense acquisition process and maintaining defense technology advantage” and providing related recommendations^[31]. Thus far, the Section 809 Panel has released one interim report^[32] and two extensive volumes of findings and recommendations^{[33],[34]}. A third and final volume is scheduled for release in January 2019. Some of the Section 809 Panel findings are discussed below.

Unfortunately, most of the problems discussed below are not new. This paper cites reports going back as far as 1998, not because there is not more current literature, but because many of the points were as salient then as they are now. Several reports and studies draw similar conclusions. For example,^[9] quotes 1982 congressional testimony by Dr. Alice Rivlin (then the director of the Congressional Budget Office) and concludes that “[s]he could give that same testimony today, not change a single word, and still be accurate”^[9].

The current system emphasizes rigid adherence to written process and systems over measurable outcomes and speed. This is not surprising where the volume of regulations, restrictions, and documentation is so vast and acquisition personnel are not trained to operational needs^[30] because acquisition personnel focus on their area of specialty: the complex acquisition system. This emphasis leads to undesirable outcomes. For example, the “operations community is stuck with dead-end, stove-piped systems which are support nightmares and risk critical missions because, in part, the formal requirements process demands little more than that^[35].”

The Section 809 Report makes similar findings in^[32], concluding that the acquisition system “creates obstacles to getting needed equipment and services” both by making the DoD an unattractive customer to nontraditional contractors and through “suffocating bureaucratic requirements”^[32]. As a result, the panel concluded that equipment needed today “may be either unavailable^[32] to the department or egregiously tardy, leading to genuine threats to the nation’s security”^[32].

Additionally, the complexity of the system is increasing, cost is increasing, and outcomes are declining.^[32] cites the 1986 Packard Report finding which essentially provided that excellence cannot be achieved with so many layers of bureaucracy. In response, the Section 809 Panel concluded that, “compared to 1986, there are far more layers at DoD, to include even larger staffs, and too many regulations to count”^[32]. The panel found that the “inescapable conclusion when viewing DoD acquisition as a whole . . . is that process wins out over results” and that “too frequently ancillary public policy objectives, often driven by statutes or executive orders, receive equal or greater priority than mission^[32].”

Reference^[9] reached a similar conclusion, finding that the “DoD’s acquisition system continues to take longer, cost more, and deliver fewer quantities and capabilities than originally planned”^[9]. Neither the Section 809 Panel nor the Defense Business Board (DBB) found fault in acquisition personnel themselves. Instead, the conclusion reached by both emphasized the unintentional nature of the bureaucratic creep swallowing efficiency and innovation within the DoD^{[32],[35]}. As stated by the DBB, the DoD acquisition system has “unintentionally evolved [to be extremely complex] over many years of well-intended policy and legislative changes”^[9].

And, while the concept of bureaucratic delay and complexity impeding acquisitions is not new, the results are magnified when applied to the cyber acquisition landscape, where accelerated technology change highlights DoD inefficiencies. Even in 1998, the DoD recognized the need for improving the speed of technology acquisitions, finding that “[t]oday, to be static is to become obsolete and at risk. Yet DoD management and oversight processes massively impede the dynamism DoD so desperately needs”^[35]. This limitation has not changed, as noted in^[9], which finds that “[c]yber and IT modernization cannot succeed under the current system due to the accelerated advances of technology and rapidly changing threats to those technologies. Cyber and IT modernization cannot succeed because the cycle times or ‘spins’ within Cyber

and IT are far shorter than the time scale used by defense acquisition processes”^{19]}.

Unfortunately, knowing that there is a problem and certain underlying causes for the problem is not always sufficient to bring about solution implementation. And, in an acquisition system that is already riddled with regulations, suggesting more regulatory change to address the problem has a high likelihood of unintended consequences. Indeed, if finding a solution was as easy as identifying the problem and a few of the underlying causes, there would not be reports dating back to 1986 describing many of the same issues the DoD acquisition system still faces today. However, as the next section discusses, the DoD is making inroads on pilot programs investigating potential solutions. Indeed, useful ideas gleaned from these efforts inform the policy recommendations discussed at the end of this paper.

B. DoD and Congress Want to Fix the System

In recent years, the DoD and Congress seem to be trying a new and innovative approach to solving the acquisition problem. Rather than just commissioning studies or rewriting regulations, the government has been implementing many different pilot programs for specific types of acquisitions. Essentially, the government is embracing innovation in the very policies that it is using to promote innovation—by trying many different things that might fail at little cost, but that will produce great benefits if they succeed. What’s more, it appears that senior leadership is encouraging maximum use of these programs. For example,^[36] states, “Our new authorities provide so many tools to be creative; using them should routinely be our default ‘fast path.’” One of these expanded authorities, Other Transaction Authorities (OTA), is discussed in more detail below.

OTAs are basically an exception to the entire acquisition system. Whenever something goes wrong, it seems that the government adds more oversight and regulations to ensure that the same thing never happens again. In turn, this additional regulation and oversight slows down everything else in the acquisition system. For this reason, it seems that some of the best solutions are the ones that simply ignore the existing system altogether.¹ OTA is one such authority. While OTAs have been around since 1994^[37], Congress increased their availability for use by expanding their applicability in 2015^[38] and authorizing simplified follow-on contracts for successful prototypes in 2016^[39]. As a result, OTAs have become a new go-to tool in the DoD and have led to rapid acquisitions of needed capability. For example, the USAF used OTA to move certain planning operations from a whiteboard to a software-based solution, saving over \$500,000 per day with only a \$2.2 million investment^[40].

While increased use of OTA seems to be one of the most hope-inspiring changes to government acquisitions in some time, recent events demonstrate that even this innovation authority is still subject to some of the same onerous oversight as more traditional methods. For example, a recent OTA award by the Department of Defense Innovation Unit Experimental (DIUx) for cloud migration services was protested before the Government Accountability Office (GAO)^[41]. Generally, GAO does not review OTA agreements. However, in this case, GAO

¹ Interestingly,^[9] suggests just that – zero-basing the entire system. As nice as it sounds to scrap all existing regulations and oversight and start over from scratch for all acquisition programs, there is a high likelihood of unintended consequences and confusion. Additionally, Congress is unlikely to endorse a solution that substantially limits congressional oversight.

expanded its jurisdiction to include review of whether an agency’s use of OTA is appropriate. This decision sets a precedent that OTA agreement awards can be reviewed by the GAO.

Moreover, this GAO decision essentially opens up all OTA awards to bid protests, even by those who were not original bidders on the OTA. And, even when GAO bid protests do not have merit, they generally delay contract award and performance by at least 100 days. Moreover, responding to a GAO bid protest is extremely time-consuming and is likely to set back all other efforts by the government organization that is responding to the protest. In his analysis of the GAO decision, military acquisition policy expert Bill Greenwalt urged the DoD to fight the decision, stating that if the decision is allowed to stand, it will “ensure that China will dominate the future military application of quantum computing, artificial intelligence and machine learning, data analytics, biotechnology, robotics and autonomous operations”^[42]. Greenwalt’s analysis is based on the willingness of innovative, nontraditional contractors to do business with the DoD if doing so means litigating “one’s way through a legal morass and hir[ing] an army of Washington consultants and lawyers to navigate through a constantly changing compliance process”^[42].²

C. Policy Considerations for Improving Cyber Acquisitions

As the above section demonstrates, the DoD has had some success in streamlining and improving acquisitions. However, there is more work to be done, and the competing priorities of efficiency and oversight will continue to make progress challenging. This section discusses three ideas that can speed up acquisitions today and that can be used to analyze proposals for changes to policy and law to determine whether they are likely to help or hinder innovation and speed up cyberspace acquisitions.

1) Manage Rather than Avoid Risk—Especially Time-Based Risks

a) What’s the idea?

Consider time up-front as a real risk (balanced with other risks the acquisition system already considers) and understand that it is better to fail fast and early when your strategy permits it. Risk cannot be fully avoided, so it must instead be managed. Moreover, mitigating every single risk at the expense of speed is not actually a safe option—it is just a very slow failure. This idea is central to^[30], which states, “The current bureaucratic approach, centered on exacting thoroughness and minimizing risk above all else, is proving to be increasingly unresponsive”^[30]. This idea is also identified in^[9], which finds that “[m]ultiple layers of legislation and DoD internal reforms have had the unintended consequence of orienting the process to avoiding mistakes rather than timely delivery of warfighter capabilities at a reasonable cost.”

b) What can we do today?

The good news is that there is nothing in existing regulations that explicitly requires that DoD acquisitions be slow and risk-averse. Indeed, there are high-performing organizations

² The DoD Inspector General is also investigating a different DIUx purchase in an after-the-fact audit^[43]. However, this type of audit might be preferable to increased oversight up-front as it allows DoD leadership to fairly assess acquisition risks in a way that does not slow down the acquisition efforts. Nothing that the DoD Inspector General has done here appears to have interfered with the aggressive acquisition schedule achieved by DIUx^[44].

within the DoD that move quickly within the existing regulations. One example of this is the Special Operations Forces Command (SOCOM). While the SOCOM acquisition model is widely believed to operate on different principles than the rest of the DoD, this belief is largely unfounded^[45]. Instead, SOCOM culture emphasizes speed of delivery within its acquisition process. Additionally, SOCOM “accepts more risk in program execution than is typical of the larger services”^[45]. This is at least in part due to the overall small size of most SOCOM projects. Indeed, James Geurts, former SOCOM acquisition executive, is quoted as saying, “Velocity is my combat advantage. Iteration speed is what I’m after, because if I can go five times faster than you, I can fail four times and still beat you to the target . . . That’s really what we’re going after here”^[45]. The USAF seems to be encouraging this as well. A recent memo to the acquisition workforce states, “Prototyping makes discovery your friend, allowing smart risk-taking and design exploration prior to subsequent procurement and fielding decisions. So it’s okay to fail here—fully or partially—because subsequent steps provide a safety net. As long as the risk versus reward of pursuing Y makes sense, you’re ready for the next step”^[36].

c) What should we consider in the future?

Future policy should go further to emphasize risk management rather than risk avoidance. Training and policy should emphasize the tailoring of acquisition strategies to balance risk appropriately to the overall goal and budget. Additionally, a policy should differentiate between by-law requirements and policy requirements so that waivers can be sought as quickly and efficiently as possible when a particular effort would benefit from an exception to policy. As emphasized in^[30], the DoD “is committed to changes in authorities, granting of waivers, and securing external support for streamlining processes and organizations” and policy should be written to encourage such requests^[30].

2) Delegate Authority to the Lowest Reasonable Level

a) What’s the idea?

Aggressively delegate authority to the lowest reasonable levels and design programs to be smaller and thus allow lower delegation. Decision-makers who are closest to the requirements are likely to be in the best position to evaluate available options and strategies and manage overall risk. Additionally, decision-makers at lower levels are more accessible if changes to the acquisition strategy are needed or if requirements change. Not delegating means that people who do not really “get” the problem are often in charge of leading the procurement. This leads to rigidity in requirements. While certain requirements might be considered “nice to have” in the field, they can be treated as deal-breakers for very senior leaders who are leading the overall acquisition.

b) What can we do today?

Senior leaders often have the discretion to delegate but choose not to do so. To enact these

changes today, senior leaders should aggressively delegate within the limits of existing policy. Decision-makers at lower levels should seek delegation from their leadership. Once again, the SOCOM acquisition culture provides a good example. In February 2018, SOCOM acquisition executive James H. Smith explained, “We’ve been fortunate to have an amazingly consistent leadership philosophy for the last 20 years: Clearly communicate our expectations for risk management and empower the team to make decisions at the appropriate level”^[45]. The rest of the DoD should follow that example.

c) What should we consider in the future?

While Congress has created many flexible authorities and funding mechanisms, they are often held only at the highest level of the Services; not delegated or available to lower-level decision-makers; and, thus, inaccessible to operational commanders. Congress could include a requirement that new authorities be delegated to lower levels. Additionally, law and policy could be crafted to carve out clear and mandatory exceptions to oversight and review requirements for certain types of small projects. The Section 809 Panel offered three suggestions for a more agile structure: 1. “[R]epeal statutorily mandated offices”; 2. “[E]liminate military service- and departmental-level oversight that is not value-added”; and 3. “[R]eorganize the acquisition enterprise from program-centric to portfolio-driven”^[34].

Finally, Congress and senior leaders are hesitant to eliminate policies that offer oversight into lower-level efforts and safeguards that lower the risk of fraud or simply bad decisions. However, Congress and policymakers should consider implementing oversight mechanisms, such as post-award audits, that do not interfere with efficiency and innovation. While these mechanisms have the disadvantage of not being able to prevent harm from specific acquisitions, they boast more accurate data rather than speculation.

3) Treat Different Problems Differently

a) What’s the idea?

While on its face this idea might sound tautological, it is not. The recognition that different requirements have different risks and need different acquisition approaches is not ingrained within the DoD. Interestingly, from 1965 through 1996, DoD IT purchases were treated differently than other requirements^[46]. However, beginning in 1996, IT acquisition policies were consolidated with non-IT policies, ironically for the purpose of streamlining the process^[46]. The end result is that the DoD purchases software in the same way that it purchases fighter jets, submarines, and janitorial services, and this process can take “7–10 years from planning to delivery”^[47]. This finding was echoed by the Section 809 Panel, which found that “[t]he acquisitions system is inflexible and takes a one-size-fits-all approach. Dissimilar products or services are acquired using the same processes”^[33]. And, even though acquisition policy is designed to be tailored, studies have shown that “there is a long-standing reluctance to deviate from standard weapon system acquisition processes, and acquisition personnel are not trained or led to differentiate the unique aspects of IT acquisition”^[46].

These distinctions go further than just IT versus traditional weapon systems. Within IT itself, there are nuanced differences—for example, the distinction between traditional IT acquisitions and support to cyber operations. As explained by the DBB, while traditional computer applications are “created to perform a function,” cyber capabilities “act on and change the functioning of software and hardware”^[9]. Accordingly, cyber capability development “is to traditional software acquisition as writing a book is to buying a book”^[9]. There are also fundamental differences between acquiring hardware and acquiring software because software generally requires frequent updates and patching, while hardware is largely static after purchase.

b) What can we do today?

Today we can take advantage of existing permissions to tailor acquisitions based on requirements, avoid treating template documents as mandatory, and ask for waivers to mandatory policies that are not value-added for the particular acquisition. For example, ^[48] makes it clear that acquisition teams should assume that strategies and procedures that are “in the best interest of the Government[,]...not addressed in the FAR, [and] not prohibited by law” or policy represent a “permissible exercise of authority.” This idea is supported by^[36], which states, “The key is common-sense tailoring to the needs of your prototype and potential subsequent procurement.”

c) What should we consider in the future?

Many of the current priorities for reform are seemingly contradictory. For example, in October 2017, Secretary of Defense Jim Mattis sent guidance to all DoD personnel highlighting three lines of effort to enable the DoD to “remain the world’s preeminent fighting force”^[49]. The final line of effort, which was directed at DoD business reforms, included several efforts, such as developing a “culture of rapid and meaningful innovation” and protecting infrastructure^[49]. While on its face, these requirements may seem contradictory (How can you move fast if you need to ensure every minor acquisition won’t damage infrastructure?), if you apply the above principle of treating different requirements differently, they do not have to contradict each other. The bottom line is this: We cannot fix everything in one unified system. With over 300 studies and hundreds of recommendations, we must recognize that different problems need different solutions that balance different risks. Accordingly, future reform efforts should more explicitly address differing risk profiles, and blanket prohibitions or requirements which apply to all DoD acquisitions should be avoided or eliminated whenever possible. ♥

ACKNOWLEDGMENT

The authors would like to thank Steven Anderson for his vision and leadership, which led to this collaboration out of which our efforts arose.

NOTES

1. Anthe C. et al, "Microsoft Security Intelligence Report." Vol. 21. Redmond VA: Microsoft Corporation, 2016.
2. Microsoft Security, "Microsoft Security Intelligence Report." Vol. 23. Redmond VA: Microsoft Corporation, 2018
3. Council on Foreign Relations, "Cyber Operations Tracker". <https://www.cfr.org/interactive/cyber-operations>, 2018.
4. Avena, E. et al, "Microsoft Security Intelligence Report." Vol. 22. Redmond VA: Microsoft Corporation, 2017.
5. "The VERIS Community Database." <http://veriscommunity.net/vcdb.html>.
6. Director of National Intelligence, "Assessing Russian Activities and Intentions in Recent US Elections." Intelligence Community Assessment ICA 2017-01D, 2017.
7. Judson, J., "US Army looks to cut typical acquisition timeline in half.", Defense News. <https://www.defensenews.com/land/2017/12/07/army-looks-to-cut-typical-acquisition-timeline-in-half/>, December 7, 2017.
8. Under Secretary of Defense, Acquisition, Technology, and Logistics, "Performance of the Defense Acquisition System: 2016 Annual Report", Washington, DC: Department of Defense, October 24, 2016.
9. Defense Business Board, "Linking and Streamlining the Defense Requirements, Acquisition, And Budget Processes." Report FY12-02. April 09, 2012.
10. Rogers, M., "Statement of Admiral Michael S. Rogers Commander United States Cyber Command Before the House Committee on Armed Services Subcommittee on Emerging Threats and Capabilities.", <https://docs.house.gov/meetings/AS/AS26/20150304/103093/HHRG-114-AS26-Wstate-RogersM-20150304.pdf>, March 4, 2015.
11. Adams E., "How on God's Green Earth is the B-52 Still in Service?" Wired. <https://www.wired.com/2016/04/gods-green-earth-b-52-still-service/>, April 19, 2016.
12. Sari A, "Legal Aspects of Hybrid Warfare." Lawfare, <https://www.lawfareblog.com/legal-aspects-hybrid-warfare>, . October 02, 2015.
13. Ma A., "China has Started Ranking Citizens with a Creepy 'Social Credit' System - Here's What You Can Do Wrong, and the Embarrassing, Demeaning Ways They Can Punish You.", <https://www.businessinsider.com/china-social-credit-system-punishments-and-rewards-explained-2018-4>, April 08, 2018.
14. Mistreanu, S., "Life Inside China's Social Credit Laboratory." Foreign Policy. <https://foreignpolicy.com/2018/04/03/life-inside-chinas-social-credit-laboratory/>, April 03, 2017.
15. FireEye, "Suspected Chinese Cyber Espionage Group (TEMP.Periscope) Targeting U.S. Engineering and Maritime Industries.", <https://www.fireeye.com/blog/threat-research/2018/03/suspected-chinese-espionage-group-targeting-maritime-and-engineering-industries.html>, March 18, 2018.
16. PwC, "Uncovering a new sustained global cyber espionage campaign." <https://www.pwc.co.uk/issues/cyber-security/data-privacy/insights/operation-cloud-hopper.html>, 2017.
17. Counter Threat Unit Research Team, "BRONZE BUTLER Targets Japanese Enterprises.", <https://www.secureworks.com/research/bronze-butler-targets-japanese-businesses>, October 12, 2017.
18. Zetter, K., "Revealed: Yet Another Group Hacking for China's Bottom Line." Wired, <https://www.wired.com/2016/06/revealed-yet-another-chinese-group-hacking-countrys-economic-bottom-line/>, June 14, 2016.
19. Keppler, N., Freifeld K., and Walcott J., "Siemens, Trimble, Moody's Breached by Chinese Hackers, U.S. Charges.", <https://www.reuters.com/article/us-usa-cyber-china-indictments/siemens-trimble-moodys-breached-by-chinese-hackers-u-s-charges-idUSKBN1DR26D>, November 27, 2017.
20. Graff, G. M., "DOJ Indicts 9 Iranians for Brazen Cyberattacks Against 144 US Universities.", <https://www.wired.com/story/iran-cyberattacks-us-universities-indictment/>, March 23, 2018.
21. BBC, "Iran blamed for Parliament cyber-attack.", <https://www.bbc.com/news/uk-41622903>, October 14, 2017.
22. Muncaster, P., "CopyKittens: Report Details Possible Iranian Threat Group.", <https://www.infosecurity-magazine.com/news/copykittens-a-new-report-details/>, July 25, 2017.
23. Lancaster, T., "Muddying the Water: Targeted Attacks in the Middle East.", <https://researchcenter.paloaltonetworks.com/2017/11/unit42-muddying-the-water-targeted-attacks-in-the-middle-east/>, November 14, 2017.
24. Mercer, W. and Rascagneres P., "Korea in the Crosshairs.", <https://blog.talosintelligence.com/2018/01/korea-in-crosshairs.html>, January 16, 2018.
25. Wikipedia, "WannaCry ransomware attack.", https://en.wikipedia.org/wiki/WannaCry_ransomware_attack, Accessed November 30, 2018.
26. Symantec, "Ransom.Wannacry.", <https://www.symantec.com/security-center/writeup/2017-051310-3522-99>, May 24, 2017.
27. Fire Eye, "North Korean Actors Spear Phish U.S. Electric Companies.", <https://www.fireeye.com/blog/threat-research/2017/10/north-korean-actors-spear-phish-us-electric-companies.html>, October 10, 2017.

NOTES

28. The Economist Intelligence Unit, "Ascending Cloud: The Adoption of Cloud Computing in Five Industries.", <https://perspectives.eiu.com/technology-innovation/ascending-cloud-adoption-cloud-computing-five-industries-0>, March 01, 2016.
29. Pew Research Center, "Mobile Fact Sheet.", www.pewinternet.org/fact-sheet/mobile, February 5, 2018
30. Mattis, J., Summary of the 2018 National Defense Strategy. <https://dod.defense.gov/Portals/1/Documents/pubs/2018-National-Defense-Strategy-Summary.pdf>, 2018.
31. US Congress, "Performance of incurred cost audits (§ 863) & Modifications to the advisory panel on streamlining and codifying acquisition regulations (§883)." National Defense Authorization Act for Fiscal Year 2018. <https://www.congress.gov/115/plaws/publ91/PLAW-115publ91.pdf>, 2018.
32. Lee, Hon. Deidre A. et al, Advisory Panel on Streamlining and Codifying Acquisition Regulations Interim Report. Arlington, VA: Section 809 Panel. https://section809panel.org/wp-content/uploads/2017/05/Sec809Panel_Interim-Report_May2017_FINAL-for-web.pdf, 2017.
33. Ahern, D. G. et al, Report of the Advisory Panel on Streamlining and Codifying Acquisition Regulations Vol. 1. Arlington, VA: Section 809 Panel. https://section809panel.org/wp-content/uploads/2018/01/Sec809Panel_Vol1-Report_Jan18_FINAL.pdf, 2018.
34. Ahern, D. G. et al, Report of the Advisory Panel on Streamlining and Codifying Acquisition Regulations Vol. 2. Arlington, VA: Section 809 Panel. https://section809panel.org/wp-content/uploads/2018/06/Sec809Panel_Vol2-Report_June18.pdf, 2018.
35. O'Hern Jr., W. L., Defense Science Board Task Force on Open Systems, Washington DC: Defense Science Board. <https://apps.dtic.mil/dtic/tr/fulltext/u2/a358287.pdf>, 1998.
36. Roper Jr., W.B., "Memorandum for the Acquisition Workforce, Seven Steps for Incorporating Rapid Prototyping into Acquisition." <http://acqnotes.com/wp-content/uploads/2018/05/Air-Force-Incorporating-Rapid-Prototyping-into-Acquisition.pdf>, 2018.
37. US Congress, "Authority of the Advanced Research Projects Agency to carry out certain prototype projects (§845)." National Defense Authorization Act for Fiscal Year 1994. <https://www.congress.gov/103/bills/hr2401/BILLS-103hr2401enr.pdf>, 1993.
38. US Congress, "Amendments relating to authority of the Defense Advanced Research Projects Agency to carry out certain prototype project (§812)." Carl Levin and Howard P. "Buck" Mckeen National Defense Authorization Act for Fiscal Year 2015. <https://www.congress.gov/113/plaws/publ291/PLAW-113publ291.pdf>, 2014.
39. US Congress, "Amendments to other transaction authority (§815)." National Defense Authorization Act for Fiscal Year 2016. <https://www.gpo.gov/fdsys/pkg/PLAW-114publ92/pdf/PLAW-114publ92.pdf>, 2015.
40. Wallace, M., "The U.S. Air Force learned to code-and saved the Pentagon millions." July 05. <https://www.fastcompany.com/40588729/the-air-force-learned-to-code-and-saved-the-pentagon-millions>, July 5, 2018.
41. U.S. Government Accountability Office, "Decision in matter of Oracle America Inc. (B-416061).", <https://www.gao.gov/assets/700/692327.pdf>, May 31, 2018.
42. Greenwalt, B., "GAO Decision Threatens US Military Dominance; Reject It.", <https://breakingdefense.com/2018/06/gao-decision-threatens-us-military-dominance-reject-it/>, June 27, 2018.
43. Inspector General, "Audit of Defense Hotline Allegations Concerning Tanium Software (Project No. D2018-D000CU-0124.00.", <https://media.defense.gov/2018/Jul/02/2001938140/-1/-1/D2018-D000CU-0125.000.PDF>, April 2, 2018.
44. Business Wire, "World Wide Technology Wins First-Ever DIUx Contract to Deliver Endpoint Management Services to Federal Government Agencies.", <https://www.businesswire.com/news/home/20171031005858/en/World-Wide-Technology-Wins-First-Ever-DIUx-Contract>, October 31, 2017.
45. Capobianco, J., "Strengths and Myths of What Makes Special Operations Forces Acquisition Special.", https://www.army.mil/article/205259/strengths_and_myths_of_what_makes_special_operations_forces_acquisition_special, May 14, 2018.
46. Campbell, W. H. et al., Achieving Effective Acquisition of Information Technology in the Department of Defense. <https://www.nap.edu/download/12823>, 2010.
47. Schoeni D., "Long on Rhetoric, Short on Results: Agile Methods and Cyber Acquisitions in the Department of Defense." Santa Clara High Technology Law Vol. 1 Issue 3: Article 1, <https://digitalcommons.law.scu.edu/cgi/viewcontent.cgi?article=1596&context=chtj>, January 1, 2015.
48. Federal Acquisition Regulation. n.d. "Statement of guiding principles for the Federal Acquisition System (§1.102)." https://www.acquisition.gov/far/html/Subpart%201_1.html#wp1130779.
49. Mattis, J., "Memorandum for All DoD Personnel." <https://dod.defense.gov/Portals/1/Documents/pubs/GUID-ANCE-FROM-SECRETARY-JIM-MATTIS.pdf>, 2017.
50. Yasin, R., "Military seeks faster cyber acquisition turnaround", Federal News Network, <https://federalnewsnetwork.com/cyber-exposure/2018/04/military-seeks-faster-cyber-acquisition-turnaround/>, April 23, 2018.

SESSION

◆ 3 ◆

United by Necessity: Conditions for Institutional Cooperation against Cybercrime

Jobel Kyle P. Vecino

*Charles and Louise Travers Department of Political Science
University of California, Berkeley
Berkeley, California, USA*

ABSTRACT

Cybercrime continues to grow despite ongoing remediation efforts at the state and international level. The ease of access to commit cybercriminal activity beyond one's borders makes this an international issue. Examining the cooperative schemes utilized in intergovernmental institutions such as the European Union (EU) Agency for Law Enforcement and Cooperation (Europol) illuminates possible conditions that encourage states to cooperate to fight cybercrime. Testing these conditions shows that the preexistence of an institution in a related issue area serves as the strongest driver of cooperation within an international institution against cybercrime.

Keywords— cybercrime; cybersecurity; Europol; institutions; international cooperation

I. INTRODUCTION

The problem of cybercrime continues to grow internationally; according to estimates, it will cost businesses an average of \$6 billion per annum globally through the year 2021^[1]. Some states have greater capabilities to handle cybercrime than others. In some cases, multinational corporations and academic research institutions wield stronger cybercrime mitigation capabilities than some states. The ubiquitous nature of cybercrime also makes it onerous for any one state to fight cybercriminals alone. Recently, national law enforcement agencies began to participate in newly-formed international institutions focused on cybercrime mitigation; Europol serves as one example. What qualities or conditions drive states to cooperate within these institutions to fight cybercrime? I seek to identify these qualities or conditions in order to draw policy implications that will encourage further cooperation among states in the realm of international security.

This paper analyzes three contentions. The first is that law enforcement agencies of different states are more likely to cooperate with one another if institutional avenues for cooperation already exist. This paper refers to this type of cooperation as “iterative cooperation.” Second, law enforcement agencies are more likely to cooperate within an organization to remedy a lack of, and inability to develop, domestic technical expertise in fighting cybercrime. This paper categorizes this type of cooperation as “cooperation by substitution,” in that the states utilize the institution’s capacities in lieu of their own due to an inability to develop those capacities. Third, if the majority of cooperative actions through organizations such as Europol can be characterized as capacity building, states cooperate within the institution to establish self-sufficiency in anti-cybercrime operations. This paper refers to this type of cooperation as “cooperation for self-reliance.” This paper capitalizes on the existence of Europol as a case study and data gathered from law enforcement officials and agencies throughout Europe to demonstrate that iterative cooperation through prior interactions represents the most important driver in what compels states to cooperate within an institution against cybercrime.

A. Europol and the European Cybercrime Center (EC3): An Overview

Europol is an EU agency headquartered in The Hague, Netherlands. It primarily concerns itself with assisting member states in fighting crime and terrorism by providing member state law enforcement agencies with a mechanism for the facilitation of secure intelligence exchange, primarily concerning internal security matters^[2]. Europol also coordinates cross-border anti-crime and anti-terrorist operations with member states’ law enforcement agencies and interfaces with outside partners, collects open-source intelligence and intelligence procured from publicly-available sources, and creates analyses from both intelligence provided by member states and intelligence collected by the agency^[3]. All participating states are members of the EU. Non-EU-member state partnerships are either considered “operational” or “strategic.” Operational partnerships allow for information exchange between partners and Europol, including the exchange of personal data. Operational partners include Australia, the United States, and the International Criminal Police Organization (Interpol)^[4]. EU partners can access more of Europol’s services than non-EU partners can, with participating EU member states having the most access.

Member status in Europol is dependent upon state ratification of EU regulations relating to home and justice matters^[5]. However, participation in the organization is noncompulsory for EU member states. Europol does not have the power to mandate participation; if one state decides to share its intelligence on cybercrime, it does not have the political authority to force all other member states to also share their intelligence. Therefore, many of the actions undertaken by Europol member states within the context of the organization are entirely voluntary. Policy plans known as European multidisciplinary platforms against criminal threats dictate Europol’s policy objectives and help determine which targets the organization pursues and the kinds of operations it chooses to undertake^[6]. Utilizing Europol as a platform for cooperation

does involve adopting predefined policy procedures and objectives that may not line up with a member state's chosen policy objectives. However, states have the ability to influence these policy objectives if they choose to provide input into their formation and adoption^[7]. This makes Europol a useful case study for analyzing conditions that lead to anti-cybercrime cooperation without some form of hierarchical enforcement.

This paper in particular focuses on participation within Europol's EC3, which provides many of Europol's base intelligence sharing and analysis functions, specifically for the purpose of fighting cybercrime^[8]. With regard to technical provisions, the institution provides tools and technical analysis to aid in investigations against cybercriminal activity, such as malware analysis, technical capability development, and the ability to decipher passwords with some success^[9]. EC3 may also provide member states and member state police agencies with funding as well as educational support in the form of training and seminars^[10]. Finally, EC3 (through Europol) also holds relationships with private firms^[11]. This paper refers to Europol and EC3 as the same entity (Europol) as Europol houses EC3, membership does not vary between the two, and Europol member states and EC3 staff have access to other Europol functions and vice-versa.

II. CONTEMPORARY WORK ON CYBERCRIME COOPERATION

A. In Search of a Definition

Before examining cooperation against cybercrime, the term "cybercrime" must first be defined to shed light on the nature of the problem. Elaine Fahey writes that a "comprehensive definition of 'cybercrime' for EU law has not been found in secondary law"^[12]. She goes on to utilize law professor Jonathan Clough's definition of cybercrime: "offences against computer data and systems but also more broadly, to include offences committed with the help of computer data and systems"^[13]. Fahey establishes cybercrime as a subset of cybersecurity, alongside cyberterrorism, cyberespionage, and cyberwar. Because tools utilized for cybercriminal activity are so widespread, states are constantly challenged to mitigate cybercrime on a massive scale. Annegret Bendiek and Andrew L. Porter present a competing definition. They define cybercrime as crime in cyberspace, including "theft of intellectual property, the extortion based on the threat of DDoS [Distributed Denial-of-Service] attacks, fraud based on identity theft, and so on"^[14]. However, they complicate this definition by including a "cyber-vandalism" category separate from cybercrime, which includes hackers defacing websites on the internet. Under Fahey's definition, the latter falls under the umbrella of cybercrime. For the purposes of this paper, Fahey's definition is the most appropriate, as it is all-encompassing, and Europol characterizes cybercrime similarly in its threat assessments^[15].

B. Cooperative Schemes and Institutional Choice

Because Europol consists of many member states but holds no authority over those states, classifying Europol as an intergovernmental organization (IGO) is appropriate; however, dis-

cerning the type of IGO provides greater insights into how states are compelled to cooperate within its auspices. Using Felicity Vabulas' and Duncan Snidal's classifications, Europol could be described as a formal IGO (FIGO), an organization established by a formal treaty (as a provision of the Treaty of Lisbon) which consists of three or more members and contains a formal secretariat to handle administrative duties^[16]. Thus, cooperation that focuses on Europol is subject to the same conditions that drive states to cooperate within FIGOs generally.

Kenneth Abbott and Snidal cite that two features of FIGOs make them attractive to states: centralization and independence^[17]. Centralization refers to the idea that institutional tasks are handled by a singular focal entity^[18]. In the case of Europol, these tasks include technical analysis, intelligence dissemination, public-private partnership facilitation, and operation coordination. Centralization facilitates the pooling of these activities as transaction costs and logistical overhead can be reduced through the use of the organization's staff, allowing all member states to share some of the burden and reap the reward of Europol's technical expertise or intelligence reports^[19]. Abbott and Snidal also suggest centralization allows for easier management of joint production activities, which in this context could constitute anything from the production of common anti-cybercrime policy to the coordination of joint anti-cybercrime operations^[20]. The independence of Europol also allows for the neutral distribution of funds and dissemination of intelligence through the organization. Both centralization and independence enable organizations to handle a large volume of work and manage complex operations, the benefit of which, given the scope and intricacy of cybercrime, cannot be understated.

But why choose to augment an existing formal institution instead of creating an institution *de novo*? Vinod Aggarwal provides a framework^[21], later co-opted by Jupille and Snidal, that prompts states to choose an existing institution to be the primary forum for cooperation to meet some goal, unless no existing institution fits the issue area that cooperation is meant to address^[22]. States can either utilize these institutions as-is or modify them in such a way that they meet the criteria necessary to address the new problem^[23]. When EC3 was first established within Europol, the specialization of Europol's functions to deal specifically with cybercrime could be seen as an example of institutional change – a pan-European institution that focused on cybercrime analysis and mitigation explicitly did not exist, but a pan-European institution that focused on crime in general did exist. Therefore, when the time came to establish an institution through which anti-cybercrime cooperation could be focused, it made sense to give an organization focused on cooperation against crime the responsibility to also facilitate cooperation against cybercrime. This is an example of nested substantive issue linkage, as cybercrime and crime at-large clearly display intellectual coherence. As an EU agency, states can see that greater cooperation against cybercrime within Europol's context also works toward the larger goal of stability within the EU^[24]. Since substantive issue linkage also leads to the creation of a stable issue area and generally stable institutional arrangements^[25], it is no surprise then that a formal institution was expanded as formal institutions are, by virtue of the overhead

required for their establishment, very stable relative to other arrangements. Such an increase in responsibilities also befits the rational institutional design conjecture that as the severity of the collective action problem increases, the issue scope of the organization increases^[26]; given that cybercrime continues to grow in size and severity and every state remains susceptible to it, any organization assigned to support the mitigation of crime in general must increase its scope to include and specifically focus on cybercrime.

The aforementioned framework also suggests that Europol's use by states is dependent on whether it holds the status of a focal institution, an institution which is "widely accepted as a 'natural' forum for dealing with a particular cooperation problem"^[27]. Decision costs and uncertainty about the world drive states to choose to utilize an existing institution and its current functions. As a state considers choosing from a group of institutions, augmenting a new institution, or creating a new institution, uncertainty increases with each of these choices, respectively. Therefore, the "use of a focal institution is usually the least costly resolution" and, as long as "actors are risk averse," they "promote safer strategies of use and selection"^[28]. The importance of being recognized as a focal institution is echoed by Benoît Dupont, who finds in his network analysis on international cybercrime cooperation that some organizations attempt to outmuscle each other due to duplicate focuses, producing separate and competing networks of cooperation, with one network consisting of members exclusive from others^[29]. As a collective action problem becomes more severe, institutions should attempt to be more inclusive in their membership^[30]. Joining competing networks put states at a disadvantage as disparate membership across institutions weakens the ability of states to mitigate cybercriminal activity emanating from or in relation to a state within a competing institution, increasing the severity of the problem. Either most actors cooperate within one organization against cybercrime or they risk feeding the problem. Thus, a key assessment for the iterative cooperation hypothesis focuses on whether states consider Europol the focal institution for fighting cybercrime.

C. Material Conditions for Cybercrime Cooperation

In contrast to the idea that the perception of an institution drives states to cooperate within it, states could be driven by more material concerns, which would support the hypothesis that states cooperate with Europol to fight cybercrime to compensate for functional shortcomings that they cannot develop on their own immediately (cooperation by substitution). Bjorn Müller-Wille presents a framework that argues that "expanded co-operation within [Europol] would make sense if it added value to the fight against crime in general"^[31]. Such cooperation must either produce something state agencies cannot produce alone, generate better intelligence than any agency could produce alone, or produce intelligence that state agencies cannot willingly or acceptably produce for political reasons^[32]. Based on these criteria, a state should only be expected to cooperate within an international intelligence organization if there are tangible benefits, such as intelligence that is not reproducible by any single state's crime or intelligence agencies. Müller-Wille surmises that most of the information passing into Europol

was produced by state intelligence agencies and could theoretically be shared with other states without the use of Europol; hence, the advantages of expanded cooperation within Europol seem unclear^[33]. States may also stray from cooperating within an organization due to the centralization of power in a specific region or institution^[34]. Taking these concerns into perspective leads to the belief that states would not engage in the usage of an international institution in a context where national crime agency functions are duplicated. However, this would only be the case if Europol's singular function was to provide intelligence sharing. As stated before, Europol also provides training; technical support and expertise; and pivotally, partnerships with private firms through public-private partnerships. The potential to access these functions and partnerships drives states to cooperate within Europol against cybercrime.

Bendiek and Porter characterize EU cybersecurity policy as a multi-stakeholder structure, emphasizing public-private partnerships. The authors express that anti-cybercrime policy must focus on bringing governmental and nongovernmental actors together as partners. They argue that the current division of responsibilities among civil defense, military defense, and law enforcement sectors in regard to cybersecurity and, by extension, cybercrime, have faltered. There exists far too much cross-pollination of threats and responsibilities for any one sector to handle these threats on their own^[35]. In practice, this informs the nature of cooperation between entities against cybercrime – interactions between states and state institutions arise as these institutions allow for cooperation among these stakeholders. These interactions progress toward formalized institutions – the authors specifically cite the example of Europol as a step toward international coordination against cybercrime^[36]. Because private firms are now responsible for a large portion of public-facing critical infrastructure in Europe, including health care and energy, these firms are now targets for cybercriminals. Moreover, private firms such as information and communications technology (ICT) companies, including Microsoft and Symantec, have expertise and tools in fighting cybercrime that some states do not^[37]. As such, their inclusion in cooperative networks is essential to organizations' attempts to foster effective anti-cybercrime cooperation^[38].

There is some skepticism toward the effectiveness of public-private partnerships within the context of formalized agreements. Tatiana Tropina argues that states should continue to establish informal relationships with private firms, as the establishment of uniform compliance procedures could hinder the effectiveness of these private firms as partners against cybercrime^[39]. Raphael Bossong and Ben Wagner disagree with Tropina and insist that formalized agreements support the effectiveness of public-private partnerships^[40]. However, through the application of a cross-cutting analysis, they find that public-private partnerships are often only rhetoric, and cooperation of this kind is not usually in the interest of private firms, therefore leading states to push toward regulating industry organizations^[41]. Whatever the effectiveness of public-private partnerships and whether firms believe it to be in their interest to cooperate with states, it is clear that states hold the potential of having private partners in fighting cybercrime in high regard, and therefore would be compelled to cooperate with an organization

through which those relationships could be exploited. Thus, states that do not have a high level of rapport with domestic ICT partners seek to augment their lack of relationships by cooperating within an institution such as Europol, which does have established partnerships with prominent, private ICT firms.

Domestically, a wide breadth and depth of nongovernmental partnerships and ICT sector size expansion require considerable time and investment to cultivate. Due to these costs, states could consider increases in partnerships and ICT sector size to be unachievable. Therefore, states seek access to institutions with a growing capability to fight cybercrime. This can be seen as another example of centralization. Previous work in rational institution design has shown that as uncertainty about the world increases, institutional centralization also increases^[42]. As stated earlier in the discussion on the definition of cybercrime, all it takes is the use of a computer system in a malicious manner; anyone who can utilize a computer proficiently becomes theoretically capable of cybercriminal acts, which effectively increases uncertainty. Even if this capability is centralized within the institution itself and these capacities cannot be transferred over to the states, states can choose between having no capabilities and utilizing the institution's capabilities. Clearly the latter choice provides more utility. Thus, in establishing whether states cooperate within an institution with the intention of substituting an institution's capabilities for their own, it is first important to determine whether adequate domestic resources in the form of the technology sector and available partnerships exist.

D. Types of Anti-Cybercrime Cooperation

The significance of capacity building can be drawn from the choices states face when prompted with an institutional bargaining game. Aggarwal defines institutional bargaining games as bargaining games that consist of the types of goods that could provide some utility related to the issue area in question; the actors' individual situations, including their position in the international order, their domestic forces, and elite preferences within the state; and the presence or absence of institutions where bargaining would take place^[43]. Institutional bargaining games result in different payoffs for different actors, which leads actors to attempt to strengthen their own positions^[44]. When prompted with an institutional bargaining game, the actor (usually a state) can choose between three choices: they can attempt to alter the goods involved, they can alter their or their opponents' individual situations, or they can choose to alter an institution or create a new institution. This section focuses on the second option, where states attempt to alter their individual situation. In this context, the bargaining game is cybercrime mitigation, the institutional context is Europol, and the goods in question are Europol's operational support capabilities against cybercrime and its capacity building activities. States then cooperate within Europol in order to utilize the institution's capacity building abilities so that the state will eventually no longer need to utilize Europol's capabilities to fight cybercrime. Thus, this hypothesis supposes that states are cooperating to develop anti-cybercrime capabilities such that the states can eventually become self-reliant in the fight against cybercrime (cooperation

for self-reliance). What distinguishes cooperation for self-reliance from the type of cooperation discussed in the previous section (cooperation by substitution) is that the former focuses on states building capacities in the immediate term through support from the institution within which the state is cooperating, whereas the latter focuses on the use of the institution's capacities in lieu of the state's inability to develop similar capacities.

The framework for assessing cooperation for self-reliance draws primarily from Benoît Dupont's work on the international governance of cybercrime. Dupont maps interactions between states and organizations in the context of cybercrime to specific classifications^[45]. He provides five categories of anti-cybercrime cooperation^[46]:

- ◆ Capacity building;
- ◆ Information sharing;
- ◆ Regulatory and legal activities;
- ◆ Criminal investigations and intelligence collection; and
- ◆ Lobbying.

The overwhelming majority (74.5 percent) of initiatives Dupont includes in his dataset involves capacity building, while information/intelligence exchange characterizes 49 percent of these initiatives^[47]. This finding also supports what some policymakers claim about cybercrime – capacity building remains the most important action in cybercrime mitigation^[48]. However, Dupont professes that these connections do not show the intensity of the cooperation between states and nongovernmental organizations (NGOs) in fighting cybercrime or the intention behind their cooperation. He also goes on to state that data focused on methodologically similar, bilateral initiatives involving cooperation under Europol would produce significantly different results^[49].

Since this paper focuses on cooperation against cybercrime within Europol, it is prudent to test Dupont's findings against this gap in the data. If states are driven to cooperate within an international organization primarily by a desire to develop their own abilities to fight cybercrime, then Europol's primary functions in facilitating intelligence sharing and providing operational coordination and support should not factor into cooperative actions against cybercrime heavily. In other words, a confirmation of the cooperation for the self-reliance hypothesis suggests that states want and generally seek to go it alone in the fight against cybercrime, and most cooperate within institutions in order to reach a point of independence. If this were to occur, they would no longer be affected by the threat of cybercrime as they were before they began cooperating within the institution. In the language of institutional bargaining games, at the point of self-reliance, states successfully change their individual situation and, therefore, their payoff structure within the game. While this assertion runs contradictory to the operational nature of Europol's activities, it is nevertheless important to assess this hypothesis in order to ascertain whether the desire to build capabilities effectively drives state police agencies to cooperate within institutions against cybercrime.

III. METHODOLOGY

This paper tracks a different variable or set of variables for each hypothesis. For the first hypothesis, iterative cooperation, I utilize interview responses and policy data to show whether Europol is seen as a focal institution. For the second hypothesis, cooperation by substitution, I utilize a combination of survey data and interviews to measure how much interaction states have with domestic ICT partners. I also measure the ICT sector size in each state by measuring ICT employment as a percentage of total employment within each Europol member state and compare each country's differential to the mean percentage in order to ascertain the size of each state's ICT sector relative to a central tendency. A percentage of ICT employment is utilized to estimate ICT sector size as opposed to absolute employment numbers in order to normalize the size of each state's ICT sector relative to other member states; utilizing absolute employment numbers results in misleading data due to the population differentials across states. These two variables measure both the reality of interactions and the potential for partnerships that state law enforcement agencies can have with private firms, and therefore characterize whether a state needs to act through Europol to interact with private firms and NGOs or seek out foreign technical expertise. Finally, for the third hypothesis, cooperation for self-reliance, I measure several variables, including the amount of funding a state police agency received and the amount of training requested from Europol in order to capture the number of interactions states had with Europol that can be categorized as capacity building. Also included is data collected from interviews which categorize the frequency and importance of capacity building activities (namely, training and funding) from the point-of-view of Europol officials.

The primary limiting factor of this methodology is the lack of data available from state law enforcement agencies on their activities within Europol. Of the 28 member states that were asked to participate in the qualitative survey, only one (the United Kingdom) gave responses. Of the 28 member states that were asked to participate in the quantitative survey, only one (Denmark) responded. The United Kingdom and Germany both purported to not have the necessary information to answer the quantitative questionnaire. This makes it incredibly difficult to draw strict conclusions from these findings as the lack of data limits the variance required to validate the results. Nevertheless, even with the lack of data, valuable insights can still be gleaned from the results collected.

IV. RESULTS

The following section discusses findings from interviews conducted with Europol's Head of Strategy, Philipp Amann; an interview conducted with the United Kingdom's National Cyber Crime Unit (NCCU); and survey data collected from a questionnaire given to Denmark's National Cybercrime Center (NC3). The survey consisted of nine multiple choice questions focusing on various topics, including funding from Europol for anti-cybercrime operations; frequency of interactions with Europol in the context of anti-cybercrime operations; frequency of interac-

tions with domestic and international, nongovernmental technology partners; and one free-response question focusing on agencies' capabilities in comparison to Europol's. The evidence also includes data collected from EU policy documents.

A. Identifying Europol as a Focal Institution

To measure whether Europol is seen as a focal institution, a combination of data was collected from policy analyses, interviews, and survey data. The EU's overall cybersecurity strategy cites Europol "as the European focal point in the fight against cybercrime"^[50]. The strategy explicitly assigns the responsibility of facilitating anti-cybercrime cooperation among states and cooperation between states and private or nongovernmental stakeholders to Europol^[51]. These statements leave no ambiguity that Europol carries the distinction of being considered a focal institution, at least from the point-of-view of the EU itself. By extension, Europol is undoubtedly seen as a focal institution against cybercrime from the point-of-view of many policymakers.

From the perspective of the institution, Europol does not directly inform a member state that its protections against cybercrime require improvement unless the state in question asked Europol for an assessment^[52]. Member states participate, including the sharing of open-source reports, malware, and other forms of data, on a voluntary basis^[53]. Should a member state choose not to share its intelligence, Europol cannot force a state to share that intelligence. As for reasons why a member state would not cooperate with Europol, member state law enforcement agencies are often either unaware of or ignore the resources Europol can provide^[54]. In fact, Europol officials are aware that member states have law enforcement agencies that are producing tools and materials that the organization has already produced^[55]. Europol officials see this as law enforcement agencies across member states being unaware of what Europol can provide those agencies, and therefore do not reach out to the institutions as much as they could^[56]. Survey data collected from the Danish NC3 reinforces this supposition; the center remarked that only up to a fifth of anti-cybercrime operations in the most recent year involved direct operational support from Europol^[57].

While the perceived lack of use by state police agencies suggests that states do not view Europol as a focal institution for cybercrime mitigation, further elaboration about the nature of the problem of cybercrime actually suggests that Europol is viewed as a focal institution for cybercrime mitigation by member states. In a comment at the end of the survey, NC3 stated that "the resources and capability of the member states...holds [sic] back the common process. Cyber [crime] has to be prevented and fought from an international perspective"^[58]. Furthermore, rather than pursuing policy-based prescriptions to bring agencies into the fold, Amann suggests that Europol needs to do a better job of advertising and reaching out to law enforcement agencies^[59]. The choice to attribute the perception that Europol lacks usefulness to lack of outreach rather than tying it to a need for hierarchical structure indicates either an unwillingness to establish a more hierarchical structure or a belief that a more hierarchical structure is unnecessary. Even with the voluntary nature of state crime agencies' relationship with the in-

stitution, Amann remarked that the member states do utilize Europol effectively^[60]. This statement, coupled with the statement from NC3 regarding the need to fight cybercrime from an international perspective, leads to the conclusion that the international nature of cybercrime gives states the impetus to place a premium on platforms for international anti-cybercrime operations, such as Europol.

B. Measuring Agency Use of Europol to Substitute Capacities

To assess whether Europol is used by states to substitute a lack of capability, a combination of interviews, survey data, and domestic ICT employment sector size data is utilized to determine whether a state's law enforcement agency perceives its available capabilities to be up to par with Europol's and whether the potential for increased partnership and capabilities exists. A measurement of these variables illustrates whether states perceive that Europol's available capabilities and partnerships within the context of mitigating cybercrime are more valuable than the state's domestic capabilities and partnerships.

Europol's operations consist of three primary categories. These categories include operational support, including intelligence sharing, analysis, and on-the-ground support; education and awareness training; and coordinating or taking part in multilateral/joint actions. Intelligence sharing serves as the primary day-to-day work that Europol undertakes^[61]. Much of this intelligence sharing occurs on the Secure Intelligence Exchange Network Application (SIENA), a platform through which law enforcement agencies from Europol's member states, Europol officials, and third parties with cooperation agreements with Europol can communicate with and disseminate intelligence to other partners or to Europol itself^[62]. Europol also conducts malicious software (malware) analysis through the Europol Malware Analysis System (EMAS)^[63]. Member state agencies can submit a piece of malware and Europol employees can conduct forensic analysis on the malware to produce conclusions and support a member state in its investigation or active operation. Member states have access to the Digital Forensics and Mobile Laboratory, which mines data from hard drives and mobile phones, and Europol's password decryption platform^[64]. Lastly, Europol interfaces with outside partners, including Interpol and third-party states, as well as nongovernmental partners, including private firms, accepting information from them, including internet protocol (IP) addresses, and consulting nongovernmental partners in an advisory capacity^[65]. When asked whether NC3 could claim equivalent anti-cybercrime capabilities to those of Europol, the agency responded, "No"^[66]. The NCCU stated that capabilities across member states varied widely and, at times, bilateral interactions with partners with similar capabilities resulted in more fruitful interactions; however, bilateral relationships lacked the ability to pool resources from other member states or construct the "big picture" pertaining to the issue at hand^[67].

Europol also maintains relationships with public-private partners for operational and advisory purposes. Private firms and NGOs provide Europol with intelligence, including IP addresses

of potentially compromised or potentially suspicious computers^[68]. Private firms and NGOs are also utilized in an advisory capacity through membership with an advisory board^[69]. Most member states are thought to hold their own relationships and partnerships with private firms and NGOs, but these are not tracked by Europol. Thus, the relationship between member states and EC3 is not at all hierarchical, despite the fact that institutional policy drives the direction of the relationship^[70]. The NCCU remarked that business and reputational costs often stand in the way of forming partnerships with private firms. However, private firms seem to be willing to share more information on some types of attacks, such as DDoS attacks, due to the lower reputational risks associated with them in comparison to attacks that disclose user data^[71].

While all of this illustrates that Europol has considerable capabilities of which member states can take advantage and that these capabilities encourage states to engage in cooperation with Europol against cybercrime, survey data illustrates that member states might already have comparable capabilities. Table 1 shows the results of a survey answered by NC3 with respect to the proportion of interactions the agency has with ICT partners both through and outside of Europol as well as the Danish ICT sector size compared to the average Europol member state sector size.

Danish Cybercrime Interactions*	
Category	Percentage
Private sector partners who also have partnerships with Europol	1-20%
Private technology partners through Europol	1-20%
Private sector partners that are also domestic partners	41-60%
EU state police agencies that occurred through Europol	21-40%
2016 national ICT sector employment percentage compared with average Europol member state 2016 ICT sector employment (SD = 1.2)	+0.6%

*All from 2017 unless otherwise noted

TABLE 1

These results indicate that the overwhelming majority of cybercrime operations in the Danish case do not require direct operational involvement from Europol. Denmark clearly has above-average domestic technology partnerships and available domestic technological prowess; most of NC3’s interactions with private partners occur outside of Europol, and around half of these interactions are with domestic, private firms, which eliminates the need to interact with them through an international organization in the first place by virtue of their domesticity. Most interactions with other Europol member states’ police agencies occurs outside of the organization. Even the Danish ICT sector size is one-half standard deviation above the average Europol-member ICT sector size – a medium-sized difference from the average Europol member state ^[72]. These results also indicate that there exists a potential for greater utilization of domestic partnerships and technical expertise in comparison with other member states.

C. Measuring Agency Use of Europol to Build Capacity

To measure the frequency and importance of capacity building activities to Europol, it is important to first know Europol's available capacity building activities. The dissemination of training, funding, and technical tools can be considered a capacity building activity. Much of the institution's educational outreach and operational support focuses on establishing a baseline level of expertise among member states to ensure effective cross-border cooperation^[73]. Europol also provides funding to member state agencies to implement policy objectives; this funding can also be used to implement joint, international projects proposed by member state agencies^[74]. Free anti-cybercrime tools, such as forensic analysis tools developed through the FREETOOL project, are also provided to member states^[75]. Training to utilize these tools is provided through Europol.

Europol officials find that capacity building activities hold a relatively low frequency and importance in comparison to other Europol functions. Amann ranks the following cooperative actions against cybercrime in order of importance from least to greatest: education and prevention outreach, intelligence sharing and operational support, and joint actions and operations. Amann also ranked the three types of cooperation in terms of frequency from least to greatest: education and prevention outreach, joint actions and operations, and intelligence sharing and operational support.

With these results, Dupont's finding that capacity building makes up the overwhelming plurality of cooperative interactions against cybercrime^[76] comes under scrutiny. This complicates the cooperation for self-reliance hypothesis. If capacity building only includes education and prevention outreach, then when examining the metrics of importance and frequency, capacity building is seen as both least important and least frequent. If operational support (in particular, intelligence sharing and analysis) can be categorized under capacity building, then capacity building becomes both most important and most frequent^[77]. However, operational support does not include common actions associated with capacity building, such as education. Admittedly, Amann emphasized that the differences in importance among these three actions are minimal, the relationships among the three are close, and each type of cooperation is often tied to another type of cooperation^[78]; the NCCU also emphasized this^[79]. Sometimes officers are sent from member state crime agencies to work on specific cases if necessary^[80]. There exist ample opportunities for states to request operational support, although intelligence sharing does make up the bulk of the day-to-day work. However, capacity building activities seem to be in sparse supply.

Results from the questionnaire given to member-state police agencies also seem to indicate that capacity building does not characterize cooperation within the organization. Survey responses from NC3 with respect to the agency's interactions with Europol strictly pertaining to capacity building activities show that the agency does not utilize Europol very much to build

capacity: the agency requested no funding and only two instances of training in the most recent fiscal year.

These figures correspond accordingly with the statements from Europol officials on the frequency of cybercrime-related training. It must be noted that Europol does not provide many instances of training per year^[81] and, therefore, numbers pertaining to training may be relatively low no matter what; however, the amount of requested funding is telling. Funding can be used to develop new technologies, hire new staff, provide training, and invest in new projects, all of which are clearly capacity building activities. Given that the previous sections have illustrated that NC3 finds cooperation with Europol incredibly important in fighting cybercrime, the fact that the agency requested no funding in the 2017 fiscal year shows that capacity building must not matter much in the calculus of that state's national law enforcement agency.

V. ANALYZING CONDITIONS FOR COOPERATION

The first hypothesis is tested by demonstrating whether states viewed Europol as a focal institution in cybercrime mitigation; if states considered Europol a focal institution in cybercrime mitigation, then, by Aggarwal's framework, iterative cooperation drives cooperation within Europol against cybercrime. When the decision was made to expand into the realm of cybercrime, Europol's preexisting structure may have given it the ability to establish its capabilities and reputation to a point that supersedes the capabilities and reputation of member-state police agencies. Europol's preexistence is an important detail to note; Europol was established in 1998, but did not establish a dedicated cybercrime operations unit until 2013^[82]. The establishment of the organization predates many of the member-state cybercrime agencies, only some of which, such as the Greek agency, predate the establishment of Europol^[83].

While Europol's cybercrime center postdates many of the member state agencies' cybercrime centers, states do not seem to feel the need to deviate from Europol's preestablished framework. If there already exists an organization that can serve as a niche for a form of cooperation, as in the case of Europol and EU-wide crime response, states require less overhead to be convinced to engage in new forms of cooperation. The remark made by the NC3 indicates that Europol's known reputation and ability entice states to approach the organization with some degree of confidence. This lines up with the perception that Europol is a "focal" institution against cybercrime.

In testing the second hypothesis, ICT employment data for each state was collected alongside survey data that measured a state police agency's involvement with domestic ICT partners (table 1). If a state's ICT sector size was small compared to the average Europol member state ICT sector size or the state police agency had weak involvement with ICT private firms and NGOs, then that state should be more driven to cooperate within Europol. When combining the ICT employment percentages compared to the average EU employment percentage, the Danish response to the survey was illuminating. According to the results, Denmark had above-average

ICT employment as a percentage of total employment when compared to other Europol member states^[84]. Only up to 20 percent of NC3's interactions with nongovernmental technology partners occur through Europol^[85]. Around half of the agency's interactions with nongovernmental technology partners occur domestically; these do not require interaction with Europol to access^[86]. Prima facie, all of these data points suggest that such a state should be less dependent on Europol's potential opportunities for access. Nevertheless, it seems that even a relatively small need to fight potential cybercrime threats internationally results in a willingness to engage in cooperation within the institution, regardless of the number of problems those activities can solve. While Denmark did not have a small ICT sector size relative to the average Europol member state sector size and had frequent interactions with technology partners outside of Europol, this did not change NC3's professed willingness to cooperate within the institution.

Furthermore, NC3's perception that its capabilities do not match Europol's and the survey results are at odds. It seems clear from the data that the idea that Europol needs to provide most of the necessary partnerships to member states to encourage cooperation does not hold water. Again, this might point to states' and state police agencies' views on the nature of the problem of cybercrime – this is an issue area for which agencies perceive there is no limit to increased support and expertise; however, this increased support and expertise do not necessarily amount to the wholesale substitution of Europol's cybercrime mitigation capabilities with domestic ones. Therefore, while it may allow states to increase their abilities to fight cybercrime, cooperation in the name of substituting capabilities only provides marginal improvement in some cases and serves more as a secondary driver toward state involvement within Europol than as a primary one. This leads to the conclusion that an intrinsic property of the problem, the international nature of cybercrime, serves as a primary motivator behind states' willingness to cooperate within an institution to fight cybercrime; in addition, other potential avenues for mitigation, specifically domestic avenues, are not enough to make a state's police agency feel secure.

Testing the third hypothesis involves identifying whether international cooperation within Europol focuses on capacity building; if a large proportion of cooperation does focus on capacity building, then states are driven to cooperate within the institution to build a sustainable, domestic, anti-cybercrime apparatus. As noted from the interview with Amann, each cooperative action is classified according to Dupont's categories^[87] to ascertain whether international cooperation against cybercrime focuses on capacity building.

Table 2 maps the categories Dupont presents in his work to the types of operations available through Europol. Clearly, these operations do not cleanly fall into the different categories. For example, as an open-source (free-to-use) project, the development of the FREETOOL project can be considered an instance of capacity building to allow member state police agencies to augment their cybercrime analysis capacity. In contrast, tools such as EMAS are only useful if other states share their malware through the system. However, both allow member states to build up their intelligence concerning malware. Furthermore, Amann characterized the use

of such tools not as capacity building, but as operational support, placing technical forensics analysis tools under the category of law enforcement operations^[88]. This overlap makes it difficult to provide a discrete category for each type of cooperation. Given that intelligence sharing makes up most of Europol’s day-to-day work, it seems reasonable to conclude that the exchange of information trumps all of the other categories in frequency. This conclusion is not necessarily predicated upon the inclusion of technical forensics analysis tool development, as SIENA still constitutes the bulk of intelligence report sharing. Therefore, if capacity building only encompasses funding, education, and capability development, then capacity building comes in third behind information exchange and law enforcement operations, respectively. Since capacity building only makes up a relatively small amount of cooperative measures that occur within Europol, cooperation for self-reliance seems to be a weak driver in encouraging states to cooperate within Europol against cybercrime.

Europol Classification of Anti-Cybercrime Activities	
Category of Action	Action/Operation
Capacity building	<ul style="list-style-type: none"> • Training and educational services • Monetary funding • Technical forensics analysis tool development
Exchange of information	<ul style="list-style-type: none"> • Intelligence exchange through SIENA • Technical forensics analysis tool usage
Law enforcement operations	<ul style="list-style-type: none"> • Investigations supported by Europol personnel • Joint investigations between member states • Technical forensics analysis tool usage
Lobbying	<ul style="list-style-type: none"> • Ability to influence Europol policy objectives

TABLE 2

Based on these findings, it is reasonable to posit that while capacity building does play an important role in anti-cybercrime cooperation, states may not focus on it if an organization is capable of facilitating more direct means of engaging potential threats. NC3’s survey responses (table 1) are very telling in this regard. The center did not request funding for anti-cybercrime operations in the 2017 fiscal year. However, the center also noted that up to 40 percent of interactions with other EU member-state crime agencies required interaction with the agency through Europol, and up to 20 percent of anti-cybercrime operations required the direct involvement of Europol^[89]. Despite neither of these interactions making up the majority of Europol’s types of operations, they still occur at regular enough frequency to be considered the primary work of Europol. Based on this evidence, the desire to build capacity only has a minimal-to-moderate effect on states’ cooperation within an institution to fight cybercrime.

One confounding variable that arose from the data collected through interviews and surveys is the cultural role of police in cybercrime investigations. Amann suggested that several Europol

member states have different cultural attitudes toward policing that affect their willingness to cooperate internationally with other law enforcement agencies or with nongovernmental partners. He brought up the example of the Netherlands, where many of the banks have close partnerships with anti-cybercrime initiatives and policing agencies. In addition, Dutch banks interface with anti-crime task forces to disseminate information to other banks and law enforcement representatives in the same room^[90]. These partnerships may not be tolerated by citizenry of other member states due to cultural and social views on privacy and police activity in those member states. The variance in legal frameworks across these countries also factors into whether these types of cooperative relationships are possible. The NCCU noted that this is a large challenge in regard to working within the institution^[91].

Another confounding variable that was brought up in the interview was the size of countries' bureaucracies. Citing Estonia, Amann noted that the country itself is small in population and does not have the same degree of bureaucratic complexity as larger member states, such as Germany and France. The lack of bureaucratic complexity leads to a reduction in formal structures in comparison with larger countries, leading to a smaller amount of people taking on a larger amount of responsibilities. This increases the responsiveness between government officials of smaller countries and Europol at the cost of higher barriers to establishing relationships with Europol when government officials first take office^[92]. In contrast, the Netherlands contains many formalized structures for partnerships with Europol, which creates a different approach to and platform for cooperation. Bureaucratic turnover also creates problems. The constant turnover of senior management in Europol member states leads to a lack of institutional memory among government staff and policymakers^[93]. This turnover may result in a new staff that does not know how to harness Europol resources effectively and efficiently.

VI. CONCLUDING THOUGHTS AND NEXT STEPS

Given the evidence presented in this piece, the strongest driver for participation in Europol is iterative cooperation. Europol's prior space within the realm of international police agency cooperation seems to have spurred states to engage in cooperation with other states through the organization and with Europol personnel, even if states had already established a cybercrime unit that predated EC3. Contributing to this willingness to cooperate also seems inherent to the problem of cybercrime; that is, effective mitigation must be international in scope.

Cooperation by substitution and cooperation for self-reliance, on the other hand, are weaker drivers. As seen in the case of Denmark, an above-average ICT sector size in terms of the percentage of employment does not lessen the value that the state's cybercrime unit places on Europol's utility in fighting cybercrime. Observations on the types of support that Europol gives also seem to focus readily on operational support and information exchange, effectively supplanting capacity building as the most frequent and important type of interaction. Again, it seems that reputation and ability play directly into how states act within Europol. The

organization's structure and services lend themselves to direct support to law enforcement operations. The ability to provide known, effective services can be construed as a precondition to states cooperating within an IGO on an operational basis.

More data from other Europol member state police agencies must be taken into account before drawing further policy implications. The current version of this project only observes two states, which both have a higher-than-average technology sector size in terms of ICT employment percentage^[94]. The next step would be to see whether data obtained from member states with a lower-than-average technology sector size would provide similar results to those of the states examined so far. Furthermore, there exist no competing IGOs or NGOs that have codified intelligence-sharing agreements and anti-cybercrime capabilities to the extent that Europol has. Therefore, it is difficult to discern whether the organization is seen as a focal institution due to a lack of available competition. The lack of a competing agency without Europol's reputation cannot be tracked to measure its comparative utilization, weakening the ability to establish a direct causal link between Europol's existence and its image as a "focal" institution.

Nevertheless, the preconditions of reputation and known competence must be taken into account as important considerations should IGOs and NGOs want to encourage international members to cooperate, whether addressing cybercrime or some other matter of international security. In his interview, Amann summed up the biggest factor in one word: "trust." This is not just trust in one's partners, however; it is trust that cooperation leads to successful operations. This indicates that the overhead necessary to convince states to cooperate is considerable, but, once that overhead has been established, states no longer need much convincing.🔒

ACKNOWLEDGEMENTS

I would like to thank Vinod Aggarwal, Amy Gurowitz, and Andrew Reddie of the University of California, Berkeley, for advising me throughout this project. I would especially like to thank Andrew, whose continued support over the past two years led me to formulating and pursuing this topic in the first place. I would like to thank Philipp Amann at Europol and Paul Timmers for speaking with me as I wandered through Europe attempting to obtain data on cybercrime cooperation. I would also like to thank Tobias Hofmann at the University of Utah, whose feedback alongside Andrew's led me toward the institutional design frameworks. Finally, I would like to thank the Center for Long-Term Cybersecurity at the University of California, Berkeley, for allowing me to utilize its space while working on this project.

NOTES

1. Nick Eubanks, "The True Cost of Cybercrime for Businesses," *Forbes* (blog), July 13, 2017, <https://www.forbes.com/sites/theyec/2017/07/13/the-true-cost-of-cybercrime-for-businesses/>.
2. European Parliament, "REGULATION (EU) 2016/794 OF THE EUROPEAN PARLIAMENT AND OF THE COUNCIL of 11 May 2016," *EUR-Lex - Access to European Union Law*, May 5, 2016, <http://eur-lex.europa.eu/legal-content/EN/TXT/HTML/?uri=CELEX:32016R0794&from=EN>.
3. Philipp Amann, Interview with Philipp Amann, Head of Strategy, European Cybercrime Center, In-Person, January 8, 2018.
4. "Operational Agreements," Europol, accessed April 29, 2018, <https://www.europol.europa.eu/partners-agreements/operational-agreements>.
5. Denmark was the last state to re-join Europol after Danish voters rejected a 2015 referendum that would have allowed Denmark to opt-in on EU home and justice matters on a case-by-case basis. A separate agreement was eventually struck between the EU and Denmark that allowed continued use of Europol in 2017.
6. "EU Policy Cycle - EMPACT | Organised Crime | Europol," accessed April 29, 2018, <https://www.europol.europa.eu/crime-areas-and-trends/eu-policy-cycle-empact>.
7. NCCU Research, Interview with United Kingdom National Cyber Crime Unit, Text, February 21, 2018.
8. "European Cybercrime Centre - EC3," Europol, accessed April 29, 2018, <https://www.europol.europa.eu/about-europol/european-cybercrime-centre-ec3>.
9. NCCU Research.
10. Amann.
11. "Symantec and Europol Strengthen Cooperation in Joint Fight against Cybercrime | Europol," accessed April 29, 2018, <https://www.europol.europa.eu/newsroom/news/symantec-and-europol-strengthen-cooperation-in-joint-fight-against-cyber-crime>.
12. Elaine Fahey, "The EU's Cybercrime and Cyber-Security Rule-Making: Mapping the Internal and External Dimensions of EU Security," SSRN Scholarly Paper (Rochester, NY: Social Science Research Network, January 24, 2014), <https://papers.ssrn.com/abstract=2384491>.
13. *Ibid.*
14. *Ibid.*
15. "Internet Organised Crime Threat Assessment (IOCTA) 2017."
16. Felicity Vabulas and Duncan Snidal, "Organization without Delegation: Informal Intergovernmental Organizations (IIGOs) and the Spectrum of Intergovernmental Arrangements," *The Review of International Organizations* 8, no. 2 (June 1, 2013): 193–220, <https://doi.org/10.1007/s11558-012-9161-x>.
17. Kenneth W. Abbott and Duncan Snidal, "Why States Act through Formal International Organizations," *The Journal of Conflict Resolution* 42, no. 1 (1998): 3–32.
18. Barbara Koremenos, Charles Lipson, and Duncan Snidal, *The Rational Design of International Institutions* (Cambridge, UNITED KINGDOM: Cambridge University Press, 2003), <http://ebookcentral.proquest.com/lib/berkeley-ebooks/detail.action?docID=255150>.
19. Abbott and Snidal.
20. *Ibid.*
21. Vinod K. Aggarwal, *Institutional Designs for a Complex World: Bargaining, Linkages, and Nesting* (Cornell University Press, 1998).
22. Joseph Jupille and Duncan Snidal, "The Choice of International Institutions: Cooperation, Alternatives and Strategies," *SSRN Electronic Journal*, 2006, <https://doi.org/10.2139/ssrn.1008945>.
23. *Ibid.*
24. Aggarwal.
25. *Ibid.*
26. Barbara Koremenos, Charles Lipson, and Duncan Snidal, *The Rational Design of International Institutions* (Cambridge, United Kingdom: Cambridge University Press, 2003).
27. Jupille and Snidal, "The Choice of International Institutions."
28. *Ibid.* See also Koremenos, Lipson, and Snidal.

NOTES

29. Benoît Dupont, “La gouvernance polycentrique du cybercrime : les réseaux fragmentés de la coopération internationale,” *Cultures & Conflits*, no. 102 (August 8, 2016): 95–120, <https://doi.org/10.4000/conflits.19292>.
30. Koremenos, Lipson, and Snidal.
31. Björn Müller-Wille, “The Effect of International Terrorism on EU Intelligence Co-Operation,” *JCMS: Journal of Common Market Studies* 46, no. 1 (January 1, 2008): 49–73, <https://doi.org/10.1111/j.1468-5965.2007.00767.x>.
32. Ibid.
33. Ibid.
34. Zahid Jamil, “Global Fight against Cybercrime: Undoing the Paralysis,” *Georgetown Journal of International Affairs*, 2012, 109–20.
35. Ibid.
36. Ibid.
37. Dupont.
38. Bendiek and Porter.
39. Tatiana Tropina, “Public–Private Collaboration: Cybercrime, Cybersecurity and National Security,” in *Self- and Co-Regulation in Cybercrime, Cybersecurity and National Security*, SpringerBriefs in Cybersecurity (Springer, Cham, 2015), 1–41, https://doi.org/10.1007/978-3-319-16447-2_1.
40. Raphael Bossong and Ben Wagner, “A Typology of Cybersecurity and Public-Private Partnerships in the Context of the EU,” *Crime, Law and Social Change* 67, no. 3 (April 1, 2017): 265–88, <https://doi.org/10.1007/s10611-016-9653-3>.
41. Ibid.
42. Koremenos, Lipson, and Snidal.
43. Aggarwal.
44. Ibid.
45. Dupont.
46. Ibid.
47. Ibid.
48. Heli Tiirmaa-Klaar et al., “Botnets, Cybercrime and National Security,” in *Botnets*, SpringerBriefs in Cybersecurity (Springer, London, 2013), 1–40, https://doi.org/10.1007/978-1-4471-5216-3_1.
49. Dupont.
50. European Commission, “Cybersecurity Strategy of the European Union: An Open, Safe and Secure Cyberspace,” European External Action Service, July 7, 2013, https://eeas.europa.eu/archives/docs/policies/eu-cyber-security/cybsec_comm_en.pdf.
51. Ibid.
52. Amann.
53. NCCU Research.
54. Amann.
55. Ibid.
56. Ibid.
57. NC3, Interview with Danish National Police Cyber Crime Center, Text, March 16, 2018, 3.
58. Ibid.
59. Amann.
60. Ibid.
61. James Igoe Walsh, *The International Politics of Intelligence Sharing* (Columbia University Press, 2010), <https://doi.org/10.7312/wals15410>.
62. “Secure Information Exchange Network Application (SIENA) | Activities & Services | Services & Support | Information Exchange | Europol,” accessed April 29, 2018, <https://www.europol.europa.eu/activities-services/services-support/information-exchange/secure-information-exchange-network-application-siena>.
63. Amann.

NOTES

64. NCCU Research.
65. Amann.
66. NC3.
67. NCCU Research.
68. Amann.
69. Ibid.
70. Ibid.
71. NCCU Research.
72. Jimmie Leppink, Patricia O’Sullivan, and Kal Winston, “Effect Size – Large, Medium, and Small,” *Perspectives on Medical Education* 5, no. 6 (December 2016): 347–49, <https://doi.org/10.1007/s40037-016-0308-y>.
73. Amann.
74. Ibid.
75. “FREETOOL v2.0 | UCD Centre for Cybersecurity & Cybercrime Investigation,” accessed April 29, 2018, http://www.ucd.ie/cci/projects/current_projects/freetool2.html.
76. Dupont.
77. Amann.
78. Ibid.
79. NCCU Research.
80. Amann.
81. Ibid.
82. European Commission, “Cybersecurity Strategy of the European Union: An Open, Safe and Secure Cyberspace.”
83. Philip Chrysopoulos, “Greek Police Transfers Cyber Crime Unit Chief, Then Repeals Decision, Then Transfers Him | GreekReporter.Com,” February 18, 2016, <http://greece.greekreporter.com/2016/02/18/greek-police-transfers-cyber-crime-unit-chief-then-repeals-decision-for-now/>.
84. European Commission, “Eurostat,” Eurostat: Your key to European statistics, August 11, 2016, <http://ec.europa.eu/eurostat/web/digital-economy-and-society/data/main-tables>.
85. NC3.
86. Ibid.
87. Dupont.
88. Amann.
89. NC3.
90. Amann.
91. NCCU Research.
92. Amann.
93. NCCU Research.
94. European Commission, “Eurostat.”

Feed the Bears, Starve the Trolls

Demystifying Russia's Cybered Information Confrontation Strategy

Nina A. Kollars, Ph.D.

*United States Naval War College
Strategic and Operational Research Department
Newport, RI, USA*

Michael B. Petersen, Ph.D.

*United States Naval War College
Strategic and Operational Research Department
Newport, RI, USA*

ABSTRACT

This paper seeks to establish an explicit connection between Russian strategic information operations theory and the execution of Russian cyber operations. These operations are part of a larger strategic construct in the Russian lexicon known as “information confrontation” – a concept that is deeply embedded in Russian strategic thought and official doctrine. Furthermore, within the information confrontation concept, the Russians posit an essential distinction between technical and psychological effects. Using this distinction, we attempt to introduce analytical clarity to the study of Russian activities in the cyber domain. Specifically, within the technical/psychological distinction, we find that Russian operations that tend toward the latter tend to be less sophisticated and conducted at some level of remove from direct control by the regime, while the former clearly demonstrates what we refer to as “organizational sophistication.”

Keywords—cyhacking, organizational structure, Russian strategy, cyber, advanced persistent threat, information operations, Russia, resources, doctrine

I. INTRODUCTION

The flood of fevered reports on Russia’s election meddling, malware assaults, and mysterious hacking teams are fundamentally disorienting. It can make stepping back to assess Russian strategic lines of effort and the “who” and “what” of Russian assets in play seem like a fool’s errand. But there is a well-established, strategic and organizational logic that underlies all of these activities. What might be called “cybered information confrontation” is at the center of a Russian strategic concept known as “New Type Warfare,” an intellectual construct embraced by Russia’s military leadership that posits in part that the exploitation of information offers Russia a key asymmetric advantage.

This is a work of the U.S. Government and is not subject to copyright protection in the United States. Foreign copyrights may apply.

II. RUSSIAN STRATEGY AND CYBERED INFORMATION CONFRONTATION

Russian political and military leadership believe that their country is locked in an existential contest with the West. However, to the Russian mind, the very rules of this struggle have changed. The essential separation between peacetime and wartime no longer exists and, while the threat of military force is still an important component of strategy, it has receded in favor of nonmilitary measures. Instead, global competition with the West has become a contest for who can best exploit the nonmilitary aspects of conflict to the greatest strategic gain. In the words of General Valery Gerasimov, the Chief of Russian General Staff, “The role of nonmilitary means of achieving political and strategic goals has grown, and, in many cases, they have exceeded the power of force of weapons in their effectiveness”^[1]. Foremost in Russian strategy among these nonmilitary aspects of conflict is the notion of “information confrontation” [informatsionnoye protivoborstvo].

While this concept can encompass open propaganda, state-sanctioned news outlets, and other activities, cybered information confrontation is the critical component of Russia’s competitive efforts. It nests within Russian strategic and military thinking as both concept and enabler. It also operates in peacetime and wartime, and its tactics range from now widely known information operations to sophisticated hunts for and exploitation of network vulnerabilities, up to and including the achievement of kinetic effects in the real world. While more work remains to be done, from a command and control perspective, there appears to be a spectrum along which these functions operate, from minimal state control to extraordinarily sophisticated operations requiring strict state organization and orchestration. Along this spectrum, and in most cases, Moscow is able to achieve what it views as a level of plausible deniability by either exploiting proxies or by embedding its operations in the deeply secret world of intelligence operations.

Weaponizing information is a key aspect of Russia’s competitive strategy. Indeed, information confrontation is the red thread running through every arena of strategic competition with the West. It is a strategy that seeks to exploit information in political, cultural, social, economic, religious, military, and other spheres. Information can be exploited for tactical and strategic gain, destroyed, planted, distorted, stolen, and manipulated. These techniques of course have historical precedent in the Soviet Union and the Cold War, but current measures go beyond Soviet traditions in that they exceed mere psychological operations, and information dominance has replaced military mass in the minds of Russian strategists and policymakers as the center of gravity in a modern conflict.

Cybered information confrontation can take many paths to many goals. In some cases, the goal is merely to inject doubt in the institutions of an adversary state, to paralyze decision-making, and/or to debilitate democratic processes^[2]. This may either be an end in itself or may also be part of a broader enabling campaign to achieve more specific strategic gains. In other cases, cybered information confrontation can seek out and exploit weaknesses in network and physical infrastructure, again, either as an end in itself or to enable wider operations.

These ideas are fundamental to Russian military and political strategy. For example, Colonel V. N. Gorbunov and Lieutenant General S. A. Bogdanov, two of Russia's most influential military strategists, write that "Weakening a country marked as a target of aggression today (and also in the long run) is possible by internal weakening of the state in all respects, including the taking of informational, psychological, moral, climatic, and organizational measures..."^[3]. Undermining an adversary state's ability to govern, either in peacetime or wartime, is therefore both an end and an enabling tool.

This notion received its fullest expression in a 2015 article in the Russian Bulletin of the Academy of Military Science by then-chief of the Main Operations Directorate of the Russian General Staff General-Lieutenant Andrey Kartapolov. Kartapolov outlined the concept of "New Type Warfare," which encompasses political methods to bring about changes in the policies of other states; political efforts to prepare the battlefield for military action; and, if necessary, high-technology conflict. The ultimate goal of New Type Warfare is to reduce the adversary's military strengths via other means. "Nonstandard forms and methods that will make it possible to level the enemy's technological superiority are being developed for the employment of our Armed Forces," he wrote. In this case, "nonstandard forms and methods" include cybered information confrontation as a tool for achieving a broader end^[4].

Intrinsic to New Type Warfare is the concept of the initial period of war (IPW). Information superiority – that is, controlling the flow and content of information – is the essential element of IPW. The key, according to Russian strategists A. V. Serzhantov and A. P. Martoflyak, is "information warfare measures undertaken in advance to achieve political aims without resort to armed force, and then...cultivat[ing] a favorable response from the world community to the use of armed force"^[5]. Information confrontation in IPW is used to reduce public faith in national institutions and make target nations ungovernable by undermining their leadership and key infrastructure. Ultimately, for Kartapolov, "the employment of independent actions and methods for a new type war makes it possible to achieve military results ... without the employment of one's own armed forces." Thus, in this formulation, cybered information confrontation is both an end and a means of achieving strategic success.

It should be noted, however, that these ideas are also partially the result of Russia's conventional military and economic inferiority to the West, and its search for asymmetric solutions to this challenge. This basic idea of finding cheap asymmetries against adversaries is deeply embedded in the highest levels of the Russian military and political hierarchy. No less a figure than Vladimir Putin himself has stated that "We must take into account the plans and directions of development of the armed forces of other countries... Our responses must be based on intellectual superiority, they will be asymmetric, and less expensive"^[6]. Likewise, in his seminal article on New Type Warfare, Kartapolov noted that "the features of preparation and conduct of new-type warfare are being fully used, and 'asymmetric' means of confronting the enemy are being developed." Cybered information confrontation is therefore a tactic designed to

short-circuit the West's military superiority by avoiding expensive and bloody, kinetic conflicts as well as achieving strategic gains by exploiting the information domain. In case a conflict were to erupt, the use of cybered information confrontation could help exploit vulnerabilities and level the playing field.

Broadly speaking, cybered information confrontation has two components in the Russian formulation: "informational-technical" and "informational-psychological." Information-technical measures tend to involve computer network operations, such as attack, defense, espionage, and exploitation^[7]. Information-psychological measures are attempts to either change people's beliefs in favor of Russian strategic objectives or to sow dissent among adversary nations to the point that decision-making is hamstrung. Moscow employs these measures in both peacetime and wartime.

The most basic and well known of these two approaches is the information-psychological approach. At the most simplistic level, Russian agencies utilize ostensibly private armies of trolls to manipulate with a certain level of plausible deniability the narrative of particular stories in an adversary country. The most infamous of these is, of course, the Internet Research Agency (IRA), which flooded the United States with fake news stories during the 2016 presidential election. Official Moscow attempted to maintain a degree of separation from this operation by using its connection to Yevgeny Prigozhin, the St. Petersburg restaurateur-cum-oligarch known as "Putin's Chef" who bankrolled the IRA with a portion of the billions of dollars paid to him through a food-service contract with the Russian military.

Russia also exploits the work of semi-autonomous, patriotic hackers and hacker organizations such as CyberBerkut. This loose network of "hacktivists," named after Berkut, the now-disbanded Ukrainian police force that became well known for its violent tactics against Euromaidan protesters in 2014, is, according to the Defense Intelligence Agency, a front organization for state-sponsored cyber activities in Ukraine.^[8] CyberBerkut generally focuses its efforts on low-level harassment and propaganda campaigns, such as distributed denial-of-service (DDoS) attacks, website defacement, and disinformation campaigns, but has more recently been involved in email hacking schemes^[9].

Campaigns like those conducted by the IRA and CyberBerkut are possible because the distinctions between the state and the private sector in Russia have blurred almost to the point of irrelevance. Particularly under Putin, institutional boundaries have become porous, allowing private citizens and organizations to conduct sanctioned state activities and allowing the state to mine society for autonomous assets to carry out state functions. This is part of a broader trend of deinstitutionalization in Russia, in which the boundaries between private and state, civilian and military, and legal and illegal are quickly disappearing, if they ever existed at all. In Russia this encourages a blending of these institutions in an effort to achieve strategic gains^[10].

Information-technical operations tend to be aimed at more specific targets and involve more

malicious intent than simple psychological operations do. Depending on the sophistication and the strategic aims of a given operation, the organizations carrying out these activities may be associated with or directly a part of Russian intelligence organizations. The intrusions on the Democratic National Committee servers perpetrated by Cozy Bear and Fancy Bear, which are affiliated with the Foreign Intelligence Service (SVR) and the Main Directorate of the General Staff of the Armed Forces of the Russian Federation (GRU), respectively, are only the most well-known and least sophisticated examples of information-technical operations. Much more sophisticated and worrisome is the malware Ouroboros, which, when installed on a network, gives its developers full and covert access to all of the files on that network, and Crash Override, which Wired magazine called “the most evolved specimen of grid-sabotaging malware ever observed in the wild”^[11]. Given their complexity and sophistication, these malware are widely believed to be products of Russian intelligence services.

Operations within the psychological and technical domains exist along a spectrum. On one end are the straightforward information-psychological operations designed to influence opinion. On the other are the malicious information-technical operations that are capable of real-world effects. In between lie operations ranging from covert observation to the exfiltration of information to network control. To be sure, these operations can overlap and influence each other. For example, data exfiltrated in the course of an espionage campaign that uses advanced persistence techniques can, and likely will, be leveraged as part of a psychological operation over time.

III. ORGANIZATIONAL SOPHISTICATION

Russia’s overall domestic hack capacity is relatively high given its emphasis on applied mathematics and computing well prior to college. This, combined with a proliferation of online tools that enable simple attacks like DDoS and website defacement, provide ample opportunity, low resource requirements, and highly permissive environments for low-end, unsophisticated, “flash mob” style disruption. This foundational resource base of potential hack types is part of why Tim Mauer refers to Russia as a country that “sanctions” its proxy hack community in regional engagements in Estonia, Georgia, and Ukraine^[12]. Simultaneously, Russia develops new malware and regularly conducts industrial espionage campaigns as well as cybered operations on physical infrastructures. So how can we meaningfully analyze this elusive and illusive set of agents and behaviors? And what can it tell us about their strategic priorities, risk acceptance, and approaches to cyber operations? From the perspective of defense, cyber-attacks may all appear to blend together. But there are distinct stability and resource costs that separate the technical and the psychological.

While we may not be able to actually identify and count Russia’s hack army, and while we cannot know, with certainty, what zero-day and malicious software will appear in its arsenal, we can think about resources, skills, and platforms. That is, we can ask what organizational

support structures are required to maintain particular lines of effort. The development of advanced malware like Ouroboros and Crash Override requires time, space, and resources. To deploy the malware, an operation needs effective intelligence, higher-level coordination with commander's intent, and political top cover. Assuming that there are dedicated anti-hacking and malware efforts, all elements of complex attacks also need consistent care and feeding in order to produce their intended effect. In this sense, sophistication matters at the organizational level beyond sheer technical savvy.

Organizational sophistication can be thought of as the overall sum of an array of resources, coordination, procedures, and practices^[13]. Highly sophisticated organizations provide individuals with an internal environment that supports consistently clear patterns of function. Those patterns may be tacit or explicit, but they are stable. In particular, we would expect to see a high degree of sophistication in environments where teamwork across different roles is a regular occurrence (both internally but also, potentially, externally).

In articulating this notion of sophistication, we want to be careful to say that we are not attempting to establish any necessary relationship among success, efficiency, and even effectiveness and organizational sophistication. Nor is it the case that a high degree of internal organizational sophistication necessarily means that the organization can coordinate well with other entities. Rather, what we are pointing to is that some kinds of cyber/information operations appear to require more or less organizational sophistication than others. In the Russian case, the organizational sophistication demonstrated appears to break roughly along the range of the psychological and technical aspects of the Russian strategic approach.

A. Information-Psychological

The capacity to conduct broad-based information operations does not in and of itself demonstrate an expansion of an adversary's capabilities. Despite continued journalistic hand-wringing regarding Russian social media and information meddling campaigns^[14], the organizational resources necessary to sustain behaviors like those exhibited by the IRA, let alone CyberBerkut, are decidedly shallow. That is, the necessary skill and sophistication level of these entities need not be particularly high to make these groups disruptive. To even refer to their social media activities as "hacking" is an abuse of the term. Using false messages to disrupt publics isn't even social engineering (hacking the human rather than the machine to bypass security)^[15]. In short, as any two-year-old can demonstrate, it does not take much skill or sophistication to break things.

However, this lack of sophistication may also result in high resiliency against efforts to stop or defeat the operations. As a question of skill and resources, media disinformation is not a complex endeavor in the contemporary era. Creating faux content and performing the relatively mindless work of creating fake accounts to generate clicks are labor that requires, at best, some degree of ability in the target country's language and a terminal connected to the internet. Thus, organizations which produce disinformation as a state-sponsored service have

no necessary need to establish long-term internal stabilization structures.

From what we know of the IRA's fly-by-night structure, the work was seasonal at best: the organization used ad-hoc hiring practices and a willingness to corral and pay the labor^[16]. For this, a regime can easily outsource the work – as it did with Yevgeny Prigozhin, the Kremlin-linked oligarch and former hot dog salesman in St. Petersburg^[17]. As we have already noted, Prigozhin bankrolled the IRA by using a portion of the billions of dollars provided by the Russian Government for food service for the military. The stability of such funding can ebb and flow as strategic need dictates. With low technical barriers to entry, the labor pool is deep and personnel need little training or support. In the Russian case, this simply amounts to an ability to write, click, or elevate noxious messages on already user-friendly platforms like Twitter and Facebook.

Similarly, “patriotic hackers” with high prestige levels, like CyberBerkut, wade in markedly unsophisticated waters, both technologically as well as organizationally. Generally, groups like these are the most loosely affiliated with state efforts. Patriotic/hacktivist agents' capabilities require little-to-no coordination beyond what Tim Mauer defines as sanctioning – the permission to operate against a regime's adversaries^[18]. Certainly, the group has garnered global notoriety for successfully blocking public access to a few German Government websites in 2015 and, more recently, “leaking” unverified documents linking Ukrainian political leaders and laundered funds to Hillary Clinton's 2016 campaign^[19]. Nevertheless, TrendMicro's analysis of the group's membership and internal squabbling dynamics provides unexpected levity.

According to data previously available on Pastebin in 2015, the menace known as CyberBerkut has at least 4 active members ranging in age from 24 to 38 years-of-age. The group's most active member is “Mink,” who also goes by the name “Zac Olden.” Mink previously set up a fake website intended to mimic a legitimate online store that sells Australian (specifically Tasmanian) jewelry beads^[20]. Mink was also the leader of “retribution network,” the site for which has lagged or gone offline entirely, as has his previous fake site. The group's instability became clear in 2014 when a fallout between Mink and two other members led Mink to “doxx” his own colleague's “MDV” and “artemova” in Pastebin posts. Later, in October of 2014, after the apparent doxxing, a second CyberBerkut Twitter account, “@cyberberkut2,” was created.

The misalignment and frequent interruptions of the group's activities, coupled with its relatively weak technical capacity, reveal a high-prestige group with no reliable resources, stability, or real infrastructure. Its stop-start net presence and hacking behavior suggest a tiny membership footprint with limited support. If CyberBerkut can be called an organization, it is one with a nearly immeasurably small level of sophistication. While we do not doubt that there may be pro-Russian hacking groups with greater degrees of organizational complexity, this one serves as a reminder of the limitations and ephemeral nature of the volunteer group dynamic.

We should note here that while the proxy work of the IRA and CyberBerkut offer the Russian Government a certain level of deniability, the risk in exploiting these actors is that the more deniability they have, the less control the government has over their activities. This may result in unsanctioned operations which are carried out for narrow, parochial reasons instead of national strategic gain, but which may nevertheless be destabilizing. Further, the fractious nature of an organization like CyberBerkut makes it an unreliable proxy for the government. Because Moscow emphasizes deniability over control in these operations, the likelihood of these actors conducting operations that aggravate their tacit supporters is higher than if they were under strict government oversight.

Overall, it appears that Moscow has assessed a relatively low risk of reprisal from information-psychological measures and low-level technological operations like DDoS attacks. Reliance on cheap, unsophisticated proxies such as the IRA and CyberBerkut carries, despite the state's tenuous control, almost no risk. Sanctions imposed on individuals like Prigozhin, (whose reaction was a shrug and a "Now I'll stop going to McDonald's") and the declaration of a few Russian intelligence officers in the U.S. as persona non grata (and whose positions may by now have already been backfilled) impose almost no cost. There is almost no serious consequence in response to these activities, demonstrating that there is likewise almost no strategic risk taken on by Moscow in its use of proxies to conduct information-psychological measures.^[21]

B. Information-Technical

In contrast to the organizational simplicity of Russian information psychological operations, the Russian approach to technical operations shows evidence of a much deeper bench of cyber agents that demonstrate team-based technical collaboration in design, execution, and support. In other words, there is likely a highly sophisticated organization (or a number of them) in the background – a system with consistent resources, stability of platform, and continuity of personnel with role-specific skill sets. The Fancy Bear and Cozy Bear hacking teams are two well-known examples of long-term, malicious agents that conduct technically sophisticated attacks globally. But, more importantly, any advanced persistent threat (APT) group is a likely suspect for high organizational sophistication, given its emphasis on long-term operations and continued curation of new potential targets. Regarding the APT attacks attributed to Russia, it may be less important to discern which Russian hacking team is responsible for a particular attack ^[22] than it is to ask whether the attacks themselves suggest that a sophisticated organization is behind them.

To wit, the espionage tool kit named "Ouroboros" ("Turla" or "Snake") and the industrial control system malware "Crash Override" ("Industroyer"), which appeared in 2016, are two of the most advanced pieces of malware to have emerged in recent years. Both cases suggest long-term planning, support, and dedicated development of breach and exploit processes.

Russian meddling in secure government systems and critical infrastructure attacks through the development of sophisticated malware are consistent components of the Russian technical approach. Ouroboros' evolutionary roots date well prior to its February 2014 christening in media coverage of the Ukraine attack during the ouster of Viktor Yanukovich^[23]. Ouroboros stands as one of the longest-running continuously evolving malware platforms of its kind. As early as 2006, security research firms obtained malware samples known generically as "Agent.BTZ." Agent.BTZ has been found on U.S. Government military systems as well as other military systems globally. In the private sector, as firms individually dissected and traced the malware, they began to give the generic label their own names, including "Snake," "Sengoku," and "Snark"^[24]. Ouroboros' meagre roots evolved over time into a highly sophisticated attack system that continues to plague government and industry alike. Ephemeral and less professional groups are unlikely to maintain this level of fortitude in sustaining the evolution of this malware.

In 2016, Crash Override infrastructure attacks on Ukrainian electrical grids were not in themselves particularly noteworthy. After all, the Ukrainians have been suffering electrical grid attacks leveraged by Russian attackers since 2015, and the Ukrainian electrical grid is supported by a series of analog backups, so the damage was more limited^[25]. What was noteworthy about Crash Override was that the attack platform was modular. That is, the malware was specifically constructed so that it could be adapted to other systems, not simply Ukrainian electrical systems^[26]. Orchestrating an attack on a power grid need not require any particular level of organizational sophistication. Designing malware that can be adapted to future conditions and attacks speaks to long-term planning, persistence, and flexibility at a minimum, and the opportunity to experiment with the tools elsewhere and in other contexts^[27].

Another potential indicator of sophistication that is specific to cyber operations is the emergence of "false flag" operations – the emulation of tactics, techniques, and procedures (TTPs) of another malign actor in order to pin an attack on them. The Olympic Destroyer attack disabled critical Olympics information technology systems and left behind a forensic signature that mimicked that of the North Korean hacking team Lazarus Group^[28]. It is one thing to copy code, but another entirely to know another agent so well that you attempt to mimic its TTPs. This also suggests that the attackers actively analyze the behaviors of other threat actors operating in this domain. Though attribution to a specific Russian ATP is open to debate, political analysts argue that the timing of the false flag attack strongly aligns with Russian sentiments^[29]. Security experts at Kaspersky also indicate that the attacker who perpetrated the Olympic Destroyer attack held its capacity in reserve, suggesting that the group may be withholding its capacity for another attack in the future^[30]. The false flag operations and the holding of capacity in reserve suggest an organization that intends to persist and continue operations into the future.

The available evidence is scant, but it appears that Russian political leadership may believe that these more advanced technical operations carry much greater strategic risk. If this is true, tighter state control of a more sophisticated organization than CyberBerkut, for example, would be merited. Grid hacking malware could result in the deaths of foreign citizens, especially the more vulnerable aged and infirm. Operating covert malware designed to exfiltrate information or take over systems requires professional espionage tradecraft measures. If these cybered espionage measures were directly attributed to Russia or if the Russian Government were to lose control of these capabilities, the blowback would potentially be enormous. Operating such sophisticated programs may force a reliance on more professional, and professionalized, organizations, such as the GRU's Fancy Bear and the SVR's Cozy Bear. Embedding these programs deeply in Russia's intelligence establishment, therefore, allows for better risk management and more reliable and consistent, evolving operations, while still maintaining a level of deniability.

All of these agents, attacks, and malware demonstrate clear evidence of high levels of organizational sophistication. They require strategic leadership; political cover; consistent funding; stable platforms; skilled technicians; and the kinds of resources that point to concerted, clear efforts by Russian organizations to move competition in the cyber domain farther than the far more simplistic information-psychological operations can.

IV. IMPLICATIONS AND FUTURE RESEARCH

How can organizational sophistication analyses matter to U.S. national security policy – particularly in a time when the leading stories of the year are almost entirely about cheap, low-cost, disruptive information operations? Thinking about organizational sophistication redirects our thinking away from the “weapon” and toward a state's intentional development and maturation of capabilities. To be clear, while information operations can and likely do have effects, the Russian case demonstrates where stability, control, and funding are prioritized. The intentional development of a highly-skilled set of hacking crews who can both breach and exploit U.S. systems is consistent with behaviors we would expect to be deployed in both peacetime and wartime efforts. This distinction may matter when a nation is working through responses to cybered operations – namely, which aspects of Russian-supported operations the United States should consider as offensive actions that necessitate offensive counters, and which operations fall below such triggers and necessitate domestic resilience-building measures. In brief, it may help draw clearer conclusions as to who should respond and how.

While it may not be the case that organizational sophistication necessarily breaks along the psychological/technical divide, the case here is that it does. The military and the intelligence community are traditionally tasked with addressing the damage wrought by technical attacks that produce physical effects or result in the loss of national security secrets, but these government organizations cannot do so in response to all attacks. Conversely, it remains unclear just exactly how or why a bot campaign run prior to an election necessitates a response via

offensive operations. However, the sophistication of Russian information-technical operations demonstrates some degree of measurable and documentable, political intent. Particularly, the longer timelines of operations with similar patterns of behavior in a coordinated cyber campaign make it justifiable to conduct counteroffensive and even offensive operations.

Conversely, those operations that lack organizational sophistication also demonstrate a lower capacity for traceable direct mechanisms, lower commitment to sustained effort, and less direct control by a regime. Under such conditions, the response should be internal rather than offensive. That is, in the absence of clear, long-term organizational development by an adversary, the mechanism for security may be increased domestic regulation of social media platforms, creating more resilient communications networks, and investing resources in civilian cyber education and hygiene. This is not to say that such information operations do not pose a fundamental threat to the Nation and its democratic processes. If the proposed mechanisms and their effects in disrupting democracy are found to be effective, then these operations certainly do pose such a threat. But the degree to which this is a concern for foreign operations by the military and the intelligence community must be much more aggressively clear than is the case currently.

The genuine concern, in the eyes of the authors, in the case of Russia should be for technical operations, not simply because the technological sophistication levels are high, but because the organizational requirements to maintain the style and methods demonstrated in the most recent Russian attacks on Ukrainian infrastructure suggest tight coordination and planning that only a sophisticated organization can provide. Specifically, there is sufficient evidence both in the orchestration of attacks as well as in the platforms and resources utilized to necessitate stable, consistent organizational structures that endure over time. That is, the discernment of the distance from or the nature of the relationship to the state may be more important in understanding the strategic goals and possible persistence of these activities than direct identification of who is employed by, sponsored by, or even permitted to act as part of the approach.

Furthermore, less sophisticated information-psychological operations may be more resilient and more resistant to measures designed to defeat them. Information-psychological efforts draw on a massive labor pool and an informal network, so efforts to defeat them at the source are mere games of whack-a-mole, and efforts to defeat them at home run the risk of becoming dangerously undemocratic. This being the case, the investment in researching and countering these operations, particularly in terms of thinking offensively, may not be a wise one. Government and social media corporations can and should be vigilant, calling out and removing disinformation efforts, but disinformation and low-level harassment campaigns are ultimately almost impossible to eliminate. The only other option may be in developing means to spread truthful information and news to local populations in Russia. The United States has apparently made a policy decision to avoid this, despite the fact that it does so in places like Iran, North Korea, and elsewhere^[31]. Finally, information disruption through online social media campaigns

is poised to become an even more common endeavor since the cost is so low. We have already seen numerous efforts, not simply by states, but also by rebel groups and terrorist organizations, to drive and influence via these platforms. If we have not already witnessed it, we will increasingly see the rise of “the rest” – of small states and non-state actors making these platforms even noisier^[32].

In summary, it is the opinion of the authors that research can and should focus on understanding the strategic goals, structure, resources, and ideas specifically tied to Russian information-technical operations. It is our opinion that the psychological component is not only more difficult to control as a function of offensive or non-domestic efforts, but that there is not anything particularly unique about the ability to influence populations through social media. Thus, the psychological efforts are likely to be leveraged by weak and strong adversaries both symmetrically and asymmetrically^[33]. The general noisiness of such low-end efforts makes understanding the unique lines of Russian effort more difficult. In contrast, following the resource and stability needs of mature technical efforts would likely yield more meaningful, specific insights as pertains to Russia-specific concerns.

This is not to suggest that U.S. agencies should match or mirror Russian efforts per se. But a clear-eyed assessment of where and just how much resourcing is being directed by an aggressive adversary can help shape our own policies regarding where and how our strategic trade-offs are positioned. Specifically, the current paralysis exhibited by the Department of Defense’s counters to Russian cybered moves is partially about which moves should be understood as heartburn, and which as a heart attack. We have posited here that more clarity between these actions should rest on the sophistication of the organization that underlies the action, rather than the activity itself. In this way, the United States and its partners will be able to develop and ensure that standards are met for hardening critical infrastructure against cyber intrusions and attacks with an eye toward risk management, rather than seeking the unattainable goal of 100-percent security. To be certain, much of this effort is currently left to the private sector to manage. In addition, a better understanding of the organizational structure behind malicious technical operations, their purpose, their motivation, and their intended effect would allow us to develop deterrence measures as well as timely and appropriate responses to attributable attacks. 🛡️

NOTES

1. Valery Gerasimov, "Znachie nauki v prognozirovanii," ["The value of science in prediction"], *Voenno-promyshlennii kur'er* [Military-Industrial Courier], Feb. 27, 2013.V.N. Gorbunov and S.A. Bogdanov, "Armed confrontation in the 21st century," *Military Thought*, 1 (2009), 21-22. Emphasis in original.I. S. Jacobs and C. P. Bean, "Fine particles, thin films and exchange anisotropy," in *Magnetism*, vol. III, G. T. Rado and H. Suhl, Eds. New York: Academic, 1963, pp. 271–350.
2. Paul, Christopher, and Miriam Matthews. "The Russian "Firehose of Falsehood" Propaganda Model." RAND Corporation (2016). Accessed September, 17, 2018. https://www.rand.org/content/dam/rand/pubs/perspectives/PE100/PE198/RAND_PE198.pdf.
3. A.V. Kartapolov, "Uroki voennykh konfliktov, perspektivy razvitiya sredstv i sposobov ih vedeniya. Prjamyje i neprjamyje dejstvija v sovremennykh mezhdunarodnykh konfliktah," ["Lessons of military conflicts, prospects for the development of means and methods for delivering them, direct and indirect actions in contemporary conflicts,"], *Vestnik Akademii Voennykh Nauk* [Bulletin of the Academy of Military Science] 9 (2015), 2. See also Timothy Thomas, "The Evolution of Russian Military Thought: Integrating Hybrid, New Generation, and New Type Thinking," *Journal of Slavic Military Studies*, Vol. 29, No. 4 (2016).R. Nicole, "Title of paper with only first word capitalized," J. Name Stand. Abbrev., in press.
4. A.V. Serzhantov and A.P Martoflyak, "Modern military conflicts," *Military Thought*, 2 (2009), 88.M. Young, *The Technical Writer's Handbook*. Mill Valley, CA: University Science, 1989.
5. Vladimir Putin, "'Soldat est' zvanie vysokoe i pochetnoe' [The rank of 'soldier' is honorable and respected], Excerpts from the Annual Address to the Federal Assembly of the Russian Federation," *Krasnaya Zvezda* [Red Star], May 11, 2006.
6. Defense Intelligence Agency, *Russia Military Power: Building a Military to Support Great Power Aspirations* (Washington, DC: Defense Intelligence Agency, 2017).
7. Timothy, L. T. "The Russian Understanding of Information Operations and Information Warfare." *The Information Age Military*. Accessed September, 17, 2018. <http://www.au.af.mil/au/awc/awcgate/ccrp/thomas.pdf>.
8. Bing, Chris, "Hacker group 'CyberBerkut' returns to public light with allegations against Clinton," *Cyberscoop*, undated, <https://www.cyberscoop.com/cyberberkut-returns-hillary-clinton/>, accessed July 25, 2018.
9. Mark Galeotti, *Hybrid War or Gibrinaya Voina? Getting Russia's Non-Linear Challenge Right* (Mayak Intelligence: Prague, 2016), 48-50.
10. G. Data Security Labs, *Uroburos: Highly complex espionage software with Russian roots*, https://public.gdatasoftware.com/Web/Content/INT/Blog/2014/02_2014/documents/GData_Uroburos_RedPaper_EN_v1.pdf, accessed July 25, 2018.
11. Maurer, Tim. *Cyber Mercenaries*. Cambridge University Press, 2018.
12. There exists a robust literature in business management whose route of inquiry traces organizational structure and its outcomes on production. While there are clearly differences between the factors that produce successful corporate structures, and effective government organized/sponsored entities, the core insights align reasonably well enough to carry the concept over. See for example: Teece, David J. "Strategies for managing knowledge assets: the role of firm structure and industrial context." *Long range planning* 33, no. 1 (2000): 35-54.
13. Bing, Chris. "Russian hacker group 'CyberBerkut' returns to public light with allegations against Clinton." *Cyberscoop*. July 12, 2017. Accessed July 25, 2018.
14. Honan, Mat. "Social Engineering Always Wins: An Epic Hack, Revisited." *Wired*. June 03, 2017. Accessed July 25, 2018. <https://www.wired.com/2014/01/my-epic-hack-revisited/>.
15. Chen, Adrian. "What Mueller's Indictment Reveals About Russia's Internet Research Agency." *The New Yorker*. February 20, 2018. Accessed July 25, 2018. <https://www.newyorker.com/news/news-desk/what-muellers-indictment-reveals-about-russias-internet-research-agency>.
16. "Inside the Internet Research Agency's Lie Machine." *The Economist*. February 22, 2018. Accessed July 25, 2018. <https://www.economist.com/briefing/2018/02/22/inside-the-internet-research-agencys-lie-machine>.
17. Maurer, Tim. *Cyber Mercenaries*. Cambridge University Press, 2018.
18. Bing, Chris. "Russian Hacker Group 'CyberBerkut' Returns to Public Light with Allegations against Clinton." *Cyberscoop*. July 12, 2017. Accessed July 25, 2018.
19. "Hacktivist Group CyberBerkut Behind Attacks on German Official Websites - TrendLabs Security Intelligence Blog." *Simply Security News, Views and Opinions from Trend Micro, Inc*, 21 Jan. 2015, blog.trendmicro.com/trendlabs-security-intelligence/hacktivist-group-cyberberkut-behind-attacks-on-german-official-websites/.

NOTES

20. Reuters, "Russian businessman Prigozhin dismisses new U.S. sanctions: RIA," March 15, 2018, <https://www.reuters.com/article/us-usa-russia-sanctions-prigozhin/russian-businessman-prigozhin-dismisses-new-u-s-sanctions-ria-idUSKCN-1GR2G7>, accessed July 27, 2018.
21. GREAT. "OlympicDestroyer Is Here to Trick the Industry." Securelist - Kaspersky Lab's Cyberthreat Research and Reports, Kaspersky, 8 Mar. 2018, securelist.com/olympicdestroyer-is-here-to-trick-the-industry/84295/.
22. Jones, Sam. "Cyber Snake Plagues Ukraine Networks." Financial Times. March 07, 2014. Accessed July 28, 2018. <https://www.ft.com/content/615c29ba-a614-11e3-8a2a-00144feab7de>.
23. "The Snake Campaign." BAE Systems | International. January 2016. Accessed July 28, 2018. <https://www.baesystems.com/en/cybersecurity/feature/the-snake-campaign>.
24. Zetter, Kim. "Inside the Cunning, Unprecedented Hack of Ukraine's Power Grid." Wired. June 03, 2017. Accessed July 27, 2018.
25. Brocklehurst, Katherine. "CRASHOVERRIDE – First Malware Platform Designed to Take Down Electric Grids." Belden. October 9, 2017. Accessed July 27, 2018. <https://www.belden.com/blog/industrial-security/crashoverride-first-malware-platform-designed-to-take-down-electric-grids>.
26. Greenberg, Andy. "How An Entire Nation Became Russia's Test Lab for Cyberwar." Wired. April 13, 2018. Accessed July 27, 2018. <https://www.wired.com/story/russian-hackers-attack-ukraine/>.
27. Jackson Higgins, Kelly. "Olympic Destroyer's 'False Flag' Changes the Game." Dark Reading, Information Week, 8 Mar. 2018, www.darkreading.com/attacks-breaches/olympic-destroyers-false-flag-changes-the-game/d/d-id/1331222.
28. Greenberg, Andy. "'Olympic Destroyer' Malware Hit Pyeongchang Ahead of Opening Ceremony." Wired. February 22, 2018. Accessed July 28, 2018. <https://www.wired.com/story/olympic-destroyer-malware-pyeongchang-opening-ceremony/>.
29. GREAT. "OlympicDestroyer Is Here to Trick the Industry." Securelist - Kaspersky Lab's Cyberthreat Research and Reports, Kaspersky, 8 Mar. 2018, <https://securelist.com/olympicdestroyer-is-here-to-trick-the-industry/84295/>.
30. Thomas M. Hill, "Is the U.S. Serious About Countering Russia's Information War on Democracies?" Brookings Institution, <https://www.brookings.edu/blog/order-from-chaos/2017/11/21/is-the-u-s-serious-about-countering-russias-information-war-on-democracies/>, accessed July 28, 2018.
31. Limbago, Andrea "Smaller Nation State Attacks: A Growing Cyber Menace." Threatpost | The First Stop for Security News. July 18, 2018. Accessed July 28, 2018. <https://threatpost.com/smaller-nation-state-attacks-a-growing-cyber-menace/134061/>.
32. Ibid.

Beyond the United Nations Group of Governmental Experts

Norms of Responsible Nation-State Behavior in Cyberspace

Major General (Ret.) John A. Davis
VP, CSO (Federal) Palo Alto Networks

Charlie Lewis
MAJ, U.S. Army Reserve

While the September 2015 meeting between President Xi of China and President Obama of the United States seemed like a tipping point for norms in cyberspace, the United Nations Group of Governmental Experts (UNGGE) has been developing a useful set of norms for responsible conduct among nations in cyberspace for years. Although consensus was difficult to establish along the way, as it almost always is between nations, the Xi-Obama meeting started the process of establishing a broader agreement on a set of norms that was later endorsed by the Group of Seven and Group of 20. The endorsed norms followed previous agreements and focused on information sharing, cooperation, protection, and avoiding malicious activities within a state's borders, as well as human rights violations. States were to avoid using their territory for attacks against technologies or critical infrastructure, abstain from disrupting supply chain security, and refrain from using cyber means to harm other states. However, the UNGGE norms effort wavered during 2017 when several key countries backed away from the original agreement for a variety of reasons ranging from inability to enforce it to concerns around its effect on future operations.

Despite the struggles of previous norms efforts, opportunities exist to reframe norms around peacetime activities. This paper proposes five peacetime norms of behavior that responsible nation-states should strive to achieve. Responsible nation-states are those that act rationally, participate in other international norms and organizations, and have not demonstrated violations of other nations' sovereignty. The five proposed norms are designed to accomplish the following objectives:

- 1) Contribute to an improved, common, international understanding at the technical, operational, and policy levels of cyberspace activities
- 2) Reinforce positive and careful control and oversight of cyber activities
- 3) Bring additional responsible partners to the effort in more effective ways
- 4) Reduce risks and chances of misinterpretations that lead to mistakes and escalation

The following sections define each norm, provide examples, and discuss opportunities for implementation.

NORM #1

Responsible nations should be more transparent about what they are doing in cyberspace and why they are doing these things.

Applicable to law enforcement; homeland security; and, especially, the militaries of responsible nations, the goal of this norm is to increase transparency, not establish total transparency. If the majority of nations' actions were transparent, this would lead to greater trust and improve cooperation and teamwork on issues of common interest. To increase transparency, a responsible state can take actions that range from announcing the development of cyber forces to publishing a cyber strategy and overall goals. Law enforcement and homeland security can also discuss prohibited activities against which they protect. Increased transparency, however, is not a requirement for, or even within, an intelligence agency's DNA, which is why these organizations are excluded from this norm.

A previous example of increased transparency is the development of coalitions to address conflict, as was done in response to Saddam Hussein's invasion of Kuwait. The international community witnessed an illegal act, established transparency regarding objectives, and eventually launched a counter-invasion to free Kuwait. The United States spoke openly about the creation and structure of its cyber force and demonstrated when it was operational. The U.S. military distributed white papers about the establishment of the Cyber Mission Forces under U.S. Cyber Command and each of the Service Cyber Component Commands and briefed not only government and military partners of the U.S. around the world, but also countries such as Russia and China. These papers and briefings included information about the force composition, its purpose, its missions, and how it would be accountable and controlled by responsible oversight. Furthermore, the U.S. military publicly declared that it was conducting cyber operations against the Islamic State of Iraq and Syria in 2016. While not disclosing any classified information, these efforts demonstrated the U.S. military's increased transparency with not only other partners, friends, and allies around the world, but also competitors and potential adversaries.

Transparency, however, can be a hard goal to achieve. Typical norms, like maritime and space law, were derived by consolidating years of mutual activities and laws. They were built after years of documented and understood conduct; this was not the case with cyber norms. Moreover, for transparency norms to succeed, major actors need to participate, which is unlikely. Despite these concerns, one dynamic that makes increased transparency possible is the increasingly lower bar for classification of all things related to cyber. There are open, even public discussions today that simply could not have occurred only a few years ago. Additionally, recent public examples of greater transparency in threat attribution include the North Korean

attack against Sony Pictures Entertainment; the Iranian distributed denial-of-service attack on the U.S. financial sector; and, most recently, Russian interference in the 2016 presidential election. There is a good reason to increase clarity, accuracy, and transparency by bringing these activities into the light of law enforcement; domestic security; and, especially, uniformed military operations to contribute to a reduction in uncertainty and an increase in stability.

NORM #2

Responsible nations should establish and enforce standardized procedures for effective oversight of military, law enforcement, and homeland security cyber operations.

Standards for bureaucratic oversight provide the layers of decision-making to ensure that norms and other requirements are met in cyberspace. Furthermore, procedural oversight includes risk management assessment and control procedures that contribute to the following five effective outcomes.

- 1) **First** is domestic and foreign policy oversight from a competent authority as established by the nation so that adequate consideration is given to the potential impact on both domestic and foreign reactions to the implementation of a cyber activity if it is discovered.
- 2) **Second** is technical oversight, which includes a “technical gain versus loss” assessment to address the unintended consequences resulting from the discovery of the technical capability and its use against other targets or the nation that used it in the first place. In addition, this is also a “technical assurance assessment”, which provides low, medium, and high assurance levels that the capability will produce technical outcomes or effects as intended and not produce unintended consequences, such as escalation or cascading effects.
- 3) **Third**, operational oversight with appropriate responsibilities, accountability, and command and control procedures that verify positive control within an authorized chain of command reinforces these risk management processes.
- 4) **Fourth** is intelligence oversight, including an “intelligence gain versus loss” assessment, which provides the consequences of exposure to and potential loss of intelligence sources and methods and the resulting insight if the cyber operation or capability is discovered or revealed.
- 5) **Fifth** is legal oversight, including two types of legal review that provide an assessment for both the capability and the operation as it applies to either the International Law of Armed Conflict or other applicable domestic and international laws and agreements.

Responsible nations applied these oversight norms during the post-Cold War era and trusted others to do the same. Nuclear treaties, the law of armed conflict, and an understanding about the effect of their use has resulted in a minimal threat from responsible nations, and may also explain why the international community signed a treaty to prevent Iran from developing its own nuclear weapons. Oversight for cyber operations is much more difficult to ascertain. While

the United States lays out its various legal codes in its military cyberspace manual, Joint Publication 3-12, it is still looking to adjust the approval process for cyberspace operations. Other nations as well may have different sets of controls over their cyberspace operations during peacetime, as made evident by the Chinese use of civilian hackers.

Many believe this norm should apply to intelligence operations as well. Notably, most nations' significant cyber capabilities began within their own national and military intelligence organizations for the purpose of espionage. In many cases, the reckless use of intelligence cyber activities can significantly complicate the cyber environment, making it increasingly difficult to determine intentions, and can lead to misperceptions, miscalculations, and mistakes in cyberspace that might "spill over" into the physical world in an unwarranted escalation. There is definitely a case to be made for addressing espionage activities in cyberspace within the norms discussion. However, perhaps the topic of intelligence cyber operations and activities is something to be addressed separately due to the likelihood that its inclusion in an open discussion would significantly complicate nations' ability to make progress.

NORM #3

Responsible nations should share cyber threat intelligence on criminal and terrorist threats of common interest.

Information sharing and alerting about terror threats and large criminal operations is standard amongst states. Within cyberspace, however, there is much less openness, as it could potentially give away operations. Instead of withholding information, responsible nations should establish and enforce effective information sharing programs and platforms that are automated and format-standardized to account for the speed and scale of today's modern criminal and terrorist cyber threats. These cyber threat intelligence and information sharing programs should be focused on cyber threat indicators of compromise along the cyber threat life-cycle steps as well as contextual information. However, a certain level of sanitization is required. These reports should not include personally identifiable information; protected health information; intellectual property content; or other types of information that create surveillance, privacy, and liability issues. Cyber threat information sharing should be done government-to-government through appropriate diplomatic, law enforcement, domestic security, intelligence, and military channels. In addition, responsible nations should encourage sharing programs and platforms between government and industry and among industry entities as appropriate to national and international laws and agreements. The result of increased and effective information sharing as described is to help reduce the "noise-to-signal" ratio so that responsible nations are able to better focus on what is important and not be confused or distracted by the ever-increasing amount of cybercriminal and terrorist activity that might cloud an already confusing cyber landscape and contribute to misinterpretation, miscalculation, mistakes, and inadvertent escalation.

This norm currently exists in the signals intelligence world under the United Kingdom-

United States of America agreement among the United States, the United Kingdom, Canada, Australia, and New Zealand. Established to codify information sharing principles that occurred during World War II, the agreement leveraged that success to create an information sharing practice between the British Empire and the United States. The agreement not only shows how effective information sharing occurs, but also demonstrates how to adapt it for new technologies, as the partnership still exists today.

Opponents of information sharing rely on the same argument that proponents of transparency do—providing information may give away trade secrets or cause malicious state actors to change their methods to avoid capture. In addition, the example cited is the result of success in World War II and occurred during a time of liberal institutional growth and trust. Today, however, a lack of the same trust is more evident, causing some to question the agreement's effectiveness. The U.S. Cybersecurity Information Sharing Act of 2015, which attempted to reduce these concerns, demonstrated an increase in the collective ability to chase down common enemies and reduce noise in cyberspace.

NORM #4

Responsible nations should encourage and incentivize increased industry participation in the development and enforcement of these and additional norms of responsible behavior in cyberspace.

Industry owns, operates, and maintains the vast majority of the underlying infrastructure and technology of cyberspace, yet the norms discussion has traditionally involved government only, as in the case of UNGGE. Industry's involvement would make the norms more practical and effective, partly because industry better understands the role that government should play in the digital environment. Many contentious issues today, such as mandatory backdoors for law enforcement, counterterrorism, and intelligence purposes; restriction of cross-border data flows; private-sector hack back; and supply chain risk management warrant industry's involvement. The Australian Strategic Policy Institute has done some excellent research on a greater role for industry in the development of cyberspace norms, highlighting the success of the United States' consortium while developing a structure for trusted information flow within Australia. Additionally, the Carnegie Endowment for International Peace has taken a detailed look at how to more effectively apply norms that could impact global stability in financial markets and the international monetary system by not manipulating or damaging financial institutes' data. Many companies have taken positions on the technology industry's role in cyberspace norms, and some have attempted to join the cause to establish greater protection from cyber threats.

Global incentives and trust can be difficult to form. Sharing ideas and secrets in a transparent manner can create opportunities for malicious actors to conduct reconnaissance. A violation of this trust or even the perception of a lack of trust may end any cooperation between international industry and government.

NORM #5

During peacetime, responsible nations should NOT deploy loosely controlled third-party actors and organizations to engage in cyber activities.

The use of surrogates, front companies, “technical research” organizations, criminal entities, moonlighters, and even patriotic hackers limits government control over actions and can violate the transparency and trust created by the previous four norms. These types of actors and organizations increase uncertainty, reduce stability, and lack the oversight and control discussed in norm #2. They are driven by an assortment of high-risk motivations and increase the chance of a miscalculation in attribution, as described in norm #3, which could result in an unacceptably high risk of escalation, especially during times of high tension. The prevention of the use of these actors increases the likelihood of the other norms succeeding. Unfortunately, the world has seen the increased use of loosely controlled third-party entities by nation-states. This is an alarming trend because the risk of a mistake happening or an unsanctioned action being perpetrated by someone with a personal grievance is growing exponentially, and all responsible nations should share a common interest in preventing these events from occurring.

The above norms of responsible nation-state behavior in cyberspace, supported by the increased involvement of global industry, are designed to accomplish improvements to contribute to an improved international understanding, reinforce positive and careful control and oversight of cyber activities, and more effectively encourage the participation of responsible partners. However, questions remain about the degree to which these norms are feasible. The U.S. Government and an increasing number of U.S.-based, private-sector cybersecurity companies not only think that the norms will work, but are increasingly and actively pursuing each of norms proposed in this paper. The U.S. military has already led the way on the first two proposed norms. Additionally, the U.S. Congress focused its Cyber Information Sharing Act of 2015 on the third and fourth norms, and U.S. law enforcement, domestic security, intelligence, and even military organizations are implementing many cyber threat intelligence and information sharing programs with an increasing number of international and industry partners. The United States is leading by example in the effort to establish norms of responsible behavior. The United States should be willing to engage with other great nations to broaden this effort, make these norms an international standard, and improve upon them in a progressive manner. 🇺🇸

NOTES

1. Garrett Hinck, "Private-Sector Initiatives for Cyber Norms: A Summary," *Lawfare*, June 25, 2018.
2. The Department of Defense, *The DoD Cyber Strategy*, April 2015.
3. James Van De Velde, "Why Cyber Norms are Dumb and Serve Russian Interests", *The Intercept*, June 6, 2018.
4. Department of Defense, *Joint Publication 3-12*, Department of Defense, June 2018.
5. Guest Blogger for Net Politics, "When China's White-Hat Hackers Go Patriotic," *Council on Foreign Relations*, retrieved from <https://www.cfr.org/blog/when-chinas-white-hat-hackers-go-patriotic>.
6. Excluding the five eyes consisting of the United States, Great Britain, Canada, Australia, and New Zealand.
7. UKUSA Agreement Release 1940-1956, retrieved from <https://www.nsa.gov/news-features/declassified-documents/uku-sa/> on 10/9/2018.
8. Van De Velde.
9. Liam Neveill, "Cyber Information Sharing: Lessons for Australia", *Australian Strategic Policy Institute*, May 2017.
10. Tim Maurer, Ariel Levite, George Perkovich, "Toward a Global Norm Against Manipulation the Integrity of Financial Data," *Carnegie Endowment for International Peace*, March 27, 2017.
11. Hinck.

SESSION

◆ 4 ◆

A Model for Evaluating Fake News

Dr. Char Sample

*ICF Inc., L.L.C.
Columbia, MD*

Dr. Connie Justice

*Purdue School of Engineering & Technology
Indiana University - Purdue University
Indianapolis, IN*

Dr. Emily Darraj

*Cyber Security Department
Capitol Technology University
Laurel, MD*

ABSTRACT

“Fake news” (FN) is slowly being recognized as a security problem that involves multiple academic disciplines; therefore, solving the problem of FN will rely on a cross-discipline approach where behavioral science, linguistics, computer science, mathematics, statistics, and cybersecurity work in concert to rapidly measure and evaluate the level of truth in any article. The proposed model relies on computational linguistics (CL) to identify characteristics between “true news” and FN so that true news content can be quantitatively characterized. Additionally, the pattern spread (PS) of true news differs from FN since FN relies, in part, on bots and trolls to saturate the news space. Finally, provenance will be addressed, not in the traditional way that examines the various sources, but in terms of the historical evaluations of author and publication CL and PS.

Keywords—fake news; computational linguistics; pattern spread; provenance; trust

I. INTRODUCTION

The term “fake news” (FN) was officially ushered into the lexicon when the Oxford Dictionary added the term in 2017^[1]. While the term is frequently used and definitions vary, the problem of deceptive data is serious and exposes a profound and underlying flaw in information and network security models. This flaw is trust in entities without verification of the content that they exchange.

“Trust but verify”^[2] is an old proverb that, until recently, resulted in trust at the expense of verification. Trust in journalists historically resulted from the reputation of the journalist as well as the news organization (publisher). However, publisher reputations of news organizations can vary widely, and the line between news and entertainment continues to

© 2019 Dr. Char Sample, Dr. Connie Justice, Dr. Emily Darraj

blur^[3]. The journalistic integrity of news organizations, while an interesting discussion, is not the focus of this effort; however, defining, measuring, and characterizing fact-based news is.

Our historical method of placing trust in reporters and news organizations is under attack^[4]. When a reporter can be discredited for \$50,000^[4] and a news story can be staged for \$200,000^[4], the facts within their context must be preserved and protected. Protection begins with understanding of the value asset that is to be protected. In the case of news, the assets include the story (data and metadata) as well as the reporter and publisher.

Insurance companies rely on statisticians to determine the value of items that they insure^[5], allowing for reasonable prediction of repairs and replacements. Data in general could benefit from a similar model, and news data specifically needs an immediate solution that is accurate and efficient.

In some cases trust was assumed without any evidence of trustworthiness (e.g., Facebook and Twitter), resulting in large groups receiving news from social media sites^[6]. In other cases, trust is granted based on reputation, as is the case with news sites^[7]. In all cases, the changing role of the news media due to the internet results in a rush to deliver news first.

Any solution to FN must consider the full scope of information or the “totality of information”^[8]. The customization of the fake narratives and the targeted delivery demand that an effective solution fuse non-technical disciplines with traditional technical responses. The attacks may originate from any source, although the Russian-based attacks are quite sophisticated^[9] and have gathered quite a bit of attention.

There are aspects of the Russian approach that warrant inclusion into the framework, even if the implementation becomes uniquely Western. The Russian term “protivoborstvo” describes the intentionally created rhetorical game that is foundational to FN; this rhetorical game can partially be addressed using CL and machine learning (ML), illustrating one example of interdisciplinary fusion.

FN and deceptive information campaigns can be thought of as opening shots in future information conflict that supports hybrid warfare^[10,11]. This framework may provide guidance for countering deceptive information campaigns. We have no cyber equivalent for “trust but verify” (yet). The purpose of this paper is to introduce a framework for evaluating FN. This framework may provide the cyber equivalent of trust but verify^[12] for FN. In addition to countering Russian FN efforts, this framework provides a foundation for examining the quality of data and may assist analysts in evaluating other news stories or events.

II. BACKGROUND

Falsehoods and deception in political discourse are a long-standing problem in an industry where words matter. Deception and propaganda have a long history: the Trojan Horse serves as an example of one of the earliest deceptions^[12]. The internet makes possible the ability to deliv-

er deceptive messages to a larger audience and social media data made possible the customization of deceptive data^[13]. Data science techniques performed by Cambridge Analytica^[14] made possible the rapid customization of messaging. Chatbots exacerbated the problem through the use of artificial intelligence (AI) software that could dynamically adjust to and manipulate user responses^[15].

There are many ways that facts can be distorted, resulting in altered perceptions, but there are a limited number of ways that facts can remain faithful to their original creation. This provides an entry point into the solution. Thus, attempting to model deceptive data is similar to attempting to model malware or any other host of cybersecurity problems. Deception is unbounded; therefore, attempting to model or predict deceptive data is difficult and subject to continual change. Facts, however, are constrained, allowing for more accurate modeling. While deceptive data may have several common features, all of these features should be examined in the context of the factual data that the deception is designed to conquer.

In order for deceptive data to be effective, the data should elicit an emotional response^[16]; otherwise, the data would be quickly forgotten. The response does not always need to be strong; this avoids the suspicion of hyperbole. Trust must be gained regarding deceptive data, and while trusted users can shorten the time required, a little initial skepticism is normal. Russian deceptions build the FN foundation by offering an alternative view or narrative that is designed to sound reasonable^[17-19].

A. What is Propaganda?

Propaganda is information or ideas that are spread by a group, such as a government, with the goal of influencing a targeted group's or person's opinions through the omission of facts or by secretly emphasizing only one narrative of the facts^[20]. Oftentimes deliberately used to control, influence, or change the cognition of the targeted group, propaganda entwines fundamental elements of psychology and technology in service of the goal. Psychological aspects of propaganda include campaigns to win the minds, means, and measures of message distribution, which requires a behavioral science understanding of message creation and application^[21]. The past is understood through the information that was recorded and left behind by the scribes of the particular time. Chronicles and annals provide contextual understanding of the past; however, these writings contain the biases of the scholars, historians, clergy, rulers, and ordinary citizens in local communities^[22].

In order for propaganda to be effective, the source or purveyor needs to fully understand the values of its target audience, thus rendering the target's intellect ineffective. The most accomplished propagandist discerns and plays on its target's values, morals, needs, or fears^[16]. This goal can be achieved subtly or overtly based on the values of the target.

Taylor suggests that the earliest form of propagandist imagery occurred in the Neolithic Age^[23]. The use of war propaganda may be found in Neolithic cave paintings, where imagery

carved in the wall commemorated battles. The carvings on cave walls illustrated clans' victorious battles; clans made them celebrate their victories and intimidate other tribes^[23].

Ancient Greece also offers some of the first examples of propaganda. Speech was utilized for conveying persuasive messages. Ingram provides the example of Confucius' writings, the *Analects*, which were used to persuade^[16]. The men who read these writings were supposed to live a more meaningful existence. From Ancient Greece to Alexander the Great to the pharaohs of Egypt, propaganda was a weapon of choice to change targets' cognition. Egyptian pharaohs' propagandist messages were prestige, nobility, and imperial legitimacy exemplified by grandiose architecture^[24].

The Roman Empire largely influenced civilization, reaching into Italy, the Mediterranean, Britain, North Africa, Portugal, and the Persian Gulf. Dating back to 48 B.C., Gaius Julius Caesar (Julius Caesar), father of Caesar Augustus (Augustus), the first emperor of Rome, used political manipulation to win the support of the people^[25]. Julius Caesar wrote war memoirs chronicling the achievements of the civil war between Gaul and Pompey (Gnaeus Pompeius Magnus) and the spoils of victory. Caesar sent runners to deliver war memoirs to be read before a crowd in a public area as the battle progressed^[25]. This action showed the recognition of both message craft and delivery speed since the common people, or plebeians, were not literate, and Caesar knew he needed their support. This early form of propaganda resulted in Julius Caesar being heralded as a hero^[25].

Patriotism was tied to the military, and only Roman citizens could be members of the military force; thus, the fighting force was respected and feared by all. In addition, Julius Caesar had a strong reputation for looking after the Roman legions. Caesar's focus on public opinion and strong concentration on providing for his military forces was a major reason the memoirs were successful, and the strong public opinion paved the way for Augustus^[25,26]. Augustus used writings as a means to deliver public information; these writings manipulated the events to tell stories from Augustus' perspective. Statues, monuments, and coins were also used to spread the image of Augustus as a strong military leader, a statesman, and peacekeeper^[25].

In addition to the spoken word, propagandist messaging was also accomplished through imagery, and this form of messaging remains popular today. In the early Stone Age, depictions of war were carved on cave walls; later, they were drawn on paper or scrolls. As propaganda matured, messaging was imprinted on clothing incorporated through imagery. This included stunning regalia and insignia-laden outfits^[21].

Eighteenth-century propagandists successfully used political cartoons and caricatures to directly communicate with their intended audiences^[16]. The caricatures and prints were biased in nature and oftentimes made fun of or poked individuals. The convergence of humor and politics in this new approach was well received.

Propaganda continued to be a means of influence in times of conflict. During the Civil War, cartoons became a popular propaganda medium. Animated movies and political and military cartoons became an attractive means for distributing propaganda^[27]. Propagandist cartoons can be divided into two categories: cathartic and ad justice^[28].

A cathartic cartoon was successful when the message convinced people that they had nothing to fear from the enemy. An ad justice cartoon was designed to spur action and could be considered successful when the message inspired voluntary enlistment in the Union forces, for example. One famous cartoon was Thomas Nast's "Compromise with the South," from September 1864. This propagandist-cartoonist used the symbols of his trade to guide the audience toward a certain predisposed objective; his cartoons condemned the idea of compromise by emphasizing the lives that had been sacrificed for the cause^[28].

According to Hinkleman, Hitler's *Mein Kampf* is considered an advanced work on the use of propaganda as a way to collect large numbers of supporters^[28]. The book masterfully pulled the audience into accepting only the author's views as true and shifted blame for previous failures away from himself to the other Germans leaders, thus perpetuating a victim mentality^[27,28]. *Mein Kampf* appealed directly to emotions formed from values and biases rather than logic. The blending of anti-Semitism and nationalism provided a way for Germany to survive—through anti-Semitism as a form of nationalism or love of country.

Hinkleman observes that Hitler believed that good propaganda targeted emotion and not intelligence or the facts. Hitler played on the hatred and despair felt by lower-class Germans. By preying on these Germans' poor economic status and fear of being unable to provide food and clothing for their families^[28], Hitler elevated emotions using reason. Combining Hitler's emotional elevation with Western societies' need to assign blame^[29] and cultural mores^[30] regarding uncertainty, avoidance, and fear of the unfamiliar, the Jewish population, along with other non-Aryan groups, were assigned the blame for the economic problems of Germany.

The United States used radio and movies to disseminate propaganda during World War II. Both Japan and the U.S. held competitions to create patriotic theme songs. Each soldier serving in the U.S. military was issued a songbook containing songs such as "Anchor Aweigh" and "Marine Corps Hymn." The songbook was part of standard issue^[31]. Japan, in an interesting twist, used American songs as propaganda to make American GIs homesick and weaken the American forces. Furthermore, the Japanese Government forbade the playing of American music at home^[31].

The Japanese were portrayed negatively in the media by focusing on physical characteristics such as crossed eyes and bucked teeth. The Japanese were referred to as "Japs," "back-stabbing monkeys," and "sneaky yellow rats"^[31] in an attempt to dehumanize them, with the goal of shaping behavior and inciting desired actions^[16].

Music with propagandist lyrics serves to convey meaning to the intended audience. Propagandist music instills a general feeling or emotion and, with the proper message, serves as a mechanism for the transmission of propaganda. Music, in general, makes messaging easier to retain due in part to patterns and repetition^[33].

Similarly, movies are a natural medium for propaganda. A moviegoer becomes a type of “hypnotized person” vulnerable to suggestions presented by a film^[31]. High-quality visual and audio design serves to reinforce the message^[32,33].

The message creation aspects of propaganda continue to evolve, growing more sophisticated and polished as technology improves and knowledge grows. As topics, phrases, and various aspects of crafting a message change, the important thing to remember is that the goal remains the same. Of equal importance is that a message has no value until it has been delivered, received, and interpreted.

The technological and behavioral aspects are loosely grouped as “message delivery.” Advancements in communication, military strategy, and technology and fluctuating partisan-elite rapport and populace contribute to the changing landscape of message delivery^[16]. According to Ingram, scholars and scientists in modern times study, determine, and understand propaganda campaigns and techniques and equate said campaigns to daily societal issues^[16].

“Falsehoods fly, and the truth comes limping after it”^[34]. Beginning with the runners used by Julius Caesar to deliver false messages as battles raged and continuing through to the written press and, more recently, images that travel at line speed, falsehoods continue to fly. Meanwhile, the facts surrounding an event take time to be researched and identified

With the invention of the printing press and print engravings, propagandists were able to print their messages on a mass scale. After 1880, messages were further impacted with the inclusion of photographs^[21]. Photographs could be staged or real, and the black-and-white images, eventually becoming full-color, made a real impact on targets’ cognitive perception. Eventually, motion pictures—first presented in black-and-white and, later, in color—captured society’s attention^[21].

During the literary age, propaganda was produced through pamphlets, newspaper articles, advertisements, flyers, billboards, and any other means that could alter or change an individual’s cognitive perception. Later, satirical caricatures and cartoons were used for target audiences. Propaganda campaigns utilized a new visual element which proved to be quite successful^[16].

In addition to literary campaigns and structural campaigns, propaganda messages have also been waged through radio, satellite, and broadband communications. During the Vietnam War, radio airwaves were laden with propaganda both for the United States and Vietnam^[35]. Radio stations, including The Voice of America, the BBC, Radio Free Europe, and Radio Liberty, transmitted both attributed and unattributed messages to their targeted audiences^[16].

Twentieth-century propagandists utilized multiple means of delivery for their messages,

using advertising and other techniques to convey the intended message to targets^[21]. All of these methods were asynchronously delivered. In the 21st century, technological advances in communications, computers, networks, smartphones, and the internet of things make a broader landscape available to propagandists and enable media saturation. For example, social media's role in the Arab Spring resulted in a new trusted news source for users. The rise of social media has made this new landscape more user-friendly and, perhaps more importantly, more trustworthy.

More recently, internet usage introduced a 21st-century feature: the ability to rapidly disseminate deceptive data both asynchronously and synchronously. Initial dissemination relied heavily on bots and trolls to establish a starting point. Once the starting point had been established, the dissemination reached the targets directly through the trusted channels of social media and social media trust relationships. Furthermore, the timing of the release of deceptive data took advantage of the inability to rapidly discern truth, allowing the falsehoods to fly. This strategic timing release of deceptive information is also known as weaponized information^[36]. This timed, mass release of weaponized information gives sources more control over the spread than they enjoyed previously. This synchronous component relies on a mixture of "true believers" (also known as "useful idiots") acting as trolls, paid trolls, and AI-controlled chatbots.

The new landscape continues to grow and the amount of information available in this new environment is so rich that a new discipline, data science, has emerged. New technology utilizing data science techniques allows for more accurate target identification and continuous bombardment with specially crafted messages from trusted or quasi-trusted sources. The volume of these messages that reinforce values can effectively alter the target's perception. When the targeted user seeks to verify the content of a message, a large number of similar messages are returned, and the target now knows that other people share the same values and beliefs.

B. Countering Propaganda

Research into CL shows that news can be accurately separated into truth, falsehoods, and satire through the analysis of linguistic features^[37]. The credibility toolkit provides the ability to assess news articles along the axes of reliability and objectivity as well as potential social media communities that might be interested in the content of the article^[38]. The toolkit provides visualization tools to assist in interpretation. Thus, CL may offer a means of performing preliminary tagging of a news article for rapid evaluation of that article's veracity. In addition to CL, a reputation analysis and PS may also offer valid insights that assist in the evaluation of a news story's veracity.

The initial response to FN relied on fact checking through sources such as Snopes^[39], PolitiFact^[40], and other fact-checking sites. This method has worked well for years, but is time-intensive and easily overwhelmed with the volume of FN stories that are generated throughout the course of a campaign. Fig. 1 provides an example of a hashtag associated with a fact-based

narrative and fig. 2 illustrates the fact-based narrative overlaid with the hashtag associated with the fake news narrative^[41]. Thus, fig. 2 shows the fact-based narrative being easily overwhelmed by the fake narrative.

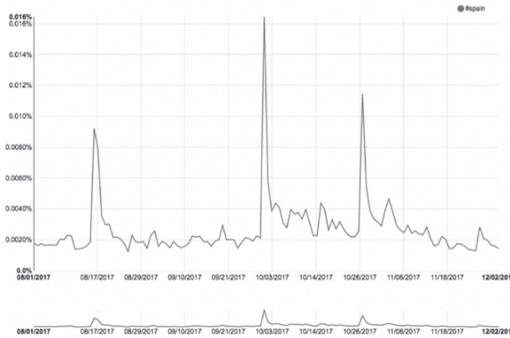


Figure 1: Hashtag associated with fact-based narrative

Fig. 1 shows the activity on #spain from August 1, 2017, to December 1, 2017, with the actual vote taking place on October 1, 2017, the highest of the three peaks^[25].

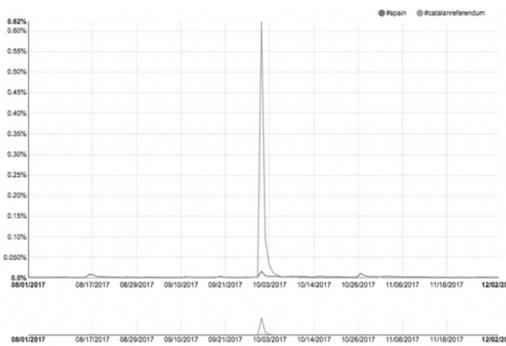


Figure 2: Fact-based hashtag overlaid with fake hashtag

The steep, peaked line in fig. 2 shows the rapid and intense injection of the #catalanreferendum hashtag associated with the fake narrative in the same time window as #spain. Notice the high volume and the very short time-line for the fake narrative^[41]. By introducing the fake narrative so close to the election (much like the introduction of Clinton’s emails^[42] and Macron’s emails^[43]), the target has little to no time to respond; thus, the information at this point is weaponized and active.

The volume associated with fake narratives is problematic because if the reader attempts to look up the story, a large number of the same narratives will be returned, thus validating the fake narrative to the reader. It takes time for fact-checking sites to perform research and post their findings, and when the source of the fake story is a friend, other close relationship biases are at play. Readers who are unable to determine the veracity of a news story oftentimes lack the time and resources to look up the story in question; instead, they rely on mental shortcuts such as biases^[44] and source reputation to determine trustworthiness^[45].

The reliance on reputation has been exploited in two ways. The first is through the use of popular social media applications, where, through the targeting of groups, multiple trusted entries become access points for individual members into society-at-large. Consider social media's original mission of bringing together like-minded people to share information in the spirit of friendship and fellowship^[46,47]. These platforms have also provided a channel for the distribution of FN since the content on these sites is also promoted as news stories. These same sites, such as Twitter, became trusted news sources due to their role in the Arab Spring^[48].

While social media sites have recently come under fire for the distribution of FN, these same sites were commended for their role in the Arab Spring. The second manner in which reputation has been exploited is the discrediting of reporters. A recent report on FN revealed that a journalist could be discredited for \$50,000 and a news event, such as a protest, can be staged for \$200,000^[4]. When these events are considered collectively, the use of reputation analysis becomes problematic. Furthermore, reputation analysis is vulnerable to the flux problem that has plagued domain name system servers.

The issue of trusting sources is complex and long-standing, as is the history of verifying trust. The handshake was one of the earliest examples of verifying trust^[45]. Referred to as "data fidelity," the verification of trust in the virtual environment is more difficult^[49,50]. A solution to this is proposed in detail in section III of this paper.

Old models may serve as inspiration in the design of the newer models for FN evaluation. Blind trust without verification of the information that is disseminated has been exploited. Trust in news sources continues to be manipulated. Untruths not only spread fast, but automated bots can persuade doubtful readers through interactive dialogue. The purveyors of propaganda have carefully profiled their targets, values, and beliefs before crafting their messages. The old adage about bringing a knife to a gunfight can now be replaced with bringing a gun to a bot fight.

III. PROPOSED MODEL

The model for evaluating FN relies on three areas: CL, PS, and source provenance (SP). Each area will be discussed in greater detail in the following subsections. CL findings can feed the PS and SP, while PS also can feed SP.

CL → PS ; CL & PS ⊆ SP

A. FN Content Analysis Using Computational Linguistics (CL)

There are several unique characteristics associated with FN, such as the size of the story in relation to the headline^[4] and the use of descriptive words and other features^[16]. As noted earlier, attempting to model the numerous characteristics of the ever-changing, deceptive data is not an efficient method, but modeling the facts or ground truth data (GTD) is.

This study requires researchers to enhance the existing RPI software with the intent of expanding the rating of the software to place event stories (from AP, Reuters, and Bloomberg) as GTD or μ in the distribution. As stories become embellished, the markers increase, resulting in the growing value of deviations. We will use a number of other unreliable sources to assess which markers are especially useful to track. Conversely, when stories omit key pieces of information, the markers will decrease or unrelated events will be used to fill the space; these unrelated events are oftentimes distractors which use a technique commonly referred to as “whataboutism”^[4]. Finally, traits such as repeating the same point three times^[4] will be considered a deviation measure. Fig. 3 depicts the proposed scoring scale.

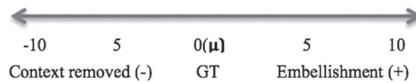


Figure 3: Scoring scale for CL output

The proposed scoring scale uses the AP/Reuters/Bloomberg results as the basis for scoring (0) or μ . The deviations from this in either direction reflect the deviations from the original event. A deviation score of -10 would indicate the story was likely taken so far out of context that the story is no longer recognized. Similarly, a story with a +10 deviation score would indicate that the story was so radically embellished that the original event may no longer be the central theme of the story.

The resultant score is an average of the criteria scores obtained from parts of speech, tone, ratios, etc. The overall score becomes a label used for both storage and signal identification. The storage label becomes relevant for comparisons used in SP, and can be useful when a story originates without the GT sources. The second use for the overall CL score, signal identification, is explained in the PS subsection.

B. FN in Motion: The Pattern Spread (PS)

Earlier background discussion highlighted both temporal and volume aspects of FN distribution. From Julius Caesar’s runners to print media to radio, television, and, more recently, the internet, the speed of distribution has minimally kept pace with news and, in some cases, outpaced news^[51]. Before the advent of the internet, mass propaganda delivery tended to be asynchronous in nature, but with the fusing of AI and data science, chatbots can be deployed at a large scale, allowing for interactive dissemination at scale. All events leave traces, and digital events are no different. While some stories follow a meme-like pattern^[52], the FN PS differs; this will be further discussed in future work.

The PS of FN offers an opportunity to revisit and reexamine aspects of signal processing. The noise level on the internet is very high, making the signal more difficult to identify. This noisy environment provides an opportunity for the identification and labeling of news stories using the techniques described in CL. The labeling should result in a picture of the environment resembling a pre-painted, paint-by-numbers picture in which the various labeled items ideally form clusters.

Fig.1 and fig. 2 show the difference in the spread of a fact-based hashtag and a FN-associated hashtag. The indiscriminate use of bots and trolls resulted in the obvious signal, but China has shown that a different signal can be as effective^[4]. Nevertheless, in the cases of Russia and China, automated software was used, and software is most efficient at patterned repetition.

Understanding the PS of a fact-based narrative is the first signal that requires identification. Once removed from the noise field, the remaining clusters can be identified and each deviation from the fact-based narrative can then be characterized numerically. Numeric characterization can be used when performing analytics in SP processing.

By performing the operations in the specified order of CL preceding PS, an additional benefit may possibly be had in the identification of the interactive behavior patterns. Ideally, 21 general PS signals should be initially detected. Naturally, ML algorithms will need to be evaluated to determine which one offers the greatest accuracy.

Detection of bots in general and chatbots specifically will require additional analysis. In addition, determining the chatbot signals in chatbots designed to increase amplification in efforts to persuade will likely require AI^[53]. In those cases, once identified, the dialogue can be analyzed through some of the CL techniques addressed previously. This particular aspect of FN is a separate but related effort.

C. Source Provenance (SP)

Data provenance has a long history of research that precedes the introduction of FN. Data provenance details data origination and the process through which the data arrived into the system^[54]. According to this definition, data provenance can be likened to quality assurance processes surrounding software development where, once again, the job of reliability or, in this case, the veracity component of examining the content remains ignored. This area can also be revisited and the solution can be more robust than with digital ledgers^[55,56], where entities can collude to lie.

By maintaining archive data on fact-based news, FN, and the values associated with these stories, additional information can be extracted on authors and publishers. Publishers' and authors' reputations can be manipulated in an attempt to decrease^[4] or increase their credibility. Thus, reputation analysis becomes problematic. However, by maintaining the first two elements of CL and PS and associating those values over time with both authors and publications, new patterns will emerge.

In the simplest cases, instances in which reporters are artificially discredited will be easily detected when examining the body of an author's or publisher's work. Additionally, temporal analysis can detect trends in the same body of works indicating a trend toward FN or fact-based narratives. If authors move to different publishers or publishers change names in an attempt to hide bad reputations, the characteristics of their previous work remain, allowing for the matching of emerging entities to existing bodies of work found in the archive.

This final area of FN determination relies heavily on the first two areas being developed. Both areas are in the early stages of development, so quite possibly other features will become relevant in the SP area. In spite of the lack of details, the basic concept can be drawn, recognizing that changes will be incorporated.

IV. DISCUSSION

The proposed model, while not perfect, offers a robust approach that can be easily modified or presented in an easily understood manner. Most credible news stories will likely fall within two standard deviations of the fact-based event reported. By focusing on the wording of fact-based narratives and characterizing these narratives, a certain robustness is built in for Byzantine behaviors which may arise as the propagandists attempt to tailor messages to match the rules of fact-based narratives. This may be less problematic as the text would have to be less emotionally appealing, possibly resulting in lower efficacy.

A more significant problem with this model would likely be related to linguistic traits and slang expressions across languages, cultures, subcultures, and tribes. The potential to inaccurately score an article is present. This work would benefit from the involvement of other experts, including linguists and social scientists..

A more interesting and potentially more challenging scenario revolves around the improvement of automated behaviors to more accurately reflect human behaviors. This may affect the PS component of the model. Historically speaking, the behavior of presenting propaganda first may also offer insight into detection. As FN pattern signals become better understood, temporal analysis will also provide additional new insights. Finally, as the archive grows and more data analytics are performed on the archived data, the ability to distinguish the fact-based narratives from the FN narrative will likely grow in sophistication.

V. EXAMPLE CASE

Before processing can begin, rules must be examined and tested. The first processing component relies on the translation of propaganda rules into computational linguistic rules. An example of one of the rules of propaganda is that the message must appear interesting to the target and use an attention-getting distribution medium^[57]. Thus, attention-grabbing headlines complete with pictures displayed on websites and social media sites would be an example.

In English, verbs are action words and adverbs are descriptor words; these words are used to convey what happened and provide details capable of eliciting a response. Thus, the article length and the rate of adverbs may provide a possible marker as a metric deviation from fact. Of course, these alone are not sufficient, but serve as an example for illustrative purposes. Also considered but not measured in this particular example is the role of context in describing event news.

One assumption suggested that the news wire (AP News, Reuters, or Bloomberg) would report the fact-based narrative, and the model suggests that the fact-based narrative should serve as μ . The example chosen occurred in 2016, when candidate Clinton collapsed at the 9/11 ceremonies in New York City. The source sites were selected based on a Google search for “Clinton collapse 2016.” The news sites were NBC News, the Washington Post, Fox News, and the New York Post. An additional opinion piece was selected with the purpose of providing observational data on this type of publication.

$$\begin{aligned} &\text{IF (WordCount} \geq \text{APWordCount) then} \\ &\quad \text{diff} = 1 - \left(\frac{\text{WordCount} - \text{APWordCount}}{\text{APWordCount}} \right) \\ &\text{else} \\ &\quad \text{diff} = \left(\frac{\text{WordCount} - \text{APWordCount}}{\text{APWordCount}} \right) \end{aligned}$$

Figure 4: % difference equation

As expected, the AP News story word count was in the middle of the group; it ranked third of six in order of low to high word count. The word count for the smallest story was 179 words and the largest story was 1,331 words. Because only one story was selected for this effort, there are no average values for news stories and no standard deviations (σ). The results of the word count are shown in table 1. The corresponding bar chart is depicted in fig. 5. Percent-based differences were calculated using the logic displayed in fig. 4.

Source	Word Count	% Change
AP News	840	0
Connor Post	923	+10%
Fox News	179	-79%
NBC	885	+5%
New York Post	693	-17%
Washington Post	1331	+58.5%

Table 1: News article word count

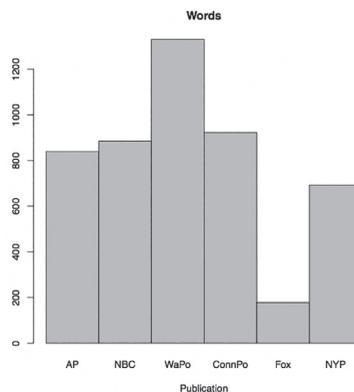


Figure 5: News article word count

The most dramatic differences can be seen with Fox News and the Washington Post. The negative differences associated with with Fox and the New York Post suggest the potential for missing context, where part of the narrative may be missing. Missing context results in the reader having to mentally complete the story by relying on existing cognitive biases. The further the measured distance from the μ in the negative direction, the greater the potential for the reader to rely on cognitive biases. The New York Post and, to a lesser extent, the Connor Post show a significant deviation.

The Washington Post showed a large deviation in the opposite direction in terms of word count, suggesting that, minimally, the publication embellishes, but, in the absence of σ , measures to determine normal variance are not yet available. Additionally, rules that separate context from propaganda would require translation into CL terms and software. Presently, much of the CL software requires modifications due to the cleaning of terms that are typically used in propaganda but are problematic for CL (e.g., “them”).

In both cases of strong deviation from the AP story, there are no measures of variance, or σ , since this is an exemplar while the research continues to determine the optimal list of weighing factors. Word count determines the positive or negative assignments and initial weight of the deviation. The weights will be modified over time as algorithms are tuned and the archive grows.

Another measure is the usage of adverbs. Because the stories are of varying lengths, the measure uses percentage values obtained by dividing the total number of adverbs in an article by the total number of words in that article. Fig. 6 shows the equation used to calculate the percentage change from the AP adverb percentages. Table 2 shows the resultant numeric differences and fig. 7 depicts the bar chart representation of the adverbs.

$$change = \left(1 - \left(\frac{\%Adverbs}{\%APAdverbs} \right) \right)$$

Figure 6: Equation to determine adverb distribution rate derivation from AP News

Source	# Adverbs	% Adverbs	% Change from AP
AP News	19	2.3	0
Connor Post	45	4.9	+113%
Fox News	3	1.7	-26%
NBC	30	3.4	48%
New York Post	32	4.6	100%
Washington Post	50	3.8	65%

Table 2: Adverb rates per article

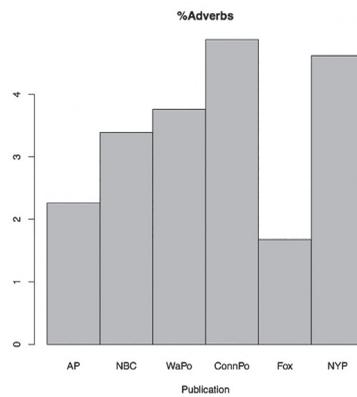


Figure 7: Adverb rates per article

Adverb usage helps to add context to the word findings. If verbs are considered action words, then adverbs are the action descriptors that are capable of adding urgency, confusion, or other emotions designed to manipulate emotional response. Descriptors the nouns and verbs that they define more dramatic.

Working off of the assumption that AP News provides the baseline, a quick examination of the values shows that the Washington Post has the most adverbs; however, when examined in the context of the number of words, the Connor Post (opinion article) and the New York Post show greater deviance from AP News. This positive deviance suggests embellishment, whereas the deviation by the Washington Post and NBC suggests a level of bias in the article. The Fox News deviation may be too small to measure for this particular example; however, when combined with the small word count, the possibility that a partial fact is being reported must be considered.

While not counted in this example, other items of interest include the punctuation deviations and paragraph sentence counts. The Connor Post and Fox both had the punctuation character “?” in their stories. This use of the “?” was of interest for two reasons. The first reason is that, by asking a question, the article provides an entry point from which it draws in readers, with the hope of engaging them in the process. The second is that most news stories use a period; thus, “?” or “!” are used for special stories that are designed to elicit a response that is most likely emotional.

One other observation that may result in a marker for opinion pieces is the number of sentences in a paragraph. In the opinion piece (Connor Post), the vast majority of paragraphs contain three or more sentences. This finding was in contrast to all of the other news articles, which typically contained single-sentence paragraphs or two-sentence paragraphs.

As the remaining criteria emerge and deviations from fact-based narratives (AP, Reuters, and Bloomberg) are determined, the differences can be averaged, creating the overall tag values for the articles. The tagged value is used to assist in defining characteristics that will be used to define training data characteristics for use in ML algorithms. Table 3 contains the overall deviation values for the six sources. The values for this table were simple averages obtained through the equal weighting of inputs (words and verbs). The overall percentage values were divided by 10 to create the measure off μ , which was used to fit the overall article scoring scale.

Source	Overall %Value	μ Distance
AP News	0	0
Connor Post	61.5%	6.15
Fox News	52.5%	-5.25
NBC	26.5	2.65
New York Post	58.5	-5.85
Washington Post	61.5	6.15

Table 3: Overall deviation measures

While some of the values suggest a high deviance, they should be considered in the context of a single story lacking a corpus of data for baselining and comparison. As mentioned earlier, separating context from propaganda terms improves the fidelity of the model, as does the tuning of algorithm weighting. One key finding was that, as suspected, the AP News story served as a good center point because, in both word count and adverb rates, there were entries above and below the AP News values.

In addition to fact-checking, trending may also be useful in tuning. The values seen in column three of table 3 represent the final value with which the news article is tagged. This assigned number can be used to tag or identify the article for observation in the larger stream of articles in the second processing phase of PS analysis. PS analysis will likely be highly dependent upon AI/ML techniques for both tagging and classification.

The remaining description is of a proposed archive where results from CL and PS can be stored and made available for additional analysis. The archive has not yet been built, but there are certain pieces of information that are of interest to this area of research. Of note, the archive is not designed to compete with existing archives; rather, the archive is designed to augment existing archives. The archive, which primarily provides historic data on meta-fields, should supplement other news archives. The archive is designed to encourage additional studies by other researchers. The design of the archive proposed here is preliminary in nature; records will most likely be stored as comma-separated value records. Table 4 provides a brief description for each of the fields.

Field	Description
Identifier	Unique record identifier
Author(s)	Vector contains names of article author(s)
Author(s)-score	Average total CL score for author's other work
Publisher	Publisher name
Publisher-score	Average publisher CL score
Links	Vector with link information to the news story and other archives
Metadata profile	Vector containing the values
Topic	Story topic and related information
Overall article score	Deviation score for the article from μ
Comparative scores	List of other CL scores for each of the components used in the overall article score
Related stories	List of related stories
Event date	Date of the news event
Publication date	Date of publication

Table 4: News archive fields

Putting together all three components of this model, the ability to evaluate any news story will ultimately be supported by all three components. The entire process is designed for both efficiency and the ability to use any single component with high assurance. Fig. 8 depicts the overall process flow.

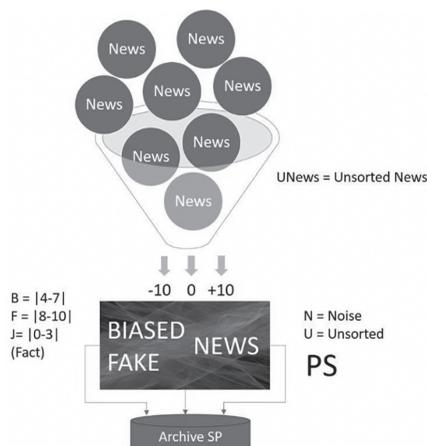


Figure 8: Overview of news processing

VI. CONCLUSION

The high efficacy and the low cost make propaganda a useful weapon in warfare. The ability to manipulate trust through various media relies on a flawed trust model that relies on object-oriented constructs^[29], resulting in a loss of context. The model presented provides contextual evaluation of news stories and offers a rapid and less subjective way to evaluate any news article and provide an objective measure of the distance between the GT (or fact-based) event and the narrative being presented.

Through the use of agreed-upon event reporting metrics, this model provides a starting point for evaluating FN in an objective manner. The ability of CL to identify FN has been shown to work on a large scale in a similar model^[37]. The PS mechanism shows promise but has not been executed on a large scale to date^[41]. The archives are being populated at several higher learning institutions, and these institutions are expanding beyond English language-based stories. The archive created for this project will augment existing archives by providing metadata characterizations and other relevant information that can add to other data mining efforts.

The ability to perform temporal analysis on the archives that are being built offers great promise because the findings can be combined with cultural and linguistic models that may ultimately identify vulnerable traits and ways in which populations can be quickly inoculated based on the identified traits. While Cambridge Analytica used data science techniques on personal data to identify potential targets, data science combined with cultural frameworks can be used for benevolent purposes.

We conclude with the observation that propaganda has been a long-standing problem with FN on the internet, elevating the effectiveness of this tool. Stand-alone point solutions run the risk of repeating the mistakes of the signature-based model that prevailed during the early days of internet security: they created a false sense of security. Therefore, the ultimate solution will likely take time and require the contextual evaluation of events. We suggest that the model presented here can meet these new and comprehensive requirements. ♥

VII. ACKNOWLEDGEMENTS

The authors would like to acknowledge the contributions of Clay Hampton; the Purdue School of Engineering & Technology, Indiana University – Purdue University Indianapolis; and Steve Hutchinson, ICF Incorporated, L.L.C.

NOTES

1. Oxford dictionaries, website, <https://www.oxforddictionaries.com/press/news/2016/12/11/WOTY-16>. 2017.
2. Collins dictionary, website, <https://www.collinsdictionary.com/woty>. 2017.
3. D. L. Altheide. "Creating fear: News and the construction of crisis". Routledge, 2017.
4. L. Gu, V.Kropotov, & F. Yarochkin. "The Fake News Machine: How Propagandists Abuse the Internet and Manipulate the Public", Trend Micro, 2017.
5. Cambridge dictionary, website, <https://dictionary.cambridge.org/us/dictionary/english/actuary> .
6. J., Gottfried, and E. Shearer. "News use across social media platforms 2016", *Pew Research Center Journalism & Media*. Available: <http://www.journalism.org/2016/05/26/news-use-across-social-media-platforms-2016/>, (2016 May 26).
7. Oxford University website, <https://medium.com/oxford-university/where-do-people-get-their-news-8e850a0dea03>, 2018.
8. M. Ristolainen. "Should 'RUNet2020' be taken seriously? Contradictory views about Cybersecurity between Russia and the West", *ECCWS conference proceedings*, Dublin, Ireland, 2017.
9. A. Entous, E. Dwsokin, and C. Timberg. "Obama tried to give Zuckerberg a wake-up call over fake news on Facebook." *Washington Post* 2017.
10. N. Verrall and D. Mason. "The Taming of the Shrewd", *The RUSI Journal*, DOI: 10.1080/03071847.2018.1445169. 2018.
11. M. Connel and S. Vogler. "Russia's Approach to Cyber Warfare", *Center for Naval Analysis*, <http://www.dtic.mil/dtic/tr/full-text/u2/1032208.pdf>, 2017.
12. R. J. Hexter. "What Was The Trojan Horse Made Of?: Interpreting Virgil's Aeneid.": 109-131, 1990.
13. A. Badawy, E. Ferrara, and K. Lerman. "Analyzing the Digital Traces of Political Manipulation: The 2016 Russian Interference Twitter Campaign." arXiv preprint arXiv:1802.04291 2018. <https://arxiv.org/pdf/1802.04291.pdf>.
14. H. Grassegger, and M. Krogerus. "The data that turned the world upside down." *Vice Motherboard* 28, 2017.
15. S. Rosenblatt. "Exacerbating our fake news problem: Chatbots" <https://www.the-parallax.com/2018/03/26/fake-news-chat-bots/>, (March 26, 2018).
16. H. Ingram. "A brief history of propaganda during conflict: Lessons for counter-terrorism strategic communications, 2016". Retrieved from <https://www.icct.nl/wp-content/uploads/2016/06/ICCT-Haroro-Ingram-Brief-History-Propaganda-June-2016-2.pdf>.
17. BBC News <https://www.bbc.com/news/world-europe-29478415>.
18. P. Pomerantsev and M. Weiss. "The menace of unreality: How the Kremlin weaponizes information, culture and money". *The Interpreter*, 22, 2014.
19. H. Volodymyr. The "Hybrid Warfare" Ontology. *Фаховe видання з економічних, філософських, політичних наук Затверджено постановами Президії ВАК України від 26 січня 2011 р. № 1*. 2016.
20. C. Sample, J. McAlaney, J.Z. Bakdash and H. Thackray. "A Cultural Exploration of Social Media Manipulators", *Proceedings of the 17th European Conference on Cyber Warfare and Security*, Oslo, Norway, pp. 342 – 341, 2018.
21. G. Jowett, and V. O'Donnell. Propaganda and Persuasion, 5th Edition. Chapter 2 – "Propaganda Through the Ages." Thousand Oaks, CA: Sage Publications, Inc. 2011.
22. R. Marlin. "Propaganda and the ethics of persuasion", in Editor (Ed)., *Book Propaganda and the Ethics of Persuasion*, 2013.
23. P. Taylor. *Munitions of the Mind: A history of propaganda from the ancient world to the present day*. Manchester, NY: Manchester University Press, 2003.
24. P.M. Taylor. *War and the Media: Propaganda and persuasion in the Gulf War*, 1992 https://books.google.com/books?hl=en&lr=&id=V9tRAQAAlAAJ&oi=fnd&pg=PR7&dq=taylor+1990+propaganda+persuasion&ots=AG3yYpVQTP&sig=9ru8pRyg-FicbOYd2J_RGKbrzERs#v=onepage&q=taylor%201990%20propaganda%20persuasion&f=false.
25. P.A. Hayward. *Factoids, Dishonesty and Propaganda in the Middle Ages*, 2018.
26. A. Pollok, and D.U. California. *Roman Propaganda in the Age of Augustus*, Dominican University of California, 2017.
27. J. Lively. "Propaganda Techniques of Civil War Cartoonists", *The Public Opinion Quarterly*, Vol. 6 (1), pp. 99-106, 1942.
28. D. Hinkleman. *Struggle for Power: By Any Means*, Create Space Independent Publishing Platform, 2016.
29. Nisbett. *The Geography of Thought: How Asians and Westerners Think Differently*. Simon and Schuster, 2010.
30. G. Hofstede, G.J. Hofstede, and M. Minkov. *Cultures and organizations*, New York, NY: McGraw-Hill Publishing, 2010.
31. W.A. Sheppard. "An exotic enemy: Anti-Japanese musical propaganda in World War I Hollywood", *Journal of the American Musical Society*, Vol. 51 (12) pp. 303-357, 2001.

NOTES

32. M. J. Marks. "Teaching the Holocaust as a Cautionary Tale." *The Social Studies* 108.4 : 129-135., 2017.
33. A. Yuhaniz, F. Zainon, Z. Ghazali, N. Man, F. M. Alipiah, and M. Y. M. Yunus. "The Influence Of Music On Memorization Performance Of Mathematics Students." *Proceedings of the ICECRS* 1.2 2018.
34. J. Swift. "The art of political lying", *The Examiner* 14, 1710.
35. J. B. Whitton. *Cold War Propaganda*, 2018.
36. E. Darraj, C. Sample and J. Cowley. "Information Operations: The use of weaponized information in the 2016 US presidential election", *Proceedings of the 16th European Conference on Cyber Warfare and Security*, Dublin, Ireland, pp. 113-119, 2017.
37. B. Horne and S. Adali. "This Just In: Fake News Packs a Lot in Title, Uses Simpler, Repetitive Content in Text Body, More Similar to Satire than Real News", *Association for the Advancement of Artificial Intelligence*. 2017.
38. Website: <http://nelatoolkit.com>.
39. Snopes website, <https://www.snopes.com>.
40. Website: <https://www.politifact.com>.
41. Website: <https://osome.iuni.iu.edu/tools/trends/#>.
42. R. Faris, H. Roberts, B. Etling, N. Bourassa, E. Zuckerman, and Y. Benkler. "Partisanship, propaganda, and disinformation: Online media and the 2016 US presidential election." (2017). https://dash.harvard.edu/bitstream/handle/1/33759251/2017-08_electionReport_0.pdf.
43. M. Burgess. "The Emmanuel Macron email hack warns us fake news is an ever-evolving beast", *Wired*, <https://www.wired.co.uk/article/france-election-macron-email-hack>.
44. M. Tanis, and T. Postmes. "A social identity approach to trust: Interpersonal perception, group membership and trusting behaviour", *European Journal of Social Psychology*, Vol. 35, no. 3, pp. 413-424, 2005.
45. J. H. Cho, K. Chan, and S. Adali. "A survey on trust modeling", *ACM Computing Surveys*, vol. 48, No. 2, Article 28, October, 2015. DOI: <http://dx.doi.org/10.1145/2815595>.
46. Facebook website: https://www.facebook.com/pg/facebookcareers/photos/?tab=album&album_id=1655178611435493.
47. J. Fox. "Why Twitter's mission statement matters", *Harvard Business Review*, 2014. <https://hbr.org/2014/11/why-twit-ters-mission-statement-matters>.
48. O. Parent and A. Zouache. "Role of peer effects in social protest. Evidence from the Arab spring." <http://erf.org.eg/wp-content/uploads/2018/02/Olivier-Abdallah.pdf>. 2017.
49. C. Sample, T. Watson, S. Hutchinson, B. Hallaq, J. Cowley and C. Maple. "Data fidelity: Security's soft underbelly", *Proceedings of the 11th International Conference on Research Challenges in Information Systems*, Brighton, UK, pp. 315 – 321, 2017.
50. M. DeLucia, S. Hutchinson and C. Sample. "Data fidelity in the post-truth era: Network Data", *Proceedings of the 13th International Conference on Cyber Warfare and Security*, Washington, DC, pp. 149-158, 2018.
51. S. Visoughi, D. Roy and S. Aral. "The spread of true and false news online.", *Science*, Vol. 359 no. 6380, pp. 1146 – 1151, 2018.
52. L. Shifman. "Memes in a digital world: Reconciling with a conceptual troublemaker", *Journal of Computer-Mediated Communication*, Vol. 18, pp. 362-377, 2013. Available at: <https://academic.oup.com/jcmc/article/18/3/362/4067545>.
53. S. Rosenblatt. "Exacerbating our fake news problem: Chatbots, March 26, 2018. Available at: <https://www.the-parallax.com/2018/03/26/fake-news-chatbots/>.
54. P. Buneman, S. Khanna, and W. Tan. "Why and Where: A Characterization of Data Provenance", In *International conference on database theory* pp.316-330. Springer, Berlin, Heidelberg.
55. S. H. Ammous. "Blockchain technology: what is it good for?" Available at: http://capitalism.columbia.edu/files/ccs/working-page/2016/ammous_blockchain_technology_.pdf.
56. M. Staples. "Blockchain is useful for a lot more than bitcoin", *The Conversation*, 2016. Available at: <http://theconversation.com/blockchain-is-useful-for-a-lot-more-than-just-bitcoin-58921>.
57. PsyWarrior Website: <http://www.psywarrior.com/Goebbels.html>.

Strategic Cyber: Responding to Russian Online Information Warfare

Matthew J. Flynn, Ph.D.

ABSTRACT

The success of the democratic world and its citizens depends to a great extent on recognizing one's strategic advantages. Secure on this high ground, a nation can dictate interstate strategic competition in favor of U.S. national security. In cyberspace, that advantage rests on defending and advancing a U.S. ideological advantage inherent in that platform. The quality of openness ensures the unfolding of confrontation well short of armed conflict and winning this war matters most to those seeking to erode U.S. strategic ascendancy. This paper follows Russia's progression in its effort to reverse its unfavorable situation in cyberspace, largely by hoping to panic the United States into a series of poor policy decisions. A failure to see openness as the means to thwart this cognitive offensive all but hands Russia a victory. Reversing this outcome stands to blunt cyber tensions from giving rise to a means of setting conditions for a *fait accompli* and a military clash of arms. With this end in mind, there is much reason for optimism at the strategic level of such a war in cyberspace.

INTRODUCTION

Perhaps no state has grasped the implications of cyberspace to foster political activism more than Russia. In 2007, and again in 2008, popular expression online helped propel Russia into conflict with its neighbors, first in Estonia in the Baltic region, and then in Georgia to the southeast. In both cases, the power of internet access challenged the Russian Government's ability to dictate events. By 2014, strongman Vladimir Putin no longer feared the unintended consequences of that platform and could in fact look to capitalize on that technology to spur unrest in other countries, an effort that climaxed with the hack of the U.S. Presidential election in 2016. Even so, Russia remains at a severe disadvantage in cyberspace because that domain, while a new arena, reinforces an old military truism—it is best to enjoy the strategic high ground in any conflict. Russian actions in cyberspace reveal a state trying to achieve this favorable dynamic and almost succeeding with the unwitting

This is a work of the U.S. Government and is not subject to copyright protection in the United States. Foreign copyrights may apply.

help of the United States. This paper exposes Russia's effort to reverse its strategic weakness in cyberspace by restricting internet access out of fear that a community of users there can threaten the legitimacy of centralized government within Russia. The Kremlin's attempts to curb this online presence should serve as a reminder of the importance of supporting the existing U.S. cyber policy of defending and advancing an open internet to hold onto the strategic high ground in cyberspace.^[1]

CYBER IDEOLOGY

For Russia, controlling online access is less about shaping the battlespace for the next war and more about accepting the ideological showdown that the internet imposes upon restrictive governments. This cognitive struggle unfolds below a threshold of violence coming at the hands of armed conflict that usually serves to define war. Russia seized upon this construct to better position itself globally in the ether of cyberspace. It did so, however, only after a painful trajectory that witnessed online users threatening the authority of the state. In fact, regimes hostile to representative governmental norms had to weather the changes stemming from these cyber rebellions and then learn how to discredit them. This reaction made clear the tangible threat that openness poses to nations fearing the quality of shared space, producing an online community which clearly embraced the democratic values of connecting people, sharing information, and doing so free of oversight from governing bodies. That dynamic ensured that online connectivity became a means of challenging authoritarian regimes through cyberspace.^[2]

This analysis covers three main events to evidence the Russian trajectory to combat this threat and find safe footing in cyberspace. The cyber wars first in Estonia in 2007, then Georgia in 2008, yield to an examination of Russian efforts at home, and then in Crimea and Ukraine. This progression not only underscores the ideological dimensions of the stand-alone cyber war, but also stresses the lack of awareness of this dynamic by the United States. The piece ends by stressing the strategic ascendancy the United States enjoys in the cyber domain and offers some suggestions for maintaining this advantage. In this way, the analysis turns state competition in cyberspace into a valued context, revealing how a cognitive cyber offensive can expand user access in cyberspace and help usher in a new era of containment that, as in the Cold War, confronts U.S. adversaries with the losing proposition of thwarting basic human values. This ideological aspect of cyberspace places foes of openness on the defensive, impeding war long before it escalates to a *fait accompli* campaign settled primarily with conventional forces.

The Russian progression of waging war in cyberspace recasts a portion of the familiar narrative of Russian online actions seamlessly interfacing with Russian military efforts. In fact, those addressing Russian actions in Estonia and Georgia note the gap between action and effect, even as they validate that coordination.^[3] This conclusion greatly overstates Russia's efficacy, as is made clear below. Too often, cyber connectivity worked against Russian authorities

at home, but that outcome simply goes unacknowledged by those weighing the military implications of Russia's use of cyberspace. Yet the concern about divisions at home is prominent in the scholarship examining Putin's effort to maintain his power; this point is also made plain in this analysis. Omitting this context skews any understanding of Russian fear as a motive for acting in cyberspace, a failure that warps U.S. policy efforts to counter the threat. In particular, Russia's hack of the U.S. 2016 Presidential election has prompted a U.S. defensive effort in cyberspace, surrendering the use of cyberspace as an attack vehicle. To regain the strategic high ground derived from cyber rebellions requires a conscious effort by U.S. decision-makers to ensure that a free exchange of online messaging gets into the cyberspace of those seeking to thwart this end.

One stops short of labeling this online political activism a revolution because that word suggests outcome more than process, and a focus on process is key. From this viewpoint, a revolution births a movement while rebellions merely embrace a possible change, something that may or may not come to pass. Cyber rebellions point to realities in cyberspace that could lead to an ideological gain for states embracing openness. The term recalls what once was and emphasizes the imperative to get it back. Reminding U.S. policymakers that those opposing democracy face a threat from this medium is the main purpose here, and one best seen in the Russian response to this threat. Getting cyber right goes a long way to validating current U.S. policy as defending and advancing openness. As one journalist recently wrote, cyber is a "perfect weapon" to fray combustible civic bodies, although that individual was referring to liberal societies.^[4] The United States must carry that fight to the Russian body politic by fostering a global online community. That task suffers as the United States retreats from demanding openness online, best seen in the alarming tendency of experts to call for a new cyber strategy to better serve U.S. interests. That pronouncement implicitly accepts cyber sovereignty and accedes to the hope of U.S. adversaries to enforce national borders in cyberspace and thereby blunt the impact of connectivity.^[5] A look at Russia's struggles with online activism underscores the need for openness in order to enable nations welcoming an online exchange to profit from the ideological utility of the cyber domain.

ESTONIA: CAUGHT BY SURPRISE

The Russian attack on Estonia seemed far short of an act of war and looked to consist only of a cyber disruption and nothing more. There is no evidence of a congruent purpose, such as a ground attack, and this cyber incident most likely substituted for such retaliation. In this sense, the cyber territory mitigated conflict by offering a new outlet for expressing a foreign policy grievance. Many Russians certainly felt that Estonia had authored such an affront when in late April 2007, after considerable public debate, the state sanctioned the removal of a statue commemorating Soviet dead who had fought to liberate Estonia from Nazi control during the Second World War. To most Estonians, the statue represented Russian occupation, not liberation. Moreover, citizens of the Baltic nation believed that the statue served as a rallying point

for extremists among Estonia's considerable number of ethnic Russians, which total a quarter of the country's population. As the removal became imminent, radicals in that group helped foment riots in Estonia's capital, Tallinn.^[6] These disruptions ceased after a few days and, by April 30, the statue was installed at the Tallinn Military Cemetery.

Outrage may not have gone further than this had it not been for the internet. The Russian-language blogosphere and online Russian forums fueled popular discontent to the point of encouraging all Russians so offended—those in Estonia and beyond its borders—to take matters into their own hands and strike back online.^[7] There, concerned citizens could find prompts to launch ping-flooding and malformed queries to enable them to execute an “attack” on Estonia and conceivably shut down the internet in many of its cities.^[8]

The response was rapid and overwhelming. Soon, this Russian popular front reduced Estonian bandwidth, crashing the websites of numerous government ministries and a few major banks. Notably, the attacks avoided power grids and water supply facilities, although the attacks demonstrated the potential to do just that.^[9] The harmful traffic intensified on May 9, the day that marked the anniversary of the end of Russia's involvement in World War II. Specialists in the employ of the Estonian Government curbed this flow by ordering some victims to unplug, thereby imposing a “self-blockade” on Estonia.^[10] The incidents dissipated shortly thereafter, although a few more waves occurred in subsequent days. During the three weeks of attacks, most Estonians experienced some service interruption. In this respect, the spontaneous Russian initiative appeared to have met its goal of disrupting the “most wired nation in Europe,” as *WIRED* Magazine labeled that country due to its purposeful reliance on cyberspace.^[11]

If the Russian intent was clear, the motives were less so. What was gained by the attack? What had been achieved? Yes, some Estonians could not function normally for a number of days, but the Estonian authorities did not return the statue to the town center. Still, Russian pride had been assuaged, and this satisfied the main purpose. The Russian citizenry had employed a “cyber riot” to lash out and avenge a wrong and had done so without violence.^[12] A popular protest had rebuked a neighbor, and one could not but acknowledge what it was: a true expression of democracy. This was all the more true because government sanction of the event did not come to the fore. In this instance, attribution was unclear, but only at the end of the chain. Certainly, populism had sent Russians to their computers, where a “hacktivist” community assisted their efforts. But were the hacktivists working at the behest of the state? This was not clear, nor has the Russian Government ever claimed responsibility, with one Russian statesman publicly denouncing the attack as “cyber-terrorism.”^[13] This label is most telling, not in underscoring the challenges of attribution, real as they are, but in stressing this incident as one of democratic activism. This ideological purpose had become plain in a country hardly known for its democratic tradition. In fact, the opposite had been the norm: authoritarianism has plagued Russia's history, from czars to Communist thugs, and even to the current appearance of *imperium* at the highest levels of government in the person of Vladimir Putin.^[14]

In defiance of this history, the internet had enabled Russian citizens to achieve what they had not been able to do over hundreds of years, which was demonstrate an outgrowth of popular expression independent of any government control. It was a phenomenal moment.

No one noticed. Russian failures to achieve more instances of this success are perhaps understandable. The Russians could not dig themselves out from under the weight of their history, so a healthier democracy was not forthcoming. While the state boasts a large and talented pool of hackers brandishing tremendous technological prowess, that capability appears to lie outside of any shared ideological purpose.^[15] The technology can stand on its own. If this view is endemic to a hacker mentality, and if this view is indicative of an ideological purpose that is more instinctive than institutional in Russians, the strike on Estonia leaves Russia pioneering cyber warfare as an ideological weapon, but not realizing this is so.

Ironically, the same can be said of Western powers. The ideological purpose of advancing democracy globally is a long-standing concern of Western nations, particularly the United States. This goal was advanced by Russia when attacking Estonia, yet the West saw no such success—only fear of motives. Russia attacked Estonia to probe the North Atlantic Treaty Organization's (NATO's) response when it came to a battle to control cyberspace.^[16] With cyber supremacy, a conventional military strike could follow.^[17] These were valid concerns and required close attention, given the attribution issues, for it remained unclear whether a popular movement could execute a distributed denial-of-service (DDoS) attack—one which needed a legion of hijacked computers turned into botnets to succeed.^[18] Only a carefully coordinated attack could marshal this resource to its greatest effect. So, the question still looms: Was the Russian Government behind the attacks? Here, the West's old fear of Soviet secrecy arose anew. Had the security arm of the Russian Government, the Federal Security Service of the Russian Federation—having taken over for its Cold War version, the KGB—orchestrated the attacks in league with criminal organizations?^[19]

Should this be true, what transpired in Estonia meant that the Russian Government had tested a tool of espionage that lay very close to an act of war, should the intent be overtaken by popular elements online. In this sense, some deniability made sense to ensure that cyber disruption did not appear to have been sanctioned by the state, for that admission could raise tensions, which could lead to an outbreak of warfare on the ground. But deniability raised another unsettling question: What if the Russian Government could not control criminal elements within the state, and they had acted independently? Here was a dangerous precedent: private actors taking matters into their own hands. But to what end? What gain would criminals enjoy in this instance? Since answers were not clear, thinking rests in larger part on the ideology of the attack—that even criminals agreed to salvage some national pride and take part in the strike. But the thinking has not gone that far. The implications of a democratic impulse sweeping Russia, pulling criminals in that direction, and resulting in a patriotic cyberattack—and such a spectacle blooming overnight, without state involvement—went unacknowledged in the West.

GEORGIA: A DANGEROUS SEQUEL

The cyber war in Estonia remained a muted affair, solely an online confrontation. Still, the fact remained that renegade online fronts had sparked this crisis by unleashing cyberattacks on Estonia, and did without the involvement of the Russian Government, even if they were given its tacit approval and, later, its encouragement. This meant that openness had hit a threshold where state control could not curb public discontent expressed online. This democratic movement assumed uncertain dimensions within the Russian state as growing authoritarian rule faced spontaneous challengers.

This experience in Estonia helped pull Russia into another confrontation the following year, this time with the country of Georgia in the Caucasus region. The former Soviet republic had asserted its independence in 1991 in the wake of the collapse of the U.S.S.R. But two territories within that state, South Ossetia and Abkhazia, mustered a counter action, and Russians living within those territories separated themselves from Georgia. An uneasy standoff ensued, with Georgia maintaining the right of control there, even as Russians in both places looked to Moscow. In July 2008, the separatists in South Ossetia launched a series of missile raids on nearby Georgian villages. Georgia retaliated with ground forces on August 7. The Russian military immediately responded and quickly engaged Georgian troops the next day. Further Russian attacks came in Abkhazia. In 5 days, Russian assistance meant that Georgia was cast out of both South Ossetia and Abkhazia, causing it to lose some territory. A *détente* was reached, with Russia backing the two territories independently of Georgian rule.

Russian online activity preceded the ground attack by a day, initially in something of a trivial fashion, as Russian actors in cyberspace defaced websites of the Georgian state, including doctoring images and likening Georgian President Mikheil Saakashvili to Adolf Hitler.^[20] Observers correctly highlighted the more serious elements of the cyberattack, such as striking out at Georgian Government websites, the banking system, news outlets, and online discussion forums as an aggressive means of isolating the country from outside contact.^[21] These actions helped a Russian media blitz justifying the legitimacy of the Russian ground attack. Strategic messaging also spoke to Russian success in impacting the command and control of Georgian forces.^[22] In a week, Russia had pioneered a new way of fighting by teaming cyber capabilities with a conventional attack.^[23]

The timing of the cyberattacks to coincide with the Russian ground attack indicated a carefully coordinated strategy. But cyber actions were underway a month before the ground attack. One could view this as necessary reconnaissance to prepare the cyber offensive and then the ground attack.^[24] That probing certainly suggested a looming attack, all but forfeiting surprise and alerting the target to its danger. In this light, the Russian ground offensive on August 8 did not represent a planned date of attack, but a point of no further recourse other than to attack in order to take advantage of the very real cyber disruption already ongoing in Georgia and soon to be readily apparent to the outside world. While long expecting a confrontation with Georgia,

Russian leadership was caught off guard by the timing of the hostilities.^[25] Russian planners had to incorporate the cyber element into the offensive both to gain military advantage and to head off the potential of a public presence online to impede those plans and take things in an unwanted direction.

The success in controlling the online elements was mixed. Youth groups again went into action and, in the name of patriotism, targeted specific websites. Much of this traffic had Russian sponsors co-opting this online movement.^[26] More telling was the suspicion that criminal organizations answered the Russian Government's call to action and engaged in the familiar DDoS attack.^[27] The Russian Government disavowed these actors, again taking advantage of attribution difficulties to disguise the fact that the government had been blindsided by the chaos unfolding online. Even if the Russian Government was teaming with such actors, the need to have to look to such unreliable online partners risked throwing Russian military plans into disarray. The message here was not that Russia had unleashed a devastating military attack, but that its online community was impacting the foreign policy actions of a government forced to keep pace with this new online offensive. This became more visible when cyberattacks continued after the cessation of ground operations as the Russian online community took the lead, using forums, blogs, and websites.^[28]

Despite efforts at control, Russian cyberattacks could not stop Georgians from blogging, detracting considerably from the Russian effort to enjoy information dominance over the battlefield.^[29] Other failures to isolate the cyber battlefield threatened to escalate the conflict. Most significantly, Georgia, in response to the cyberattacks, shifted access to a server based in the state of Georgia in the United States without U.S. Government approval. The Russian aggressors online followed them there.^[30] Now, a border dispute in the Caucasus region threatened to include an offensive cyber action on American soil. How should the United States respond? Furthermore, Georgia was pursuing membership in NATO and a Russian attack could have triggered a response from that organization, thereby escalating the local conflict. But NATO had not responded in that fashion when Estonia, an alliance member, experienced its cyberattack, with member states deciding that the strike did not amount to an attack.^[31] The possibility of NATO taking action in the case of Georgia over mere cyber events was remote, but a ground attack could have provoked a different response.

The last thing Russia wanted was a clash with NATO.^[32] Disarray after the Soviet Union's demise left Russian military forces in marked decline both in quality and capability. The First Chechen War exposed these shortcomings, and not much had changed almost 2 decades later.^[33] Still, a naval action as well as air assets accompanied the attack on Georgia, and this joint force spoke to some Russian vitality of arms. However, the need to supplement the district forces with outside specialized units further stoked the fear that a military action could flounder, given the poor state of Russian arms. Those planning the attack employed overwhelming numbers, with 30,000 Russian troops—double the size of the Georgian military.^[34] The clash

that followed needed to be brief to avoid triggering adventurism in countries along Russia's periphery. The struggle in Chechnya had devolved into an ugly guerrilla war that included acts of terrorism in Russia itself, plunging the Russian home front into discord. Having another protracted border dispute on its hands in 2008 could well have crippled Russian efforts to recover from the 1991 collapse.

The good news of a very short, limited conflict in Georgia was dampened by more troubling developments when the cyber element of the clash was considered. A close look at the cyber events surrounding the Russian attack on Georgia presented observers with a Russian reaction to online activity it strove to control, rather than a carefully planned test of future war in the hands of a sophisticated Russian army. Theoretically, the two purposes could coexist. Russia could test its ability to use cyberattacks in conjunction with conventional force. A formal Russian military response followed a barrage of online attacks on Georgia, signaling an evolutionary step in warfare, as kinetic force teamed with cyber actions designed to prepare the invasion. This synergy certainly defined events in Georgia, but alarmed Western observers then missed the key, related significance of that episode. The foreign policy goals of the Russian state would be set by its government, not by popular mandates online enabled by a handful of computer adventurers. That activism smacked of populism in far too clear a way to be tolerated. Russian intervention in Georgia cemented resolve among the leadership to seize control of the patriotic hackers so markedly unrestrained in this domain.

To achieve this end, the risk of a larger war was worth it. For Russia, the main struggle was heading off democratic movements in neighboring territories. Georgia had endured this fate in late 2003 with the Rose Revolution that brought Saakashvili to power. A year later, vast public protests deposed the leader of Ukraine during the Orange Revolution, and a year after that, the leader of Kyrgyzstan with the Tulip Revolution. Putin's antipathy for Saakashvili underscored his determination to humble all instances of these "Color Revolutions."^[36] By 2008, Putin, now prime minister, had helped orchestrate Russian military action against Georgia, but that strike failed to topple his rival and, in at least one way, made matters worse. Encouraging separatists abroad invited such dissidence to spill over into Russia.^[37] An activist cyber element compounded that risk and blunting that online presence to help shore up the homeland would come next.

CONSOLIDATION: 2008–2014

The events in Estonia stress that openness had fostered a rogue element within Russian politics that acted by its own compass and initiated Russian cyber actions against the Baltic state. What transpired in Georgia just over a year later reflects the Russian Government's endeavor to tailor online realities in favor of state authority, with imperfect results. After 2008, this aim became Putin's aim. Having given up the presidency, he looked to stay in charge in a state which ostensibly curtails such permanence. Operating in elite circles, he overwhelmed his peers in government, manipulating state offices and the personnel holding those offices.^[38]

Such politicking was an obvious step away from robust democracy, as was the next, related effort. The public also had to accept a strongman, or, at least, centralized power in one office. But in this case, the Russian inclination to gravitate to personality rather than process and favor authoritarianism collided with online capabilities offering to blunt this sentiment. The ability of Russia to change from the old ways to the new came face-to-face with the openness that defines cyberspace.

Putin already felt threatened by public demonstrations that, in his view, had helped West Germany absorb East Germany, starting a reaction that eventually destroyed the Soviet Union.^[39] Indeed, protests had surfaced in Russia as he plotted his return to the presidency in 2012. On December 10, 2011, Russians rallied against fraudulent parliamentary elections during the Snow Revolution. On May 6, 2012, large crowds protested Putin's pending inauguration as president the next day with the March of Millions.^[40] Once regaining that office, Putin cracked down on such groups within Russia. The Duma allowed the targeting of foreign groups that had accepted outside money. Russian Government spokesmen tied any protests to Western influence coming from organizations such as the United States Agency for International Development and nongovernmental organizations, and so-called liberal outlets were harassed by government operatives.^[41]

The internet age complicated matters because opposition groups within Russia enjoyed an online presence, a sign that traditional adherence to government decree was suspect in the extreme. Social media played a leading role in posing an internal threat, helping independent organizations manipulate people into street demonstrations.^[42] To rebuff what was no less than Western interference in Russia's internal affairs, the Kremlin had to act: Internet use had to be controlled; dissent relabeled slander and libel and, therefore, a criminal act; and websites blacklisted, then blocked.^[43] These measures underscored Putin's desperation to crack down on the internet, something he publicly labeled no more than a Central Intelligence Agency project in April 2014.^[44] When National Security Agency (NSA) contractor Edward Snowden released NSA-classified information starting in June 2013, he exposed some of that agency's online surveillance efforts and helped Putin justify his actions.^[45]

The Russian Government's ability to shore up things at home still did not address how connectivity aided what in Putin's mind amounted to fifth columns that imposed a democracy beholden to Western interests on Russia's neighbors.^[46] Putin responded with his own Color Revolutions. To this end came a government-led campaign extolling a pure Russian identity based on true Russian cultural values. Russia could go on the offensive by the means of a "Eurasianism" ideology announcing values as the key weapon to reasserting a Russian-led heartland.^[47] Russia had its own story to tell in this regard: it was a nation long beleaguered by Western threats and actions. In this respect, the information battleground was a key asset: a means to sow discord within states by reinforcing prejudice and bias among diverse populations that would rally to Russia because of a shared persecution.

Russia brazenly tested this approach by orchestrating a takeover of Crimea in the name of supporting an indigenous revolt of ethnic Russians against Ukrainian rule in February 2014. Regardless of widespread dissension and a tangible groundswell in favor of Crimea joining the Russian Federation, the action of Russian military forces (minus uniform markings and identification) proved decisive.^[48] Ukraine faced the prospect of armed confrontation with para-Russian forces and chose not to engage. The ensuing information campaign by Russian authorities merely announced the supposed proclivity of Crimea to seek separation from Ukraine and then demand annexation to Russia. These two outcomes came to pass rapidly and, by March 2014, Ukraine had lost control of that province. Putin then proclaimed a triumph of nationalism and the Russian public accepted the results as a measure of ancient Russian suzerainty in the region at the expense of Western interference.

Ukraine's renewed, internal, political turmoil had opened the door to this Russian adventurism in Crimea. Ukraine's Euromaidan reaction of February 2014 deposed the current president, who favored closer ties with Moscow. Putin countered with the conquest of Crimea, making clear that these popular movements now faced the prospect of Russian intervention, including the use of ground forces.

Continuing to pressure Ukraine, Russia at first repeated the Crimean pattern of supporting internal forces willing to engage in violence to challenge Ukrainian rule. In the Donbas in eastern Ukraine, a region that is home to a large Russian population, Putin supplied arms to dissidents and at times committed Russian paramilitary forces, foisting a battle onto the again-reluctant Ukrainian Government. As this struggle continued over an extended period of time, Russia appeared unwilling to seek annexation and, instead, hoped that that state could come under the umbrella of Russian influence, if not in declarations of subservience, then in the unsettled notion of state security.^[49] To this end, Russia turned to a sophisticated cyber effort—one tangibly more physical than seeking an information-operation success by inciting internal dissent. Instead, Russian cyberattacks successfully targeted the Ukrainian power grid in December 2015. That act forced three distribution centers offline for several hours, impacting 220,000 residents.^[50] This strike represented strategic cyber power, but Russian forces unleashed tactical actions as well. Malware helped Russian-backed rebels in the Donbas to attain the “locational data” of Ukrainian artillery and target those units for destruction.^[51] Altogether, the concept that Russia sought to test—cyber capabilities in conjunction with acts of war—gained much credence. In the seams between cyberspace and ground conflict came an effort to enable a physical means of disruption on the ground which coexisted alongside the same effort of physical disruption via cyberspace. That challenge indeed left its victims in the grip of a seemingly perpetual assault that reminded nations to think twice about embracing connectivity as a means of feeding a global democratic inevitability.

STRATEGIC ASCENDANCY: RUSSIA ON THE ATTACK

In Russian hands, a deliberate effort to curb any notion of a shared online space hosting a community of users to achieve a more enriched body politic came by conducting cyberattacks alongside deploying paramilitary force in neighboring countries. But a longer reach was needed to impact world events in which the confrontation would be strictly cognitive. One could not simply stand by and receive the daily offensive from those enjoying connectivity in cyberspace. Blunting cyber activities to secure internal and even regional consolidation was one priority and turning openness on its head the other. It would not be a big leap to use cyber to sow doubt in an adversary's national sovereignty well beyond Russian borders. The primary target was obvious. The online threat had to be met at its source, and this meant eroding the standing of the creator of this platform, the United States.

It did not take much to cast American confidence in the democratic process in stark relief to a technological age that exposed that very sentiment as obsolete. Unleashing an army of trolls that dispenses fake news has almost done the trick of getting the United States to distrust and question an open internet. The distribution of disinformation, sowing of doubt in trusted institutions, and injection of paranoia into the American body politic were not new. What was new was the willingness of the American public to accept these efforts as proof of untrustworthy online interaction, of seeing only a nemesis in cyberspace. Openness became the foremost casualty of now-suspicious interactions in cyberspace, as had to be the case from Russia's point of view to offset the strategic ascendancy inherent in the very act of being online. Exchanging and sharing information among internet users became more a worry, less a right.

This success meant a Russian strategic high ground in cyberspace, which was no small accomplishment, given the threat the platform had posed to an authoritarian Russian state. The nemesis of cyber rebellions at present appears quiescent. Putin again stood for election in 2017 and, according to media reports, won by overwhelming mandate. His success points to very little political opposition or unrest within Russia, suggesting that the potential for online activism is well under control. Moreover, the hack of the U.S. Presidential election indicates that Putin has learned his lessons well and authored his own form of cyber rebellion within U.S. borders designed to undermine democracy.^[52] The response in the United States to better defend cyberspace means a retrenchment from openness and a further gain for the Russian strongman. In seeking greater online security, Americans no longer press the advantage of an open internet giving a voice to political expression. In abandoning the ideological high ground in cyberspace, U.S. officials offer Russia a much sought-after reprieve from facing political rancor and agitation in a nation that otherwise does not allow such dissent. That discourse is, of course, the hallmark of democracy, not a call for oversight, as Putin would have the world believe. It seems that too many Western leaders must relearn this basic lesson in representative government and protect the right to information and what amounts to virtual assembly online.

The United States must serve as a measuring stick for the rest of the world and then reap a concomitant benefit from the ideological dimensions of cyberspace. In that scenario, Russia would again be forced to play defense.

States endorsing political plurality merely have to defend and advance openness to blunt the Russian cognitive offensive in cyberspace. As was the case during the Cold War, an ideological struggle between authoritarianism and liberalism has again become central to U.S.-Russian relations.^[53] When Cold War parameters help shape cyberspace, a new period of containment emerges as a means of defending openness in that domain. This cognitive stand online cements ideology as paramount in conflict by ensuring an arena of shared values that challenge authoritarian rule, a success that would mean the strategic initiative lies in Western hands, or the hands of those who favor openness. Better still, openness is already U.S. policy, is already endorsed by the private sector that does so much to shape that domain and is already a means of delivering a nonviolent offensive in a war no less imperative to win than a physical war in other domains. This recognition means that U.S. efforts must not await a ground war to team with cyber capabilities and thereby present a familiar picture of war. Rather, one must embrace the ongoing online ideological struggle to maintain the permanent, strategic advantage of openness in cyberspace. From that strategic high ground, one can go on the attack in cyberspace by offering a universal appeal to an online global commons that serves democracy. ♥

NOTES

1. The US cyber policy defending and advancing openness rests on a series of public documents. See “The International Strategy for Cyberspace,” White House, May 2011; Department of Defense, “Strategy for Operating in Cyberspace,” July 2011; Department of Defense, “Cyber Strategy,” April 2015; Department of State, “Cyber Strategy,” 2016, and from the Trump administration, Presidential Executive Order, “Cyber Security,” May 11, 2017.
2. John Arquilla highlighted this potentiality early on. See Arquilla and David Ronfeldt, “Cyberwar is Coming!” *Comparative Strategy*, Vol. 12, No. 2 (Spring 1993): 144,145. Martin C. Libicki said something similar when examining what he called “friendly conquest,” or ideologically driven conflict over values online. See Libicki, *Conquest in Cyberspace: National Security and Information Warfare* (New York: Cambridge University, 2007), 4, 230. Many authors withhold judgment on whether mere connectivity can force authoritarian regimes to liberalize given the success of strongmen using the internet to augment state oversight thus far. See Jay Blumler and Stephen Coleman, *The Internet and Democratic Citizenship: Theory, Practice and Policy* (New York: Cambridge University, 2009), 9; Larry Diamond, “Liberation Technology,” *Journal of Democracy*, Vol. 21, No. 3 (July 2010): 70; and recently, Vincent Mosco, *Becoming Digital: Towards a Post-Internet Society* (Bingley: Emerald Publishing, 2017), 14.
3. For Estonia, see Gadi Evron, “Battling Botnets and Online Mobs: Estonia’s Defense Efforts During the Internet War,” *Georgetown Journal of International Affairs*, Winter/Spring 2008: 123; Robert A. Miller and Daniel T. Kuehl, “Cyberspace and the ‘First Battle’ in 21st-century War,” *Defense Horizons*, No. 68 (September 2009): 3; Richard A. Clarke and Robert K. Knake, *Cyber War: The Next Threat to National Security and What to do About It* (New York: HarperCollins, 2010), 16; *International Cyber Incidents: Legal Considerations* (Tallinn, Estonia: Cooperative Cyber Defense Center of Excellence, 2010), 23; Stephen Herzog, “Revisiting the Estonian Cyber Attacks: Digital Threats and Multinational Responses,” *Journal of Strategic Security*, Vol. 4, No. 2 (Summer 2011): 51; and Alison Lawlor Russell, *Cyber Blockades* (Washington, DC: Georgetown University, 2014), 86. For Georgia, see *Cyber Attacks Against Georgia: Legal Lessons Identified* (Tallinn, Estonia: Cooperative Cyber Defense Center of Excellence, 2008): 9-10, 12; *International Cyber Incidents*, 75-76; Clarke and Knake, *Cyber War*, 20; Paulo Shakarian, “The 2008 Russian Cyber Campaign Against Georgia,” *Military Review* (November-December 2011): 63-64; Russell, *Cyber Blockades*, 105.
4. David E. Sanger, *The Perfect Weapon: War, Sabotage, and Fear in the Cyber Age* (New York: Crown, 2018), xv-xvi.
5. Advocates of a cyber Westphalian norm do the most damage in this regard, appealing to the establishment of national borders in Europe in 1648, a development that fed authoritarian rule. See Chris C. Demchak and Peter Dembrowski, “Rise of a Cybered Westphalian Age,” *Strategic Studies Quarterly* (Spring 2011): 35, 37.
6. *International Cyber Incidents*, 16; Evron, “Battling Botnets and Online Mobs,” 122; and Russell, *Cyber Blockades*, 75.
7. *International Cyber Incidents*, 23; and Herzog, “Revisiting the Estonian Cyber Attacks,” 51.
8. *International Cyber Incidents*, 20; Russell, *Cyber Blockades*, 75; and Evron, “Battling Botnets and Online Mobs,” 123.
9. *International Cyber Incidents*, 21; Herzog, “Revisiting the Estonian Cyber Attacks,” 52.
10. Russell, *Cyber Blockades*, 79.
11. Joshua Davis, “Hackers Take Down the Most Wired Country in Europe,” *Wired Magazine*, Issue 15.09, August 21, 2007.
12. Evron, “Battling Botnets and Online Mobs,” 123.
13. Herzog, “Revisiting the Estonian Cyber Attacks,” 54; Evron, “Battling Botnets and Online Mobs,” 124. The Russian Deputy Press Secretary Dmitry Peskov called the attack an act of terrorism. See Russell, *Cyber Blockades*, 82.
14. William Zimmerman, *Ruling Russia: Authoritarianism from the Revolution to Putin* (Princeton, NJ: Princeton University, 2014), 2. The Russian people acquiescing to authoritarianism is in Richard Pipes, “Flight from Freedom: What Russians Think and Want,” *Foreign Affairs* 83, 3 (May-June 2004): 15; and Gregory Feifer, *Russians: The People Behind the Power* (New York: Twelve, 2014), 8.
15. Michael Connell and Sarah Vogler, “Russia’s Approach to Cyber Warfare,” *CNA* (March 2017): 10.
16. Herzog, “Revisiting the Estonian Cyber Attacks,” 55; Häly Laasme, “Estonia: Cyber Window into the Future of NATO,” *Joint Force Quarterly*, Issue 63, 4th Quarter (2011): 60.
17. E. Lincoln Bonner III, “Cyber Power in 21st Century Joint Warfare,” *Joint Force Quarterly*, Issue 74, 3rd Quarter (2014): 103.
18. A botnet refers to computers marshaled together via the internet and answerable to an individual who forwards transmissions usually without the owner’s awareness. A DDoS or distributed denial of service attack uses compromised systems and often botnets to overwhelm a targeted system.
19. Andrei Soldatov and Irina Borogan, *The New Nobility: The Restoration of Russia’s Security State and the Enduring Legacy of the KGB* (New York: Public Affairs, 2010), 238, 3.

NOTES

20. Shakarian, "The 2008 Russian Cyber Campaign Against Georgia," 64; *Cyber Attacks Against Georgia*, 7-8.
21. *International Cyber Incidents*, 70.
22. Miller and Kuehl, "Cyberspace and the 'First Battle' in 21st-century War," 2, 5.
23. Miller and Kuehl, "Cyberspace and the 'First Battle' in 21st-century War," 1, 2.
24. Miller and Kuehl, "Cyberspace and the 'First Battle' in 21st-century War," 3; *International Cyber Incidents*, 69.
25. Andrei Illarionov asserts that Russia acted on a plan in place for ten years. See Illarionov, "The Russia Leadership's Preparation for War, 1999-2008," in *The Guns of August: Russia's War in Georgia*, eds. Svante E. Cornell and S. Frederick Starr (New York: Routledge, 2015), 50. For Russian leadership being surprised by the timing of hostilities, see Ariel Cohen and Robert E. Hamilton, *The Russian Military and the Georgian War: Lessons and Implications*, Strategic Studies Institute, US Army War College, Carlisle, PN (June 2011), 22, 23.
26. "Russia/Georgia Cyber War – Findings and Analysis," Project Grey Goose, Phase I Report (17 October 2008): 6-8.
27. Shakarian, "The 2008 Russian Cyber Campaign Against Georgia," 64; and Russell, *Cyber Blockades*, 109-110.
28. *International Cyber Incidents*, 68, 71.
29. Paul A. Goble, "Defining Victory and Defeat: The Information War Between Russia and Georgia," in *The Guns of August: Russia's War in Georgia*, eds. Svante E. Cornell and S. Frederick Starr (New York: Routledge, 2015), 191.
30. Stephen W. Korns and Joshua E. Kastenber, "Georgia's Cyber Left Hook," *Parameters* (Winter 2008-2009): 60; and *International Cyber Incidents*, 70, 77.
31. Häly Laasme, "Estonia: Cyber Window into the Future of NATO," *Joint Force Quarterly*, Issue 63 (4th Quarter 2011): 60.
32. Herzog, "Revisiting the Estonian Cyber Attacks," 53.
33. Zoltan Barany, *Democratic Breakdown and the Decline of the Russian Military* (Princeton, NJ: Princeton University, 2007), 95, 99-100.
34. For the Russian order of battle, see Cohen and Hamilton, *The Russian Military and the Georgian War*, 10; and Pavel Felgenhauer, "After August 7: The Escalation of the Russia-Georgia War," in *The Guns of August: Russia's War in Georgia*, eds. Svante E. Cornell and S. Frederick Starr (New York: Routledge, 2015), 166-167.
35. Lincoln A. Mitchel, *The Color Revolutions* (Philadelphia, PN: University of Pennsylvania, 2012), 113.
36. Ronald D. Asmus, *A Little War that Shook the World: Georgia, Russia, and the Future of the West* (New York: Palgrave Macmillan, 2010), 179.
37. Goble, "Defining Victory and Defeat," 190. See also Angela Stent, *The Limits of Partnership: US-Russian Relations in the Twenty-First Century* (Princeton, NJ: Princeton University, 2014), 101, 115; Fiona Hill and Clifford G. Gaddy, *Mr. Putin: Operative in the Kremlin* (Washington DC: Brookings Institution Press, 2013), 343; and Roger N. McDermott, "Learning from Today's War: Does Russia Have a Gerasimov Doctrine?" *Parameters*, Vol. 46, No. 1 (Spring 2016): 99, 101.
38. For elites dominating Russian political fortunes and Putin manipulating such factions, see Vladimir Gel'man, *Authoritarian Russia: Analyzing Post-Soviet Regimes Changes* (Pittsburgh, PA: University of Pittsburgh, 2015), xiv, 150; and Arkady Ostrovsky, *The Invention of Russia: From Gorbachev's Freedom to Putin's War* (New York: Viking, 2015), 8.
39. Hill, *Mr. Putin*, 363; and Steven Lee Myers, *The New Tsar: The Rise and Reign of Vladimir Putin* (New York: Alfred A. Knopf, 2015), 48.
40. Mischa Gabowitsch, *Protest in Putin's Russia* (Polity, 2017), 8, 38.
41. Hill, *Mr. Putin*, 348; and Richard Sakwa, *Putin Redux: Power and Contradiction in Contemporary Russia* (New York: Routledge, 2014), 169, 181.
42. Hill, *Mr. Putin*, 349; Stent, *The Limits of Partnership*, 100, 101.
43. Andrei Soldatov and Irina Borogan, *The Red Web: The Struggle Between Russia's Digital Dictators and the New Online Revolutionaries* (New York: PublicAffairs, 2015), xi.
44. Hill, *Mr. Putin*, 349.
45. Myers, *The New Tsar*, 441.
46. Hill, *Mr. Putin*, 348.
47. Myers, *The New Tsar*, 445-446.

NOTES

48. Mary Ellen Connell and Ryan Evans, “Russia’s ‘Ambiguous Warfare’ and Implications for the US Marine Corps,” CNA (May 2015): 9.
49. Mark Galeotti, “Hybrid, Ambiguous, and Non-linear? How New is Russia’s ‘New Way of War’?” *Small Wars and Insurgencies*, Vol. 27, No. 2 (2016): 285.
50. Michael Connell and Sarah Vogler, “Russia’s Approach to Cyber Warfare,” CNA (March 2017): 20.
51. “Use of Fancy Bear Android Malware in Tracking of Ukrainian Field Artillery Units,” Crowdstrike Global Intelligence Team, December 22, 2016; then UPDATED (corrected): “Use of Fancy Bear Android Malware in Tracking of Ukrainian Field Artillery Units,” Crowdstrike Global Intelligence Team, December 22, 2016. March 23, 2017. <https://www.crowdstrike.com/resources/reports/idc-vendor-profile-crowdstrike-2/>.
52. Connell and Vogler, “Russia’s Approach to Cyber Warfare,” 24.
53. Michael McFaul, *From Cold War to Hot Peace: An American Ambassador in Putin’s Russia* (New York: Houghton Mifflin Harcourt, 2018), x-xi.

Fake News, (Dis)information, and the Principle of Nonintervention

*Scope, limits, and
possible responses to cyber
election interference
in times of competition*

Annachiara Rotondo

*Department of Political Sciences
University of Campania Luigi Vanvitelli
Caserta, Italy*

Pierluigi Salvati

*Department of Political Sciences
University of Naples Federico II
Naples, Italy*

ABSTRACT

In the era of asymmetrical conflicts, information and communication technologies (ICT) play an essential role due to their importance in the manipulation and conditioning of public opinion.^[1] Several threats are linked to the use of ICT but, in terms of interstate, strategic competition, one of the main dangers is represented by so-called “cyber election interference” (i.e., cyber election-meddling activities carried out by foreign states to influence the electorate of a target state through the diffusion of “fake news” or “alternative truths,” principally via the media and social networks (Facebook, Twitter, YouTube, etc.)). The aim of this paper is to clarify whether and when this kind of interference constitutes a breach of international obligations—in particular, of the principle of nonintervention in the internal affairs of a state—and to envisage possible lawful responses under international law by states targeted by said interference.

Keywords— cyber election meddling, international law, principle of nonintervention, options for response.

I. INTRODUCTION

Although the interference of foreign states in the electoral processes of other states has certainly occurred in the past,^[2] some recent elections and crucial referenda^[3] have brought a particular feature of this phenomenon to the attention of the international community—namely, so-called “cyber election interference.”^[4] This expression does not refer herein to the physical destruction of or tampering with equipment or electoral systems, or to the modification of the results through malware aimed at causing irregular recounting of votes.^[5]

Nor does it refer to operations of mere cyber-intelligence collection (i.e., aimed at gathering information on electoral processes and which do not seem to have *per se* characteristics of unlawfulness).^[6] The reference herein is rather to nondestructive phenomena with a

© 2019 Annachiara Rotondo, Pierluigi Salvati.

persuasive scope—that is, campaigns of (dis)information promoted by foreign states aimed at surreptitiously influencing the vote in another state through the diffusion of “fake news” or “alternative truths,” principally via the media and social networks (Facebook, Twitter, YouTube, etc.). The growing number of episodes of interference in said terms against fundamental electoral processes by foreign states makes it relevant to address the question of whether these activities constitute a breach of international law and, in particular, of the principle of nonintervention in the internal affairs of a state, and to envisage possible lawful responses.

II. CYBER ELECTION INTERFERENCE AND THE PRINCIPLE OF NONINTERVENTION IN THE INTERNAL AFFAIRS OF A STATE

Cyber election meddling can be defined as a cyber operation resulting in subtle campaigns of (dis)information.^[7] aimed at influencing the electoral vote and its outcome through the spread of fake news with a view to affecting the political and institutional system of the target state. In this case, foreign intervention takes the form of activities that are more or less nuanced and not always attributable, undermining the correct formation of the will of the target state in the definition of its own government apparatus; its institutional structure; and, consequently, the determination of its policies. This represents a potential violation of the principle of non-intervention in the internal affairs of a state, since the electoral process is the highest and most significant moment of expression of domestic jurisdiction.

The principle of nonintervention is a principle of general international law^[8] and has been constantly affirmed in the Resolutions of the United Nations General Assembly,^[9] with particular reference to the “sovereign and inalienable right of a State freely to determine its own political...system, to develop its international relations...without outside intervention, interference, subversion, coercion or threat in any form whatsoever,”^[10] and with specific reference to electoral processes (“the principle of...non-interference in the internal affairs of any State should be respected in the holding of elections”).^[11] However, said principle has often been linked to the (more restricted) principle of the prohibition of the use of force, leading some scholars to sustain a substantial overlap between them, as far as to consider the former as essentially absorbed by the latter.^[12]

The scope of the principle of nonintervention has been further examined by the International Court of Justice (ICJ) in the judgment *Military and Paramilitary Activities in and against Nicaragua (Nicaragua v. United States of America)*. Therein, the court clarified the notion of “unlawful intervention,” on the one hand by delimiting its extent to matters of the target state’s domestic jurisdiction,^[13] and on the other hand by identifying the use of methods of coercion regarding these matters as its defining characteristic.^[14] Therefore, in its statement, the court identified in coercion—«*which defines, and indeed forms the very essence of prohibited intervention*»—the parameter to affirm the unlawfulness of an episode of interference. Although the ICJ has observed that said element is *ipso facto* subsistent in the case of the use of force,^[15] however, it

did not intend to reduce the hypothesis of coercive intervention exclusively to the use of force, which is albeit considered paradigmatic of the phenomenon. Nonetheless, by omitting further examples,^[16] the court did not contribute either to understanding how coercion can concretize under the threshold of the use of force or whether it is necessarily constituted by a wrongful act (or the threat of a wrongful act).^[17] Therefore, the wording of *Nicaragua* does not seem to be particularly effective in identifying further hypotheses of coercive intervention falling below the threshold of art. 2(4) of the UN Charter, as is the case with the (dis)information campaigns which, by their very nature, do not involve the use of force.

Some authors assert that coercion could be recognized not only in the exercise (or the threat) of a wrongful act, such as the use of force, but also in the forced modification of the «*normal or natural or expected course of events*».^[18] This approach is absolutely relevant for a broader interpretation of the concept of coercion beyond the paradigm provided by the ICJ in *Nicaragua* as it disconnects said notion from the threat or implementation of an unlawful act^[19] by anchoring it to a “neutral” element (i.e., the achievement of a fact which, without the foreign intervention, would not have occurred: it would be precisely the modification of the natural course of events which would make the aforementioned intervention “coercive”). Also the Group of Expert Editors (GEE) of the Tallinn Manual 2.0 on the International Law Applicable to Cyber Operations (hereinafter “Tallinn Manual 2.0”) found that a coercive act “must have the potential for compelling the target State to engage in an action that it would otherwise not take,”^[20] so decoupling the concept of coercion from the commission of a wrongful act and linking it to the constraint for the target state to act in a way in which it would not have otherwise acted. Therefore, interpreting in this sense the concept of coercion, activities aimed at influencing the determination of political choices impacting on electoral processes might result in a coercive interference and, thus, a violation of the principle of nonintervention.

On this point, the Tallinn Manual 1.0 on the International Law Applicable to Warfare (hereinafter “Tallinn Manual 1.0”) already clearly stated that “cases in point [i.e., coercive] are *the manipulation by cyber means of public opinion elections* [emphasis added], as when online news services are altered in favour of a particular party, false news is spread...”^[21] This approach is undoubtedly more suitable for extending the scope of coercion beyond the silences of *Nicaragua*, and is particularly relevant with reference to the cyber operations under examination. Indeed, election meddling in the terms under discussion could result in a coercive interference «designed to deprive another State of its freedom of choice...to force [the] State to act in an involuntary manner or involuntarily refrain from acting in a particular way.»^[22]

In this case, a key element of coercion seems to be identified in the covert nature of the foreign interference. The target state would find itself, in fact, in a situation of coercion “unknownst to it” (i.e., without knowing it was being manipulated). The unlawful intervention—consisting of influencing the sentiments of the populace with a view to determining the results of elections—would, in this case, materialize as constraint through induction in that the same

state would be led to take fundamental choices without having determined them autonomously and freely, thus resulting in a coercive modification of the normal or natural course of events.

A different approach based on an interpretation of coercion in terms of scales and effects achieved by the foreign intervention would lead to similar, although not identical, results. The classical doctrine has, in fact, dwelt on the “dimensions of consequentiality” which define coercion and has identified as relevant «the importance and number of values affected, the extent to which such values are affected and the number of participants whose values are so affected,»^[23] which Professor Watts transposed *mutatis mutandis* into the framework of cyber operations and translated into the «nature of State interests affected...the scale of effects the operation produces in the target State, and the reach in terms of number of actors involuntarily affected...»^[24]

In the case of cyber election interference, all of these “dimensions” seem to be achieved. The free and sovereign determination of the political and institutional apparatus, and consequently of national and foreign policies, appears to be a primary interest of the state which is affected by foreign meddling. Moreover, said activity may reach, through the widespread diffusion of fake news via the media and social networks, most of the electorate, influencing its orientation in a decisive way, therefore causing it to act (i.e., to vote) on the basis of false information, which results in a manipulation of its determinations. As for the outcomes of interference, the GEE of the Tallinn Manual 2.0 affirmed that the scale of effects produced could not be limited in terms of desired results, since the violation of the principle in question does not require the intervention to be successful; therefore, simply forcing the electoral process may amount to a breach of the principle of non-intervention, not being necessary to the successful pursuit of the objective set by the foreign state.^[25] In reality, the question of the outcome of the interference remains a debatable issue. It depends on what one considers a coercive act with the potential to compel, or solely an act which effectively compels the target state to engage in a course of action that it would otherwise not undertake.

Of course, the overall approach above should not be overestimated. It is evident that not any hypothesis in which one state pushes another to act differently to how, in the absence of its intervention, it would otherwise have acted may represent coercive interference. In fact, for the purpose of the configurability of coercion, it is necessary that the target state is, in fact, “forced” (i.e., it has no other choice or option),^[26] which mostly translates into its unawareness of being manipulated, it not being sufficient or relevant that the same has consciously modified its behaviour simply because it considers it to be advantageous (or to avoid a disadvantage).

Therefore, cases of foreign influence, such as a public campaign promoted by a foreign state aimed at inducing another state to act in a determined way (e.g., to ratify a treaty) or the endorsement of a foreign leader in favor of the election of a candidate through the media^[27] cannot be considered a violation of the obligation to abstain from interfering with the internal affairs of a state. In these cases, in fact, the character of coercivity is lacking. And even cyber

operations aimed at influencing a state to comply with an international obligation would not constitute a violation of the principle of nonintervention inasmuch as the subject matter is not among those in which the state «is permitted to decide freely» under *Nicaragua*^[28] since the international obligation externalizes *ipso iure* compliance beyond the scope of the domestic jurisdiction.^[29]

Consequently, it is not easy to achieve a unitary reconstruction of a regime of foreign intervention aimed at meddling in elections through the spread of fake news, but it is necessary to carry out a holistic check on a case-by-case basis.^[30] Therefore, the interference of a foreign state in the electoral process of another state may result in different legal qualifications, depending on the activities carried out. Thus, cyber election interference resulting in propaganda; the dissemination of real news; or, on the contrary, fake news, in order to influence foreign political and electoral processes will be subject to a different regime, depending on the existence and the degree of coercion.

For this reason, for example, the lawfulness of public propaganda activities promoted by a foreign state has been affirmed: publicity, in fact, excludes the element of coercion, and thus such activity—although it may represent an unfriendly act—cannot be said to be wrongful, at least with respect to the prohibition of interference, unless a different prohibition at the level of a specific rule is provided.^[31]

III. QUESTIONS OF ATTRIBUTION

In order to result in a breach of the duty of non-intervention, cyber election interference must be attributable to a foreign state. In fact, attribution is an indispensable element in order to consider a determined act as an internationally wrongful act, as provided for by art. 2 of the draft Articles on the Responsibility of the States for Internationally Wrongful Acts (ARSIWA),^[32] which reads that, “There is an internationally wrongful act of a State when conduct consisting of an action or omission: (a) is attributable to the State under international law; and (b) constitutes a breach of an international obligation of the State.” As a matter of law, the burden of attribution has to be resolved on a case-by-case basis under strict adherence to the principles provided for under chapter II of ARSIWA. Therefore, cyber election interference can be attributable to a foreign state if mainly carried out by an organ of said state (art. 4); persons or entities exercising elements of governmental authority of said state (art. 5); organs placed at the disposal of a state by another state (art. 6); or a person or group of persons acting on the instructions of, or under the direction or control of, that state in carrying out the conduct (art. 8).^[33]

A formal attribution to a state organ under art. 4 of ARSIWA would represent the most direct ascription of the alleged interference to a foreign state, as it would be possible to trace back the intervention and attribute it, even if it was carried out *ultra vires* (i.e., beyond the responsibility assigned to said organ),^[34] and even in the case of *de facto* organs. For example,

a report released in 2017 by the Central Intelligence Agency, Federal Bureau of Investigation, and National Security Agency under the auspices of the Office of the Director of National Intelligence analyzed the “influence campaign” allegedly conducted by Russia in order to meddle in the 2016 U.S. presidential election and assessed that it was approved at the highest level of the Russian Government. In particular, the report denounced the participation of the main Russian intelligence service (the Main Directorate of the General Staff of the Armed Forces of the Russian Federation) as well as the direct involvement of President Vladimir Putin, who ordered said campaign, according to the report.^[35] In this case, the U.S. intelligence agencies have clearly attributed these activities to Russia, although they did not provide evidence in order to avoid identifying their sources, thus allowing the (alleged) offending state to reject charges.

However, attribution to a state organ may be, in practice, complex because often such activities are carried out by foreign secret services, and so are difficult to trace. Even when they are manifestly attributable to foreign state organs, their formal ascription to the foreign government concerned in terms of international responsibility is a further step which is not always taken by the target state.^[36] For example, even though Special Counsel Robert Mueller’s office identified Guccifer 2.0 as a Russian intelligence officer in the light of forensic determination and indicted him for crimes related to the alleged hacking of the Democrats in 2016,^[37] the response by U.S. authorities against the Russian Government was limited, after some hesitation, to a mere public accusation, which did not result in any consequence under the international law of responsibility.

Very often, cyber election interference is carried out by non-state actors acting “on the instructions of, or under the direction or control of,”^[38] a foreign power to interfere with the target state’s political system—e.g., in the case of the Internet Research Agency (IRA), a Russian company allegedly linked to Moscow and accused by the U.S. of having hired hundreds of “trolls” to post fake news and socially divisive content on social media, such as Facebook, Twitter and YouTube, and share it among millions of people.^[39] In cases like this, attribution to a foreign state under art. 8 of ARSIWA lends itself to further and more complex problems. Indeed, if the concepts of “instruction,” “direction,” and “control” are broadly meant to be understood as disjunctive,^[40] therefore potentially broadening the scope of attribution, the degree of control required in attributing an act committed by non-state actors to a foreign state must be identified when the state in question “directed or controlled the specific operation,” and “the conduct complained of was an integral part of that operation.”^[41] Yet, in this case, it may amount to “effective control” in the terms outlined by the ICJ in *Nicaragua*^[42].

However, in most cases, neither the effective control test nor the different “overall control test”^[43] developed by the International Criminal Tribunal for the former Yugoslavia (ICTY) in the case *Prosecutor v. Duško Tadić (Appeal Judgement)*—which lowered the standard of attribution—represent a sufficient solution, as both of them require a level of control and evidence regarding non-state actors which is hard to establish in relation to cyber election interference.

Moreover, in most cases, a sure attribution of cyber interference is not possible because of purely technical problems: in fact, hackers' activities, as well as the active perpetrators of foreign interventions, can hardly be traced. Also, the identification of the origin of internet protocol routings, spoofing, and other cyber means, as well as possible similarities among malware used in hacking or spreading fake news involving a determined foreign state,^[44] can be considered a clue but not decisive evidence in attributing a cyber operation to said state. Even if the target state is successful in linking a determined cyber operation to a foreign state-owned infrastructure, this does not allow the target state to conclude definitively either that such cyber action effectively originated from that place or that such identification can be considered more than an indication that the state of origin may be involved with the interference. This is because non-state actors, or other states interested in muddying the waters, may have acquired control over such infrastructure.^[45]

Therefore, using classical standards of proof may often result in the failure to attribute these kinds of operations to a specific foreign state.^[46] In all of these cases (i.e., when attribution is not certain in legal terms), said activities could not amount to an internationally unlawful act, lacking one of the two essential conditions provided by art. 2 ARSIWA.

The question seems to be often faced by target states as a matter of fact, and therefore mainly subjected to standards of reasonability, resulting in accusations of cyber meddling which respond to prevailing political purposes and do not translate into a manifest accusation against the foreign state of having committed an "international wrongful act."^[47] In these hypotheses, electoral intervention may be considered at the least to be an unfriendly act, without entailing the international responsibility of the acting state.

IV. CYBER ELECTION MEDDLING: OPTIONS FOR RESPONSE UNDER INTERNATIONAL LAW

Even though international responsibility arises simply from the commission of an internationally wrongful act by a state, if the injured state aims to seek cessation of the conduct or to obtain reparations, it has to react through mechanisms provided by international law. This is because a lack of response may have legal consequences, such as the loss of the right to invoke responsibility, as is the case in waiver or acquiescence.^[48]

International law offers several options for response, the choice of which is not driven by the rule of international law, but which depends on the overall balance of the opportunities and purposes of the target state. Particularly, in the case of cyber election meddling, the choice of response is strictly connected to the possibility of confirming the violation of the principle of non-interference and to the ability of the target state to attribute the violation to another state.

A. Waiver or Acquiescence

The practice shows that often, even in the presence of strong suspicions allowing attribution, states sometimes choose not to react at all.^[49] The option of non-reaction against a wrongful act

configures the hypothesis of implicit waiver or acquiescence, which can represent a feasible option for an injured state. This is because the target state which has reached the proof of attribution of the cyber violation committed may not want to reveal the same in order to protect its sources and intelligence means. Inactivity seems also to respond to the will of the target state to carry out the same interference activity in turn, on the basis of a *tu quoque* practice, and not contribute to forming an express, prohibitive rule.

However, it is necessary to underline that the option of waiver precludes any claim for reparation, as does the option of acquiescence. Obviously, a waiver is considered effective only if given in a valid manner, thus excluding all cases in which states express a waiver under the coercion of another state, or because of the existence of a material error.

Equally, acquiescence, as pinpointed by the ICJ in the *Certain Phosphate Lands in Nauru* case, determines the loss of the state's right to invoke responsibility.^[50] Consequently, if the target state opts for non-reaction, it has to consider that it is excluding every future possibility to act against the perpetrator of the violation through instruments provided by international law (doctrine of estoppel).

B. Countermeasures

When cyber election meddling violates the principle of nonintervention in internal affairs and is attributed to a foreign state, the target state may resort to countermeasures, which are those actions constituting a breach of an international obligation— a breach of treaty law or of customary international law—that have to be considered lawful because the state involved has been itself victim of a wrongful act. Under art. 2, lett. a of ARSIWA, any activity which constitutes a breach of an international obligation implies the responsibility of a state when undertaken by one of the parties cited therein.^[51] In these cases, the target state can react by resorting to countermeasures within the limits expressly provided by international law (i.e., the principle of proportionality and the sole aim of inducing the responsible state to desist its ongoing unlawful conduct, thus excluding other aims, such as punishment). Moreover, under art. 52 of ARSIWA, countermeasures shall be terminated as soon as the state has complied with its obligations.

However, countermeasures do not seem an often-practicable option in the context of cyber election meddling because of the existing disconnect between the general requirements of international law in terms of attribution and the practical necessities of states targeted by cyber operations.^[52]

On the one hand, international law requests the respect of discipline on international responsibility, which requires attribution of the wrongful act to another state in order to allow the injured state to resort to countermeasures as well as exhort the offending state to fulfill its obligations, notify its intent in responding to countermeasures, and negotiate.^[53] On the other hand, states need to promptly respond to cyber election interference to protect their interests, economies, citizens, and territories in order to avoid, or at least contain, negative consequences.

The result is that, to date, there has been no state reaction in the form of a real countermeasure against cyber violations. In addition, even when it is possible to identify the exact location where the cyber operation originated, investigative activities often require the assistance of the authorities of the state where the interference was launched,^[54] and this assistance is not necessarily provided.^[55]

Furthermore, in carrying out cyber operations, states generally use subjects who, even after investigation, frequently remain anonymous, a further circumstance which prevents the injured state from resorting to countermeasures.^[56]

C. Retorsions

In the context of uncertainty around the international legal status of cyber election interference, retorsions can play an important role for states which aim to respond to preserve their interests and rights without resorting to wrongful conduct.

Indeed, measures of retorsion (i.e., “unfriendly” conducts which are not inconsistent with any international obligation of the State engaging in it even though may be a response to an internationally wrongful act”)^[57] amount to acts which may be considered wrongful only in a political and moral sense.^[58]

An example of retorsion in the field of cyber election meddling was the declaration of *persona non grata* made by the U.S. Department of State with regard to 35 Russian intelligence operatives in response to aggressive Russian cyber activities during the last U.S. presidential election.^[59] In this case, since international law does not oblige states to maintain relations with other states, the declaration of *persona non grata* and the following expulsion of intelligence operatives constituted a mere unfriendly act.

Because of its characterization, retorsion seems to be the most practicable, legal, functional response in case of non-attributable cyber election interference.

The so-called “active defence strategies,” consisting of cyber operations—including those of a preventative nature—that the target state may resort to without having previously attributed interference to another state, can be considered a form of retorsion.^[60]

Such activities can be allowed, owing to the fact that these kinds of measures should never reach the threshold of unlawful conduct as they are limited to striking the systems from which the attack has been launched in order to avoid damage within the territory of the target state.^[61] Indeed, as they are focused on impending damage by the incoming cyber operations, they should be considered an instrument available to states to ensure the integrity of their territories and the security of their population within the exercise of their sovereign powers.^[62]

The lawfulness of retorsion depends on the relation between means and ends which, if imbalanced (e.g., when a state interrupts the supply of vital goods to another state only with the aim of exercising coercion in matters of its domestic jurisdiction), pushes the retorsion beyond the threshold of lawfulness.

V. CONCLUSIONS

The complex and uncertain legal qualification of cyber activities resulting in electoral meddling, mostly due to their hard attribution, represents a serious concern for states which have their electoral processes targeted.

The last Group of Seven summit held in Canada in June 2018 heavily stressed the danger posed by attempts on the part of foreign actors to weaken democratic societies and institutions by undermining their electoral processes through “malicious, multi-faced and evolving tactics [which] constitute a serious strategic threat.”^[63]

However, the same legal uncertainty can be considered an opportunity for target states which can obtain strategic advantages through cyber counteroperations aimed at containing collateral effects without assuming international responsibility. This perspective should not necessarily be considered negative because, as previously demonstrated, nothing impedes target states from reacting through international legal mechanisms which do not result in internationally wrongful acts—more precisely, retorsions, which seem to be a functional tool in terms of results.

In fact, above all, if carried out in the cyber domain, retorsions are able to reach results analogous to those achievable through countermeasures which, in turn, would put the target state which wants to respond to the cyber election interference at serious risk of violating international law.

If states aim to reduce their vulnerabilities and contrast cyber threats such as cyber election meddling, the strategy cannot be solely based on legal responses. The identification of further preventive efforts, in terms of strengthening cyber defense capabilities to protect electoral processes, is a topic that merits further discussion.

On this point, it is to be noted that some states are strengthening their electoral systems with the cooperation of their respective intelligence organizations and with a view to avoiding foreign interference in future elections. For example, Australia has formed an ad hoc task force (the Electoral Integrity Task Force, or EITF) to guard its electoral process against foreign cyber interference. The organization involves the efforts of multiple agencies, with a particular attention given to strengthening precautionary measures. Led by the Home Affairs Department and involving the Australian Security Intelligence Organisation, the Australian Federal Police, the Department of Finance, and the Australian Electoral Commission, the EITF aims to avoid foreign interference in elections. In addition, the government of Australia has decided to adopt an ad hoc legislation with the aim of preventing foreign electoral meddling.^[64]

Even the European Union (EU) has developed a strategy to counter propaganda and disinformation: in 2015, the Council of the EU tasked the High Representative for Foreign Affairs and Security Policy to submit an action plan on strategic communication,^[65] which led to the establishment within the EU's External Action Service of a unit (the European Strategic Communication Task Force, or StratCom) that challenges foreign (mainly Russian) disinformation campaigns.

To date, StratCom has been organized into three units—StratCom East, South, and Western Balkans—even though the main body is represented by StratCom East. StratCom East identifies, analyses, and raises awareness about pro-Kremlin disinformation; it is aimed at increasing public awareness of disinformation activities by foreign powers and improving the EU’s capacity to anticipate and respond to such challenges. In September 2017, a website was launched that features a database of over 3,000 cases of disinformation, gives an overview of the latest fake news stories published, and explains how trolling and manipulation in media really work.

More recently, in January 2018, the European Commission set up a High-Level Expert Group (HLEG)^[66] to contribute to the development of an EU-level strategy in facing the spread of fake news. In March 2018, the HLEG published a report suggesting a multidimensional approach to countering disinformation based on five pillars, consisting of concrete and inter-dependent actions ranging from enhanced transparency to the promotion of media and information literacy.^[67] Nevertheless, despite all of these efforts, the European Parliament recently urged the EU to increase its resilience to Russian propaganda.^[68]

A significant contribution to contrasting the spread of fake news aimed at influencing the electorate could also come from the most popular media and social networks, which should strengthen their internal tools for verifying the authenticity of news and profiles. Even on this point, new initiatives seem to have been undertaken,^[69] although the concrete tools available to be used in contrasting disinformation raise questions in their own right (e.g., the protection of the right to freedom of expression).

The fact is that cyber phenomena are not purely legal in nature, so to fully understand and, consequently, contrast them, states have to think in terms of integrated strategies which cannot avoid the involvement of international law, but, at the same time, must require the active intervention of other disciplines. 🍷

NOTES

1. As pinpointed by Stephanie Bellier “*Asymmetric warfare seeks to convert the enemy’s strength into weakness, and is, therefore, especially focused on manipulating information and communication [...] asymmetrical strategies aim more to influence and to change minds than to conquer*”; see S. Bellier, *Unilateral and Multilateral Preventive Self-Defense*, 58 Me. L. Rev. 508 (2006), p. 509.
2. See D. H. Levin, *When the Great Power Gets a Vote: The Effects of Great Power Electoral Interventions on Election Results*, *International Studies Quarterly*, Vol. 60, Issue 2 (2016), p. 189 ff.; see also D. Corstange and N. Marinov, *Taking Sides in Other People’s Elections: The Polarizing Effect of Foreign Intervention*, *American Journal of Political Science*, 56 (2012), p. 655 ff.
3. Cases of alleged foreign interference have been reported in the US and France presidential elections, Dutch and German elections, as in the 2016 Brexit and Italian constitutional referenda; for an overview, see P. Baines and N. Jones, *Influence and Interference in Foreign Elections*, *The RUSI Journal*, 163 (2018), 12.
4. The term ‘interference’ and ‘intervention’ are used in the present paper interchangeably, without a juridical implication, unless otherwise specified. The Tallinn Manual 2.0 on the International Law Applicable to Cyber Operations, Cambridge: Cambridge University Press, (2017) uses the term ‘interference’ in reference to acts which lack the requisite of coerciveness, while the term ‘intervention’ refers to acts that have coercive effects.
5. On the topic, see amplius M. Roscini, *Cyber Operations and the Use of Force in International Law*, Oxford University Press (2014), p. 45 ff. See the Tallinn Manual 1.0 on the International Law Applicable to Warfare, Cambridge: Cambridge University Press (2013), p. 54, where the International Group of Experts «unanimously concluded that some cyber operations may be sufficiently grave to warrant classifying them as an ‘armed attack’ within the meaning of the Charter»; see also Tallinn Manual 2.0, supra, p. 415; contra, Jaqueline Van De Velde, *The Law of Cyber Interference in Elections*, (May 15, 2017), p. 29.
6. M. N. Schmitt, “Virtual” Disenfranchisement: Cyber Election Meddling in the Grey Zones of International Law, *Forthcoming in Chicago Journal of International Law*, (2018), p. 21; see also Tallinn Manual 2.0, supra, p. 168.
7. Van De Velde, supra, p. 8
8. *Military and Paramilitary Activities in and against Nicaragua (Nicaragua v. United States of America)*, Merits, International Court of Justice (ICJ), 27 June 1986, par. 202; see also *Corfu Channel Case (United Kingdom v. Albania)*; Merits; International Court of Justice (ICJ), 9 April 1949, par. 35; Declaration on Rights and Duties of States, annexed to A/RES/374 (IV), Art. 3.
9. See e.g. Declaration on the Inadmissibility of Intervention in the Domestic Affairs of States and the Protection of Their Independence and Sovereignty (A/RES/20/2131); Declaration on Principles of International Law concerning Friendly Relations and Co-operation among States in accordance with the Charter of the United Nations (A/RES/25/2625); Declaration on the Inadmissibility of Intervention and Interference in the Internal Affairs of States (A/RES/36/103); Respect for the principles of national sovereignty and non-interference in the internal affairs of States in their electoral processes (A/RES/50/172).
10. A/RES/36/103, supra, Art. 2(b).
11. See A/RES/44/147 and A/RES/50/172.
12. B. Conforti, *Diritto Internazionale*, Naples (2015), p. 270.
13. *Nicaragua*, supra, para. 205: «A prohibited intervention must [...] be one bearing on matters in which each State is permitted, by the principle of State sovereignty to decide freely».
14. *Ibid.*: «Intervention is wrongful when it uses methods of coercion in regard to such choices, which must remain free ones».
15. *Ibid.*: «The element of coercion [...] is particularly obvious in the case of an intervention which uses force, either in the direct form of military action, or in the indirect form of support for subversive or terrorist armed activities within another State».
16. The ICJ affirmed to outline only those aspects of the principle of non-intervention which were relevant to the solution of the dispute; see *Nicaragua*, supra, para. 205.
17. On the point, see Jens D. Ohlin, *Did Russian Cyber Interference in the 2016 Election Violate International Law?*, 95 *Texas Law Review* (2017), p. 1589.
18. Amplius, Robert Nozick, *Coercion*, in S. Morgenbesser et al. (Eds.), *Philosophy, Science, and Method: Essays in Honor of Ernest Nagel*, St Martin’s Press (1969), p. 447.
19. On the topic, see Michell Berman, *The Normative Functions of Coercion Claims*, 8 *Legal Theory* 45 (2002).
20. Tallinn Manual 2.0, supra, p. 319.
21. Tallinn Manual 1.0, supra, p. 45.
22. Tallinn Manual 2.0, supra, p. 317.

NOTES

23. Myres S. McDougal and Florentino P. Feliciano, *International Coercion and World Public Order: The General Principles of the Law of War*, 67 Yale L. J. (1958), p. 782.
24. S. Watts, *Low-Intensity Cyber Operations and the Principle of Non-Intervention*, in Jens D. Ohlin et. al. (Eds.), *Cyber War: Law and Ethics for Virtual Conflict*, Oxford, (2015), p. 257.
25. Tallinn Manual 2.0, supra, p. 322; e.g. the Council of Ministers of the Organization of African Union deplored «the attempts [emphasis added] by some foreign interests, through the [...] manipulation of the media to interfere in and influence the outcome of the elections» in Zimbabwe in 2000; see Decision on the Developments in Zimbabwe, CM/Dec544.(LXXII).
26. The Tallinn Manual 1.0, supra, p. 43, states that «[...] it is clear that not every form of political or economic interference violate the non-intervention principle [...] It is clear that not all cyber interference automatically violates the international law prohibition on intervention: interference pure and simple is not intervention».
27. Ohlin, supra, p. 1588.
28. *Nicaragua*, supra, para. 205.
29. Tallinn Manual 2.0, supra, p. 317.
30. *Ibid.*, p. 319: «A few Experts, however, argued that it is impossible to prejudge whether an act constitutes intervention without knowing its specific context and consequences. For them, the context and consequences of a particular act that would not normally qualify as coercive could raise it to that level».
31. E.g. Art. 19(2)(d) of the United Nations Convention on the Law of the Sea (UNCLOS) provides that «Passage of a foreign ship shall be considered to be prejudicial to the peace, good order or security of the coastal State if in the territorial sea it engages in any of the following activities: [...] (d) any act of propaganda aimed at affecting the defense or security of the coastal State».
32. Draft Articles on Responsibility of States for Internationally Wrongful Acts, with commentaries – ARSIWA (2001) Text adopted by the International Law Commission at its fifty-third session, in 2001, and submitted to the General Assembly as a part of the Commission's report covering the work of that session (A/56/10). The report, which also contains commentaries on the draft articles, appears in the Yearbook of the International Law Commission, 2001, vol. II, Part Two, as corrected.
33. Further provisions under the ARSIWA on attribution seem to be here less relevant in practice.
34. ARSIWA, supra, Art. 7.
35. *Assessing Russian Activities and Intentions in Recent U.S. Elections*, report released by the ODNI on 6 January 2017, available at www.dni.gov.
36. E.g. President Trump has long refused to acknowledge Russia's meddling in U.S. elections; reported in www.pbs.org.
37. See also, e.g. the Joint Statement from the Department of Homeland Security and Office of the Director of National Intelligence on Election Security of 7 October 2016 where the U.S. Intelligence Community affirms «to be confident that the Russian Government directed the recent compromises of e-mails from US persons and institutions, including from US political organizations [...] These thefts and disclosures are intended to interfere with the US election process. Such activity is not new to Moscow—the Russians have used similar tactics and techniques across Europe and Eurasia, for example, to influence public opinion there. We believe, based on the scope and sensitivity of these efforts, that only Russia's senior-most officials could have authorized these activities».
38. ARSIWA, supra, Art. 8.
39. See U.S. Special counsel indictment in the case *United States of America v. Internet Research Agency LLC et al.*, available at www.justice.gov.
40. ARSIWA, supra, Art. 8 para. 7 of Commentary.
41. ARSIWA, supra, Art. 8 para. 3 of Commentary.
42. *Nicaragua*, supra, paras. 86 and 115; under the 'effective control' standard elaborated therein, the ICJ required that «For this conduct to give rise to legal responsibility of the United States, it would in principle have to be proved that that State had effective control of the military or paramilitary operations in the course of which the alleged violations were committed».
43. In the judgment *Prosecutor v. Duško Tadić (Appeal Judgement)*, 1999, para. 145, the ICTY stated that the requisite degree of control by the Yugoslavian «authorities over these armed forces required by international law for considering the armed conflict to be international was overall control going beyond the mere financing and equipping of such forces and involving also participation in the planning and supervision of military operations».

NOTES

44. E.g. the malware found on Democratic National Committee computers seem to be the same as used by hacking groups allegedly linked to Russia intelligence services, codenamed APT 28/Fancy Bear and APT 29/Cozy Bear; reported in S. Biddle, *Here Is the Public Evidence Russia Hacked the DNC – It’s Not Enough*, in *The Intercept*, 14 December 2016.
45. Tallinn Manual 2.0, *supra*, p. 91.
46. The ICJ has not developed a standard of proof for the attribution of internationally wrongful act, assessing each dispute by case-by-case approach; the lack of case-law related to cyber interference issues does not provided for useful elements to determine *ad hoc* principles.
47. E.g. President Obama, when announcing actions against alleged Kremlin-backed cyber interference during the 2016 Presidential elections, affirmed that «these actions [...] are a necessary and appropriate response to efforts to harm U.S. interests in violation of established international norms of behavior»; see Statement of the President on Actions in Response to Russian Malicious Cyber Activity and Harassment, available at www.whitehouse.gov.
48. ARSIWA, *supra*, p. 119.
49. E.g. in the *Stuxnet* case, Iran did not react even if the media worldwide attributed the attack to the United States.
50. *Certain Phosphate Lands in Nauru (Nauru v. Australia)*, Preliminary Objections, Judgment, I.C.J. Reports 1992, para. 32: «*The Court recognizes that, even in the absence of any applicable treaty provision, delay on the part of a claimant State may render an application inadmissible*».
51. Tallinn Manual 1.0, *supra*, p. 31: «*Any cyber activity undertaken by the intelligence, military, internal security, customs, or other State agencies will engage State responsibility under international law if it violates an international legal obligation applicable to that State*».
52. «*The technology inherent in cyberwarfare makes it nearly impossible to attribute the attack to a specific source or to characterize the intent behind it. Furthermore, acts of cyberwarfare occur almost simultaneously. A legal system that requires a determination of the attacker’s identity and intent does not account for these features of the digital age. The current international paradigm therefore limits the options available to states, making it difficult to effectively respond without risking a violation of international law. Restraining a state’s ability to respond will encourage rogue nations, terrorist organizations, and individuals to commit increasingly severe cyberattacks*», M. Hoisington, *Cyberwarfare and the Use of Force Giving Rise to the Right of Self Defense*, *Boston College International & Comparative Law Review*, Vol. 32 (2009), p. 452.
53. Countermeasures presuppose attribution as stated in Art. 49 ARSIWA, reading that «*An injured State may only take countermeasures against a State which is responsible for an internationally wrongful act in order to induce that State to comply with its obligations*» and Art. 51 ARSIWA, reading that «*Countermeasures are limited to the non-performance for the time being of international obligations of the State taking the measures towards the responsible State [and...] Countermeasures must be commensurate with the injury suffered*».
54. R. A. Clarke and R. K. Knake, *Cyberwar*, Harper Collins Publisher, New York (2010), p. 215.
55. «*Although states can trace the cyberattack back to a computer server in another state, conclusively ascertaining the identity of the attacker requires an intensive, time consuming investigation with assistance from the state of origin [...] This attribution problem locks states into the response crisis*», M. J. Sklerov, *Solving the Dilemma of State Responses to Cyberattacks: a Justification for the Use of Active Defenses Against States Who Neglect their Duty to Prevent*, *Military Law Review/Vol. 201* (2009), pp. 7 and 8.
56. Tallinn Manual 1.0, *supra*, p. 31: «*States may contract with a private company to conduct States cyber operations. Similarly, States have reportedly called upon private citizens to conduct cyber operations against other States or targets abroad*».
57. ARSIWA, *supra*, p. 128 of Commentary.
58. M. N. Schmitt, *supra*, p. 25.
59. In the past, retorsions included all forms of retaliation by a State against another in response of all kind of unwelcome acts by the latter. Nowadays, this concept is limited only to those actions which do not interfere with the target State’s rights under international law; *amplius* T. Giegerich, *Retorsion*, in R. Wolfrum (Ed.), *The Max Plank Encyclopedia of Public International Law*, Oxford (2012), p. 976.
60. On the point, see M. Hoisington, *supra*, p. 453, according to which the international community should promulgate a list of “critical national infrastructure” whose violation via cyber-attack would authorize the State upon whose territory the infrastructure lies to respond via active defense measures without incurring in international legal responsibility but, above all, without the loss of time involved in identifying the author of the attack.

NOTES

61. «Active defense measures, however, use offensive means in order to defend against and neutralized a threat. The purpose of using a cyber counterattack is to stop a specific, immediate, or ongoing cyber threat rather retaliate with a strategic purpose. It is offensive action for a defense purpose», C. Lotrionte, *Active Defense for Cyber: A Legal Framework for Covert Countermeasures*, in J. Carr (Ed.), *Inside Cyber Warfare*, O'Reilly Media (2011), p. 274.
62. T. Giegerich, *supra*, p. 980.
63. Charlevoix Commitment on Defending Democracy from Foreign Threats, G7 Summit, Charlevoix, 9 June 2018, available at www.g7.gc.ca
64. reported in www.reuters.com.
65. European Council meeting (19 and 20 March 2015) – Conclusions, para. 13.
66. The HLEG consisted of 39 members coming from academia, journalism, press and broadcasting organizations, online platforms as well as civil society and fact-checking organization; see 'A multi-dimensional approach to disinformation' - Report of the independent High-level Group on fake news and online disinformation, presented on March 2018 and available at www.ec.europa.eu.
67. *Ibid*, p. 20.
68. reported in www.europarl.europa.eu.
69. see e.g. Mark Zuckerberg, 'Protecting democracy is an arms race. Here's how Facebook can help', 4 September 2018, reported in www.washingtonpost.com.

SESSION

◆ 5 ◆

Defense Support to the Private Sector

*New Concepts
for the DoD's National
Cyber Defense Mission*

Jason Healey

*School of International and Public Affairs
Columbia University
New York City, New York, United States*

Erik B. Korn

*Army Cyber Institute
U.S. Army
West Point, New York, United States*

ABSTRACT

A primary mission of the Department of Defense (DoD) remains defending the nation in cyberspace, a function which has until this point has been oriented around the traditional Defense Support of Civil Authorities (DSCA) framework. However, conceptual confusion as to the most effective mechanisms for DoD support during national cyber emergencies has generated a perpetual “fog” that restricts the frameworks optimal employment. This paper examines the typical forms of DoD cyber support currently employed, and presents four additional pillars for consideration. These proposed pillars highlight the potential value of the DoD’s defined role and functionality as a supporting command to the private sector during national cyber emergencies. Furthermore, this paper recommends new, adaptable structures and defined roles that can serve as a model for the DoD’s future composition, disposition, and employment in cyberspace when called upon to defend the nation. Because the private sector is on the front lines of the conflict, a new model of Defense Support to the Private Sector (DSPS) needs consideration.

Keywords— Department of Defense, U.S. Cyber Command, Defense Support of Civil Authorities, Supported Command, Supporting Command, Dowding System, Persistent Engagement, Defensive Cyber Operations Response Action.

I. INTRODUCTION

The DoD has a central mission to “defend the nation” in cyberspace, a mission which has focused on DSCA, and rightly so. After all, almost all cyberattacks are not attacks on the nation, so the Department of Homeland Security (DHS) will often have the lead. It is homeland security, not homeland defense.

© 2019 Jason Healey

The contribution of Erik B. Korn is the work of the U.S. Government and is not subject to copyright protection in the United States. Foreign copyrights may apply.

But the DoD has significant capability and is regularly called in to provide support. The four main pillars of such DoD support are relatively well known: information sharing and collaboration; “away teams” and other post-incident support to U.S. critical infrastructure companies which have been attacked; counteroffensives to disrupt adversary operations against the United States; and the direct monitoring and defense of networks belonging to U.S. critical infrastructure companies. These types of support are not often so clearly described and, while the first two are relatively straightforward, the last two are controversial.

This paper examines these typical forms of support and takes on the conceptual confusion that surrounds the defense of the Nation. Much of the confusion comes from scenarios that are not sufficiently extreme, so that the roles of DHS and the DoD are still intertwined. To break out of this grey conceptual fog, it is necessary to imagine, as a thought experiment, the role of the DoD in the conceptual clarity of a black-and-white scenario of a true cyber war targeting the private sector, and then work down from there into the fog. Treating the DoD role in such a cyber war as “support to civilian authorities” is missing the point, as the military would have a direct role in fighting the adversary. In addition, civil authorities do not need support, but the private sector does. Given that the private sector is not just the main target of the adversary, but has significant capabilities of its own, the DoD role in defending the Nation is in many ways the “supporting command.” This method suggests four additional pillars of support: private-sector call for fire support, coordination of multi-stakeholder defensive actions, response-support forces, and private-sector access to the entire intelligence cycle. Together, these can be a new approach: “Defense Support to the Private Sector” (DSPS).

II. DEFENSE SUPPORT OF CIVIL AUTHORITIES (DSCA)

“[D]uring a natural disaster, like a hurricane, military troops and helicopters are often used by ...[the Federal Emergency Management Agency] to help deliver relief. In a similar vein, the military’s cyber capabilities will be available to civilian leaders to help protect the networks that support government operations and critical infrastructure. As with all cases of military support to civilian authorities, these resources will be under civilian control and used according to civil laws”^[1].

–Then-Deputy Secretary of Defense William J. Lynn III

The cyber response is only part of the larger National Response Framework (NRF), a whole-of-nation approach for unified response actions for emergencies and natural disasters, of DHS’s Federal Emergency Management Agency (FEMA). The NRF is the central strategy for local, state, tribal, private, and federal entities in conducting joint operations during national emergencies^[2]. The DoD is specified in the NRF as a resource authorized for commitment to domestic emergencies upon approval of the secretary of defense or when directed by the president^[3]. The NRF is primarily for physical emergencies, like hurricanes or earthquakes, while the National Cyber Incident Response Plan (NCIRP) is only for cyber incidents (an incident which had both

cyber and physical consequences would invoke both—one reason why DHS is a natural choice for national incident response).

Federal Government cyber response is centered on DHS, which has the statutory mission of ensuring cybersecurity through a better understanding of the U.S. risk posture and “reducing or mitigating vulnerabilities, threats, and the potential consequences from cybersecurity incidents”^[4]. Per Presidential Policy Directive (PPD) 41 from 2016, DHS is the nominated lead for “asset response activities” (as compared to investigative and intelligence activities, which are the responsibility of the Federal Bureau of Investigation (FBI) and the Office of the Director of National Intelligence, respectively) and oversees the bulk of the federal response to cyber incidents of national significance^[5]. When a “significant cyber incident affects critical infrastructure owners and operators” and may “reasonably result in catastrophic regional or national effects on public health or safety, economic security, or national security,” the government forms a Cyber Unified Coordination Group (UCG) as “the primary method for coordinating between and among Federal agencies in response to a significant cyber incident as well as for integrating private-sector partners into incident response efforts”^[6]. Though PPD-41 does not mention the DoD, it would participate in a Cyber UCG as an additional participant.

The NCIRP defines the various responsibilities, capabilities, and coordination efforts for a national response to cyber incidents and, unlike PPD-41, explicitly details DoD responsibilities in the event of a national cyber incident^[7]. Securing the DoD Information Network and civil authorities’ organic assets is a primary responsibility, but the NCIRP also includes details on providing support to civil authorities when requested to do so through lead federal agencies or when directed to do so by the president^[8]. These supporting structures are just a few of the resources that manage the civil-military support relationship in times of national crisis. Of course, the DoD has significant capabilities for responding to cyber incidents, not least of which are those at U.S. Cyber Command (USCYBERCOM) and the National Security Agency (NSA). One of the DoD’s key missions is for it to “be prepared to defend the United States and its interests against cyberattacks of significant consequence”^[9].

Since DHS has the overall lead, the DoD’s cyber defense of the Nation is typically rooted in the larger framework of DSCA. There are a variety of authorities, joint doctrine publications, and federal response plans that oversee the support relationships among the DoD, civil authorities, and industry during disasters. The DoD maintains an inherent role in bolstering civil authorities during national emergencies as well as the responsibility to provide necessary support in the event of a domestic emergency. The Stafford Act and Economy Act constitute a legislative structure that provides state governments and federal agencies a mechanism with which to request DoD support when organic capabilities and resources become overwhelmed during an emergency^[10]. U.S. Code (U.S.C.) also specifies authorities for the support relationship between the DoD and civilian entities. Specifically, Title 32 and Title 10 directly permit DSCA, an affiliation generally characterized by DoD reinforcement of civilian entities in response to “domestic emergencies, law enforcement support, and other domestic activities”^[11].

This legislative foundation has been further developed with joint military doctrine such as Joint Publications (JP) 3-27, “Homeland Defense,” and JP 3-28, “Defense Support of Civil Authorities,” as well as previously mentioned federal response action plans like DHS’s NRF and NCIRP. JP 3-27 explains the different roles of the responsible commands and clarifies the missions of homeland security, homeland defense, and DSCA; homeland defense involves “defending against traditional external threats or aggression...and against external asymmetric threats” that are outside the scope of homeland security and related DSCA tasks^[12].

During DSCA operations, the military typically assumes a supporting role that is subordinate to the designated lead federal department or agency^[13]. Titles 32, 10, and 14 of the U.S.C. sanction support from the National Guard, active duty forces, and the United States Coast Guard in the event of national emergencies^[14]. DoD Directive 3025.18 further expands on the DSCA request process in accordance with sections 1521, 1535, and 9701 of U.S.C. Title 31^[15]. JP 3-27 also further stipulates additional guidance for joint operations in support of homeland defense.

III. CURRENT PILLARS OF DEFENSE SUPPORT

Despite the general strength of the DSCA framework, according to a panel at a 2018 strategy symposium run by USCYBERCOM, “there is little consensus on what it means to defend the Nation and its interests in cyberspace, or on what role the Department of Defense should be for this mission”^[16]. Just how should the DoD and USCYBERCOM go beyond DSCA for homeland defense?

There have been four main pillars of support: information sharing and collaboration; “away teams” and other post-incident support to U.S. critical infrastructure companies which have been attacked; counteroffensives to disrupt adversary operations against the United States; and direct monitoring and defense of networks belonging to U.S. critical infrastructure companies. The first two are far more straightforward than the last two, and there are actually far more ways that the DoD can defend the Nation, as this paper discusses in the next section.

A. Information Sharing and Collaboration

DoD efforts (such as the Enduring Security Framework) to share information on threats and vulnerabilities and collaborate with the private sector and other government agencies to reduce the threats and vulnerabilities have been important mechanisms. These operate at levels well below homeland defense and focus more on threat reduction before an event than response once an incident has begun^[17].

B. Post-Incident Support

Perhaps the most-used mechanism is the DoD’s support to other federal departments after a major incident occurs against (typically) a company that is part of the country’s critical infrastructure. The FBI has Cyber Action Teams at all 56 of its field offices, which will “travel around the world” within 48 hours “to assist in computer intrusion cases”^[18]. DHS also has

such “fly-away teams” that can deploy with the FBI for incidents which are not just crimes, but have a larger homeland-security nexus, such as attacks against major critical infrastructure companies^[19]. DHS and the FBI somewhat routinely call in DoD capabilities to assist; in at least one case, when Google suffered a severe intrusion by China, it reached out directly to the NSA for a “secure tailored solution” which brought in the FBI and DHS^[20].

C. Shooting Back

The DoD, of course, has unique authorities, beyond those of the FBI and DHS, and, when directed, “the U.S. military may conduct cyber operations to counter an imminent or on-going attack against the U.S. homeland or U.S. interests in cyberspace...to blunt an attack and prevent the destruction of property or the loss of life”^[21]. The National Cyber Mission Teams were created for just this homeland-defense eventuality. Such an order, though, has rarely if ever been given, even during known attacks from nation-state adversaries, such as the 2012 distributed denial-of-service (DDoS) attacks by Iran against the U.S. financial system, when “the Obama administration rejected an option to hack into the adversary’s network in Iran and squelch the problem at its source”^[22]. As the next section will discuss, there is far more that can be done to develop this pillar.

D. Monitoring and Direct Response

General Keith Alexander, when he was Commander of USCYBERCOM, opined that “within the United States, I do not believe that’s where Cyber Command should or will operate”^[23]. However, he wanted to improve his ability to monitor and defend the banking sector by installing government “surveillance equipment on their networks” to detect attacks using NSA’s “secret sauce” of threat signatures^[24]. The plan did not proceed, though the idea of direct monitoring and protection of private-sector assets does live on in some corners. At the 2018 US-CYBERCOM strategy symposium, one cyber general asserted that if companies “want to meet us halfway,” they must agree to allow the military to monitor their networks, even when those companies spend hundreds of millions on cybersecurity^[25]. Indeed, joint cyber doctrine opens the possibility that “National-level CPT [Cyber Protection Team] support can be extended to defend non-DOD mission partner or critical infrastructure networks when ordered” by the secretary of defense^[26].

This most controversial of the pillars is worth additional exploration. On one hand, the DoD directly defends U.S. territory; on the other, cyberspace is not the same as physical territory, and it is not always clear that the DoD has the authority or even superior capabilities. Despite these limitations, it is often the default assumption of military cyber defenders that, to defend the Nation, they must take control of the assets themselves. For example, Mark Young in 2010 wrote, “there is little that the DoD could do if the attack came across a commercial network,” but a national cyber doctrine and processes could smooth coordination with the private sector “when the networks to be protected by the Cyber Command belong to a commercial entity”^[27].

These mechanisms could “address the concerns” of commercial network service providers “to allow a U.S. government organization, such as the Cyber Command, to operate on their networks” for defense purposes^[28].

IV. EXPANDING DOD SUPPORT IN THE BLACK-AND-WHITE CLARITY OF CYBERWAR

There are several reasons it is hard to determine the appropriate role for the DoD in defending the nation in cyberspace. Identifying these reasons can help develop additional policy responses.

One of the most critical differences between cyber conflict and conflict in the air, land, sea, and space is that “it is non-state actors, not governments, which typically are decisive in cyber defense...[o]nly uncommonly are governments able to bring the superior resources of their unwieldy bureaucracies in enough time to decisively defend against attacks”^[29]. Companies like Microsoft, Verizon, and FireEye have massive security budgets and tremendous agility and routinely change the “terrain” of cyberspace to stop attacks. They are overly burdened with deciding if they have the legal authority to conduct defensive measures; as private entities, they are permitted all which is not specifically restricted—the opposite of what applies to the U.S. Government.

Banks like JPMorgan Chase spend over \$500 million on cybersecurity with complex networks^[30]. USCYBERCOM only has a limited set of resources and experienced personnel, so it is not clear how it could effectively monitor such networks or help defend them, even if asked to do so. It is like defending a labyrinth: unless you are on the network for long periods of time, you do not know the terrain well enough to defend it. Fortunately, as will be argued shortly, it is not clear that USCYBERCOM’s homeland defense mission depends on such on-site defense.

Another critical difference between cyber and conflict in the other domains is that there is constant contact between adversaries, creating an environment of “persistent engagement.” Some of these incidents, such as Chinese commercial espionage or attacks on critical infrastructure like the finance sector, can be classified as major national security threats—and, indeed, President Barack Obama declared a “national emergency” to deal with them^[31]. This can lead to the recommendation that since the DoD is the part of the Federal Government that deals with national security threats, it should be engaged now in the defense of critical infrastructure networks. Even when that recommendation is rejected (for reasons such as the DoD does not have enough capability to act so routinely and DoD presence is not wanted by the affected companies), the way out of the conceptual fog is usually framed from the bottom up: envisioning scenarios a bit (or a lot) worse than today’s and then trying to determine the appropriate role for the DoD and its relationship to DHS and the private sector.

This approach can be useful, but only goes so far when caught up in a conceptual fog. As in any fog, turning up the high beams on your headlights only shows you more grey. In most scenarios that are based in some worse version of today, DoD and DHS authorities will still

be intertwined, and the private sector will still be hesitant regarding a lead role for the DoD. To break out of this grey conceptual fog, it is necessary to imagine the role of the DoD in the conceptual clarity of a black-and-white scenario of a true cyber war and then work down from there into the fog.

Treat this as a thought experiment only—perhaps such a cyber war is impossible—but, to set the scene, imagine that an adversary nation-state is using cyber capabilities to kill thousands of American citizens. More attacks are coming every day. What is the DoD’s role in this obvious homeland-defense scenario?

Treating the DoD role in such a cyber war as “support to civilian authorities” is missing the point: “For most contingencies, the usual DoD role of support to civil authorities will apply. However, in the event of a high-end attack, the DoD will likely need to take the lead role”^[32]. The republic is at war, and the American people and the president would expect the DoD to be at the forefront of defense. But in such high-tempo operations, USCYBERCOM will certainly not have the resources to deploy CPTs to defend specific critical infrastructure-sector companies; it will likely be having to use every last person to defend the DoD and the U.S. Government and take the fight to the enemy.

So what else can the DoD and USCYBERCOM do to help win in this cyber-war thought experiment? What might be part of a DSPS project? There are several different mechanisms that can enable the expansion of DoD defense of the Nation: private-sector call for fire support, coordination of multi-stakeholder defensive actions, response-support forces, and private-sector access to the entire intelligence cycle. In each case, these measures are not just useful for high-end cyber warfare, but far down into the grey-zone conflicts of today.

V. PRIVATE SECTOR CALLS FOR FIRE SUPPORT

As part of the cyber-war thought experiment, further imagine that the finance sector reports that cyberattacks will turn into a financial crisis unless specific adversary command and control (C2) servers are not attacked and taken offline in three hours.

In one sense, this is a normal Defensive Cyber Operations Response Action (DCO-RA) mission in which “actions are taken external to the defended network or portion of cyberspace without the permission of the owner of the affected system [which] may include actions that rise to the level of use of force, with physical damage or destruction of enemy systems”^[33]. Yet there are currently no channels through which USCYBERCOM can receive such private-sector calls for fire or through which such calls can be validated. The banks collectively making the request through official channels are under direct attack by an adversary choosing to target the U.S. by attacking them online. They are the Forward Edge of the Battle Area of the war and their request for fires should be taken just as seriously as if it had come through a combatant command. In cyber conflict, the private sector is the supported command. This will prove much easier for sectors such as finance, which has hired many cyber veterans and has a formal governance structure to make official and time-sensitive requests.

There is already some evidence of such ties, though they are informal. The Financial Systemic Analysis & Resilience Center (FSARC) is sharing malware indicators and other information with USCYBERCOM where “this intelligence is independently evaluated and, if appropriate, Cyber Command *responds under its own unique authorities*”^[34].

VI. COORDINATING MULTI-STAKEHOLDER DEFENSIVE ACTIONS

The DoD can work toward supporting the synchronization of defensive actions and establish a joint battle rhythm between the Federal Government, private-sector industries, and additional civil authorities. What might be needed is a cyber equivalent of the Dowding system, the British system for detecting inbound bombers during the Battle of Britain and providing direct defense^[35]. The network of sensors, operations centers, and communications acted as a central nervous system for situational awareness of all available information and control defenses. However, in stark contrast to conflict in other domains, it may be the private sector which controls the main tempo, with the DoD supporting it.

In a notional, high-end cyber war, the current mechanisms for coordinating defensive actions would quickly become swamped. The DHS National Cybersecurity and Communications Integration Center (NCCIC) is the main operational coordination body, a “central location where a diverse set of partners involved in cybersecurity and communications protection coordinate and synchronize their efforts [and] coordinate national response to significant cyber incidents in accordance with the National Cyber Incident Response Plan”^[36]. The NCCIC also connects with FEMA’s NRF for cyber-physical incidents and coordinates with DoD operations centers, including USCYBERCOM. However, NCCIC has suffered persistent staffing and technical training issues, and would be challenged to work at the scale of a cyber war with many separate attack campaigns^[37]. For example, when responding to just one past campaign, the Conficker worm, the DHS team not only played no decisive role, but, when it needed to brief the White House, simply used the slides of the private-sector, Microsoft-funded Conficker Working Group, substituted its own logo, “and classified it to boot”^[38]. The DoD and USCYBERCOM may have better staffing and capabilities but would also have difficulty scaling quickly. They also do not have the visibility or connections with industry to coordinate the defense of private-sector networks.

There are already many private-sector response organizations. One presidential advisory committee composed of technology executives developed a report with a full set of recommendations for sector “mobilization” that notes that “the vast majority of enterprise incidents are resolved with the support and collaboration” of companies and trust groups, such as Network Service Provider Security (NSP-SEC) and information sharing and analysis centers^[39]. Indeed, in most incidents:

“[T]he fundamental incident management actions occurred through private sector collaboration or mobilization at a [small] scale, limited to a group of actors that had the technical competence and ability to develop and propose appropriate mitigations to address the core

vulnerability. This group is distinct from the affected community, which constitutes those end users with the responsibility for managing the actual manifestations of the consequences of the attack”^[40].

The Federal Government simply has a less decisive role than non-states. Even as far back as the 2007 attacks on Estonia, NSP-SEC, “comprised of technical experts of various network provider companies,” was sent to Estonia to help coordinate defensive efforts with international telecommunication carriers and “mitigated [these] down to fairly low levels over the course of the next seven hours”^[41]. The spirit of the group focuses on immediate action: “If something needs to be taken down, it needs to be taken down, and there isn’t time for argument ... that’s understood upfront [within NSP-SEC]”^[42]. Another alliance of technology companies, the Industry Consortium for Advancement of Security on the Internet, created a Unified Security Incident Response Plan (USIRP) for its membership (which includes Microsoft, CISCO, Intel, Amazon, and Oracle) so that they can “trigger a USIRP event; share critical information about it; and work together effectively on a coordinated response”^[43]. The Cyber Threat Alliance coordinates responses between many threat intelligence teams, such as at Palo Alto Networks and CISCO, to generate a common threat picture^[44]. Within the critical infrastructure sectors, there are many groups handling various aspects of response. Just the finance sector has three groups: FSARC, the Financial Services Information Sharing and Analysis Center (FS-ISAC, of which one of the authors has been vice chair), and the Financial Services Sector Steering Committee (FSSSC).

Cyber defense has long been recognized as a team sport or, rather, a multi-stakeholder effort, with distributed responsibilities. The main goal of the coordination of all of these defensive efforts as well as the integration of DCO-RA response missions and outright offensive attacks from the DoD is not unity of command centered on USCYBERCOM or NCCIC, but *unity of effort*, *unity of action*, and loose coordination to keep independent groups working toward the same goal. It may be counterproductive to insist that “clear chains of command for a high-end contingency...be established between the civil authorities and the DoD,” or that “private sector cyber security expertise” should be “working under government direction and control in connection with high-end contingencies or in direct support to the ISPs [internet service providers] and grid operators”^[45].

Unity of effort through multi-stakeholder coordination would mean that the DoD would not be able to synchronize offensive and defensive efforts as well as if it controlled them both, but this is a small loss to achieve better synchronization *across all* defense, in both the public and private sectors. Efforts to build such a multi-stakeholder Dowding system based on a unity of effort and support to the private sector would be useful at levels well below full cyber war.

The DoD (and the rest of the Federal Government) cannot and should not lead these efforts, but do need to support them. For example, in the “event an incident surpasses industry’s mitigation ability,” then “industry would want recommendations or direction on the priorities

for...recovery”—that is, a political decision on national security priorities^[46]. Industry may also need a “comprehensive, legal, and operational framework,” as it would be “operating on a catastrophic” footing, far beyond business as usual^[47].

VII. SECTOR-WIDE RESPONSE—SUPPORT FORCES

During high-tempo cyber warfare against the United States, DoD CPTs deployed to directly monitor and protect private-sector networks would only get in the way. However, there may be a role for the DoD, possibly through a new kind of Cyber Support Team (CST), to support the private-sector response process, rather than helping to defend private-sector networks.

To return to the thought experiment of cyber warfare against the private sector, imagine again a massive attack against the finance sector. Sector-wide incident response is handled by groups such as FSARC, FS-ISAC, and FSSSC, typically on conference calls every few hours. These calls cover technical and intelligence issues (usually at the more operationally focused FS-ISAC) as well as top-level policy issues, such as whether the markets will be able to remain open (at the more senior FSSSC). Overwhelmingly, the same people on these calls handling sector-wide response are the same executives overseeing response within their own financial institutions. They are very thinly spread, with some limited 24/7 capability, and if an incident lasts more than a few days, the system may break.

One of the authors (Healey) led the coordination of these calls for the FS-ISAC. What could have been useful was a few, competent, company-grade or senior non-commissioned officers to give more organizational depth and staying power to the response. These officers could help run the response playbook, keep track of the dozens of details needed for a successful response, and provide much-needed continuity and stability to the process. Such officers do not have to be highly trained DoD cyber ninjas and do not necessarily even need much knowledge of the affected sector (though these knowledge and skills could be useful). They only need to be capable responders—the kind of officers which exist in great numbers in all services.

VIII. PRIVATE-SECTOR ACCESS TO THE ENTIRE INTELLIGENCE CYCLE

Intelligence cooperation between the Federal Government and the private sector is improving—especially with more cleared individuals in critical infrastructure sectors and companies, which have hired former intelligence professionals—but it is still far behind the level which might be required in a notional cyber war. Too often, companies (even in key sectors) are only included at the tail end of the intelligence cycle—dissemination. They receive tear-line reports of declassified and watered-down reports. Sometimes, select executives are given a “special one-day, top-secret security clearance” which “scare[s] the bejeezus” out of them^[48]. But with private-sector companies on the Forward Edge of the Battle Area, they should not just be receiving reports; they should be active in all phases of the intelligence cycle, especially in the submission of requirements for the collection and clarification of analysis and the provision of

feedback^[49]. This would primarily be the responsibility of the Director of National Intelligence, but, as the NSA has had a lead role in such activities in the past, much would fall onto the DoD's shoulders, especially in wartime.

The downsides of this kind of support are obvious: there are currently few ways for a sector to validate any requests or feedback, few if any mechanisms for passing requests and feedback from the private sector, and a major gap between sectors in the sophistication of intelligence consumers. As with the potential to calls for fires, the finance sector is perhaps a natural place to start, with many cyber and intelligence veterans and a formal governance structure in place.

IX. RECOMMENDATIONS: TO DEFEND THE NATION, SUPPORT THE PRIVATE SECTOR

The DoD possesses unique tools and resources for DSPS. However, large gaps remain.

A recent Government Accountability Office (GAO) report identified some of the challenges and shortcomings in the DoD's current approach and its application to cyberspace. Most glaringly, the report highlights a lack of definition in the DoD organizational roles and responsibilities for providing civil support during a national cyber incident^[50]. The DoD's C2 guidance for cyber DSCA operations is highlighted as contradictory and confusing. Additionally, conflicting delineations for U.S. Northern Command and USCYBERCOM as the supporting command to civil authorities for cyber incidents further complicates DoD guidance^[51]. With C2 being a primary component of effective military operations, the Pentagon's ability to streamline unity of command policies and processes is vital. Another area identified by GAO as a challenge is the DoD's visibility of capabilities within National Guard cyber units, a limitation that currently impedes timely and effective support to civil authorities^[52]. Furthermore, GAO's recent findings of DoD delinquency in the maintenance of a repository of Guard capabilities for each state must be rectified quickly for this option to work effectively^[53]. These deficiencies can be debilitating and limit the DoD's ability to provide support to industry and civil authorities in cyberspace.

In order to best leverage DoD cyber capabilities, the Pentagon must go even beyond these recognized gaps and recognize a new role as a supporting command to the non-state actors on the front lines of defending the Nation in cyber conflict. One important early step, highlighted by several former defense and intelligence officials, is to incorporate "establishing and exercising the procedures necessary" for cooperation for high-end crises into the memorandum of understanding between the DoD and DHS^[54]. Likewise, the National Security Telecommunications Advisory Committee report on mobilization has several recommendations, which we support, including the identification and organization of the correct public- and private-sector entities, and then conducting training and exercises "to ensure the Nation is prepared to manage a cyber-related event of national significance"^[55].

An important capability for expanded support is Reserve and Guard cyber units. The DoD's decision to fully invest in these units and their often-unique capabilities and authorities can provide a force able to build closer relationships among government, civil authorities, and industry. The individuals in these units also typically work in various sectors of industry or with other civilian entities on a daily basis. When operating under U.S.C. Title 32 at the direction of state governors, Guard cyber teams provide a unique flexibility in supporting civil authorities and sectors of industry (and are not subject to the restrictions of the Posse Comitatus Act—legislation that limits military units from operating domestically, such as working with law enforcement)^[56]. In order to address civil authority support, the DoD has already worked with the Council of Governors on the establishment of the Joint Action Plan for State-Federal Unity of Effort on Cybersecurity, which provides a collaborative framework to “expedite and enhance the nation’s response to cyber incidents” through collaboration, information sharing, capabilities, and resources^[57].

The Army National Guard and the Air National Guard have partnered to ensure cyber-team coverage of all 10 FEMA response regions to better integrate with DHS efforts and to help counter large-scale domestic cyber emergencies^[58]. This idea should be extended, with a Guard or Reserve team working with each critical infrastructure sector. For example, the Air Force Reserve or Air National Guard might work with the energy sector, as many Air Force cyber assets are in Texas, and the Army might work with the finance sector, as the Army Cyber Institute is just north of Manhattan. Each unit would be a CST, hopefully, composed of officers and enlisted personnel from the supported sector. Each unit could assist with some of the additional support pillars mentioned in this paper: developing processes for calls for fire, backstopping responses, assisting with intelligence requirements, and being better consumers of intelligence. There are some advantages, mostly in simplicity, to these CSTs being run by a single service, though, given the likely lack of qualified people, making them joint (with perhaps a single service as the lead) may make them stronger.

USCYBERCOM has created new joint headquarters for many specialized purposes, from defending its own networks to attacking those of the Islamic State of Iraq and Syria. A new, modestly sized, joint task force or joint forces headquarters might be created solely to support the private-sector fight and, to a lesser degree, work with civil authorities on homeland defense^[59]. As the parent command of the Guard and Reserve teams, it would support each sector, with responsibilities to improve operational coordination for high-end cyber incidents and warfare, though it would not conduct response actions itself. Such a headquarters might be largely staffed with Reserve and Guard personnel and located in the San Francisco Bay or Seattle areas to better coordinate with technology companies that control the high ground of cyberspace.

Regardless of whether the DoD creates new units for this purpose, it must make progress on these additional support pillars as well as help create the framework to support a cyber Dowding system. As the finance sector is perhaps the most mature, for the reasons mentioned above,

the DoD should extend its current efforts with that sector, starting with an informal discussion (including DHS and the Department of the Treasury) on how the sector might call for fire from USCYBERCOM, should that ever be required. This can serve as a basic model for the other sectors, especially those with strong governance mechanisms.

One way to support the idea of a cyber Dowding system is for the DoD to encourage, and perhaps match, DHS grants to create new organizations dedicated not to sharing information, but collaborating to respond to each kind of major incident. The goal of these Cyber Incident Collaboration Organizations (CICO) is to streamline the current response process for an incident type to provide an umbrella for making such work easier at a larger scale. As one of us wrote earlier this year:

“A Counter-Malware CICO could be built, using the lessons learned from the Conficker Working Group, for a faster, more effective response to such incidents. A Counter-Botnet CICO would be similarly global and led by the private sector, with membership including the global organizations that have had the largest role in takedowns—such as, say, Microsoft, FireEye, and the Department of Justice. The Counter-DDoS CICO would bring together the global Tier 1 service providers, content-distribution managers, and other organizations that focus on the core Internet infrastructure ... By comparison, the Counter-APT CICO might be led and funded by the U.S. government, working with the “Five Eyes” partners...and, perhaps, with representation from the Defense Industrial Base and key cybersecurity companies. Much of its work would be classified.”

Such CICOs, or similar organizations, would make the multi-stakeholder response much easier at scale, both simplifying and clarifying the role of USCYBERCOM and the larger Federal Government.

The DoD has the necessary capabilities, resources, and forces for DSPS. To achieve an effective response to domestic cyber emergencies, the Pentagon will need to understand how it can best bolster these entities as a supporting command when the call for reinforcements is received. Expanded areas of support can include core military functions, such as intelligence, C2, defensive actions, and calls for fire. The question now is whether the DoD can seize these opportunities to provide more effective support functions during significant cyber events, or if it will fall back into the trap of institutional norms where it feels compelled to take the lead. ♥

ACKNOWLEDGMENT

The authors would like to acknowledge Divyam Nandrajog and Augusta Grondquist for their research help and other support. This work was funded in part by the Office of Naval Research under the Office of the Secretary of Defense Minerva program (grant number N00014-17-1-2423).

NOTES

1. W. J. Lynn III, “Deputy Secretary of Defense Speech: Remarks on Cyber at the RSA Conference,” U.S. Government, *U.S. Department of Defense*, (February 15, 2011), <http://archive.defense.gov/speeches/speech.aspx?speechid=1535>.
2. U.S. Department of Homeland Security, “National Response Framework” (Federal Emergency Management Agency (FEMA), June 2016), 1–2, https://www.fema.gov/media-library-data/1466014682982-9bcf8245ba4c60c120aa915a-be74e15d/National_Response_Framework3rd.pdf.
3. *Ibid.*, 17–18.
4. US Department of Homeland Security, “Cybersecurity Strategy,” 15 May 2018, A-5, https://www.dhs.gov/sites/default/files/publications/DHS-Cybersecurity-Strategy_1.pdf.
5. *Ibid.*
6. The White House, “Presidential Decision Directive 41, United States Cyber Incident Coordination,” 26 July 2016, <https://obamawhitehouse.archives.gov/the-press-office/2016/07/26/presidential-policy-directive-united-states-cyber-incident>.
7. U.S. Department of Homeland Security, “The National Cyber Incident Response Plan (NCIRP),” *U.S. Government, United States Computer Emergency Readiness Team*, (2017), 4, <https://www.us-cert.gov/ncirp>.
8. *Ibid.*, 14.
9. Department of Defense, DoD Cyber Strategy, April 2015, p5, https://www.defense.gov/Portals/1/features/2015/0415_cyber-strategy/Final_2015_DoD_CYBER_STRATEGY_for_web.pdf.
10. U.S. Government Accountability Office, “Civil Support: DOD Needs to Clarify Its Roles and Responsibilities for Defense Support of Civil Authorities during Cyber Incidents” (U.S. Government Accountability Office, April 4, 2016), 5, <http://www.gao.gov/products/GAO-16-332>.
11. “Joint Publication 3-28: Defense Support of Civil Authorities” (U.S. Joint Chiefs of Staff, July 31, 2013), vii, http://www.dtic.mil/doctrine/new_pubs/jp3_28.pdf.
12. “Joint Publication 3-27: Homeland Defense” (U.S. Joint Chiefs of Staff, July 29, 2013), I-1, I-2, <https://www.hsdl.org/?view&did=742874>.
13. *Ibid.*, I-5.
14. *Ibid.*, I-7.
15. W. J. Lynn III and A. B. Carter, “Department of Defense Directive 3025.18: Defense Support of Civil Authorities (DSCA)” (U.S. Department of Defense, September 21, 2012), 3–4, https://fas.org/irp/doddir/dod/d3025_18.pdf.
16. US Cyber Command, “USCYBERCOM Cyberspace Strategy Symposium Proceedings, 2018,” p7, <https://www.cybercom.mil/Portals/56/Documents/USCYBERCOM%20Cyberspace%20Strategy%20Symposium%20Proceedings%202018.pdf?ver=2018-07-11-092344-427>.
17. T. Gjelten, “Cyber Briefings 'Scare The Bejeezus' Out Of CEOs,” National Public Radio, 9 May 2012, <https://www.npr.org/2012/05/09/152296621/cyber-briefings-scare-the-bejeezus-out-of-ceos>.
18. Federal Bureau of Investigation, Cyber Crime, <https://www.fbi.gov/investigate/cyber>.
19. Department of Homeland Security, “Industry Offerings, Products, and Services,” https://www.dhs.gov/sites/default/files/publications/DHS-Industry-Resources_4.7.edits_.pdf.
20. E. Nakashima, “Google to enlist NSA to help it ward off cyberattacks,” *The Washington Post*, 4 February 2010, <http://www.washingtonpost.com/wp-dyn/content/article/2010/02/03/AR2010020304057.html>.
21. DoD Cyber Strategy, p8.
22. E. Nakashima, “U.S. rallied multinational response to 2012 cyberattack on American banks,” *The Washington Post*, 11 April 2014, https://www.washingtonpost.com/world/national-security/us-rallied-multi-nation-response-to-2012-cyberattack-on-american-banks/2014/04/11/7c1fbb12-b45c-11e3-8cb6-284052554d74_story.html?utm_term=.45920d5ece84.
23. N. Shachtman, “Military’s Cyber Commander Swears: ‘No Role’ in Civilian Networks,” *Brookings Op-Ed*, 23 September 2010, <https://www.brookings.edu/opinions/militarys-cyber-commander-swears-no-role-in-civilian-networks/>.
24. S. Harris, *@War: The Rise of the Military-Internet Complex*, Houghton Mifflin Harcourt, 2014, Chapter 10.
25. Comment from non-for-attribution participant, a general officer on the “Defend the Nation” panel, US Cyber Command Strategy Symposium, 15 September 2018. Also see, C. Bing, “Inside 'Project Indigo,' the quiet info-sharing program between banks and U.S. Cyber Command,” *CyberScoop*, 21 May 2018, <https://www.cyberscoop.com/project-indigo-fs-isac-cyber-command-information-sharing-dhs/>.

NOTES

26. Department of Defense, Joint Publication JP 3-12, Cyberspace Operations, 8 June 2018, pII-8, https://fas.org/irp/doddir/dod/jp3_12.pdf.
27. M. D. Young, "National Cyber Doctrine: The Missing Link in the Application of American Cyber Power," *Journal of National Security Law & Policy*, (4:), http://jnslp.com/wp-content/uploads/2010/08/12_Young.pdf. p186.
28. Ibid.
29. J. Healey, ed, *A Fierce Domain: Cyber Conflict, 1986-2012*, CCSA, 2013, p22.
30. S. Morgan, "Why J.P. Morgan Chase & Co. Is Spending A Half Billion Dollars On Cybersecurity," *Forbes*, 30 June 2016, <https://www.forbes.com/sites/stevemorgan/2016/01/30/why-j-p-morgan-chase-co-is-spending-a-half-billion-dollars-on-cybersecurity/#269503c12599>.
31. The White House, Executive Order -- "Blocking the Property of Certain Persons Engaging in Significant Malicious Cyber-Enabled Activities," 1 April 2015, <https://obamawhitehouse.archives.gov/the-press-office/2015/04/01/executive-order-blocking-property-certain-persons-engaging-significant-m>.
32. F. D. Kramer, R. J. Butler, and C. Lotrionte, "Cyber and Deterrence: The Military-Civil Nexus in High-End Conflict" (Brent Scowcroft Center on International Security: Atlantic Council, January 2017), p14, http://www.atlanticcouncil.org/images/publications/Cyber_and_Deterrence_web_0103.pdf.
33. JP 3-12, pII-4.
34. C. Bing, "Inside 'Project Indigo', emphasis added.
35. G. Rattray and J. Healey, *Chapter: Categorizing and Understanding Offensive Cyber Capabilities and Their Use*, Proceedings of a Workshop on Detering Cyberattacks: Informing Strategies and Developing Options for U.S. Policy (Washington, D.C: National Academies Press, 2010), p94, <https://www.nap.edu/read/12997/chapter/8>.
36. US-CERT Webpage, "National Cybersecurity and Communications Integration Center," accessed 25 July 2018.
37. Department of Homeland Security Inspector General, DHS' Efforts to Coordinate the Activities of Federal Cyber Operations Centers," Spotlight OIG-14-02, October 2013, https://www.oig.dhs.gov/assets/Mgmt/2014/OIG_SLP_14-02_Oct13.pdf.
38. M. Bowden, *Worm: The First Digital World War*, Grove Press, 2011, p180.
39. National Security Telecommunications Advisory Committee, "Report to the President on Information and Communications Technology Mobilization," November 2014, p6, <https://www.dhs.gov/sites/default/files/publications/NSTAC%20-%20Information%20and%20Communications%20Technology%20Mobilization%20Report%2011-19-2014.pdf>.
40. Ibid.
41. J. Healey, ed. *A Fierce Domain*, p71, quoting Bill Woodcock of the Packet Clearing House, and NSP-SEC member who helped mitigate the attacks on Estonia.
42. Ibid., p71.
43. Industry Consortium for Advancement of Security on the Internet website, <https://www.icas.org/current-activities/>.
44. Interview with Neil Jenkins, Cyber Threat Alliance, 17 July 2018.
45. Kramer, Butler, and Lotrionte: *Cyber and Deterrence*, p2.
46. NSTAC Mobilization Report, pp8,12.
47. Ibid., p13.
48. T. Gjelten, "Cyber Briefings 'Scare The Bejeezus' Out Of CEOs," *National Public Radio*, 9 May 2012, <https://www.npr.org/2012/05/09/152296621/cyber-briefings-scare-the-bejeezus-out-of-ceos>.
49. M. Lowenthal, *Intelligence: From Secrets to Policy*, Fifth Edition, Sage Copress, 2012, pp68-69.
50. U.S. Government Accountability Office, "Civil Support."
51. Ibid., 15.
52. U.S. Government Accountability Office, "Defense Civil Support: DOD Needs to Identify National Guard's Cyber Capabilities and Address Challenges in Its Exercises" (U.S. Government Accountability Office, September 6, 2016), <http://www.gao.gov/products/GAO-16-574>.
53. Ibid., 13-14.
54. Kramer, Butler, and Lotrionte: *Cyber and Deterrence*, p14.
55. NSTAC Mobilization Report, p31.

NOTES

56. K. M. Donovan, "Expanding the Department of Defense's Role in Cyber Civil Support," Defense Technical Information Center (DTIC) (NORFOLK VA: National Defense University Joint Advanced Warfighting School, June 17, 2011), 53, <http://www.dtic.mil/docs/citations/ADA545641>.
57. "Joint Action Plan for State-Federal Unity of Effort on Cybersecurity" (Council of Governors, July 15, 2014), 1, <https://www.nga.org/files/live/sites/NGA/files/pdf/2014/1407CouncilofGovernorsCyberJointActionPlan.pdf>.
58. J. Soucy, "National Guard Set to Activate Additional Cyber Units," United States Army, [Www.army.mil](http://www.army.mil), (December 9, 2015), http://www.army.mil/article/159759/National_Guard_set_to_activate_additional_cyber_units.
59. C. A. Hopes, "The Challenges of Defense Support of Civil Authorities and Homeland Defense in the Cyber Domain," Defense Technical Information Center (DTIC) (Newport RI: Naval War College, Joint Military Operations Department, May 20, 2013), <http://www.dtic.mil/docs/citations/ADA583525>.

Borders in Cyberspace

*Strategic
Information Conflict
since 9/11*

Michael Warner

“The idea of degrading the opponent's information flow and, conversely, to protect or improve our own, has gained reasonably widespread acceptance and has resulted in important applications.”

– Thomas P. Rona, *Weapons Systems and Information War*, 1976^[1]

The Cold War ended in 1991 with the Soviet Union extinct and the United States perhaps the most powerful country in history, at least in relative terms. President Bill Clinton suggested at his 1993 inauguration that conflict had become an isolated phenomenon of extremists fighting against world order, disrupting nations and peoples but holding no real hope of accomplishing anything positive.^[2] The end of the Cold War seemed to have restored respect for sovereignty grounded in international law. History had “ended” and the world had turned toward liberalism—but not wholly.

The Westphalian ideal that sovereign powers should manage their internal affairs without outside interference had always been honored more in the breach, at least outside of Europe. In the 1990s, however, a new doctrine dawned—that strong nations had the right and, indeed, the duty to collaborate under the auspices of international bodies in order to stop widespread atrocities and humanitarian disasters—with force, if necessary, and even inside the sovereign borders of states unable or unwilling to halt the depredations.

The notion that international law and institutions could be used to justify and potentially even require interventions by military coalitions against autocratic regimes keeping order (however brutally) on their own territory disturbed some prominent United Nations (UN) members, especially Russia. International law of that stripe could potentially find a way around sovereignty to let liberal coalitions foment an insurrection against autocrats—and then use the regime's suppression of the revolt as a pretext under the UN (or some other body) to intervene.

© 2019 Michael Warner

This would have repercussions for international relations, the internet, and every user connecting online.

A Freedom Agenda

“The best hope for peace in our world is the expansion of freedom in all the world.”

– President George W. Bush at his Second Inaugural, 2005^[3]

UN Secretary-General Kofi Annan in March 2000 issued a report that, perhaps to his surprise, would quietly frame much of the dialogue over international relations in the decade to come:

Few would disagree that both the defence of humanity and the defence of sovereignty are principles that must be supported....But surely no legal principle – not even sovereignty – can ever shield crimes against humanity. Where such crimes occur and peaceful attempts to halt them have been exhausted, the Security Council has a moral duty to act on behalf of the international community. The fact that we cannot protect people everywhere is no reason for doing nothing when we can. Armed intervention must always remain the option of last resort, but in the face of mass murder it is an option that cannot be relinquished.^[4]

The doctrine that the Secretary-General articulated would soon be dubbed the “responsibility to protect.” Dictators and one-party states feared it. Their resistance to it had to be indirect or muted, however, while the United States remained the world’s preeminent military power and worked in concert with allies. President Bush had implied that certain nations should be wary of such notions in his January 2002 State of the Union address, mentioning North Korea, Iraq, and Iran, and insisting that “[s]tates like these and their terrorist allies constitute an axis of evil, arming to threaten the peace of the world.”^[5]

Saddam Hussein’s regime in Iraq survived barely a year after Bush’s speech. The British and Americans argued that they already possessed a warrant for intervention from the UN Security Council’s 1991 demand for Iraqi disarmament and its call for “such further steps as may be required...to secure peace and security in the area.”^[6] Their coalition assault on Iraq in March 2003 resulted in the destruction of Saddam’s regime in just three weeks. The Iraq War then paid one nearly immediate bonus—it convinced Libyan dictator Muammar Qaddafi, long a thorn in Europe’s side, to abandon his chemical weapons in late 2003.^[7] Other states drew the opposite lesson about weapons of mass destruction: North Korea and Iran soon accelerated their nuclear efforts. And in Iraq and Afghanistan, insurgencies arose to bleed the coalition occupiers and complicate their potentially Sisyphean efforts to rebuild those societies.

The years that followed thus saw varied efforts to deter or weaken Western power and resolve to impose international standards of rights in particular sovereignties. Even the possibility of synchronized, regime-changing warfare haunted the dictators. Such strength emboldened

democratic reformers in Ukraine (the Orange Revolution), Burma (the Saffron Revolution), Lebanon (the Cedar Revolution), and other lands who trusted America's commitment to what President Bush called his "freedom agenda."^[8] As Bush stated at his second inauguration, the United States applauded such revolutions. Bush stated America would "seek and support the growth of democratic movements and institutions in every nation and culture, with the ultimate goal of ending tyranny in our world."^[9]

To survive, the dictators had to adapt. One of the most creative in doing so would be Russia's President Vladimir Putin, who took the time to explain what he was doing when he spoke to the annual Munich Security Conference in February 2007. Russia wanted cooperation, particularly in arms control, Putin insisted, but his speech nonetheless struck an ominous tone. No state however powerful could build a "unipolar world" in modern times, he explained. Yet that did not stop some parties from wanting such an international order, and in this quest they had caused "new human tragedies and created new centres of tension." Putin left little doubt whom he blamed for the "almost uncontained hyper use of force—military force—in international relations, force that is plunging the world into an abyss of permanent conflicts." After all, it was "first and foremost the United States" that had "overstepped its national borders in every way. This is visible in the economic, political, cultural, and educational policies it imposes on other nations."^[10]

The United States had accomplices in this work, Putin hinted. International law had become an instrument of the strong, who showed disdain for its principles and independent legal norms. Such overreach was "extremely dangerous" because it had created a situation in which "no one feels safe." Indeed, "no one can feel that international law is like a stone wall that will protect them"—hence, the race by "a number of countries to acquire weapons of mass destruction." The nations of Europe had helped to erode the rule of law and had begun working to isolate Russia, imposing "new dividing lines and walls...that cut through our continent." There were instruments like the Organization for Security and Cooperation in Europe and "non-governmental organisations" financed and controlled from afar for "interfering in the internal affairs of other countries." Groups like these were busily "imposing a regime that determines how these states should live and develop." Now Russia would go its own way, or at least work with "responsible and independent partners" in constructing "a fair and democratic world order that would ensure security and prosperity not only for a select few, but for all."^[11]

Putin's speech in Munich previewed the tensions that would emerge over the next decade. Moscow now possessed the resources and will to act on the hitherto academic critiques of Western dominance that Putin had echoed in Munich. In the years since taking over from the garrulous democrat Boris Yeltsin, Putin had consolidated power, strengthening a handful of oligarchs, suppressing independent media outlets, and rigging the political system to keep himself in command. Most dictatorships sooner or later quarrel with their neighbors, even if such frictions do not always lead to war, and Russia was no different. Massive denial-of-service

attacks against Estonian cyberspace briefly crippled the government of Estonia in 2007 after the Estonians moved a Soviet-era war memorial in a gesture that Moscow deemed disrespectful. The disruption of Estonia—a member of the North Atlantic Treaty Organization (NATO) and the European Union (EU)—drew no blood. Nonetheless, a senior EU official was quoted in an article in *The Guardian* just after the attacks as saying, “Frankly it is clear that what happened in Estonia in the cyber-attacks is not acceptable and a very serious disturbance.”^[12] Russian forces tangled with Georgian troops the following year, this time over the status of two disputed provinces. Moscow sought to teach a lesson to Georgian President Mikheil Saakashvili and Russia’s troops advanced to within 40 miles of Georgia’s capital before the Kremlin signed a ceasefire. Afterward, President George W. Bush professed to liking Saakashvili but described him to Putin as “hot-blooded.” “I’m hot-blooded, too,” retorted Putin. “No, Vladimir,” Bush observed. “You’re cold-blooded.”^[13]

President Barack Obama’s new administration in 2009 sought to turn Putin’s energies toward more constructive channels. Hillary Clinton, the new secretary of state, promised a “reset” of bilateral relations, dealing constructively with the Russians where mutual interests converged, showing firmness to “limit their negative behavior,” and “engaging consistently with the Russian people themselves.”^[14] That last element—reaching the peoples of Russia and other dictatorships—would become a cornerstone of American foreign policy during President Obama’s first term, as Secretary Clinton later explained in her memoir. Autocracies increasingly sought to shield their subjects from the Internet to decrease U.S. and Western influence, Clinton lamented:

Around the world, some countries began erecting electronic barriers to prevent their people from using the internet freely and fully. Censors expunged words, names, and phrases from search engine results...One of the most prominent examples was China, which, as of 2013, was home to nearly 600 million internet users but also some of the most repressive limits on internet freedom. The “Great Firewall” blocked foreign websites and particular pages with content perceived as threatening to the Communist Party.^[15]

This was information conflict that targeted the populace, Clinton suggested. She pushed the State Department to counter such restrictions—for instance, by training citizen activists around oppressive regimes to employ cyber tools that could “protect their privacy and anonymity online and thwart restrictive government firewalls.” By 2011, she noted, “we had invested more than \$45 million in tools to help keep dissidents safe online and trained more than five thousand activists worldwide, who turned around and trained thousands more.” Clinton herself visited one of these workshops that year in Lithuania, figuratively on Russia’s doorstep.^[16]

The Internet, as many in the West hoped, became a powerful tool for dissent. Iranian repression would be seen by millions in 2009 with the shooting death in Tehran of a young protester, Neda Agha-Soltan, captured on cell-phone video, uploaded online, and shared via Twitter and Facebook.^[17] Iranian authorities crushed widespread protests that year but emerged from the crisis badly shaken. Another long-ruling regime in Tunisia, by contrast, would not survive similarly popular unrest facilitated by social media the following year. When Tunisian strongman Ben Ali tried to suppress social media sites, the leaderless but surging protests against repression and corruption turned to text messaging on nearly ubiquitous cell phones as the organizing tool.^[18] Mass protests against the rule of yet another dictator, Hosni Mubarak of Egypt, soon followed the Tunisian example. Mubarak left office less than a month after Tunisia's Ben Ali fled in January 2011. "Thanks to the internet, especially social media, citizens and community organizations had gained much more access to information and a greater ability to speak out than ever before," reflected Secretary Clinton in her memoirs.^[19]

A brief but tumultuous "Arab Spring" emerged from these upheavals and swept across the Middle East, with protests in Algeria, Bahrain, Jordan, Kuwait, Morocco, Oman, Sudan, Yemen, and beyond. Dictatorships elsewhere saw they had to respond. They did so clumsily at first, trying to close down internet service providers or block social media sites. The smarter ones, like Iran, quickly learned to hunt on the web in order to develop a meaningful understanding of where their adversaries were, what they did, and where they were headed. "The new technologies allow us to identify conspirators and those who are violating the law, without having to control all people individually," boasted Iran's top policeman, Esmail Ahmadi-Moghaddam, in early 2010.^[20] No countries saw more violence, however, than Libya and Syria, both ruled by secular Arab dictators and oppressed for decades by pervasive police states. Both regimes turned their militaries on protesters, who rebelled and found arms and courage to defend themselves, pitching both nations into civil war.

Libya proved an early test of the Kofi Annan's "responsibility to protect" doctrine in March 2011. With the African Union condemning the violence and the Arab League voting to impose a No-Fly Zone over rebel-held territory to deter Qaddafi's avenging tanks, the Security Council passed (with Russia and China abstaining) a resolution finding that the "deteriorating situation" constituted "a threat to international peace and security." With this justification for intervening in an internal Libyan crisis, the council authorized "all necessary measures" short of foreign occupation to protect Libyan civilians.^[21] The resulting military intervention followed almost immediately in now-classic fashion, with U.S.-led airstrikes and countermeasures to suppress Libyan air defenses and permit NATO aircraft to pound Qaddafi's armor and artillery (under Operation UNIFIED PROTECTOR). An unnamed adviser to President Obama described the American role in the Libya campaign to *The New Yorker* as "leading from behind."^[22] Qaddafi's regime shrank to nothing over the following summer, with the dictator himself cornered and killed in October 2011.

Syria would be a much tougher problem. Libya would shape the Syrian conflict that opened in 2011. NATO's intervention had caused uncharacteristic public disagreements among Russian leaders. Putin, then serving as prime minister (and thus officially not the chief executive of the Russian state), alleged Western hypocrisy in attacking Qaddafi's regime while tolerating other dictators: "When the so-called civilized community, with all its might, pounces on a small country, and ruins infrastructure that has been built over generations – well, I don't know, is this good or bad?"^[23] His ostensible boss, President Dimitri Medvedev, shunned such rhetoric and had declined to veto the Security Council resolution authorizing "all necessary means" in Libya. The NATO effort still looked to Moscow like a campaign to depose Qaddafi, however, and the Russians felt they could take no such risks with Syria, Russia's only ally in the Middle East (with ties dating back to the Cold War). Moscow thus opposed any Security Council action aimed at Syria's Bashar al-Assad unless it ruled out armed intervention.^[24] Russia and China cast the only dissenting votes in vetoing a Security Council resolution condemning Assad's suppression of the growing rebellion. Moscow's foreign minister complained that the resolution was "taking sides in a civil war," while the Russian ambassador to the UN alleged that the Western leaders once again were "calling for regime change, pushing the opposition towards power."^[25] Secretary Clinton, in her memoirs called the Russian and Chinese veto "despicable."^[26]

Prime Minister Putin for his part had already expressed his contempt for Clinton and her ideas. Shortly after announcing his ultimately successful candidacy to resume the presidency of Russia, which would be decided in a spring 2012 election, Putin showed his anxiety over democratic movements like the Arab Spring. Responding to popular complaints of election corruption in Russia's late 2011 parliamentary balloting, Putin blamed the disturbance on Secretary Clinton: "She set the tone for some actors in our country and gave them a signal," said Putin. "They heard the signal and with the support of the U.S. State Department began active work." Once again, he saw shadowy foreign forces dividing Russians against one another, spending vast sums of "foreign money" to influence the Russian balloting.^[27]

For the time Putin could only fume. The liberal West seemed triumphant, with its enemies and all dictators at risk. That moment would ironically prove to be the crest of a soon-receding democratic wave. Baghdad and its Shi'a government promptly turned a blind eye while the Iranian Revolutionary Guard Corps—the Praetorian Guard of Tehran's theocracy—ferried civilian airliners over Iraqi airspace to deliver troops and weapons to Assad's beleaguered regime in Syria.^[28] With Iran's military help and Russian diplomatic cover, Assad managed to hold on against the various squabbling rebel groups, and even began using chemical weapons on the insurgents in 2012.^[29] Libya meanwhile degenerated into a vicious civil war. Democracy retreated in Egypt. The successor regime to Mubarak's authoritarianism held an election won by the Muslim Brotherhood, who began imposing a different brand of Egyptian authoritarianism until they were ousted a year later by millions of protesters across Egypt and a military coup. Washington showed no inclination for military intervention in the region. Indeed, Secretary

Clinton, in contemplating the “wicked problem” that was Syria, found little willingness to arm insurgent factions or allow U.S. forces to engage. She and President Obama’s advisors felt a military solution was “impossible” and resolved to avoid “another quagmire, like Iraq.”^[30]

The diplomatic and military turn against democracy corresponded with a new boldness among autocracies and one-party states in using cyberspace operations to defend themselves from falling to the sorts of popular unrest seen in the Arab Spring. As Clinton noted above, they worked to guard their digital as well as their physical borders, erecting national firewalls, enhancing the reach and quality of internal propaganda, tightening control of state media, and floating proposals in international forums to replace the allegedly U.S.-dominated “multi-stakeholder model” of Internet governance. Perhaps just as importantly, they turned their portions of cyberspace into surveillance systems with which they could monitor internal and external challenges. So disturbed, the regimes perhaps shared little beyond an abhorrence and a fear of liberal nostrums like elections, dissent, and a free press. Ironically, the Internet soon proved to be just as powerful a support for the centralization of political power as it had been for dissent.

The Internet had endangered state control in many ways, yet, at the same time, it facilitated state surveillance on a hitherto unimagined scale and repression even beyond a state’s physical borders. Seen from the perspective of the regimes in question, such steps looked purely defensive and, indeed, necessary in a world where liberal ideals like international law could now be used, as in the cases of Kosovo and Libya, to trump the traditional, Westphalian defense of state sovereignty. A Chinese military organ, for example, implicitly rejected Secretary Clinton’s optimism about the web’s force for good; as noted by *Xinhua* in 2015:

The Chinese military’s mouthpiece newspaper has warned of the possibility of “Western hostile forces” using the Internet to foment revolution in China. “The Internet has grown into an ideological battlefield, and whoever controls the tool will win the war,” according to an editorial published in the People’s Liberation Army (PLA) Daily on Wednesday. It stressed the need for cyber security measures to ensure “online ideological safety”, euphemisms suggesting efforts to safeguard China’s mainstream ideology. “Western hostile forces along with a small number of Chinese ‘ideological traitors’, have maliciously attacked the Communist Party of China, and smeared our founding leaders and heroes, with the help of the Internet,” according to the paper. “Their fundamental objective is to confuse us with ‘universal values’, disturb us with ‘constitutional democracy’, and eventually overthrow our country through ‘color revolution’,” it added, using a term commonly applied to revolutionary movements that first developed in the former Soviet Union in the early 2000s. “Regime collapse that can occur overnight often starts from long-term ideological erosion,” it warned. The paper said the military should not only safeguard national sovereignty and security on traditional battlefields, but also “protect ideological and political security on the invisible battleground of the Internet.”^[31]

These sentiments echoed those voiced by senior Chinese military spokesmen since 2010, when China began informing American diplomats that its territorial claims in the South China Sea were now “core interests,” on par with Taiwan and Tibet in Beijing’s strategic calculus. The Americans, Chinese rear admiral Guan Youfei angrily remarked to a delegation that included Secretary Clinton, were acting like a “hegemon” and seeking to encircle China.^[32]

The key development here was something that might have seemed impossible: a merging of Information Age–technology facilitating regime propaganda and surveillance. Authoritarian, anti-liberal regimes craved external threats to justify central direction; mobilization of the citizenry; and, ultimately, repression. Such states could not abide open borders with prosperous, liberal democracies, so they sought to keep those physical and virtual borders closed—or those neighbors less free. These regimes, moreover, could now surveil their opponents’ every key-stroke. Targeting and suppression of civilian dissent were aided as well by intelligence services utilizing cyber means to attain global reach and unprecedented economies of scale. Even the poorest dictators now could acquire means to monitor dissidents on distant continents.^[33]

A Return to War

“...it is essential to have a clear understanding of the forms and methods of the use of the application of force.”

– General Valery Gerasimov, Chief of Russia’s General Staff, 2013^[34]

The Winter Olympics in 2014 opened in Sochi, Russia, showcasing some of the world’s best athletes competing for medals and honors rather than land and treasure. That year the Olympic spirit of sportsmanship did not linger, however, after the Games’ closing ceremony on February 23. Two subsequent events would soon shape global relations in the years to come. Russian troops intervened in Ukraine just days later, effectively seizing Crimea. Their intervention shook Western leaders. “You just don’t in the 21st century behave in 19th century fashion by invading another country on a completely trumped up pretext,” complained the new U.S. Secretary of State, John Kerry, when asked on a news program about Russia’s bullying of Ukraine.^[35] The 19th century looked civilized, however, compared to what happened in the Middle East. Barely a hundred days after the Olympics, fighters from the Islamic State of Iraq and the Levant (ISIL)—whom President Obama in January had called the “JV team”—burst out of Syria into western Iraq.^[36] In weeks they overran perhaps 35,000 square miles in Syria and Iraq, including Mosul, Iraq’s second-largest city, where they seized the central bank and hundreds of millions of dollars in assets. ISIL then declared itself “the Islamic State” and proclaimed it was now a worldwide caliphate to which was owed the allegiance of all faithful Muslims.^[37]

Events turned as they did in 2014 because dictators accelerated measures to protect their physical and virtual borders, keeping the democracies at a distance by building buffer zones around themselves. Russian leaders claimed aloud that this was a defensive strategy, made

necessary by the liberal West's promotion of regime change under the guise of humanitarian intervention. Indeed, one of Putin's advisors, Vladislav Surkov, had been watching for years the progress of the color revolutions. An interviewer from *Spiegel* asked Surkov in 2005 how Moscow might defend itself "against the revolutionary virus that could jump over into Russia from Georgia, Kyrgyzstan and Ukraine." Surkov responded that Russia would see no such uprising, despite the desires of some in his country. He complained of "various foreign non-governmental organizations that would like to see the scenario repeated in Russia. We understand this. By now, there are even technologies for overthrowing governments and schools where one can learn the trade."^[38]

The possibility of an Arab Spring in Russia also occurred to General Valery Gerasimov, chief of the General Staff, before he visited the Academy of Military Science in February 2013 to call on its experts to help Russian leaders adapt in a rapidly changing world. "In the 21st century," he began, "we have seen a tendency toward blurring the lines between the states of war and peace. Wars are no longer declared and, having begun, proceed according to an unfamiliar template."^[39] This lack of sharp lines between peace and war made contemporary conflicts seemingly non-linear but no less deadly, said Gerasimov:

The experience of military conflicts – including those connected with the so-called [color] revolutions in north Africa and the Middle East – confirm that a perfectly thriving state can, in a matter of months and even days, be transformed into an arena of fierce armed conflict, become a victim of foreign intervention, and sink into a web of chaos, humanitarian catastrophe, and civil war.^[40]

Gerasimov suggested to his military audience that crises like the Arab Spring might just be "typical of warfare in the 21st century." "The information space" created by global networking and mass media had opened "wide asymmetrical possibilities" for attacking a regime: "In North Africa, we witnessed the use of technologies for influencing state structures and the population with the help of information networks." Indeed, nonmilitary means of achieving strategic goals often exceeded "the power of force of weapons in their effectiveness," for "methods of conflict" such as "political, economic, informational, humanitarian, and other non-military measures" could now be "applied in coordination with the protest potential of the population." Aggressor powers bide their time, holding their armed forces in reserve until the right moment: "The open use of forces – often under the guise of peacekeeping and crisis regulation – is resorted to only at a certain stage, primarily for the achievement of final success in the conflict."^[41]

What General Gerasimov viewed as so potentially deadly was the combination by the "world's leading states" of the Information Warfare concepts derived from Thomas Rona with the new media- and diplomacy-enabled means of influencing a population ruled by the target regime. Mobile, combined arms forces, "acting in a single intelligence-information space because of the use of the new possibilities of command-and-control systems," now ensured that a victim had no respite or opportunity to counterattack. "Frontal engagements of large formations" would

be few, for the United States and others were learning to launch “[l]ong-distance, contactless actions” to defeat an adversary “throughout the entire depth of his territory.” Even powerful adversaries (and, by implication, Russia, Gerasimov hinted) could see their military advantages nullified by “the use of special operations forces and internal opposition to create a permanently operating front through the entire territory.”

Russia, suggested Gerasimov, should heed that warning and learn to conduct “activities in the information space, including the defense of our own objects.” The Russian military, he said, well understood “the essence of traditional military actions carried out by regular armed forces,” but Russian military leaders possessed “only a superficial understanding of asymmetrical forms and means”—hence, his request to the Academy of Military Science to help “create a comprehensive theory of such actions.” Conflicts in Ukraine and Syria would soon demonstrate how quickly the Russians learned.^[43]

A newly democratic Russia had once pledged (in 1994) to respect Ukraine’s borders when the post-Communist government there had returned Soviet-era nuclear weapons to Moscow’s control. Russian troops took control of Crimea in 2014, however, six days after the pro-Russian President of Ukraine, Viktor Yanukovich, fled in what Moscow had called a coup. The new, pro-Western government in Kiev hailed his flight as a liberation, calling the revolution the *Euromaidan* (after the protests that erupted when Yanukovich’s government derailed an imminent association agreement with the EU). Russian leaders insisted they had not violated the 1994 pledge, yet offered no consistent rationale for their position. Masked, Russian-speaking troops with no insignia suddenly were guarding Russian-made, heavy weapons all over Crimea. Local residents noted their alien origin and dubbed them “little green men,” a term that was quickly echoed in the Ukrainian press and beyond.^[44]

The UN Security Council soon debated the Crimea crisis. A draft resolution in March did not mention Russia but declared invalid the upcoming, Moscow-endorsed referendum in Crimea (which asked Crimeans whether they wanted Russian rule). The UN Security Council resolution also noted the international community’s “commitment to the sovereignty, independence, unity and territorial integrity of Ukraine within its internationally recognized borders.” Moscow vetoed the draft resolution, and in the Crimea referendum the following day, 97 percent of voters expressed their desire to join Russia. The Kremlin quickly granted their request, declaring its annexation of Crimea on March 18, 2014.^[45] Unlike the Iraqi annexation of Kuwait in 1990, however, this time the UN never contemplated armed intervention to restore the pre-crisis borders of Ukraine. Instead, the democracies turned to the UN General Assembly, which passed a nonbinding resolution of its own, calling on “all States, international organizations and specialized agencies not to recognize any alteration of the status of the Autonomous Republic of Crimea and the city of Sevastopol.”^[46] Russia’s Foreign Ministry called the General Assembly’s resolution counterproductive and complained that “shameless pressure, up to the point of political blackmail and economic threats, was brought to bear on a number of

[UN] member states” by Western diplomats seeking “yes” votes for the measure.^[47] Moscow’s subsequent intervention in Ukraine appeared ad hoc and driven by circumstances. After the Crimean annexation, ethnic Russians in two eastern Ukrainian districts also began agitating to join Russia, forcibly resisting Ukrainian troops and declaring their territory “New Russia” that spring. Kiev launched a counteroffensive in July, only to see it stall as the rebels gained support from units of the Russian military with armor, artillery, and anti-aircraft missiles. The missiles nullified the combat effectiveness of Ukraine’s small air force and promptly caused a major international embarrassment for Moscow when a battery of SA-11s downed Malaysian Airlines Flight 17 that July, destroying the cruising jetliner at 33,000 feet and killing all 298 people aboard.

Moscow denied responsibility and blamed Ukrainian forces, in keeping with its official disavowal of any direct role in the conflict. Russia’s misdirection from the beginning outraged European governments. Britain’s Secret Intelligence Service told its parliamentary oversight committee in late 2017, for example, that Russia had mounted a massive disinformation effort to support its actions in Ukraine and beyond:

An early example of this was a hugely intensive, multi-channel propaganda effort to persuade the world that Russia bore no responsibility for the shooting down of [Malaysian Airlines flight] MH-17 (an outright falsehood: we know beyond any reasonable doubt that the Russian military supplied and subsequently recovered the missile launcher).^[48]

Eastern Ukrainian separatists received their support from more “little green men,” who advised in all manner of military and civil matters. “We’re Russian. We’re all Russian,” quipped one in Donetsk to the BBC in April 2014. “And this land isn’t Ukraine: it’s Novorossiya - and we will defend it.”^[49] NATO, especially its eastern members, took alarm at this mostly nonviolent but effective display of force, calling it “hybrid” warfare, in which “a wide range of overt and covert military, paramilitary, and civilian measures are employed in a highly integrated design.”^[50] As in Syria, diplomatic efforts to end the conflict in Ukraine proved futile.^[51] Low-level hostilities between the Ukrainian military and Russian-backed separatists continue to this day.

As the Ukrainian conflict erupted in 2014, another crisis emerged almost simultaneously from the ongoing Syrian Civil War and its threat to Russia’s allies in Damascus. Insurgencies and even terrorists seek in their various ways to attain statehood—to overturn an existing regime or to fashion a new one from the territory of some other power. Al-Qaeda came closest to attaining global influence while not ruling its own territory, but that was while its Taliban allies ran most of Afghanistan. The chaotic conflict in Syria by 2014 had created a political and military vacuum in Syria’s eastern reaches, while the Shi’a-dominated government of neighboring Iraq alienated the Sunnis of its western districts. The American withdrawal from Iraq at the end of 2011 had ended the sustained presence of sophisticated intelligence, reconnaissance, and strike forces in the area, and now troops and vehicles could once again gather on a battlefield.

Into the vacuum stepped ISIL, which in 2013 turned its energies from fighting Assad; despite its retrograde social views, ISIL saw statehood as its best path toward the ultimate goal of a caliphate across the Muslim world. ISIL stormed over Iraq's border in early 2014, its reputation for savagery preceding it, panicking Iraqi defenders (the group tortured and executed those soldiers it caught).^[52] Its fighters seized thousands of square miles of Syrian and Iraqi territory in just weeks. By summer, ISIL had erected "a primitive but rigid administrative system" maintaining "some basic services in a highly repressive environment" and imposing its version of Islamic law on more than eight million people, including Sunnis and Shiites, along with Christians, Yazidis, Kurds, and other beleaguered minorities.^[53]

ISIL sought to make its offensive global over the next year, accepting allegiance from like-minded groups in Asia and Africa and calling for attacks in the West. Thousands of adherents from around the world journeyed to ISIL-controlled areas to fight on its behalf.^[54] ISIL's barbarity attracted adherents, yet succeeded in uniting a diverse coalition of states to oppose it in the Middle East and beyond.^[55] The United States assembled in late 2014 a Coalition of fifty-nine states and the EU to work against ISIL; its charter endorsed "a common, multifaceted, and long-term strategy to degrade and defeat ISIL" by military, diplomatic, and economic means. The Coalition's communique also noted that some participants insisted on the need for "effective ground forces to ultimately defeat ISIL" and "increased support to these moderate opposition forces which are fighting on multiple fronts against ISIL/Daesh, Al Nusrah Front, and the Syrian regime." Iraq and its neighbors cosigned the communique; Syria, Iran, and Russia did not. The U.S. military soon organized a Combined Joint Task Force in Kuwait to coordinate combat operations against ISIL. The military intervention that followed in Iraq and Syria was patterned on the model of NATO operations in Libya and Afghanistan, with advanced coalition forces mounting airstrikes and supporting commandos working with local forces, who did most of the fighting against their countrymen (and sometimes even their neighbors). The U.S. campaign began reaching into Syria in May 2015 with a Special Forces raid that killed senior ISIL leader Abu Sayyaf. Washington also hinted in August that it would defend friendly Syrian forces with airstrikes, even against Assad's troops.^[58]

Russia and Iran then worried that Assad's regime could collapse under the simultaneous (though uncoordinated) pressure from ISIL and the Coalition-backed "moderate opposition forces." Assad controlled less than a fifth of Syria's territory by the summer of 2015.^[59] The international effort to suppress ISIL thus gave Moscow a diplomatic opening to introduce Russian forces directly into the Syrian conflict. All services of Russia's military joined in the campaign that fall, mounting well publicized strikes with all of the advanced conventional arms at their disposal. Russian strategic bombers and warships firing cruise missiles saw their combat debuts as General Gerasimov and his lieutenants gained practical experience synchronizing long-range strike operations, ostensibly mounted against ISIL, but often hitting the Coalition-backed Syrian opposition instead.^[60] Moscow implicitly patterned its intervention on the

U.S.-led Coalition effort, in which the advanced militaries provided local allied forces with the intelligence, surveillance, and reconnaissance; logistics; and command and control essential for sustained, modern campaigns.^[61] With Russia's newest and most powerful weapons now frequenting Syria's crowded airspace, moreover, Coalition leaders lost whatever opportunity they might have had to impose on Assad a military solution to the Syrian Civil War.

The intervention by Russia and Iran allowed Assad to slowly reclaim Syrian cities from his opponents as the Coalition drove ISIL from Iraq and reduced its holdings in Syria. Assad's forces took Aleppo in late 2016, while the Iraqi army, with Coalition support, uprooted ISIL from Mosul in July 2017 and declared Iraqi territory ISIL-free the following December. By then the Syrian city of Raqqah, the ostensible capital of ISIL's caliphate, had already fallen to Coalition forces. ISIL had "lost nearly all of the territory they once held," explained a Combined Joint Task Force spokesman at the end of 2017, though he cautioned that ISIL was not quite finished. "We know this enemy is as adaptive and savvy as it is cruel and evil."^[62] Yet Moscow and Washington apparently agreed at this point that military victory in the Middle East was not impossible.

A Clash of Worlds?

General Gerasimov, in 2013, predicted that future conflicts would be waged in what he called the "information space." Within a few years of his speech, every shooting war also had a digital dimension. Almost every gun or missile today is employed with the aid of some digital device, even if only the cell phone that detonates the roadside bomb or the video that spurs the aspiring jihadist. Networked digital information gets the weapons and ammunition to the right place at the right time—whether such armaments reach the battlefield on tanks, fighter jets, or ships, or in men's arms—and digital technology helps to maintain and control them. At the same time, several regimes now attack opponents in cyberspace as well. The clashes over borders between the West and the various anti-liberal regimes became virtual as well as physical.

Such attacks had already begun when General Gerasimov made his prediction. Iranian hackers between late 2011 and mid-2013 attacked American financial companies, according to the indictments of seven Iranians won by the Justice Department in March 2016:

Using botnets and other malicious computer code, the individuals—employed by two Iran-based computer companies sponsored and directed by the Iranian government—engaged in a systematic campaign of distributed denial of service (DDoS) attacks against nearly 50 institutions in the U.S. financial sector.

Their coordinated attacks disabled bank websites, frustrated customers, and "collectively required tens of millions of dollars to mitigate."^[63] North Korea entered the fray the following year, attacking Sony Pictures Entertainment for releasing an otherwise forgettable satire about an assassination attempt on North Korea's dictator Kim Jong-un. Secretary of State Kerry publicly condemned North Korea's "cyber-attack targeting Sony Pictures Entertainment and

the unacceptable threats against movie theatres and moviegoers.” Kerry called the attacks “a brazen attempt by an isolated regime to suppress free speech and stifle the creative expression of artists beyond the borders of its own country.”^[64] China moved with greater discretion. In March 2015, someone attacked the website of GreatFire for hosting material that would help computer users avoid official censorship. Independent researchers at the University of Toronto’s Citizen Lab found that this new weapon rested on China’s so-called “Great Firewall”; Citizen Lab called this capability “the Great Cannon” and noted its sinister novelty:

The operational deployment of the Great Cannon represents a significant escalation in state-level information control: the normalization of widespread use of an attack tool to enforce censorship by weaponizing users. Specifically, the Cannon manipulates the traffic of “bystander” systems outside China, silently programming their browsers to create a massive [distributed denial-of-service] attack.^[65]

At least one regime has gone well beyond censorship and cyberattacks on opponents to manipulate information with cyber tools. According to the indictment of 13 Russians handed up by Special Counsel Robert Mueller’s investigation in February 2018, for instance, Moscow, soon after the Ukrainian intervention, mounted a covert campaign to get Americans arguing with one another. A Russian organization called the Internet Research Agency “as early as 2014... began operations to interfere with the U.S. political system, including the 2016 U.S. presidential election,” noted the indictment.^[66] The Russians employed a classic divide-and-conquer tactic, attacking the presidential candidates that they (along with most American experts) considered strongest while ignoring their apparently weaker challengers. Russian agents, said the indictment:

engaged in operations primarily intended to communicate derogatory information about Hillary Clinton, to denigrate other candidates such as Ted Cruz and Marco Rubio, and to support Bernie Sanders and then-candidate Donald Trump...On or about February 10, 2016, Defendants and their co-conspirators internally circulated an outline of themes for future content to be posted to [Internet Research Agency]-controlled social media accounts. Specialists were instructed to post content that focused on “politics in the USA” and to “use any opportunity to criticize Hillary and the rest (except Sanders and Trump—we support them).”^[67]

The efforts of these Russian hackers received support from leaks of embarrassing emails exfiltrated from the headquarters of the Democratic Party and released to the news media in increments to hamper Clinton’s campaign. A month before the election, the secretary of homeland security and the director of national intelligence jointly explained to the world that the “Russian Government directed the recent compromises of e-mails from US persons and institutions, including from US political organizations.” The disclosures resembled “the methods and motivations of Russian-directed efforts”; indeed, “the Russians have used similar tactics and techniques across Europe and Eurasia, for example, to influence public opinion there.”

Secretary Jeh Johnson and Director James Clapper assessed, in light of “the scope and sensitivity of these efforts, that only Russia's senior-most officials could have authorized these activities.”^[68]

As the world saw in America's 2016 election, such targeting of individuals and societies via the “information space” can have strategic effects. Cyber campaigns backed by massive arsenals looked very formidable indeed by late 2017. British leaders began discussing in public the apparently growing threat of Russian cyber and electoral disruption backed by powerful, conventional, and even nuclear forces. Prime Minister Theresa May warned in November 2017 that Moscow had “mounted a sustained campaign of cyber-espionage and disruption.”^[69] Its tactics, she claimed, “included meddling in elections and hacking the Danish Ministry of Defence and the [German] Bundestag among many others.” A few days later, Ciaran Martin, chief of Britain's new National Cyber Security Centre, accused Russia of attacking Britain's media, telecommunications, and energy sectors, and of “seeking to undermine the international system.”^[70]

American strategists recognized as well the return of great-power competition by 2018. Secretary of Defense James Mattis released his *National Defense Strategy* that January and observers immediately noted its bleak tone and its argument that “inter-state strategic competition, not terrorism, is now the primary concern in U.S. national security.”^[71] The new American strategy saw states remaining the primary locus of power in the modern world, but perhaps did not see how much states were now driven by technological and ideological influences beyond their control.

CONCLUSION

...war was now understood as a process, more exactly, part of a process, its acute phase, but maybe not the most important.

– Natan Dubovitsky, “Without Sky”^[72]

Ancient ways of mobilizing power for force and using it to scatter foes have gained new reach and impact in the last two decades, both on the battlefield and for internal security. It lies beyond the scope of this paper to explain how these new means became subject, for the sake of efficiency, to automated logical programs sorting digitized data and new concepts of international law. What the paper narrates is how that very technology opened new avenues for force and extraordinary opportunities for surveillance while new ideas of law ironically canalized conflict in a “humanitarian” direction. The question of trust remained throughout, at the level of the leader, the commander, and the individual. Can you trust those with whom you would do business? Can you trust that your computer is guarding your data or presenting you with the truth? Can you trust that international law will protect your sovereignty—or protect you from

your government? Conflict endured as regimes and organizations that could not live at peace with their own citizens ultimately could not remain at peace with their neighbors. The liberal ascendancy that President Clinton described in 1993 thus brought not peace but a long struggle for survival on the part of dictators against the ostensibly universal appeal of liberal ideals. For the foreseeable future, that struggle will proceed on physical, legal, and virtual battlefields, with the “borders” between narratives and visions—and questions of trust—cutting across geographic terrain and reaching into every nation. 🛡️

Michael Warner serves in the U.S. Department of Defense. This paper is excerpted from his upcoming book, *Twin Swords: A History of Force*, co-authored with John Childress. The opinions expressed in *this paper are the author's alone*, and do not represent official positions of the Department of Defense or any U.S. Government entity.

NOTES

1. Thomas P. Rona, *Weapons Systems and Information War*, Seattle: Boeing Corporation [for the Office of the Secretary of Defense], July 1, 1976, p. 5; accessed on February 4, 2018 at www.esd.whs.mil/Portals/.../09-F-0070-Weapon-Systems-and-Information-War.pdf.
2. William J. Clinton, Inaugural Address, January 20, 1993; accessed December 28, 2017 at <http://www.presidency.ucsb.edu/ws/index.php?pid=46366>.
3. Second Inaugural Address of George W. Bush; January 20, 2005; accessed January 22, 2018 at http://avalon.law.yale.edu/21st_century/gbush2.asp.
4. United Nations, Report of the Secretary General, "We the peoples: the role of the United Nations in the twenty-first century," March 27, 2000, p. 35; accessed January 21, 2018 at <http://unpan1.un.org/intradoc/groups/public/documents/un/unpan000923.pdf>.
5. George W. Bush, State of the Union address, January 29, 2002; accessed January 21, 2018; <https://web.archive.org/web/2011011053416/http://millercenter.org/president/speeches/detail/4540>.
6. George W. Bush, *Decision Points* (New York: Random House, 2010), p. 244. UN Security Council Resolution 687, April 3, 1991, accessed January 20, 2018 at [https://undocs.org/S/RES/687\(1991\)](https://undocs.org/S/RES/687(1991)).
7. Commission on the Intelligence Capabilities of the United States regarding Weapons of Mass Destruction, *Report to the President of the United States* [the WMD Commission Report] (Washington: Government Printing Office, 2005), pp. 251-252; accessed February 4, 2018 at <https://www.gpo.gov/fdsys/pkg/GPO-WMD/content-detail.html>.
8. Bush, *Decision Points*, p. 437.
9. Second Inaugural Address of George W. Bush; January 20, 2005; accessed January 22, 2018 at http://avalon.law.yale.edu/21st_century/gbush2.asp.
10. Speech of Russian President Vladimir Putin at the Munich Security Conference, February 10, 2007; accessed February 10, 2018 at <http://www.washingtonpost.com/wp-dyn/content/article/2007/02/12/AR2007021200555.html>; see also Robert M. Gates, *Duty: Memoirs of a Secretary at War* (New York: Knopf, 2014), pp. 154, 326.
11. Ibid.
12. Ian Traynor, "Russia accused of unleashing cyberwar to disable Estonia," May 16, 2007; accessed February 11, 2018 at <https://www.theguardian.com/world/2007/may/17/topstories3.russia>.
13. Bush, *Decision Points*, p. 435.
14. Hillary Rodham Clinton, *Hard Choices* (New York: Simon & Schuster, 2014), pp. 228, 231.
15. Clinton, *Hard Choices*, p. 548.
16. Clinton, *Hard Choices*, pp. 545, 549.
17. Nazila Fathi, "In a Death Seen Around the World, a Symbol of Iranian Protests," *New York Times*, June 22, 2009; accessed February 14, 2018 at <http://www.nytimes.com/2009/06/23/world/middleeast/23neda.html>.
18. Philip N. Howard and Muzammil M. Hussain, "Egypt and Tunisia: The Role of Digital Media," in Larry Diamond and Marc F. Plattner, eds., *Liberation Technology: Social Media and the Struggle for Democracy* (Baltimore: Johns Hopkins University Press, 2013), pp. 111-113.
19. Clinton, *Hard Choices*, p. 49.
20. "Iran's police vow no tolerance towards protesters," *Reuters*, February 6, 2010; accessed February 17, 2018 at <https://www.reuters.com/article/idUSDAH634347>.
21. UN Security Council Resolution 1973, March 17, 2011; accessed February 17, 2018 at <https://www.un.org/sc/suborg/en/s/res/1973-%282011%29>.
22. Ryan Lizza, "Leading from Behind," *The New Yorker*, April 26, 2011; accessed February 16, 2011 at <https://www.newyorker.com/news/news-desk/leading-from-behind>.
23. Ellen Barry, "Putin Criticizes West for Libya Incursion," *New York Times*, April 26, 2011; accessed February 17, 2018 at <http://www.nytimes.com/2011/04/27/world/europe/27putin.html>.
24. Steve Gutterman, "Russia says will veto 'unacceptable' Syria resolution," *Reuters*, January 31, 2012; accessed February 17, 2018 at <https://www.reuters.com/article/us-syria/russia-says-will-veto-unacceptable-syria-resolution-idUSTRE80S08620120201>.

NOTES

25. Neil MacFarquhar and Anthony Shadidfeb, "Russia and China Block U.N. Action on Crisis in Syria," *New York Times*, February 4, 2012; accessed February 17, 2018 at <http://www.nytimes.com/2012/02/05/world/middleeast/syria-homs-death-toll-said-to-rise.html> Russia and China did endorse UNSCR 2042 the following month; the new resolution called for a ceasefire and authorized observers to monitor it, but added no text supportive of an international humanitarian intervention.
"Veto on Syria sparks Arab and Western fury," Al Jazeera, February 5, 2012; accessed February 17, 2018 at <http://www.aljazeera.com/news/middleeast/2012/02/201224162422121856.html>.
26. Clinton, *Hard Choices*, p. 452.
27. David M. Herszenhorn and Ellen Barry, "Putin Contends Clinton Incited Unrest Over Vote," *New York Times*, December 8, 2011; accessed February 11, 2018 at <http://www.nytimes.com/2011/12/09/world/europe/putin-accuses-clinton-of-instigating-russian-protests.html>.
28. Michael R. Gordon, "Iran Supplying Syrian Military via Iraqi Airspace," *New York Times*, September 4, 2012; accessed February 17, 2018 at <http://www.nytimes.com/2012/09/05/world/middleeast/iran-supplying-syrian-military-via-iraq-airspace.html>.
29. Clinton, *Hard Choices*, pp. 461.
30. Clinton, *Hard Choices*, pp. 460-463.
31. This story was covered by several Western outlets. See "Internet the key front in China's battle with Western hostile forces: military paper," Reuters, May 20, 2015; accessed May 23, 2015 at <https://ca.news.yahoo.com/internet-key-front-chinas-battle-western-hostile-forces-120506329.html>. See also Sean Gallagher, "Chinese Army newspaper calls for military role in Internet culture war: Claims West and 'ideological traitors' use Internet to weaken Party's authority," *Ars Technica*, May 21, 2015; accessed May 23, 2015 at <http://arstechnica.com/tech-policy/2015/05/chinese-army-newspaper-calls-for-military-role-in-internet-culture-war/>.
32. Clinton, *Hard Choices*, p. 76. Jim Pomfret, "In Chinese admiral's outburst, a lingering distrust of U.S.," *Washington Post*, June 8, 2010; accessed February 18, 2018 at <http://www.washingtonpost.com/wp-dyn/content/article/2010/06/07/AR2010060704762.html?sid=ST2010060705111>.
33. See, for instance, Citizen Lab, Munk Centre for Global Affairs, University of Toronto, "Hacking Team and the Targeting of Ethiopian Journalists," February 12, 2014; accessed March 3, 2018 at <https://citizenlab.org/2014/02/hacking-team-targeting-ethiopian-journalists/>.
34. Gerasimov's speech was translated by Robert Coalson and reprinted in his "Top Russian General Lays Bare Putin's Plan for Ukraine," *Huffington Post*, September 2, 2014; accessed February 19, 2018 at https://www.huffingtonpost.com/robert-coalson/valery-gerasimov-putin-ukraine_b_5748480.html.
35. CBS News, *Face the Nation* Transcripts, March 2 2014; accessed February 18, 2018 at <https://www.cbsnews.com/news/face-the-nation-transcripts-march-2-2014-kerry-hagel/>.
36. At American colleges, the junior varsity (J.V.) team is the secondary, training squad for less-seasoned athletes. See David Remnick, "Going the Distance," *New Yorker*, January 27, 2014; accessed February 19, 2018 at <https://www.newyorker.com/magazine/2014/01/27/going-the-distance-david-remnick>.
37. United Nations, "Rule of Terror: Living under ISIS in Syria," Report of the Independent International Commission of Inquiry on the Syrian Arab Republic, November 14, 2014, p. 3; accessed February 24, 2018 at https://web.archive.org/web/20150204115327/http://www.ohchr.org/Documents/HRBodies/HRCouncil/CoSyria/HRC_CRP_ISIS_14Nov2014.pdf.
38. "Interview with Kremlin Boss Vladislav Surkov: 'The West Doesn't Have to Love Us'," *Spiegel*, June 20, 2005; accessed February 19, 2018 at <http://www.spiegel.de/international/spiegel/spiegel-interview-with-kremlin-boss-vladislav-surkov-the-west-doesn-t-have-to-love-us-a-361236.html>.
39. Gerasimov's speech was translated by Robert Coalson and reprinted in his "Top Russian General Lays Bare Putin's Plan for Ukraine," *Huffington Post*, September 2, 2014; accessed February 19, 2018 at https://www.huffingtonpost.com/robert-coalson/valery-gerasimov-putin-ukraine_b_5748480.html.
40. Coalson, "Top Russian General Lays Bare Putin's Plan for Ukraine."
41. Coalson, "Top Russian General Lays Bare Putin's Plan for Ukraine."
42. Coalson, "Top Russian General Lays Bare Putin's Plan for Ukraine."
43. Coalson, "Top Russian General Lays Bare Putin's Plan for Ukraine."

NOTES

44. Vitaly Shevchenko "'Little green men' or 'Russian invaders?'" *BBC*, 11 March 2014; accessed February 24, 2018 at <http://www.bbc.com/news/world-europe-26532154>.
45. UN Security Council minutes, S/PV.7138, March 15, 2014; accessed March 3, 2018 at www.securitycouncilreport.org/atf/cf/%7B65BFCF9B-6D27-4E9C-8CD3-CF6E4FF96FF9%7D/s_pv_7138.pdf.
46. UN General Assembly Resolution 68/262, March 27, 2014; accessed February 18, 2018 at https://www.un.org/en/ga/search/view_doc.asp?symbol=A/RES/68/262.
47. "Russia criticizes U.N. resolution condemning Crimea's secession," *Reuters*, March 28, 2014; accessed February 18, 2018 at <https://www.reuters.com/article/us-ukraine-crisis-un-russia/russia-criticizes-u-n-resolution-condemning-crimeas-secession-idUSBREA2R0DA20140328>.
48. Intelligence and Security Committee, *Annual Report 2016-2017* (London: Her Majesty's Stationery Office, December 2017), p. 52, accessed February 20, 2018 at https://sites.google.com/a/independent.gov.uk/isc/files/2016-2017_ISC_AR.pdf?attredirects=1.
49. Steven Rosenberg, "Ukraine crisis: Meeting the little green men," *BBC*, April 30, 2014; accessed February 24, 2018 at <http://www.bbc.com/news/world-europe-27231649>.
50. North Atlantic Council, *Wales Summit Declaration*, September 5, 2014; accessed March 3, 2018 at http://www.nato.int/cps/en/natohq/official_texts_112964.htm.
51. Organization for Security and Co-operation in Europe, "Forward Patrol Bases: Two Years on the Contact Line," September 26, 2017; accessed February 20, 2018 at <https://www.osce.org/stories/forward-patrol-bases-two-years-on-the-contact-line>.
52. United Nations Security Council, "Implementation of Security Council resolutions 2139 (2014) and 2165 (2014): Report of the Secretary-General," S/2014/696, September 24, 2014, pp. 5-6; accessed February 25, 2018 at <http://unbisnet.un.org:8080/ipac20/ipac.jsp?session=1A1SN72309111.128093&profile=bib&uri=full=3100001-!1035254-!576&ri=1&aspect=subtab124&menu=search&source=-!horizon>.
53. United Nations, "Rule of Terror: Living under ISIS in Syria," Report of the Independent International Commission of Inquiry on the Syrian Arab Republic, November 14, 2014, p. 3; accessed February 24, 2018 at https://web.archive.org/web/20150204115327/http://www.ohchr.org/Documents/HRBodies/HRCouncil/CoISyria/HRC_CRP_ISIS_14Nov2014.pdf.
54. UN Security Council, "Implementation of Security Council resolutions 2139 (2014) and 2165 (2014)," p. 5.
55. United Nations, "Secretary-General's remarks to Security Council High-Level Summit on Foreign Terrorist Fighters," September 24, 2014; accessed February 25, 2018 at <https://www.un.org/sg/en/content/sg/statement/2014-09-24/secretary-generals-remarks-security-council-high-level-summit>.
56. US Department of State, "Joint Statement Issued by Partners at the Counter-ISIL Coalition Ministerial Meeting," December 3, 2014; accessed February 25, 2018 at <https://2009-2017.state.gov/r/pa/prs/ps/2014/12/234627.html>.
57. US Department of Defense, "Statement by Secretary of Defense Ash Carter on Counter-ISIL Operation in Syria," Press Operations Release No: NR-175-15, May 16, 2015; accessed February 23, 2018 at <https://www.defense.gov/News/News-Releases/News-Release-View/Article/605506/>.
58. David Lerman, "Obama Authorizes Airstrikes to Defend Syrian Rebels If Attacked," *Bloomberg*, August 2, 2015; accessed February 23, 2018 at <https://www.bloomberg.com/news/articles/2015-08-02/obama-authorizes-airstrikes-to-defend-syrian-rebels-if-attacked>.
59. Columb Strack, "Syrian government no longer controls 83% of the country," *Jane's Intelligence Review*, August 24, 2015; accessed February 25, 2018 at <http://www.janes.com/article/53771/syrian-government-no-longer-controls-83-of-the-country>.
60. Helene Cooper, Michael R. Gordon, and Neil MacFarquhar, "Russians Strike Targets in Syria, but Not ISIS Areas," *New York Times*, September 30, 2015; accessed March 4, 2018 at <https://www.nytimes.com/2015/10/01/world/europe/russia-air-strikes-syria.html>.
61. Matthew Bodner, "Russia Shows Early Success, New Capabilities in Syria," *Defense News*, October 18, 2015; accessed March 4, 2018 at <https://www.defensenews.com/home/2015/10/18/russia-shows-early-success-new-capabilities-in-syria/>.
62. US Department of Defense [Combined Joint Task Force-Operation Inherent Resolve], "Department of Defense Press Briefing by Colonel Dillon via Teleconference From Kuwait," December 19, 2017; accessed February 25, 2018 at <https://www.defense.gov/News/Transcripts/Transcript-View/Article/1400723/departement-of-defense-press-briefing-by-colonel-dillon-via-teleconference-from/>.

NOTES

63. Federal Bureau of Investigation, “Iranians Charged with Hacking U.S. Financial Sector,” press release, March 24, 2016; accessed February 26, 2018 at <https://www.fbi.gov/news/stories/iranians-charged-with-hacking-us-financial-sector> . See also Department of Justice, Office of Public Affairs, “Seven Iranians Working for Islamic Revolutionary Guard Corps-Affiliated Entities Charged for Conducting Coordinated Campaign of Cyber Attacks Against U.S. Financial Sector,” March 24, 2016; accessed February 26, 2018 at <https://www.justice.gov/opa/pr/seven-iranians-working-islamic-revolutionary-guard-corps-affiliated-entities-charged> .
64. John Kerry, Secretary of State, “Condemning Cyber-Attack by North Korea,” press statement, December 19, 2014; accessed February 26, 2018 at <https://2009-2017.state.gov/secretary/remarks/2014/12/235444.htm>.
65. Bill Marczak, Nicholas Weaver, et al., “China’s Great Cannon,” Citizen Lab, Munk School of Global Affairs, University of Toronto, April 10, 2015; accessed February 18, 2015 at <https://citizenlab.ca/2015/04/chinas-great-cannon/>.
66. United States of America v. Internet Research Agency et al., US District Court for the District of Columbia, February 16, 2018, p. 3; accessed February 17, 2018 at https://www.scribd.com/document/371718383/Internet-Research-Agency-Indictment-pdf#from_embed.
67. Ibid, p. 17. See also Scott Shane, “These Are the Ads Russia Bought on Facebook in 2016,” *New York Times*, November 1, 2017, accessed February 19, 2018 at <https://www.nytimes.com/2017/11/01/us/politics/russia-2016-election-facebook.html>.
68. “Joint Statement from the Department Of Homeland Security and Office of the Director of National Intelligence on Election Security,” October 7, 2016; accessed February 26, 2018 at <https://www.dhs.gov/news/2016/10/07/joint-statement-department-homeland-security-and-office-director-national>.
69. “Theresa May accuses Vladimir Putin of election meddling,” *BBC*, November 14, 2017; accessed February 26, 2018 at <http://www.bbc.com/news/uk-politics-41973043>.
70. “UK cyber-defence chief accuses Russia of hack attacks,” *BBC*, 15 November 2017; accessed February 26, 2018 at <http://www.bbc.com/news/technology-41997262>.
71. Department of Defense, *National Defense Strategy*, January 19, 2018, p. 1; accessed on January 22, 2018 at <https://admin.govexec.com/media/20180118173223431.pdf>.
72. Natan Dubovitsky, [Vladislav Surkov], “Without Sky,” *Russky Pioneer* 46 (March 12, 2014); accessed February 19, 2018 at http://www.bewilderingstories.com/issue582/without_sky.html . See also Peter Pomerantsev, “Non-Linear War,” *London Review of Books*, March 28, 2014; accessed February 19, 2018 at <https://www.lrb.co.uk/blog/2014/03/28/peter-pomerantsev/non-linear-war/>.

Persistent Engagement, Agreed Competition, and Cyberspace Interaction Dynamics and Escalation

Michael P. Fischerkeller

*Institute for Defense Analyses
Alexandria, Virginia*

Richard J. Harknett

*University of Cincinnati
Cincinnati, Ohio*

ABSTRACT

Policymakers and academics have raised concerns over escalation should states adopt a more proactive cyberspace posture. The unspoken context for those fears is potential, episodic, offensive cyber operations that threaten to cause, or cause, physical damage. This narrow focus excludes an equally, if not more important, strategic space—actual, continuous, strategic competition without resort to armed attack, a space which, according to 2018 U.S. strategic guidance, poses a central challenge to national security. U.S. Cyber Command (USCYBERCOM) has described a strategic approach to cyberspace intended to counter and contest adversary gains: persistent engagement. This approach is assessed through a re-consideration of Herman Kahn's *On Escalation*. It is concluded that competitive interaction in cyberspace short of armed conflict in an agreed competition, as opposed to spiraling escalation, best explains the dynamic from persistent engagement and, consequently, prevailing concerns of escalation are unwarranted. Agreement to compete robustly short of armed conflict may be the grand strategic consequence of cyberspace.

Keywords—escalation, agreed competition, cyberspace, interaction, persistent engagement, strategy.

I. INTRODUCTION

A significant concern among policymakers and academics discussing cyber operations is a fear of escalation should states adopt a more proactive posture in cyberspace.^[1] Past policy statements and international security scholarship tend to focus narrowly on the escalation dynamics resulting from cyberattacks, or the threat thereof, which might cause physical damage or loss of life. This limited focus on *potential and episodic*, cyber-enabled

This work was funded by the Institute for Defense Analyses, Alexandria, Virginia.

© 2019 Michael P. Fischerkeller, Richard J. Harknett

crises or war scenarios excludes an equally, if not more important, strategic space—*actual and continuous*, strategic competition in cyberspace that does not reach the level of armed conflict. In 2018, U.S. strategic guidance in the *National Security Strategy of the United States of America* (NSS) shifted to emphasize the significance of this competitive space, and US-CYBERCOM prescribed a strategic approach of *persistent engagement* to contest and counter the ability of adversaries to gain strategic advantage without engaging in armed attack. This article considers this shift in U.S. guidance documents and analyzes the potential interaction dynamics in a strategic cyber environment structured by interconnectedness—constant contact—persistent engagement. In so doing, the article introduces a distinction between interaction and escalation dynamics. This article concludes that fears that persistent engagement in cyberspace will result in spiraling or uncontrollable escalation are not warranted because advantage can be gained through competitive interactions, rather than through the pursuit of escalation dominance.

This article is structured as follows. To set the context under which interaction dynamics will be considered, the first section summarizes the view of a competitive environment described in the White House and U.S. Department of Defense (DoD) 2017 and 2018 strategic guidance. This is followed by an overview of the strategic approach of *persistent engagement*—both its theoretical and conceptual foundations and its operational prescription as provided by USCYBERCOM. Next is a review of the core security studies literature on escalation dynamics—in general and specific to cyberspace. The current strategic environment is then considered in light of this scholarship, generating a set of propositions regarding the impact of persistent engagement on cyberspace interaction dynamics. The stability of these operational dynamics is then discussed, followed by a brief consideration of shifting away from the traditional “ladder” metaphor for understanding cyberspace interaction dynamics.

II. STRATEGIC ENVIRONMENT

The 2018 NSS and its complements, the *National Defense Strategy* (NDS) and the *Department of Defense 2018 Cyber Strategy*, stand in marked contrast to their predecessors in their declarations that adversaries are executing strategic campaigns short of an armed attack to secure and advance national interests. Indeed, these documents assert that the central challenge to U.S. security and prosperity is the reemergence of a long-term, *strategic* competition with revisionist and rogue regimes and actors that have become skilled at operating below the threshold of armed conflict, challenging the United States, its allies, and partners with deniable, hostile actions that seek to undermine faith and confidence in democratic institutions and the global economic system.^[2]

Cyberspace and its derivative cyber operations, in particular, have been identified as offering state and non-state adversaries the ability to wage strategic campaigns against American political, economic, and security interests without ever physically crossing U.S. borders.^[3] This view is presented most comprehensively in the 2018 Command Vision for U.S. Cyber Command, in

which adversaries are described as continuously operating against the United States below the threshold of armed conflict—demonstrating the resolve, technical capability, and persistence to undertake strategic cyberspace campaigns to weaken U.S. democratic institutions and gain economic, diplomatic, and military advantages.^{[4],[5]} What is of critical importance to note from these documents is the assessment that these operations short of armed conflict can have a cumulative impact at the strategic level: these operations can degrade or damage sources of American national power. Analytically, if this assessment is correct, it is not simply the United States that can be affected by such operations, but, in practice, all state actors reliant on cyberspace for the development and projection of national power. It is in response to this challenge that USCYBERCOM has prescribed the strategic approach of *persistent engagement*.

III. PERSISTENT ENGAGEMENT

From a security studies perspective, cyberspace may be best understood as a technically enabled operational domain with distinct features that shape particular behaviors by state actors, businesses, and even individuals. Interconnectedness is the oft-cited, but rarely embraced in strategic thinking, core structural feature. If one accepts interconnectedness as such, then fundamental international relations concepts for understanding or explaining actor behaviors come into question, such as sovereignty and territoriality, because the core condition that follows from interconnectedness is constant contact, a term referenced by USCYBERCOM to describe the cyberspace operating environment.^{[6],[7]} This condition, when coupled with the nature and substance of cyberspace—a vulnerable and resilient technological system that is a global warehouse of and gateway to troves of sensitive, strategic information—encourages persistent opportunism to access and leverage those sensitive data while simultaneously requiring states to continuously seek to secure those data and data flows from others. The combination of interconnectedness and constant contact with cyberspace’s ever-changing character both in “terrain” and in the capacity for maneuver across that terrain further encourages operational persistence and persistent engagement in order to secure and leverage critical data and data flows.^[8] When these factors are considered in sum, in operational reality, operational persistence and persistent engagement become a strategic imperative for states seeking to secure and advance their interests in, through, and from cyberspace.

This theoretical and conceptual argument for operational persistence and persistent engagement is consistent with nearly a decade of domain and operational observations by USCYBERCOM. For example, in reference to the ever-changing character of cyberspace, the *Command Vision* notes that cyberspace is where new vulnerabilities and opportunities continually arise as new terrain emerges; no target remains static; no offensive or defensive capability remains indefinitely effective; no advantage is permanent; and well-defended cyber terrain is attainable but continually at risk. And adversary offensive activities are also said to persist because opportunity costs are low, and accesses, platforms, and payloads can remain useful for extended periods.^{[9],[10]}

To operate effectively in this dynamic environment, USCYBERCOM prescribes that the United States increase resiliency, defend forward as close as possible to the origin of adversary activity, and contest cyberspace actors to generate continuous tactical, operational, and strategic advantage.^[11] They argue that a strategic approach of *persistent engagement*—described operationally as the combination of seamless resiliency, forward defending, and contesting—will compel many U.S. adversaries to shift resources to defense and reduce attacks. Moreover, *persistent engagement* is expected to allow for greater freedom of maneuver to impose tactical friction and strategic costs on U.S. adversaries pursuing more dangerous activities before they impair U.S. national power. This effort seeks to render the majority of adversary cyber and cyber-enabled activity inconsequential.

The Command Vision is absent any discussion of potential escalation risks from a strategic approach of *persistent engagement*.^[12] This is a notable omission because the document does include a section on risks and risk mitigation.^[13] Given that continuous engagement is intended to create uncertainty and cause friction, two factors often associated with increased risk of escalation, those predisposed to escalation concerns likely view this approach with alarm. Whether or not they should is a key question and the focus of the remainder of this article.

IV. BACKGROUND ON ESCALATION DYNAMICS

It is not contentious to say that modern thinking regarding escalation dynamics was introduced in the seminal work of Herman Kahn, in which he defined escalation as “an increase in the level of conflict in international crisis situations.”^[14] Starting with the assumption of some limited conflict or *agreed battle*, Kahn proposed a framework populated by three mechanisms (“ways”) in which a would-be escalator could increase, or threaten to increase, his efforts: “increasing intensity,” “widening the area,” and “compounding.”^[15] *Intensity* is described as a function of doing more of what one is already doing—using more equipment; using new equipment; attacking new targets, such as logistics; or a more “intensive increase,” such as switching to nuclear weapons or attacks on cities.^[16] *Widening the area* is described as increasing the geographical scope of the conflict. *Compounding* is described as extending the conflict to include allies or clients. Kahn’s escalation ladder was developed with a focus on the deliberate escalation in *potential, episodic* conflicts, giving primary attention to the threat or reality of force or coercion as a factor in negotiation.^[17] Stated differently, in order to explore potential escalation dynamics from the launching point of a limited conflict, Kahn assumed that pursuit of any of these three ways would be viewed as escalatory. The state that could employ these mechanisms to achieve escalation dominance could gain strategic advantage. This was all necessitated by the need to avoid all-out nuclear war.

Kahn argues that there are two basic classes of strategies that each side can use when engaged in limited conflict or *agreed battle*. One class makes use of the factors relating to particular levels of escalation in order to gain an advantage. The other uses the risks or threat of escalation or eruption from the *agreed battle*.^[18] The latter, he notes, refers to the class of deterrence strategies.

Given its foundational and enduring value, it is not surprising to find Kahn's influence in more recent scholarship on escalation dynamics that focuses on nuclear as well as non-nuclear-capable states in *potential, episodic* confrontations that involve or might come to involve the use of military force.^[19] Morgan *et alia* expand Kahn's focus of deliberate escalation to include other mechanisms: inadvertent as well as accidental escalation. Similar to Kahn's description, *deliberate* escalation is understood as being carried out with specific purposes in mind. For example, a party may deliberately escalate a conflict to gain an advantage, to preempt, to avoid defeat, to signal an adversary about its own intentions and motivations, or to penalize an adversary for some previous action.^[20] *Inadvertent* escalation is described as when one party deliberately takes actions that it does not believe are escalatory but which are interpreted as escalatory by another party to the conflict.^[21] Such misinterpretation may occur because of incomplete information, lack of shared reference frames, or one party's thresholds or "lines in the sand" of which other parties are not aware. Finally, *accidental* escalation is described as when some operational action has direct effects that are unintended by those who ordered them—for example, a weapon may go astray to hit the wrong target, the rules of engagement may be unclear, a unit may take unauthorized actions, or a high-level command decision may not be received properly by all relevant units.^[22]

Morgan *et alia* also assign Kahn's "ways" of escalating to dimensions, where the *vertical* dimension is associated with "increasing intensity" and a *horizontal* dimension is associated with "widening the area." They further equate the combination of *horizontal* and *vertical* with Kahn's "way" of *compounding*. In addition, they introduce a *political* dimension to escalation, which is described as when states adopt more extreme or unlimited objectives in crises/conflicts or, alternatively, pursue measures such as relaxing behavioral constraints that protect civilians.^[23] Like Kahn's work, the study also proposes that the class of deterrence strategies is best suited for managing an enemy's propensity for deliberate escalation—discouraging an enemy from deliberately escalating a conflict by convincing that enemy that the costs of such actions will outweigh the benefits that may be accrued through escalation.^[24] Within that class of strategies, they further argue that the key to managing risks of inadvertent escalation lies in clarifying thresholds—on all sides of a conflict.^[25] Finally, they propose that the key to mitigating accidental escalation lies in an effective command and control strategy.^[26]

V. CYBERSPACE ESCALATION DYNAMICS

Herbert Lin was an early adopter/adaptor of the Morgan *et alia* framework for cyberspace by referencing it to aid in answering how the initial stages of conflict in cyberspace might evolve or escalate and what might be done to prevent or deter such escalation.^[27] Lin also focused on how *potential, episodic* cyber conflict at any given level might be de-escalated or terminated (and what might be done to facilitate de-escalation or termination) and how cyber conflict might escalate into kinetic conflict (and what might be done to prevent kinetic escalation).^[28]

Lin's approach to responding to these questions is largely grounded in generating new sets of questions about, and challenges associated with, escalation dynamics in cyberspace. In support of his objective in writing the article, these serve as valuable checklists for national security planners and policymakers to reference in preparing for and managing a cyber-enabled crisis or armed conflict.^[29]

Martin Libicki also adopted the Morgan *et alia* framework to explain escalation risk and dynamics in cyberspace, albeit with a stronger focus on potential risk.^[30] Like Kahn and Morgan *et alia*, the context for his escalation discussion is *potential, episodic* conflicts (conflicts that involve or might come to involve military force); once a crisis has blossomed into conflict, he states, crisis management becomes escalation management.^[31] Stated differently, he focuses on the escalation risks associated with operational cyber war in which cyberattacks are carried out against targets that are considered legitimate war targets. Different types of targets are argued to carry different risks of escalation. Those outside a local conflict zone will carry one set of risks, civilian targets may carry another, dual-use another, and military and strategic targets yet another. Libicki argues that the relative severity of those risks is a function of the value the adversary places on the targets.^[32]

A similar argument is presented by Lawrence Cavaiola *et alia* in an article on escalation dynamics in a *potential, episodic, cyber-enabled* war.^[33] This effort blends Libicki's arguments into a succinct presentation, arguing that escalation could happen along three paths: horizontal, from military to civilian systems; vertical, from tactical to strategic military systems (perhaps affecting those that control nuclear weapons); and vertical, from limited civilian targeting to major civilian consequences.^[34] Similar to other studies, the primary focus is on deliberate escalation, but the potential for inadvertent and accidental escalation is also explored by considering the many unique challenges that cyberspace and cyber operations pose, perhaps the most significant being uncertainty associated with attribution and primary (and/or potential secondary or tertiary) operational effects.

In sum, Kahn's work laid the conceptual foundations for thinking about "ways" in which would-be escalators could pursue escalation dominance and thereby achieve a strategic advantage in a limited conflict. Scholars have begun to theorize what escalation dynamics may look like using similar ways in a cyber conflict. That said, there exists no "escalation ladder" equivalent, nor has there been a rich discussion of whether the "ladder" metaphor is even appropriate. This review also highlights that most of the cyberspace escalation scholarship adopt the same point of origin as Kahn (i.e., the deliberate escalation from a *potential, episodic, operational conflict* or *agreed battle*), giving primary attention to the threat or reality of force or coercion as a factor in negotiation. In addition, all also argue that the class of deterrence strategies is best for managing escalation from this starting point. Set against the empirical record of cyber operations over the past 15 years, however, it raises the question of why have we not seen a recurring escalation.^[35] Why has this remained a space dominated, instead, by competitive interaction?^[36]

VI. CYBERSPACE INTERACTION DYNAMICS AND ESCALATION IN TODAY'S STRATEGIC ENVIRONMENT

The security studies community primarily has focused on *escalation* dynamics in cyberspace at the exclusion of interaction dynamics. Kahn, however, provides a basis for their consideration by mentioning a second class of strategies for managing escalation for *agreed battle*, a class that has all but been forgotten—*making use of the factors relating to particular levels of escalation in order to gain an advantage*.^[37] This is the class of strategies into which persistent engagement appears to fit. Whereas deterrence strategies are well and commonly understood, this second class deserves further elaboration because it can play an important role in understanding cyberspace *interaction* as opposed to *escalation* dynamics. But first, the concept of *agreed battle* has to be considered in light of the current strategic environment because it will establish the strategic context for discussing this second class of strategies in the same.

As noted above, *agreed battle* is a concept rooted in factors relating to particular levels of escalation. It emphasizes that in an escalation situation in which both sides are accepting limitations, there is, in effect, an “agreement,” whether or not it is explicit or even well understood. “Thus the term does not have any connotation of a completely shared understanding, an intention of containing indefinitely with the limitation, or even a conscious quid pro quo arrangement.”^[38] Scholars who emphatically and urgently emphasize the importance of establishing cyberspace behavioral norms will see the construction of norms in this concept.^[39] Others have argued, however, that de facto norms have already been established in cyberspace by states pursuing strategic cyber campaigns that generate effects short of armed attack.^[40] In fact, the U.S. 2018 NSS, NDS, *DoD Cyber Strategy*, and *Command Vision* admit as much by stating that adversaries are continuously operating strategically against the United States short of armed conflict via strategic cyberspace campaigns to gain economic, diplomatic, and military advantages. What is important to note in Kahn’s rendering is that the “agreed” part of the battle rests on interactions between adversaries, which, despite being complex and nuanced, can come to be understood and shared between actors.^[41] He notes that states can come to recognize “what the ‘agreed battle’ is and is not, what the legitimate and illegitimate moves are, and what are ‘within the rules’ and what are escalatory moves.”^[42]

Building upon Kahn’s notion and applying it to current cyberspace campaigns and operations, open-source evidence suggests that U.S. adversaries have, through their behaviors, tacitly established an *agreed competition* in cyberspace, bounded by the operational space inclusive of and above operational restraint (i.e., inactivity) and exclusive of and below operations generating armed-attack equivalent effects.^[43] After eight years of observing the persistent operation of adversaries in cyberspace, USCYBERCOM argued that a strategic approach of *persistent engagement* was best suited for securing and advancing national interests in this *agreed competition*.^[44] This, in effect, meets Kahn’s definition of a class of strategy that makes use of the features of the particular agreed interaction space. The United States’ adoption of this strategic approach will introduce new interactions into the *agreed competition*.

A. Structural Imperatives and Strategic Incentives

The earlier introduction to the theoretical and conceptual foundations supporting *persistent engagement* argued that the interconnectedness of cyberspace creates a structural condition that generates a strategic imperative for operational persistence and persistent engagement. Presuming that states respond to this imperative, a robust, strategic competition in cyberspace should be expected. However, that same condition and those same features also generate incentives for states to limit the impact of their cyber operational effects below the threshold of armed attack. Two incentives, in particular, are that deliberate escalation to armed attack equivalence could result in a cyberspace war that would likely be of long duration; expensive; and result in few, if any, enduring strategic gains.^[45] In addition, crossing the armed attack threshold opens the door for states to legitimately bring to bear cross-domain, conventional, kinetic weapons based on an argument of self-defense.^[46] Regarding the latter, once a conflict has expanded into multiple domains, the pursuit of national interests involves very different risks, costs, and challenges. It would no longer be *agreed competition*, but conflict, and potentially war.

In addition to these strategic incentives, James Lewis has offered a thoughtful and comprehensive discussion of the political and strategic constraints states also face in deliberately escalating above the armed attack threshold.^[47] He argues that, if you consider how great powers have historically made strategic decisions about entering into conflict, resorting to operations equivalent to an armed attack in cyberspace is highly unlikely. The existential conflicts of the last century—conflicts that required mass mobilization, territorial invasion, and mass destruction (including critical infrastructure) to realize strategic ends—are not present today.^[48] States may seek to challenge the existing international order, but these are not existential challenges to any other state, and the constraints of cost and destruction induce caution in the ways and means which those challengers adopt. And so, for example, destructive attacks on critical infrastructure are more likely to appear as too risky for U.S. adversaries, of limited benefit to their goals, and perhaps irrelevant in achieving the desired strategic outcome of undermining U.S. hegemony and building regional dominance without armed conflict with the United States.^[49] This perspective is further supported empirically through an analysis of a decade of cyber disputes among rival states.^[50]

One of the main impetuses to examining escalation control in the 1960s was the recognition among theorists and policymakers that fighting all-out nuclear war overshot any advancement of national interest. So the question became how one might advance interests, despite that risk, without using nuclear weapons. It appears that a parallel logic is taking (or has taken) hold in the strategic use of cyber means. That is, if cyber means are to have unique, strategic value, it will come from operations short of armed attack equivalence that cumulatively enhance one's own power or degrade and destabilize others' sources of national power. It could be argued, therefore, that armed attack/war (traditionally involving measures of death and

destruction) with cyber means actually overshoots the strategic utility of cyber operations. That would be “eruption,” in the language of Kahn, beyond the ceiling of *agreed competition*. And that outcome would be, for rational, strategic cyber actors, a failure of strategy. And so there is a strategic rationale for seeking to gain an advantage in, through, and from cyberspace short of armed attack. Actors might decide to engage in war, but the strategic purpose of the competitive interactions in *agreed competition* is to avoid having to do so.^{[51],[52]}

If one accepts the above arguments that there are structural incentives and strategic rationales from which *agreed competition* emerged and because of which it will sustain if and when the United States adopts a strategic approach of *persistent engagement*, an entirely new strategic space that has heretofore been unexplored for *interaction* and *escalation* dynamics is laid bare.

B. Agreed Competition – Competitive Interaction

To reiterate, when discussing *agreed battle*, Kahn argues one class of strategies use the risks or direct threat of escalation beyond the *agreed battle* to gain advantage over an adversary. These range from red lines (declared deterrence) to riskier forms of brinkmanship as well as forms of Thomas Schelling’s coercive bargaining.^[53] In discussing *agreed battle*, Kahn also recognizes a second class of strategies through which advantage can be gained by leveraging the unique features particular to a level of escalation (the space between recognized rungs in Kahn’s escalation ladder). It has been argued above that in today’s strategic environment, what defines the “particular level of escalation” associated with *agreed competition* is the space inclusive of and above operational restraint and exclusive of and below effects equivalent to an armed attack. As such, the latter represents a de facto ceiling for effects in this competition. In efforts to gain advantage in this *agreed competition*, then, it can be expected that states will do so through *competitive interaction* below this ceiling.

Kahn describes three mechanisms for seeking strategic advantage through escalation: widening, compounding, and intensifying. If we operationalize how these mechanisms manifest in cyberspace and review open-source data on their occurrence, we are left wondering why we’ve not seen recurring escalation as Kahn would have expected given the prevalence of all three over the past decade. We argue it is a result of the combination of the structural and strategic features discussed above combining to produce a strategic environment in which competitive interaction is actually strategically salient; that is, one can gain an advantage without escalating, so that operations and the strategy guiding them are focused on a very different dynamic.

Employing cyber operations short of armed-attack equivalence, states are able to secure their own and degrade, usurp, or circumvent others’ national power (economic, diplomatic, military, and social cohesion) by targeting specific data, data flows or sectors, industries, and populations that are the sources of that power. *Competitive interaction* in *agreed competition*, then, can be understood as campaigns populated by cyber operations seeking, over time and space, to generate cumulative, strategic effects (i.e., to gain advantage) by targeting sources of national

power. We propose that a different set of mechanisms (from Kahn) for achieving advantage is more descriptive of the behaviors in which comprises competitive interaction: increases in scale, scope, and/or intensity.^[54] In this *agreed competition* within cyberspace, *increasing scale* can be measured as an increase in the number of systems affected, and scope as the number of actors affected or implicated as having caused an effect (we address intensity later in this article). Characterizing cyber operational behavior using these measures leads to an obvious conclusion—the class of strategies best suited for managing competitive interaction dynamics in this *agreed competition* is that which inhibits adversary efforts to increase the scale, scope, and/or intensity of cyber operations/campaigns. The strategic approach of *persistent engagement* intends to do just that through operations that maneuver seamlessly between defense and offense across the interconnected cyber battlespace to compete more effectively outside of armed conflict.^[55]

There is substantial, publicly reported evidence of specific U.S. adversaries engaging in efforts to increase the scale and scope of their activities (as described in this manner) for the last several years, with different states doing so for different reasons to address their strategic interests.^[56] China has invested a great deal of effort in targeting a range of industry and commercial enterprises in pursuit of general scientific, technical, and business information. Examples include exfiltration of data on the F-35 Joint Strike Fighter, the F-22 Raptor fighter jet, and the MV-22 Osprey. This cyber campaign, directed at contractors and agencies residing within and external to U.S. borders (a combination of increasing scale and scope), will reduce costs and accelerate the development of foreign weapon systems; enable reverse engineering and countermeasure development; and undermine U.S. military, technological, and commercial advantage.^{[57],[58]} China has also sought out more specific information through cross-sector industry cyber operations targeting personally identifiable information (PII), possibly with the objective of using these data to facilitate future “insider” cyber operations, assist in the recruitment of human intelligence assets, or identify and monitor persons of interest to the government (e.g., dissidents, foreign journalists, and/or others who may pose a threat to the Communist Party’s image and legitimacy).^[59] Russia, through its campaign of cyber operations—including those used in Russia’s war with Georgia in 2008 and those used to influence the Brexit referendum and the U.S. election in 2016—is pursuing a strategic campaign to undermine Western democracies and weaken the multilateral alliances that Russia sees opposing its future, including the North Atlantic Treaty Organization and the European Union.^[60] Finally, it has been concluded with confidence that North Korea, in efforts to mitigate the impact of international economic sanctions, has successfully subverted for significant monetary gain the Society for Worldwide Interbank Financial Telecommunication system.^[61] Those funds likely contributed to North Korea’s ability to continue investing in its nuclear enterprise, allowing it to finally cross the threshold for intercontinental ballistic delivery and thereby undermine U.S. military overmatch.

Table 1 offers a brief summary of a few strategic cyber campaigns over a two-year period characterizing operations/campaigns of increasing scale and scope and ascribes motivations for the same by advanced persistent threat (APT) groups—groups that are assessed as taking direction from a nation-state.^[62] The table includes a 2014–2016 summary of a few strategically relevant industries, the number of threat sources, ascribed objectives for the operations, and malware families.^[63] Note that the breadth of the reported industry threats and the objectives for the same cut across military, economic, and diplomatic sources of national power.

Industry	Attack Source	Objective	Malware Families (Top Three)
Aerospace & Defense	24 APT groups	Acquire intellectual property to advance domestically produced capabilities, develop countermeasures to degrade adversary military overmatch, and produce arms for sale on global market.	47% GhOstRAT 21% PcClient 13% ZXShell
Construction & Engineering	25 APT groups	Acquire intellectual property pertaining to technical innovations, expertise, and processes to develop and advance state-owned firms and to better position those firms for bids against and negotiations with foreign firms.	52% LEOUNCIA 20% LV (a.k.a. NJRAT) 13% GhOstRAT
Financial Services & Insurance	15 APT groups	Gain insight into company operations or information on potentially sensitive customers.	34% WITCHCOVEN 22% XtremeRAT 19% GhOstRAT
Government & International Organizations	9 APT groups	Gain an edge in negotiations and agreements.	49% GhOstRAT 30% ERACS 14% PHOTO
Health Care & Health Insurance	13 APT groups	Acquire PII to facilitate future “insider” cyber operations, assist in the recruitment of human intelligence assets, or identify and monitor persons of interest to the government.	49% WITCHCOVEN 32% XtremeRAT 11% ChinaChopper
Hi-Tech & Information Technology	20 APT groups	Acquire economic and technical information to support the development of domestic companies through the reduction of research and development costs.	29% GhOstRAT 26% TAIDOOR 19% POISON IVY

Table 1: Summary of 2014–2016 Cyber Threats to Industry

A second example of the increasing scale and scope is the previously referenced case of Russia’s use of cyberspace (through social media, specifically) to undermine the confidence of adversaries’ populations and leaders in their democratic institutions and alliances, respectively.^[64] In this campaign, the increasing scale was characterized by micro-targeting at scale within populations.

In all of these cases, at the individual actor level, the strategic advantage is being gained without needing to erupt out of the agreed competition space. The mechanisms of increasing scale and scope in cyberspace are best understood not as ways of leveraging escalation, but as ways of leveraging competitive interactions.

C. Cyber-Enabled Conflict – Deliberate Intensification and Escalation

It is from the point of origin of cyber-enabled crises or war that most cyberspace escalation dynamics scholarship has been written. In this context and as related to this article, this point is realized when an actor has deliberately escalated from *agreed competition* by threatening to or generating cyber operational effects that are equivalent to armed attack. Escalation in cyberspace, then, is defined as an increase from the level of *agreed competition* to conflict (which would be inclusive of Kahn's definition of an increase in the level of conflict in international relations in crisis situations).^[65] In this framework, the potential mechanism for erupting out of agreed competition is *intensifying*. Intensifying within cyberspace is characterized by campaigns and/or operations that include increases in frequency (as a function of count over time), duration, damage, hierarchical level, and visibility of effects.^[66] Intensifying may also include expanding cyber operations to other operating domains. To help ground the concept of intensifying in actual events, a few examples follow.

Intensifying is found in the Russian campaign targeting Estonia in 2007. On the night of April 26, 2007, Estonian Government websites were subject to denial-of-service (DoS) and distributed DoS (DDoS) effects. The perpetrator launched 1,000 assaults that day, increasing that number to 2,000 per hour on the second day. On May 9, the day marking the peak of the assault, the perpetrator was injecting an average of four million packets of data per second. The assaults came in waves, were delivered from up to 85,000 systems, and continued for a 23-day period.^[67]

Behavior that would be characterized as escalatory (i.e., intensifying to generate armed-attack equivalent effects—a breach of the ceiling associated with *agreed competition*) can be illustrated through two cases^[68] Perhaps the most publicized example occurred in 2010 with the deployment of Stuxnet, which caused significant damage to the Natanz Fuel Enrichment Plant.^[69] Additionally, in 2014, a report issued by Germany's Federal Office for Information Security revealed that an unnamed steel mill in Germany had suffered “massive,” though unspecified, damage when its control systems were manipulated and disrupted to such a degree that a blast furnace could not be properly shut down.^[70]

In the escalation dynamics scholarship referenced in this article, the strategic recommendation for managing deliberate escalation, in cyberspace as well as other domains, is the class of deterrence strategies. But what if such a strategy fails and an adversary deliberately intensifies in cyberspace? How can such an action be managed in cyberspace through cyber operations within *agreed competition* and beyond it? The cases cited above hint that managing such intensification and escalation is possible, since in none of them does one find extended spirals of increasing intensification or escalation. Rather, what occurred was dissipation or a move back into the *agreed competition* space, respectively, followed by a recommencing of cyber campaigns/operations whose effects were short of armed attack. In what may appear counterintuitive to conventional wisdom, the more *competitive interaction that occurs within the*

agreed competition space, the more that clarity will emerge on the demarcations of illegitimate or legitimate cyber operations and what is outside or within the “rules” of agreed competition and, thus, may or may not lead to escalation.^[71] These cases of intensification imply that the management of dynamics (rather than spiraling) is possible.^[72]

D. Cyber-Enabled Conflict - Managing Deliberate Intensification and Escalation

While we have argued there are strong, strategic rationales for not breaching *agreed competition*, there may be certain circumstances under which actors nonetheless feel compelled to do so. But even when those circumstances may arise, the unique characteristics of cyberspace and cyber operations present opportunities for actors to mitigate the likelihood that such deliberate intensification will lead to an extended breach of *agreed competition* and a spiraling escalatory dynamic. Those same characteristics, therefore, may reinforce cautiousness when considering deliberate escalation and limitations if escalation were to occur.

To begin, let us quickly and briefly set aside the notion that escalation dominance within cyberspace is a viable strategic option at this time. It is not, because dominance is not sustainable in cyberspace, given the fluidly contested and congested nature of the domain. Importantly, there is a distinction, however, between the condition of dominance and the possibility of contested superiority that might be sustained for some period of time, leading to some strategic advantage. This position has support from both a theoretical/conceptual perspective and an operational one, with the latter stated in USCYBERCOM's *Command Vision*.^[73] If cyberspace escalation dominance (or a threat thereof) is not sustainable, what management alternatives remain? The answer lies in the unique characteristics of cyberspace and cyber operations. Note that the discussion that follows applies equally well for managing *inadvertent* as well as *accidental* intensification and escalation in cyber-enabled conflict.

To reiterate, intensifying within cyberspace is characterized by campaigns and/or operations that include increases in frequency (as a function of count over time), duration, damage, hierarchical level, and visibility of effects. If an adversary chose to erupt from *agreed competition* in cyberspace (i.e., generated effects equivalent to armed attack), and the target state chose to respond with equivalent operations in cyberspace, spiraling escalation should not be assumed. One way to limit the potential for an undesired escalatory spiral would be to ensure that unintended effects through increasing scale, scope, or intensification (collateral damage) were highly unlikely. Bellovin *et alia* argue that, contrary to conventional wisdom, such precise targeting and discrimination are possible (indeed, we have already witnessed them) and cyber operations can be designed to reduce proliferation risks.^[74]

An alternative (or complementary) targeting strategy would be to select targets whose destruction, damage, or degradation was visible to only a select audience. In contrast, an alternative design strategy could be to allow for temporary degradation or damage and effects whose frequency and duration could be continuously and actively managed. All three of these

operational options could serve to reduce the risk of further deliberate or inadvertent/accidental intensification or escalation.^[75] In certain scenarios, covert cyber operations designed to generate well-directed effects that only leadership are able to detect would send a message of resolve, but may also create an environment more conducive to deintensification and non-escalation, as leadership might be more inclined toward resolution when considerations of public awareness and any associated protestations need not figure into their deliberations.^[76] Libicki discusses this aspect of visibility by offering a distinction between making the adversary look powerless versus making the United States look powerful, where the former focuses on making a challenger aware (quietly) of its vulnerabilities, and the latter focuses on demonstrating (loudly) U.S. power.^[77]

A common, current example of cyber operations that could be designed to allow for temporary degradation or damage is cyber operations targeting electrical grids. Such operations could be designed to target industrial control systems—or, specifically, supervisory control and data acquisition systems—and to disrupt power delivery, which would, in essence, hold hostage the functions which those systems support. In such scenarios, states could negotiate demands for system functionality to be restored and permanent system damage to be avoided.^[78]

Finally, cyber operations can be designed to be continuously and actively managed, thereby allowing for a constant metering of their effects. This would allow for responsive tuning, for example, of the frequency (count over time) and the duration of effects as a function of adversary behavior. Such active command and control of cyber operations could allow for agile management of cyberspace interaction dynamics as uncertainties regarding adversary intentions, objectives, and capabilities become clearer over time.^[79]

Conceptually, intensification is a necessary but not sufficient condition for escalation out of *agreed competition*. The point of the observations above is to note that operations can go beyond increasing scale and scope and not precipitate spiraling escalation, although it should be acknowledged that in the current immature state of understanding among cyber actors about the consequences of operations, being very careful about not intensifying if one does not want to escalate is prudent. In these early stages of learning about cyber interactions, the possibility of inadvertent or accidental escalation remains more likely than if we had a longer history of cyber interactions upon which to draw.

E. Agreed Competition – Inadvertent and Accidental Intensification and Escalation

Recall that *inadvertent* escalation was described as when one party deliberately takes actions that it does not believe are escalatory but which are interpreted as escalatory by another party to the conflict. In addition, *accidental* escalation is when some operational action has direct effects that are unintended. *Inadvertent* and *accidental* can be considered as modifiers for both intensification and escalation. Regarding the former, misinterpretation may occur because of incomplete information, lack of shared reference frames, or one party's thresholds of which other parties are not aware. When considered in the context of *agreed competition*, cyber operational effects from inadvertent or accidental increases in scale or scope of effects

(e.g., NotPetya) could lead to intensification and then escalation; however, the existing political context would, in large part, determine the degree to which the operations were viewed as consequential. In a period of severe crisis between adversaries, for example, inadvertent and/or accidental effects from cyber operations could subsequently lead to deliberate intensification or escalation by the targeted state or states. In the previous section, however, several unique characteristics of cyberspace and cyber operations were highlighted which an affected state could leverage to respond in a measured manner and potentially deintensify or de-escalate the situation. So it is not contradictory to note that, while states will increasingly experiment with strategically salient cyber campaigns and operations, they will likely do so in a risk-informed manner as they have done over the past decade, in part to manage the potential for inadvertent and accidental effects, while the *agreed competition* in this space remains relatively immature. In essence, one can expect the structural incentives and strategic rationales cited previously to compete short of armed attack to affect choices in an environment of unclear operations and encourage care.^[80]

F. Stability of Agreed Competition

Just as it is critical to distinguish interaction from escalation in cyberspace, it holds logically that engagement should not be defined in and of itself as instability. Questions that require significant study beyond this article are: (1) under what conditions could *competitive interaction* involving increasing scale and scope lead to deliberate intensification and, thus, the destabilization of *agreed competition* short of armed conflict; and (2) under what conditions the use of non-cyber instruments of national power may exacerbate or moderate the intended effects of cyber operations, or vice versa.

When states seek to gain an advantage in, through, and from cyberspace, the dominant dynamic in *agreed competition* is *competitive interaction*. Within the context of long-term *agreed competition*, however, the incentive for intensification could emerge if there were present an enduring and significant imbalance of *persistent engagement* between adversaries, leading to a relative shift in power between them or a relative decline of a state across the global distribution of power. This article posits that within the strategic contest of *agreed competition*, such extended or enduring imbalances of competitive outcomes leading to relative power shifts are a necessary condition for instability. Under such a condition, the declining state might see no other option but to break out of the *agreed competition* and use armed attack–equivalent operations to reverse the situation. Thus, a sustained loss of relative power would undermine the stability of *agreed competition* short of war. The structural imperative for *persistent engagement*, therefore, produces dynamics toward an equilibrium of stability since the main objective of this strategic approach is to inhibit increases in scale, scope, and intensity, which can lead to relative power loss. Instability would be a consequence of ineffective or nonexistent, persistent engagement.^[81] Operationally, restraint is structurally encouraged only when a particular state gains sustained advantage so as not to create incentives for adversaries to challenge the integrity of the *agreed competition*.

VII. INTERACTION AND ESCALATION METAPHORS FOR CYBERSPACE

Kahn noted that metaphors can be useful, but have their limitations; he took this perspective regarding his own metaphor of a ladder. The arguments presented in this article suggest that a ladder is not well suited as a metaphor for building a model of potential cyberspace interaction dynamics and escalation. There are two reasons for this conclusion. First, it has been offered that today's strategic environment is considered to be a long-term, strategic competition in which states will pursue their national interests short of war. The *agreed competition* in cyberspace, in particular, is, similarly, characterized by operations that generate effects short of armed conflict equivalence. In this strategic space, *competitive interaction* will be the predominant cyberspace dynamic as states seek to gain advantage. This dynamic is more analogous to the grappling one sees in a wrestling match in which competitors are locked in constant contact with one another while they seek to gain the initiative in the pursuit of sustained advantage.

Second, should a state deliberately choose intensification and challenge the integrity of *agreed competition*, cyberspace dynamics are unlikely to be as straightforward as an ascending ladder. Libicki offers a modification of the ladder metaphor by arguing that escalation in cyberwar—particularly cyber against cyber—is likely to be jerky rather than smooth. What may look like a carefully calibrated ladder could, in practice, end up as a hodgepodge of sticky and bouncy rungs, where sticky rungs are those from which one cannot rise and bouncy rungs are those from which one rises much farther than anticipated.^[82] This has some salience, given the lack of states' experiences in cyber-enabled conflict and the uncertainty that is a consequence of the same. However, awareness of that uncertainty demands a consideration of how best it can be managed. It was argued in the previous sections that cyberspace and cyber operations offer opportunities for the management of intensification and escalation risks associated with those uncertainties. Operations that intensify or escalate but are designed to allow for the metering of effects and/or temporary degradation or damage, for example, take account of the uncertainty the target state may have reading another's intentions and, therefore, facilitate deintensification or de-escalation.^[83] But the notion of rungs still implies a linearity biased toward intensification that we have not witnessed to date in the competitive interaction dynamics of agreed competition.

Grappling and effects management (through persistent engagement, for example) in *agreed competition* or beyond it may lead to "movements" up, down, and sideways. This *competitive interaction* may be best visualized and conceptualized as the Penrose Stairs, represented most famously in M. C. Escher's 1960 lithograph entitled *Ascending and Descending*. Experience over time might help clarify whether one is going up, down, or sideways, but cyber interactions may not be straightforward in any of those three directions consistently. As an interactive space populated by many actors with many interests, any single cyber operation will be interaction-specific. Penrose's stairs, rather than Kahn's ladder, is the better visualization of this competitive and dynamic space.

VIII. CONCLUSION

Several years ago, U.S. adversaries waded cautiously but strategically into the strategic competitive space between war and peace, perhaps most fulsomely in cyberspace. Adversaries are now pursuing aggressive, strategic campaigns in, through, and from cyberspace to gain a strategic advantage in military, economic, and diplomatic arenas. As evidenced in recent U.S. strategic guidance, the United States has recognized that it must operate persistently in this space as well if it hopes to regain the upper hand over adversaries who have been reaping the benefits of their early, strategic adaptation to cyberspace at the expense of U.S. national interests. Over the past nine years, USCYBERCOM has been both observing adversarial behavior and *learning* from it, resulting in the identification of a new strategic approach to arresting adversary gains and securing and advancing U.S. interests in cyberspace—*persistent engagement*.

Sustained, robust competition should be expected (and is occurring) in cyberspace in an *agreed competition*, and *competitive interaction* is currently, and will continue to be, the dominant interaction dynamic. If pursued strategically, persistent engagement could lead not only to reductions in the scale, scope, and intensity of adversary cyber operations/campaigns, but it may also, over time, clarify what can be regarded as being within the rules of an increasingly stabilizing *agreed competition*.

Ultimately, tacit and formal agreements to compete robustly short of armed conflict may be the grand, strategic consequence of cyberspace. This represents a different form of national security challenge of consequence that will require not just persistent engagement, but persistent study as well. 🛡️

NOTES

1. See, for example, *Cyber Warfare in the 21st Century: Threats, Challenges, and Opportunities*. Committee on Armed Services, U.S. House of Representatives, March 1 2017, <https://www.gpo.gov/fdsys/pkg/CHRG-115hrg24680/pdf/CHRG-115hrg24680.pdf>; Lawrence J. Cavaiola, David C. Gompert, and Martin Libicki (2015) “Cyber House Rules: On War, Retaliation and Escalation,” *Survival* (2015), 57:1, 81–104; David C. Gompert and Martin Libicki, “Cyber Warfare and Sino-American Crisis Instability,” *Survival* (2014), 56:4, 7–22; Jason Healy, “Triggering the New Forever War in Cyberspace,” *The Cipher Brief* (April 1, 2018), <https://www.thecipherbrief.com/triggering-new-forever-war-cyberspace>.
2. See *National Security Strategy of the United States of America* (The White House, December 2017), p. 3 and 31, respectively; *Summary of The 2018 National Defense Strategy of The United States of America* (Department of Defense, 2018), p. 2; *Summary of the Department of Defense Cyber Strategy* (Department of Defense, 2018), p.1.
3. *National Security Strategy*, op. cit., p. 12.
4. *Command Vision for U.S. Cyber Command: Achieve and Maintain Cyberspace Superiority* (United States Cyber Command, 2018), p. 3.
5. See, *Statement of Admiral Michael S. Rogers, Commander United States Cyber Command, Before the Senate Committee on Armed Services*, (May 9, 2017). https://www.armed-services.senate.gov/imo/media/doc/Rogers_05-09-17.pdf
6. See Michael Fischerkeller and Richard Harknett, “Deterrence is Not a Credible Strategy for Cyberspace,” *Orbis* (Summer 2017), 61:3, pp. 381–393.
7. *Command Vision*, op. cit., p. 4.
8. The structure of cyberspace induces both a behavioral orientation—operational persistence—and a prescriptive necessity to manage that behavior, labeled in US documents as persistent engagement.
9. *Command Vision*, op. cit., p. 4.
10. Michael Fischerkeller, *Offense-Defense Theory, Cyberspace, and the Irrelevance of Advantage* (Institute for Defense Analyses: Alexandria, VA, 2018), p. 15, fn 58. Fischerkeller refers to low barrier to entry as an operational *incentive* for operational persistence vice a strategic imperative.
11. The *Vision* describes how they would operate—maneuvering seamlessly between defense and offense across the interconnected battlespace; where they would operate—globally, as close as possible to adversaries and their operations; when they would operate—continuously, shaping the battlespace; and why they operate—to create operational advantage for the United States while denying the same to U.S. adversaries. See, *Command Vision*, op. cit., p. 5.
12. Herbert S. Lin and Max Smeets in “What Is Absent from the U.S. Cyber Command ‘Vision,’” *Lawfare*, (May 3, 2018), <https://lawfareblog.com/what-absent-us-cyber-command-vision>.
13. The two risks highlighted are the impact of continuous engagement on high-demand low-density cyber forces and a diplomatic risk associated with claims that the United States is “militarizing” cyberspace.
14. Herman Kahn (with a new introduction by Thomas C. Schelling), *On Escalation: Metaphors and Scenarios* (Routledge: London, 2017), p. 3. While developed in response to the nuclear strategic environment, in spite of the important distinctions between it and the cyber strategic environment, the value of the framework is not diminished.
15. *Ibid*, pp. 4–6.
16. *Ibid*, p.4.
17. *Ibid*, p. 15.
18. *Ibid*, p. 7.
19. Forrest E. Morgan, Karl P. Mueller, Evan S. Madeiros, Kevin L. Pollpeter, Roger Cliff, *Dangerous Thresholds: Managing Escalation in the 21st Century* (Santa Monica, CA: RAND Corporation, 2008).
20. *Ibid*, p. 20.
21. *Ibid*, p. 23.
22. *Ibid*, p. 26.
23. *Ibid*, p. 18.
24. *Ibid*, p. 22.
25. *Ibid*, p. 24.
26. *Ibid*, p. 27.

NOTES

27. Herbert S. Lin, "Escalation Dynamics and Conflict Termination in Cyberspace," *Strategic Studies Quarterly* (Fall 2012), pp. 46–70.
28. Lin also complemented the Morgan et alia framework by including another mechanism of escalation highlighted by Kahn—*catalytic*—which occurs when some third party succeeds in provoking two parties to engage in conflict (often referred to as "false flag" operations). *Ibid.*, p. 46.
29. *Ibid.*, p. 56.
30. Martin C. Libicki, *Crisis and Escalation in Cyberspace* (Santa Monica, CA: RAND Corporation, 2012). Of note, he deviates a bit from Morgan et alia by describing horizontal escalation as the successive entry of the uninvolved into war on one or both sides. This descriptions aligns with Kahn's description of *compound* escalation.
31. *Ibid.*, p. 73.
32. This point is also made by Michael Fischerkeller, "Incorporating Offensive Cyber Operations into Conventional Deterrence Strategies," *Survival* (January 2017), 59:1, pp. 103–134.
33. Lawrence J. Cavaiola et alia, "Cyber House Rules," *op. cit.*
34. *Ibid.*, p. 84.
35. Brandon Valeriano, Benjamin Jensen, Ryan Maness, *Cyber Strategy* (UK: Oxford University Press, 2018).
36. This article argues that the distinction between interaction and escalation dynamics is critically important and not merely "distinctions without a difference." See, Herman Kahn, *On Escalation*, *op. cit.*, p. xvi.
37. Arguably, this class of strategies has been overshadowed in the last 70 years by strategies of deterrence, the class of strategies that was, and continues to be, the predominant focus of U.S. strategic thought and practice.
38. Herman Kahn, *On Escalation*, *op. cit.*, fn 4, p. 3. Kahn attributes this term to Max Singer.
39. For example, Lin, Libicki, Cavaiola et alia and many policymakers repeatedly call for the establishment of such norms in cyberspace to encourage "responsible" behavior, make appropriate a strategy of deterrence, and facilitate escalation management. Also see, *Department of Defense – Defense Science Board Task force on Cyber Deterrence* (Department of Defense: 2017).
40. See, James A. Lewis, *Rethinking Cyber Security: Strategy, Mass Effects, and States* (Center for Strategic and International Studies, January 2018), Michael Fischerkeller and Richard Harknett, "Deterrence is Not a Credible Strategy for Cyberspace", *op. cit.*
41. The strategic focus on interactions reduces the importance of attribution of source, since it biases in favor of focusing in on behaviors. While relative anonymity is exploitable in cyberspace, anchoring ones' strategy on behavior rather than source offers some re-balancing in favor of the defender.
42. Herman Kahn, *On Escalation*, *op. cit.*, xiii.
43. On this topic, also see Michael P. Fischerkeller and Richard J. Harknett, "What is Agreed Competition in Cyberspace?" *Lawfare* (19 February 2019), <https://www.lawfareblog.com/what-agreed-competition-cyberspace>.
44. See, *Command Vision*, *op. cit.*, p.6., where *persistent engagement* is described as allowing the United States to compete more effectively below the level of armed conflict.
45. See, Colin S. Gray, *Making Strategic Sense of Cyber Power: Why the Sky is Not Falling* (Strategic Studies Institute and U.S. Army War College Press: Carlisle, PA, 2013), pp. 45–48 and Michael Fischerkeller, *Offense, Defense, and the Irrelevance of Advantage*, *op. cit.*, pp. 15–16.
46. See *Summary of the Department of Defense Cyber Strategy*, *op. cit.*
47. James A. Lewis, *Rethinking Cyber Security*, *op. cit.* See, specifically, Chapter 4, "Cyber Operations and Interstate Conflict," and Chapter 5, "Political and Strategic Constraints on Cyber Attack."
48. *Ibid.*, p. 27.
49. *Ibid.*, p. 28.
50. See Chapter 4 in Brandon Valeriano and Ryan C. Maness, *Cyber War Versus Cyber Realities: Cyber Conflict in the International System* (Oxford University Press: New York, NY, 2015).
51. It is interesting to ponder why much of security studies literature on cyberwar, cyber conflict, cyber deterrence, cyber crisis, and escalation has been focused on a narrow band of important, but least likely activity, while the *agreed competition* space has emerged rather unexamined.

NOTES

52. A note of caution for U.S. and western policymakers is warranted. It would be folly to think that U.S. adversaries won't attempt to dissuade the adoption of *persistent engagement* by initially responding in ways that seek to fuel the flames of fear of escalation from *agreed competition*. With this expectation, it would behoove U.S. policymakers to keep in mind the distinction recently offered between mass effects vice strategic effects. Mass effect cyber operations are intended to be visible and disconcerting but are not of strategic consequence and so their early appearance after the adoption of a more proactive cyberspace strategy should not be unexpected. Their occurrence, therefore, should not dampen policymakers' resolve or confidence in pursuing persistent engagement in cyberspace. See, James A. Lewis, *Rethinking Cyber Security*, op. cit.
53. Thomas C. Schelling, *The Strategy of Conflict* (Harvard University Press: Cambridge, MA, 1960).
54. This mechanism terminology differs from that used in our earlier Lawfare articles on agreed competition in which we merely repurposed and redefined Kahn's widening and compounding. We feel that associating these new mechanisms with agreed competition and competitive interaction will limit potential confusion that may emerge from repurposing Kahn's terms.
55. See, *Command Vision*, op. cit., p. 6. The *Vision* also notes that in form and conduct, the competition in cyberspace is one over initiative, i.e., by sustaining initiative over time through operations that can cumulatively affect relative power, strategic advantage can be realized.
56. For a chronological list of significant events, see *Center for Strategic and International Studies' Significant Cyber Events List*. https://csis-prod.s3.amazonaws.com/s3fs-public/180308_Significant_Cyber_Events_List.pdf?Szs5ZuZShjAlfgeUXRsvB5T8C76P-JR0y.
57. The reference to "within" and "external" is intended to reinforce the notion that, through cyberspace, adversaries are able to secure their own and degrade, usurp, or circumvent others' sources of national power no matter where those sources are located. See, *2016 Report to Congress of the U.S.-China Economic and Security Review Commission* (Government Publishing Office, Washington, D.C.: November 2016), p. 299. https://www.uscc.gov/Annual_Reports/2016-annual-report-congress.
58. *Statement for the Record: Worldwide Threat Assessment of the U.S. Intelligence Community, May 11, 2017*, p. 2. <https://www.dni.gov/files/documents/Newsroom/Testimonies/SSCI%20Unclassified%20SFR%20-%20Final.pdf>.
59. China is said to have been the source of 2015 cyber operations targeting the U.S. Office of Personnel Management and the health care firms Anthem, and Premera and Carefirst Blue Cross. See, *Krebs on Security: Catching Up on the OPM Breach*, <https://krebsonsecurity.com/2015/06/catching-up-on-the-opm-breach/>, and *Mandiant Consulting: M-Trends 2016* (February 2016).
60. Garrett M. Graff, "A Guide to Russia's High Tech Toolbox for Subverting US Democracy," *Wired*, (August 13, 2017). <https://www.wired.com/story/a-guide-to-russias-high-tech-tool-box-for-subverting-us-democracy/>.
61. Sean Lyngaas, "Symantec Traces Swift Banking Hacks to North Korea," *FCW* (May 31, 2016). <https://few.com/articles/2016/05/31/swift-hack-dprk.aspx>.
62. Persistent engagement follows from the structure of cyberspace and thus we should expect actors who seek advantage will turn to this strategic orientation. While this paper focuses on U.S. strategy, the operations ascribed in open-source reporting to China, Russia, and North Korea align with the expectations of persistent engagement and can be understood as variants of this strategic approach.
63. The comprehensiveness of public records of attacks and exploitations is a function of the willingness of targets to report them. Many targets, for various reasons, do not publicly disclose them nor is there a single source detailing the same. That said, general patterns of increasing scale and scope are still evident in analyses of events that have been reported. The trends data presented in this paragraph are based on industry research reports authored by FireEye Corporation and Mandiant, a FireEye company.
64. Garrett M. Graff, "A Guide to Russia's High Tech Toolbox for Subverting US Democracy," op. cit.
65. This is a modification of Kahn's definition of escalation to include escalation from agreed competition.
66. Hierarchical levels include, for example, regular hosts, Domain Name Service servers, and gateway routers.
67. Rebecca Grant, *Victory in Cyberspace*. (Air Force Association Special Report: Washington D.C., October 2007), pp. 5–7.
68. These examples exclude interactions between states already engaged in armed conflict.
69. For a comprehensive analysis of "Stuxnet," see Kim Zetter, *Countdown to Zero Day: Stuxnet and the Launch of the World's First Digital Weapon* (New York, Crown Publishers, 2014). Note that the 2011-2013 DDoS operations against Wall Street ascribed to Iran is evidence of a desire to not engage in an escalatory spiral. The DDoS attacks did not cause physical damage as STUXNET did so they were yet another instance of a cyber interaction in *agreed competition* and not a spiral escalatory response.

NOTES

70. See Kim Zetter, “A Cyberattack Has Caused Confirmed Physical Damage for the Second Time Ever,” *Wired*, 1 August 2015. <http://www.wired.com/2015/01/german-steel-mill-hack-destruction/>.
71. There need not be any necessary symmetry to the “rules” nor does *agreed competition* require initial concurrence on what is legitimate or acceptable. There are cyber actions/operations short of war that some states may seek to legitimize/delegitimize, and differing perspectives or initial ambiguity over specific types of operations introduce a potential for intensification short of escalation. “Rules” and conventions, however, will develop over the course of interactions through interactive learning and other forms of signaling, i.e., diplomatic communications. Herman Kahn, *On Escalation*, op. cit., pp. 260–263.
72. Importantly, intensification could also be managed by leveraging non-cyber instruments of national power.
73. See Herbert S. Lin, “Escalation Dynamics and Conflict Termination in Cyberspace,” op. cit., p. 68, Michael Fischerkeller, *Offense-Defense Theory, Cyberspace, and the Irrelevance of Advantage*, op. cit., and *Command Vision*, op. cit., p. 6, where it is argued that cyber escalation dominance is not sustainable and superiority is always at risk. There are those who, nonetheless, refer to cyberspace escalation dominance as a viable strategy. See, Lawrence J. Cavaiola et alia, “Cyber House Rules,” op. cit., p. 99.
74. Steven M. Bellevin, Susan Landau, and Herbert S. Lin, “Limiting the Undesired Impact of Cyber Weapons: Technical Requirements and Policy Implications,” *Journal of Cybersecurity* (March 2017), 3:1, pp. 59–68.
75. See Michael Fischerkeller and Richard Harknett, “Deterrence is Not a Credible Strategy for Cyberspace,” op. cit., pp. 390–393.
76. Such considerations in conflict resolution or bargaining scholarship are often referred to as “two-level games.” See, for example, Robert D. Putnam, “Diplomacy and Domestic Politics: The Logic of Two-Level Games,” *International Organization* (Summer 1988), 42:3, pp. 427–60.
77. An action could also be selected that serves both objectives simultaneously. See Martin C. Libicki, *Brandishing Cyberattack Capabilities* (Santa Monica, CA: RAND National Defense Research Institute, 2013).
78. Andy Greenberg, “Hackers Gain Direct Access to US Power Grid Controls,” *Wired* (September 6, 2017); ICF International (US Dept. of Energy Report), *Electric Grid Security and Resiliency: Establishing a Baseline for Adversarial Threats* (June 2016). Note that this either/or proposition cannot be offered via kinetic solutions.
79. Note that this reference to command and control differs from that discussed by Morgan et alia and Libicki. Whereas the concern here is with command and control of a specific cyber operation to actively manage escalation dynamics, their references are to the command and control of forces, writ large, to manage against unauthorized cyber operations. Forrest E. Morgan et alia, *Dangerous Thresholds*, op. cit., p. 26 and Martin C. Libicki, *Crisis and Escalation in Cyberspace*, op. cit., pp. 114–119.
80. Libicki discusses the use of narrative, rather than signaling to manage escalatory dynamics. Such an approach would align with our notion of strategic rationales for why escalation dynamics could be muted. Martin C. Libicki, *Crisis and Escalation in Cyberspace*, op. cit., Chapter 3.
81. Relative power loss can occur outside the agreed competition of cyber operations short of armed attack and also cause states to consider intensification or escalation through cyber means as an option. One might consider the use of code against Iranian centrifuges as such an example.
82. Martin C. Libicki, *Crisis and Escalation in Cyberspace*, op. cit., p. 120.
83. While these types of operations share the same strategic objective of the massively destructive operations associated with the Russia’s strategic concept of escalating to de-escalate, they do not share the same destructive result. See, Joshua Stowell, “The Problem with Russia’s Nuclear Weapons Doctrine,” *Global Security* (February 13, 2018). <https://globalsecurityreview.com/nuclear-de-escalation-russias-deterrence-strategy/>.

THE CYBER DEFENSE REVIEW

CONTINUE THE CONVERSATION ONLINE

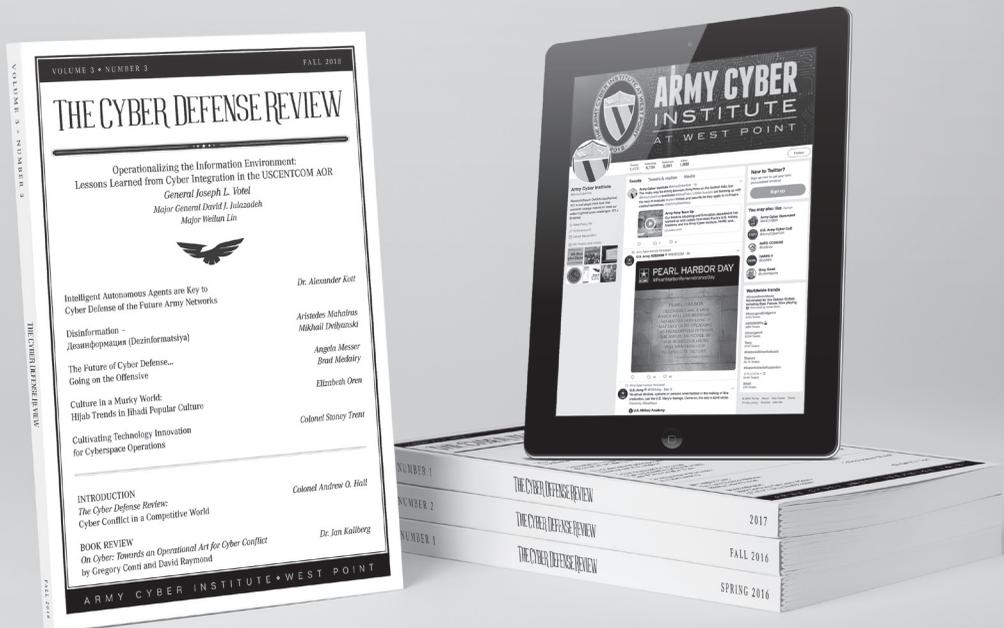
 CyberDefenseReview.Army.mil

AND THROUGH SOCIAL MEDIA

 Facebook [@ArmyCyberInstitute](https://www.facebook.com/ArmyCyberInstitute)

 LinkedIn [@LinkedInGroup](https://www.linkedin.com/company/ArmyCyberInstitute)

 Twitter [@ArmyCyberInst](https://twitter.com/ArmyCyberInst)
[@CyberDefReview](https://twitter.com/CyberDefReview)



ARMY CYBER INSTITUTE ♦ WEST POINT



THE ARMY CYBER INSTITUTE IS A NATIONAL RESOURCE FOR RESEARCH, ADVICE AND EDUCATION IN THE CYBER DOMAIN, ENGAGING ARMY, GOVERNMENT, ACADEMIC AND INDUSTRIAL CYBER COMMUNITIES TO BUILD INTELLECTUAL CAPITAL AND EXPAND THE KNOWLEDGE BASE FOR THE PURPOSE OF ENABLING EFFECTIVE ARMY CYBER DEFENSE AND CYBER OPERATIONS.