The Irregulars: Third-Party Cyber Actors and Digital Resistance Movements in the Ukraine Conflict

Conference Paper · May 2023

DOI: 10.23919/CyCon58705.2023.10182061

Margaret Smith
United States Army

Published by CCDCOE

2023

15th International Conference on Cyber Conflict: Meeting Reality

T. Jančárková, D. Giovannelli,
K. Podiņš, I. Winther (Eds.)

CYCON

IEEE
Advancing Technology
for Humanity

2023
15TH INTERNATIONAL CONFERENCE ON CYBER CONFLICT: MEETING REALITY

IEEE Catalog Number:	CFP2326N-PRT
ISBN (print):	978-9916-9789-2-4
ISBN (pdf):	978-9916-9789-3-1

COPYRIGHT AND REPRINT PERMISSIONS

[Article author(s)], [full article title]
2023 15th International Conference on Cyber Conflict:
Meeting Reality
T. Jančárková, D. Giovannelli, K. Podiņš, I. Winther (Eds.)
2023 © CCDCOE Publications

CCDCOE Publications
Filtri tee 12, 10132 Tallinn, Estonia
Phone: +372 717 6800
Fax: +372 717 6308
E-mail: publications@ccdcoe.org
Web: www.ccdcoe.org
Layout: JDF

LEGAL NOTICE: This publication contains the opinions of the respective authors only. They do not necessarily reflect the policy or the opinion of NATO CCDCOE, NATO, or any agency or any government. NATO CCDCOE may not be held responsible for any loss or harm arising from the use of information contained in this book and is not responsible for the content of the external sources, including external websites referenced in this publication.

NATO COOPERATIVE CYBER DEFENCE CENTRE OF EXCELLENCE

The NATO Cooperative Cyber Defence Centre of Excellence (CCDCOE) is a NATO-accredited knowledge hub offering a unique interdisciplinary approach to the most relevant issues in cyber defence. The heart of the CCDCOE is a diverse group of international experts from military, government, academia, and industry, currently representing 39 nations.

The CCDCOE maintains its position as an internationally recognized cyber defence hub, a premier source of subject-matter expertise, and a fundamental resource in the strategic, legal, operational, and technical aspects of cyber defence. The Centre offers thought leadership on the cutting edge of all aspects of cyber defence and provides a 360-degree view of the sector. The Centre encourages and supports the process of mainstreaming cybersecurity into NATO and national governance and capability, within its closely connected focus areas of technology, strategy, operations, and law.

The Tallinn Manual, prepared at the invitation of the CCDCOE, is the most comprehensive guide for policy advisers and legal experts on how international law applies to cyber operations carried out between and against states and non-state actors. Since 2010, the Centre has organized Locked Shields, the biggest and most complex technical live-fire cyber defence challenge in the world. Each year, Locked Shields gives cybersecurity experts the opportunity to enhance their skills in defending national IT-systems and critical infrastructure under real-time attacks. The focus is on realistic scenarios, cutting-edge technologies, and simulating the entire complexity of a massive cyber incident, including strategic decision-making and legal and communication aspects.

The CCDCOE hosts the International Conference on Cyber Conflict, CyCon, a unique annual event in Tallinn, bringing together key experts and decision makers from the global cyber defence community. The conference, which has taken place in Tallinn since 2009, attracts more than 600 participants each spring.

The CCDCOE is responsible for identifying and coordinating education and training solutions in the field of cyber defence operations for all NATO bodies across the Alliance. NATO-accredited centres of excellence are not part of the NATO Command Structure.

CYCON 2023 SPONSORS

DIAMOND SPONSORS

Microsoft

FORTINET®

paloalto® NETWORKS

TREND MICRO™

GOLD SPONSORS

CY4 GATE

ELETTRONICA GROUP

thinklogical®

A BELDEN BRAND

SILVER SPONSOR

FUJITSU

TECHNICAL SPONSOR

IEEE

Advancing Technology
for Humanity

TABLE OF CONTENTS

Introduction 1

Unpacking Cyber Neutrality 9
Scott Sullivan

The Law of Neutrality and the Sharing of Cyber-Enabled Data During International Armed Conflict 25
Yann L. Schmuki

Obligations of Non-participating States When Hackers on Their Territory Engage in Armed Conflicts 39
Marie Thøgersen

Privatized Frontlines: Private-Sector Contributions in Armed Conflict 55
Tsvetelina J. van Benthem

Business@War: The IT Companies Helping to Defend Ukraine 71
Bilyana Lilly, Kenneth Geers, Greg Rattray and Robert Koch

Evaluating Assumptions About the Role of Cyberspace in Warfighting: Evidence from Ukraine 85
Erica D. Lonergan, Margaret W. Smith and Grace B. Mueller

The Irregulars: Third-Party Cyber Actors and Digital Resistance Movements in the Ukraine Conflict 103
Margaret W. Smith and Thomas Dean

Analytical Review of the Resilience of Ukraine's Critical Energy Infrastructure to Cyber Threats in Times of War 121
Andrii Davydiuk and Vitalii Zubok

Digital Supply Chain Dependency and Resilience 141
Lars Gjesvik, Azan Latif Khanyari, Haakon Bryhni,
Alfred Arouna and Niels Nagelhus Schia

Modeling 5G Threat Scenarios for Critical Infrastructure Protection 161
Gerrit Holtrup, William Blonay, Martin Strohmeier,
Alain Mermoud, Jean-Pascal Chavanne and Vincent Lenders

Toward Mission-Critical AI: Interpretable, Actionable, and Resilient AI 181
Igor Linkov, Kelsey Stoddard, Andrew Strelzoff,
S.E. Galaitsi, Jeffrey Keisler, Benjamin D. Trump,
Alexander Kott, Pavol Bielik and Petar Tsankov

Zero-Day Operational Cyber Readiness 199
Barış Egemen Özkan and İhsan B. Tolga

AI-assisted Cyber Security Exercise Content Generation: Modeling a Cyber Conflict 217
Alexandros Zacharis, Razvan Gavrila,
Constantinos Patsakis and Demosthenes Ikonomou

Request for a Surveillance Tower: Evasive Tactics in Cyber Defense Exercises 239
Youngjae Maeng and Mauno Pihelgas

Towards Generalizing Machine Learning Models to Detect Command and Control Attack Traffic 253
Lina Gehri, Roland Meier, Daniel Hulliger
and Vincent Lenders

Human-centered Assessment of Automated Tools for Improved Cyber Situational Awareness 273
Benjamin Strickson, Cameron Worsley
and Stewart Bertram

Leveling the Playing Field: Equipping Ukrainian Freedom Fighters with Low-Cost Drone Detection Capabilities 287
Conner Bender and Jason Staggs

Russian Invasion of Ukraine 2022: Time to Reconsider Small Drones? 313
Aleksi Kajander

Weaponizing Cross-Border Data Flows: An Opportunity for NATO? 329
Matt Malone

Limits on Information Operations Under International Law 345
Talita Dias

Seeing Through the Fog: The Impact of Information Operations on War Crimes Investigations in Ukraine 365
Lindsay Freeman

From Cyber Security to Cyber Power: Appraising the Emergence of 'Responsible, Democratic Cyber Power' in UK Strategy 381
Joe Devanny and Andrew C. Dwyer

Sharpening the Spear: China's Information Warfare Lessons from Ukraine 399
Nate Beach-Westmoreland

Cyber Diplomacy: NATO/EU Engaging with the Global South 417
Eduardo Izycki, Brett van Niekerk and Trishana Ramluckan

INTRODUCTION

A lot can happen in one year. The shocking escalation of conflict in Ukraine has given rise to a daily reality of war. After decades of neutrality, Finland has joined NATO, and Sweden is on its way to doing the same. Artificial intelligence has moved from the sole purview of technologists and futurists to entertainment for the masses. NATO CCDCOE has welcomed new member states, underscoring its commitment to deliver quality training, education, and research to its constituency.

In turbulent times, it is essential to be able to pause and think. By 'Meeting Reality', which is the theme of CyCon 2023, we are invited to take stock of the many assumptions, conclusions, and forecasts made about cyberspace, technologies, and their users, both in peacetime and in times of crisis and conflict. 'Meeting Reality' is also about facing a reality we had hoped would never come again. The war in Ukraine has brought new geopolitical tensions and partnerships, tested our ideas, presumptions and established practices, and presented new challenges. It has also brought new opportunities for the application and interpretation of law, policies, and technology.

The drive to stand up to challenges, old and new, can be seen in the more than 200 submissions received in response to the CyCon 2023 call for papers. The final selection of 24 articles is hereby put to you, dear reader, representing the three traditional CyCon tracks: law, technology, and strategy/policy.

All three tracks naturally contain reflections of the events in Ukraine. The CyCon 2023 authors have contemplated topics ranging from the third-party obligations and participation of non-state actors to information operations to the implications of and for the use of drones.

The law of neutrality section opens with **Scott Sullivan** unpacking cyber neutrality in the context of the war in Ukraine. **Yann L. Schmuki** goes further and offers an analysis of the obligations and rights of neutral states applicable to the sharing of data obtained in cyberspace. **Marie Thøgersen** employs the law of neutrality perspective to examine the relationship between non-participating States and volunteer hackers based in their territory, while **Tsvetelina J. van Benthem** explores the legal implications of the involvement of the private sector, specifically digital service providers, in an armed conflict.

The extensive ICT industry engagement in Ukraine is further examined by **Bilyana Lilly**, **Kenneth Geers**, **Greg Rattray**, and **Robert Koch**, who investigate the specific products and services supplied by private companies and compile lessons learnt and recommendations for better navigation in future conflicts.

The empirical approach to the war in Ukraine continues with **Erica D. Lonergan**, **Margaret Smith**, and **Grace B. Mueller** presenting a data analysis to evaluate earlier predictions on the role cyberspace would play in the conflict. In a similar vein, a paper by **Margaret Smith** and **Dean Thomas** examines a database of content related to the IT Army of Ukraine in order to assess the latter's effectiveness as a resistance movement.

Several papers deal with the protection of critical information infrastructure – in Ukraine and beyond – and explore the prerequisites for a nation's effective resilience, including the interdependencies of supply chains and the use of new technology standards. **Andrii Davydiuk** and **Vitalii Zubok** offer an insight into the challenges faced by the Ukrainian energy sector. **Lars Gjesvik**, **Azan Latif Khanyari**, **Haakon Bryhni**, **Alfred Arouna**, and **Niels Nagelhus Schia** present case studies of six world capitals and related dependencies of infrastructural and architectural configurations, through which they demonstrate differing effects on the resilience of digital technologies at the national level. In turn, **Gerrit Holtrup**, **William Blonay**, **Martin Strohmeier**, **Alain Mermoud**, **Jean-Pascal Chavanne**, and **Vincent Lenders** analyse the technical vulnerabilities of the 5G standard and evaluate multiple threat scenarios that affect the system core and radio access.

Igor Linkov, **Kelsey Stoddard**, **Andrew Strelzoff**, **Stephanie E. Galaitsi**, **Jeffrey Keisler**, **Benjamin D. Trump**, **Alexander Kott**, **Pavol Bielik**, and **Petar Tsankov** have teamed up to offer a concept of interpretable, actionable, and resilient AI to expand the possibilities for AI use in mission-critical contexts. Military cyber operations are also at the heart of a paper by **Barış Egemen Özkan** and **İhsan B. Tolga**, which introduces a zero-day cyber readiness model.

Also looking to future conflicts, **Alexandros Zacharis**, **Razvan Gavrila**, **Constantinos Patsakis**, and **Demosthenes Ikonomou** explore the results of applying machine learning to unstructured information sources to generate structured cyber exercise content in preparation for or during a cyber conflict. **Youngjae Maeng** and **Mauno Pihelgas** examine evasive tactics in cyber defence exercises, stressing the importance of developing a robust scoring system in order to have effective exercises. **Lina Gehri**, **Roland Meier**, **Daniel Hulliger**, and **Vincent Lenders** draw on the CCDCOE's flagship exercise – Locked Shields – to analyse and propose mitigation techniques for the insufficiencies of existing machine learning models used to detect command and control attack traffic.

Benjamin Strickson, **Cameron Worsley**, and **Stewart Bertram** address the implementation challenges faced in the deployment of autonomous capabilities, including AI, through a case study of a wargaming environment, and assess the quantitative and qualitative requirements for a human-centred approach to cyber situational awareness.

Conner Bender and **Jason Staggs** take us back to Ukraine, offering a case study on drones used in the conflict and the problems caused by their easy remote identification and tracking by the adversary.

New technologies used in Ukraine, and in contemporary conflicts in general, do not escape the attention of the legal papers either. **Aleksi Kajander**, too, considers the small drones used in the conflict, this time from the perspective of the sanction mechanisms adopted by the European Union. **Matt Malone** focuses on cross-border data flows – another topic that has acquired visibility with the beginning of the war in Ukraine – to explore the security implications and opportunities 'weaponization' of these could bring for NATO countries. The final two legal papers revolve around the theme of information operations. **Talita Dias** offers a framework approach to the limits on information operations under international law, while **Lindsay Freeman** uses the reality of information operations in the Russia–Ukraine conflict as a springboard for a study of the implications of information technologies and their (mis)use for war crimes investigations.

Taking a broader strategic look, **Joe Devanny** and **Andrew C. Dwyer** examine the evolution of the United Kingdom's national cyber strategies with a particular focus on the concept of a responsible, democratic cyber power.

The final two papers of the CyCon 2023 proceedings take us beyond NATO's boundaries. **Nate Beach-Westmoreland** offers an informed assessment of the lessons to learn from the war in Ukraine by China and its information and cyber warfare strategies. **Eduardo Izycki**, **Brett van Niekerk**, and **Trishana Ramluckan** conclude the series with observations on NATO and European Union cyber diplomacy engagement with the countries of the Global South.

Despite life's turbulences, there are always certainties that we can rely on. As usual, all articles published in the proceedings have undergone a double-blind peer review. The members of the CyCon Academic Review Committee have generously taken time out of their busy schedules to help us with the final selection of papers, for which we cannot thank them enough. We have also been fortunate to be able to rely on the continued and invaluable support of the Institute of Electrical and Electronics Engineers (IEEE) and its Estonian section, without which this volume would not be possible.

Last but far from least, Liis Poolak and Jaanika Rannu have provided their usual impeccable CyCon logistics and moral support, and our colleagues (in alphabetical order) János Barbi, Henrik Paludan Beckvard, Anna Blechová, Sungbaek Cho, Sebastian Cymutta, Emre Halisdemir, Erik Ilves, Ágnes Kasper, Claire Kwan, Lauri Lindström, Bernhard zur Lippe, Liina Lumiste, Dobrin Mitev Mahlyanov, Tomomi Moriyama, Rónán O'Flaherty, Sigurður Emil Pálsson, Piret Pernik, Graham Price, Urmas Ruuto, Lisa Schauss, Lami Tagoe-Tawobola, Urmet Tomp, Grete Toompere, Ann Väljataga, and Ben Valk have kindly extended editorial assistance. Thank you.

THE EDITORS

Academic Review Committee Members for CyCon 2023:

- Liisi Adamson, NATO CCDCOE
- Maj. Geert Alberghs, Ministry of Defence, Belgium
- Maj. Vasileios Anastopoulos, NATO CCDCOE
- Lt.Col. Kraesten Arnold, Ministry of Defence, Netherlands
- Dan Black, NATO HQ, Belgium
- Henrik Paludan Beckvard, NATO CCDCOE
- Jacopo Bellasio, RAND Europe, Belgium
- Dr Bernhards Blumbergs, CERT.LV, Latvia
- Dr Russel Buchan, University of Sheffield, United Kingdom
- Prof. Thomas Chen, City, University of London, United Kingdom
- Sungbaek Cho, NATO CCDCOE
- Dr Sean Costigan, George C. Marshall Center for Security Studies, Germany
- Sebastian Cymutta, NATO CCDCOE
- Paul Darcy, University College Dublin, Ireland
- Samuele De Tomas Colatin, NATO CCDCOE
- Dr Thibault Debatty, Royal Military Academy, Belgium
- Dr Andrew Dwyer, Royal Holloway University of London, United Kingdom
- Dr Amy Ertan, NATO HQ, Belgium
- Cmdr Jacob Galbreath, NATO CCDCOE
- Dr Kenneth Geers, 2501 Research, Ukraine
- Keir Giles, Conflict Studies Research Centre, United Kingdom
- Cmdr Davide Giovannelli, NATO CCDCOE
- Shota Gvineria, Baltic Defence College, Estonia
- Maj. Emre Halisdemir, NATO CCDCOE
- Dr Jakub Harašta, Masaryk University, Czech Republic
- Jason Healey, Columbia University, United States
- Prof. David Hutchison, Lancaster University, United Kingdom

- Erik Ilves, NATO CCDCOE
- Dr Gabriel Jakobson, Altusys Corporation, United States
- Taťána Jančárková, NATO CCDCOE
- Kadri Kaska, eGovernance Academy, Estonia
- Prof. Sokratis Katsikas, Norwegian University of Science and Technology, Norway
- Dr Panagiotis Kikiras, AGT R&D GmbH, Germany
- Dr Keiko Kono, University of Copenhagen, Denmark
- Dr Csaba Krasznay, National University of Public Service, Hungary
- Lauri Kriisa, Ministry of Defence, Estonia
- Lt.Col. (ret.) Franz Lantenhammer, Germany
- Dr Lauri Lindström, NATO CCDCOE
- Dr Erica D. Lonergan, United States Military Academy at West Point, United States
- Liina Lumiste, NATO CCDCOE
- Dr Kubo Mačák, International Committee of the Red Cross, Switzerland
- Prof. Olaf Maennel, Tallinn University of Technology, Estonia
- Maj. Dobrin Mahlyanov, NATO CCDCOE
- Dr Matti Mantere, Starship Technologies, Estonia
- Dr Luigi Martino, University of Florence, Italy
- Dr Paul Maxwell, United States Military Academy at West Point, United States
- LCdr Michael McCarthy, Canadian Armed Forces, Canada
- Dr Stefano Mele, Italian Atlantic Committee, Italy
- Dr Tal Mimran, Hebrew University of Jerusalem, Israel
- Tomáš Minárik, NÚKIB, Czech Republic
- Dr Dóra Mólnar, National University of Public Service, Hungary
- Dr Jose Nazario, Fastly, United States
- Lt.Col. Gry-Mona Nordli, Norwegian Armed Forces, Norway
- Dr Sven Nõmm, Tallinn University of Technology, Estonia
- Dr Alexander Norta, Tallinn University of Technology, Estonia
- Cmdr Rónán O'Flaherty, NATO CCDCOE
- Maj. Erwin Orye, Belgian Armed Forces, Belgium
- Dr Anna-Maria Osula, Tallinn University of Technology, Estonia
- Dr Piroska Páll-Orosz, Ministry of Defence, Hungary
- Piret Pernik, NATO CCDCOE
- Capt. (N) Jean-Paul Pierini, Italian Navy, Italy
- Mauno Pihelgas, NATO CCDCOE
- Col. Dr Peter Pijpers, Ministry of Defence, Netherlands
- Dr Karl Platzer, Austrian Armed Forces, Austria

- Kārlis Podiņš, NATO CCDCOE
- Dr Radim Polčák, Masaryk University, Czech Republic
- Dr Narasimha Reddy, Texas A&M University, United States
- Lt.Col. Anastasia Roberts, NATO SHAPE, Belgium
- Dr Przemysław Roguski, Jagellonian University, Poland
- Lt.Col. Kurt Sanger, Department of Defense, United States
- Lisa Catharina Schauss, NATO CCDCOE
- Lt.Col. Massimiliano Signoretti, Italian Air Force, Italy
- Dr Max Smeets, ETH Zurich, Switzerland
- Dr Edward Sobiesk, United States Military Academy at West Point, United States
- Dr Maria Claudia Solarte-Vasquez, Tallinn University of Technology, Estonia
- Dr Tim Stevens, King's College London, United Kingdom
- Siri Strand, Norwegian Intelligence School, Norway
- Maj. Damjan Štrucl, NATO CCDCOE
- Dr Jens Tölle, Germany
- Grete Toompere, NATO CCDCOE
- Kristel Urke, University of Texas at Austin, United States
- Dr Risto Vaarandi, Tallinn University of Technology, Estonia
- Ann Väljataga, NATO CCDCOE
- Lt.Col. Berend Valk, NATO CCDCOE
- Lt. Juraj Varga, NATO CCDCOE
- Dr Adrian Venables, Tallinn University of Technology, Estonia
- Maj. Gábor Visky, Tallinn University of Technology, Estonia
- Alessandro Vitro, Council of the European Union, Belgium
- Sean Watts, United States Military Academy at West Point, United States
- Dr Laurin Weissinger, Tufts University and Yale University, United States
- Dr Christopher Whyte, Virginia Commonwealth University, United States
- Cmdr Michael Widmann, NATO MARCOM, United Kindgom
- Ingrid Winther, NATO CCDCOE
- Jan Wünsche, Swedish Armed Forces, Sweden
- Philippe Zotz, Luxembourg Armed Forces, Luxembourg

CyCon 2023 Programme Committee:

- Dr Sigurður Emil Pálsson, chair
- Dr Lauri Lindström, deputy chair
- Taťána Jančárková, chief editor of the proceedings
- Cmdr Davide Giovannelli, co-chair, law track

- Kārlis Podiņš, co-chair, technology track
- Ingrid Winther, co-chair, strategy track
- Ann Väljataga

2023 15th International Conference on Cyber Conflict
Meeting Reality
T. Jančárková, D. Giovannelli, K. Podiņš, I. Winther
2023 © NATO CCDCOE Publications, Tallinn

Unpacking Cyber Neutrality

Scott Sullivan
Harvey A. Peltier Professorship
J. Dawson Gasquet Endowed Professorship
Louisiana State University Law Center
Baton Rouge, LA, United States
ssullivan@lsu.edu

Abstract: Since the beginning of Russia's war against Ukraine, Western states have repeatedly and adamantly insisted that they would not become directly embroiled in the conflict. According to US President Joseph Biden and other leaders, the direct involvement of Western forces would inevitably result in the next world war. However, this ironclad prohibition of direct action has apparently not included cyber operations. According to the US Cyber Command, the United States has engaged in "the full spectrum" of cyber operations in support of Ukraine. At the same time, the EU has directly deployed one of its newly formed cyber rapid response teams to Ukraine to counter Russian cyber warfare.

How does the direct involvement of the US and other states in the cyber conflict fit within international legal rules regarding neutrality and co-belligerency? This article will examine what we currently know about cyber operations in the Russia–Ukraine war and filter that (admittedly limited) knowledge through competing standards of neutrality and co-belligerency. After addressing the potential implications of traditional neutrality, the article will describe how particular qualities of cyber operations pose unique challenges for the continuing viability of the legal standard of where qualified neutrality ends and co-belligerency begins.

Keywords: *neutrality, qualified neutrality, co-belligerency, Ukraine*

1. INTRODUCTION

The Western response to Russia's invasion of Ukraine has brought renewed attention to international legal rules regarding neutrality and co-belligerency.[1] Since the invasion began on February 24, 2022, the United States and other NATO countries have provided Ukraine with over $30 billion in military equipment and security assistance.[2] That assistance includes an unprecedented volume of lethal hardware.[3] Beyond hardware, numerous reports suggest, NATO states have also provided Ukraine with real-time battlefield intelligence and ongoing military training.[4]

However, this unprecedented level of support has been accompanied by two major boundaries designed to avoid a broader conflict with Russia. First, neither NATO nor its member states have directly involved themselves in the conflict.[5] To avoid "direct" conflict with Russia, the US ordered the withdrawal of all its troops in Ukraine shortly before the anticipated invasion,[6] refused requests to enforce a no-fly

[1] An international armed conflict between Russia and Ukraine was originally triggered in 2014 by Russian military operations in Crimea. *See* Michael N. Schmitt, *Ukraine Symposium – Classification of the Conflict(s)*, Articles of War (Dec. 14, 2022), https://lieber.westpoint.edu/classification-of-the-conflicts/.

[2] Jim Garamone, *U.S. Sends Ukraine $400 Million in Military Equipment*, U.S. Department of Defense (Mar. 3, 2023), https://www.defense.gov/News/News-Stories/Article/Article/3318508/us-sends-ukraine-400-million-in-military-equipment/ (outlining U.S. aid); Calin Trenkov-Wermuth & Jacob Zack, *Ukraine: The EU's Unprecedented Provision of Lethal Aid is a Good First Step*, United States Institute of Peace (Oct. 27, 2022), https://www.usip.org/publications/2022/10/ukraine-eus-unprecedented-provision-lethal-aid-good-first-step (describing EU military aid). This military aid has been augmented by tremendous macro-financial assistance intended to enable the continuation of basic governmental services and avoid a broad economic collapse of the Ukrainian economy). Decision (EU) 2022/1201 of the European Parliament and of the Council of 12 July 2022 Providing Exceptional Macro-Financial Assistance to Ukraine, 2022 O.J. (L 186) ¶ 3 (describing urgent and "sizeable risks to the macro-financial stability of [Ukraine]").

[3] *See* Michael Schwirtz, Anton Troianovski, Yousur Al-Hlou, Masha Froliak, Adam Entous & Thomas Gibbons-Neff, *How Putin's War in Ukraine Became a Catastrophe for Russia*, N.Y. Times (Dec. 16, 2022), https://www.nytimes.com/interactive/2022/12/16/world/europe/russia-putin-war-failures-ukraine.html.

[4] Julian E. Barnes, Helene Cooper & Eric Schmitt, *U.S. Intelligence Is Helping Ukraine Kill Russian Generals, Officials Say*, N.Y. Times (May 4, 2022), https://www.nytimes.com/2022/05/04/us/politics/russia-generals-killed-ukraine.html?; *EU Military Assistance Mission in support of Ukraine*, European Union (EUMAM), https://www.eeas.europa.eu/eumam-ukraine_en (training); *UK to Offer Major Training Programme for Ukrainian Forces as Prime Minister Hails Their Victorious Determination*, UK Prime Minister's Office (Jun. 17, 2022), https://www.gov.uk/government/news/uk-to-offer-major-training-programme-for-ukrainian-forces-as-prime-minister-hails-their-victorious-determination (Operation INTERFLEX - UK); *Operation UNIFIER*, Government of Canada, https://www.canada.ca/en/department-national-defence/services/operations/military-operations/current-operations/operation-unifier.html (Operation UNIFIER - Canada).

[5] To drive this home, U.S. President Joseph Biden stated that "direct conflict between NATO and Russia is World War III." Brett Samuels, *Biden: Direct Conflict Between NATO and Russia Would Be "World War III,"* The Hill (Mar. 11, 2022), https://thehill.com/policy/international/597842-biden-direct-conflict-between-nato-and-russia-would-be-world-war-iii/.

[6] Amanda Macias, *Pentagon Orders Departure of U.S. Troops in Ukraine as Russia Crisis Escalates*, CNBC (Feb. 12, 2022), https://www.cnbc.com/2022/02/12/pentagon-orders-departure-of-us-troops-in-ukraine.html.

zone,[7] and discouraged US citizens from traveling to fight alongside the Ukrainians.[8] Second, NATO-provided military equipment sent to Ukraine would fundamentally be defensive in nature.[9] Originally, this "defensive limitation" was construed to preclude the provision of weapons beyond those that might prove helpful in resisting Russian territorial advances.[10] As the conflict has progressed, the stringency of this limitation has relaxed, most notably through agreements to provide Ukraine with various combat vehicles, including more modern tanks, that are likely to be used in a Ukrainian counteroffensive.[11] However, the core of the defensive limitation—avoiding being tied to strikes inside Russian territory—persists. According to press reports, the US has "secretly modified" some of the arms it has provided to preclude long-range usage.[12] And, to date, despite concessions regarding other weapons, the US and other states have continued to refuse Ukrainian entreaties to provide weapons such as fighter jets[13] and long-range missile systems, which Ukraine might easily use beyond its territory.[14]

However, cyber operations in support of Ukraine have apparently not been as strictly contained to these boundaries. In the first weeks following Russia's invasion, reporting indicated that the US and European Union member states were directly engaged in cyber operations intended to assist Ukraine. *The New York Times* cited experts as opining that the Western-assisted pre-invasion hardening of Ukrainian

[7] Nancy Youssef, *What Is a No-Fly Zone and Why Has NATO Rejected Ukraine's Calls for One?* Wall Street Journal (Mar. 18, 2022), https://www.wsj.com/articles/what-is-no-fly-zone-ukraine-russia-nato-us-11646783483.

[8] Dan Lamothe, Alex Horton, Peter Hermann & Jonathan Baran, *Despite Risks and Official Warnings, U.S. Veterans Join Ukrainian War Effort*, Washington Post (Mar. 11, 2022), https://www.washingtonpost.com/national-security/2022/03/11/americans-veterans-ukraine-russia/.

[9] Joshua Yaffa, *Inside the U.S. Effort to Arm Ukraine*, New Yorker (Oct. 24, 2022), https://www.newyorker.com/magazine/2022/10/24/inside-the-us-effort-to-arm-ukraine.

[10] This piece understands "cyberspace operations" as "the employment of cyberspace capabilities where the primary purpose is to achieve objectives in or through cyberspace." Joint Chiefs of Staff, Joint Publication 3-12, Cyberspace Operations (2018), at vii [hereinafter Joint Publication 3-12]. This definition, while broad, avoids some of the ongoing debate surrounding the precise contours of definitions of terms such as "cyber warfare" and "cyber attacks." Tallinn Manual 2.0 on the International Law Applicable to Cyber Operations (Michael N. Schmitt ed., 2nd ed. 2017) [hereinafter Tallinn Manual 2.0], 415 (Rule 92 – Definition of cyber attack).

[11] *See, e.g.*, Jim Garamone, *U.S. $3 Billion Military Package to Ukraine Looks to Change Battlefield Dynamics*, U.S. Department of Defense (Mar. 3, 2023), https://www.defense.gov/News/News-Stories/Article/Article/3261583/us-3-billion-military-package-to-ukraine-looks-to-change-battlefield-dynamics/ (combat vehicles); David Axe, *More Ex-British Challenger 2 Tanks Are Bound for Ukraine as London Doubles Its Pledge*, Forbes (Mar. 4, 2023), https://www.forbes.com/sites/davidaxe/2023/03/04/more-ex-british-challenger-2-tanks-are-bound-for-ukraine-as-london-doubles-its-pledge/?sh=19c2277d232b.

[12] *See* Michael R. Gordon & Gordon Lubold, *U.S. Altered Himars Rocket Launchers to Keep Ukraine from Firing Missiles into Russia*, Wall Street Journal (Dec. 5, 2022), https://www.wsj.com/articles/u-s-altered-himars-rocket-launchers-to-keep-ukraine-from-firing-missiles-into-russia-11670214338.

[13] *See* Amber Phillips & Miriam Berger, *Why Washington Shut Down Poland's Offer to Give Ukraine Fighter Jets*, Washington Post (Mar. 10, 2022), https://www.washingtonpost.com/politics/2022/03/09/ukraine-poland-mig-29-fighter-jets/.

[14] The MGM-140 Army Tactical Missile System (ATACMS) is a notable example. *See* Brad Dress, *US Announces New $400 Million Ukraine Security Aid Package*, The Hill (Mar. 3, 2022) https://thehill.com/policy/defense/3882802-us-announces-new-400-million-ukraine-security-aid-package/; Shane Harris and Dan Lamothe, *Intelligence-Sharing with Ukraine Designed to Prevent Wider War*, Washington Post, (May 11, 2022), https://www.washingtonpost.com/national-security/2022/05/11/ukraine-us-intelligence-sharing-war/.

cyber defenses “cannot explain” the limited success of Russian cyber operations.[15] The report continued that US officials were “understandably tight-lipped, saying the cyber operations underway [had] been moved from an operations center in Kyiv to one outside the country” and were actively targeting “Russia’s military intelligence, to try and neutralize their activity.”[16] In a March 10, 2022, interview with the Times, Ann Neuberger, the Deputy National Security Advisor for Cyber & Emerging Technology, described Russia’s cyber operations related to Ukraine. Neuberger noted that the US was monitoring “disruptive or destructive attacks and ensuring… they can be blocked, not only in Ukraine, but blocked from spreading whether unintentionally or intentionally.”[17] According to Neuberger, a part of the US strategy was to “make it harder for attackers to conduct disruptive operations, whether that is disrupting infrastructure and more sensitive operations that I won’t get into here.”[18]

During this same period, the EU announced the deployment of a Cyber Rapid Response Team (CRRT) to Ukraine to help defend it from cyber operations that it described as “an important part of Russia’s hybrid [warfare] toolkit.”[19] Later, when asked “what direct technical assistance” was provided by the CRRT, the European Parliament responded opaquely, stating that CRRTs “are cooperating with the Ukrainian authorities to identify the support needs.”[20]

In June 2022, the head of US Cyber Command, Gen. Paul Nakasone, stated that “we’ve conducted a series of operations across the full spectrum; offensive, defensive, [and] information operations.”[21] At the Mobile World Congress in March 2023, Nathaniel Fick, the US Ambassador at Large for Cyberspace and Digital Policy, was asked whether the US was “helping Ukraine with its cybersecurity directly.”[22] Ambassador Fick’s reply noted the “close collaboration” with Ukraine and stated, “Yes, it was already being done before the war, and it continues and will continue.”[23]

The public reporting to date does not conclusively prove that Western states are providing “direct” or “offensive” cyber operations assistance to Ukraine. However,

[15] David E. Sanger, Eric Schmitt, Helene Cooper, Julian E. Barnes & Kenneth P. Vogel, *Arming Ukraine: 17,000 Anti-Tank Weapons in 6 Days and a Clandestine Cybercorps*, N.Y. Times (Mar. 6, 2022), https://www.nytimes.com/2022/03/06/us/politics/us-ukraine-weapons.html.

[16] *Id*.

[17] *Transcript, Sway: Are We Ready for Putin’s Cyber War? I asked One of Biden’s Top Cybersecurity Officials*, N.Y. Times (Mar. 10, 2022), https://www.nytimes.com/2022/03/10/opinion/sway-kara-swisher-anne-neuberger.html?showTranscript=1.

[18] *Id*.

[19] Joe Tidy, *Ukraine Deploys Cyber Rapid Response Team*, BBC News (Feb. 22, 2022), https://www.bbc.com/news/technology-60484979.

[20] Reply, Parliamentary question - E-000267/2022(ASW) (Apr. 4, 2022), https://www.europarl.europa.eu/doceo/document/E-9-2022-000267-ASW_EN.html.

[21] *See* Alexander Martin, *US Military Hackers Conducting Offensive Operations in Support of Ukraine, Says Head of Cyber Command*, Sky News (Jun. 1, 2022), https://news.sky.com/story/us-military-hackers-conducting-offensive-operations-in-support-of-ukraine-says-head-of-cyber-command-12625139.

[22] *Nathaniel Fick, U.S. Digital Policymaker: “The War in Ukraine Has Put Cybersecurity at the Forefront,”* Noticias Financieras, 2023 WLNR 7605227 (Mar. 1, 2023).

[23] *Id*.

the above affirmative indications and the refusal by Western government officials to clearly and definitively deny engaging in direct cyber operations stand in marked contrast to the clarity and line-drawing offered regarding non-cyber assistance to Ukraine.[24]

This article seeks to unpack two international legal frameworks central to assessing both cyber and "real-world" third-party state assistance to Ukraine. The first, international neutrality law, establishes the rights and responsibilities of neutral states relative to belligerents. The second, co-belligerency, encapsulates legal standards surrounding when a third-party state may be considered to have become a party to an ongoing international armed conflict (IAC). After considering each, this piece theorizes how specific attributes surrounding cyber operations might explain, even if not necessarily legally justify, the suspected differing treatment described above.

2. "NEUTRAL" STATE CYBER OPERATIONS IN THE UKRAINE–RUSSIA CONFLICT

As noted above, the precise nature and scope of cyber operations undertaken by Western states are unclear. However, even without knowing about exact operations, there is considerable reason to believe that Western states have been more open to proactively and directly using cyber means to "disrupt" and "disable" Russian attacks—cyber and potentially otherwise—in a manner that they have declined to do using more traditional military means.

General Nakasone did not specify the "offensive" or "defensive" cyber operations that the US had undertaken in support of Ukraine. However, governing cyber operations doctrine and the "persistent engagement" strategy embraced by Cyber Command suggest that such operations likely involved conduct or effects in foreign cyberspace.

The US cybersecurity strategy of "persistent engagement" suggests that, to the extent that US cyber forces are engaged in any direct activity, some of their actions are conducted within belligerent territory.[25] As Gen. Nakasone explained, persistent engagement requires that cyber forces be not merely a "response force" but also a

[24] Russian accusations that the U.S. had attacked "state institutions, the media, critical infrastructure facilities, and life support systems" via cyber operations drew a narrow denial by an NSA spokesperson that the "United States Government has not engaged in the activity described by Russia." Brad Dress, *Russia Accuses US of Leading Massive Cyber Campaign*, The Hill (Mar. 29, 2022), https://thehill.com/policy/cybersecurity/600140-russia-accuses-us-of-leading-massive-cyber-campaign/. The White House was similarly ambiguous in responding to Gen. Nakasone's comments. *Press Briefing by Press Secretary Karine Jean-Pierre*, White House (Jun. 1, 2022), https://www.whitehouse.gov/briefing-room/press-briefings/2022/06/01/press-briefing-by-press-secretary-karine-jean-pierre-june-1-2022/.

[25] *CYBER 101 – Defend Forward and Persistent Engagement*, U.S. Cyber Command (Oct. 25, 2022), https://www.cybercom.mil/Media/News/Article/3198878/cyber-101-defend-forward-and-persistent-engagement/.

"persistence force" in which cyber forces can "contest adversaries globally" and "operate against our enemies on their virtual territory."[26]

US doctrine broadly identifies different types of cyber operations undertaken by the US military.[27] These include offensive cyber operations (OCO) and defensive cyber operations (DCO). Geography and intent separate the two. Cyber operations "conducted outside of blue cyberspace [cyberspace 'protected by the US']"[28] with a commander's intent other than to defend blue cyberspace from an ongoing or imminent cyberspace threat are OCO missions."[29] However, "defensive" cyber operations assisting in the protection of Ukrainian cyberspace "from active threats" would likely, as a practical matter, also involve actions undertaken inside the territory of a belligerent (Ukraine), and to the advantage of that belligerent.

OCO and DCO both contemplate a broad range of effects likely to be felt in foreign jurisdictions that US doctrine acknowledges ultimately may "rise to the level of use of force."[30] At one end of the spectrum, exploitation actions, such as information collection, are conducted without any associated physical or cyberspace effects.[31] In contrast, attack actions create either a noticeable effect in cyberspace or lead to effects in physical domains.[32] The degree of these effects varies from minor (temporary loss of access to the system) to significant enough to constitute an international use of force.[33]

The US does not appear to be alone in involving state cyber operatives in the conflict. As noted earlier, the first EU CRRT formally requested by Ukraine was deployed in the early days of the conflict to "help Ukraine to face cyberattacks."[34] The UK's "Ukraine Cyber Programme," the existence of which had been secret to "protect its operational security," had been active since the early days of the conflict in "preventing Russian malign actors from accessing vital networks."[35] China has also been linked to

[26] Paul M. Nakasone, *A Cyber Force for Persistent Operations*, Joint Force Quarterly, JFQ 92 (1Q, 2019), at 12.

[27] Joint Publication 3-12, *supra* note 10. While specifically discussing US operations, the contours described are shared by other NATO states. *See, e.g.*, UK Ministry of Defence, Cyber Primer, Third Ed., (Oct. 2022).

[28] Joint Publication 3-12, *supra* note 10, at I-4.

[29] *Id*. at II-5.

[30] *Id*. (OCO) and at II-4 (DCO).

[31] *Id*. at II-6.

[32] *Id*. at II-7.

[33] *Id*. at II-4 (DCO-RA actions may "rise to the level of the use of force"); *id*. at II-5 (same with OCO).

[34] *See* Laurens Cerulus, *EU to Mobilize Cyber Team to Help Ukraine Fight Russian Cyberattacks*, Politico (Feb. 21, 2022), https://www.politico.eu/article/ukraine-russia-eu-cyber-attack-security-help/; *Cyber Rapid Response Teams and Mutual Assistance in Cyber Security*, Permanent Structured Cooperation (Pesco), https://www.pesco.europa.eu/project/cyber-rapid-response-teams-and-mutual-assistance-in-cyber-security/#:~:text=Cyber%20Rapid%20Response%20Teams%20(CRRTs,operations%20as%20well%20as-%20partners.)

[35] *Press Release: UK Boosts Ukraine's Cyber Defences with £6 Million Support Package*, UK Foreign, Commonwealth, and Development Office (Nov. 1, 2022), https://www.gov.uk/government/news/uk-boosts-ukraines-cyber-defences-with-6-million-support-package.

cyber operations targeting the Ukrainian government and potentially seeking to gain relevant intelligence.[36]

Further complicating the cyberspace operations picture is the widespread involvement of private actors. Western technology companies have played an instrumental role in reinforcing and expanding Ukraine's cyber capabilities. Microsoft, which has played a crucial role in identifying and disabling Russian malware threats against Ukraine, has articulated Ukraine's cyber defense as one reliant "on a coalition of countries, companies, and NGOs."[37] While much of the corporate support of Ukraine appears to reflect independent judgment, some can be tied to direct government encouragement. For example, the British government has implemented a program that facilitates Ukrainian access to commercial cybersecurity support funded by the UK, resulting in "commercial cybersecurity capabilities [controlled by Ukraine] for immediate operational effect."[38]

3. NEUTRALITY AND CO-BELLIGERENCY

The law of neutrality describes the rights and duties of belligerent and neutral states during IACs.[39] The fundamental purpose of neutrality rules is two-fold: inhibiting the expansion of hostilities and reducing the impact of armed conflict on nonbelligerent states and populations.[40] The traditional "neutral" state is essentially a bystander, potentially interested in the outcome of the conflict but unwilling to become embroiled in it.

Co-belligerency, in contrast, seeks to identify parties that, though perhaps not initially participants in the conflict, have come off the sidelines and entered the fray.[41] Between traditional neutrality and co-belligerency lies "qualified" or "benevolent" neutrality—in which neutral states are empowered to provide assistance to states that are victims of aggression.

[36] Gordon Corera, *Mystery of Alleged Chinese Hack on Eve of Ukraine Invasion*, BBC News (Apr. 7, 2022), https://www.bbc.com/news/technology-60983346.

[37] Brad Smith, *Defending Ukraine: Early Lessons from the Cyber War*, Microsoft on the Issues (Jun. 22, 2022), https://blogs.microsoft.com/on-the-issues/2022/06/22/defending-ukraine-early-lessons-from-the-cyber-war/.

[38] Nick Beecroft, *Evaluating the International Support for Ukrainian Cyber Defense*, Carnegie Endowment for International Peace (Nov. 3, 2022), https://carnegieendowment.org/2022/11/03/evaluating-international-support-to-ukrainian-cyber-defense-pub-88322.

[39] The applicability of neutrality rules to (NIACs) is contested. *See* Tess Bridgeman, *The Law of Neutrality and the Conflict with Al Qaeda*, 85 N.Y.U. L. Rev. 1186, 1211–12 (2010) (describing the difficulty of gauging "the extent to which the law of neutrality applies to NIACs"); *compare* Karl S. Chang, *Enemy Status and Military Detention in the War Against Al-Qaeda*, 47 Tex. Int'l L.J. 1, 33 (2012).

[40] *See* Stephen Neff, The Rights and Obligations of Neutrals 8 (2000).

[41] *See* Rebecca Ingber, *Co-Belligerency*, 42 Yale J. Int'l L. 67, 93 (2017).

A. Traditional and Qualified Neutrality

The traditional principles of neutrality are fundamentally clear. The 1907 Hague Conventions V (addressing neutrality in land war) and XIII (addressing neutrality in naval war) form the cornerstone of the traditional rights and obligations of neutral states. While neutral states possess rights designed to avoid adverse effects emanating from the conflict, neutrality law also imposes corollary duties of non-participation and impartiality.[42] The duty of impartiality precludes neutrals from engaging in acts that benefit one party of the conflict to the detriment of the other, including the provision of "war material of any kind," such as weapons.[43]

Western assistance to Ukraine has generally not been justified under the traditional neutrality rules, but rather under "qualified" or "benevolent" neutrality.[44] Qualified neutrality asserts that subsequent treaty and customary international law has transformed neutrality rules in a manner enabling states to retain neutral status while assisting a belligerent that was a victim of aggression.[45] At the core of qualified neutrality is the fundamental circumscription of the lawful use of force articulated in the Kellogg-Briand Pact of 1928 and the UN Charter.[46] These instruments, at least theoretically, effectively placed the lawful use of force under the exclusive authority of the UN Security Council.[47] As a result, third-party states that would have borne an obligation of impartiality under the traditional rules of neutrality would now find that same impartiality impermissible.[48]

The validity of qualified neutrality in circumstances where the UN Security Council has affirmatively acted on the question is widely accepted.[49] Its viability beyond

[42] *See* Neff, *supra* note 40, at 485.

[43] *Id*. at 496. Hague Conv. XIII prohibits the provision "in any manner, directly or indirectly, by a neutral Power to a belligerent power, of war-ships, ammunition, or war material of any kind whatever." 1907 Hague Convention XIII Concerning the Rights and Duties of Neutral Powers in Naval War, art. 6, *reprinted in* Adam Roberts & Richard Guelff, Documents on the Laws of War 61, 109 (Adam Roberts & Richard Guelff eds., 2nd ed. 1989). This prohibition has been broadly construed. *See* Manuel Rodriguez, *Operation Rubicon: An Assessment with Regard to Switzerland's Duties Under the Law of Neutrality*, 50 Int'l J. Legal Info. 82, 97 (2022).

[44] *See, e.g*., Oona Hathaway & Scott Shapiro, *Supplying Arms to Ukraine is Not an Act of War*, Just Security (Mar. 12, 2022), https://www.justsecurity.org/80661/supplying-arms-to-ukraine-is-not-an-act-of-war/.

[45] *See, e.g*., Michael N. Schmitt, *Providing Arms and Materiel to Ukraine: Neutrality, Co-Belligerency, and the Use of Force*, Articles of War, (Mar. 7, 2022), https://lieber.westpoint.edu/ukraine-neutrality-co-belligerency-use-of-force/.

[46] *See* Tallinn Manual 2.0, supra note 10, at 562.

[47] *See* Hathaway & Shapiro, *supra* note 44; Patrick M. Norton, *Between the Ideology and the Reality: The Shadow of the Law of Neutrality*, 17 Harv. Int'l L.J. 249, 251 (1976). Lawful self-defense was considered "an interim measure until the collective security mechanism of the United Nations could be organized to meet the armed aggression." *Id*.

[48] *Id*. at 251. Some have argued that neutrality rules are "obsolete as both strategic and humanitarian third-party interventions have become the norm." *See* Jide Nzelibe, *Courting Genocide: The Unintended Effects of Humanitarian Intervention*, 97 Cal. L. Rev. 1171, 1213 (2009).

[49] *See* Tallinn Manual 2.0, *supra* note 10, at 562. The Security Council might act in a variety of ways under Chapter VII of the Charter that would preclude reliance on traditional neutrality. *See, e.g*., UN Charter art 39 (authorizing Council to determine an "act of aggression").

such circumstances, however, is contested.[50] Absent UN Security Council action, a state's characterization of a belligerent as a "victim" or "aggressor" may more often reflect perceived self-interest rather than an inescapable factual conclusion.[51] This characterization problem has led states and scholars alike to voice skepticism, if not outright reject qualified neutrality in circumstances where the UN Security Council has not acted.[52]

Similarly, the scope of assistance that a state claiming qualified neutrality is empowered to offer a victim-belligerent is ambiguous. Robert Jackson, in articulating the US qualified neutrality in the early part of World War II viewed all measures of assistance "short of war" as appropriate to provide a victim of aggression.[53] The "short of war" standard itself possesses ambiguity. In operation, the "short of war" standard was limited to the provision of weapons despite increasing attacks on the US in 1941.[54] Such restraint reflects the view that US assistance justified by qualified neutrality was definitively bounded by "entry into the war a belligerent," a result that was not exclusively within American control.

Regardless of approach, violations of a state's neutral obligations do not alone extinguish neutral status.[55] Instead, violations give rise to enforcement rights of belligerents that typically do not include the use of force.[56] However, as both a doctrinal and practical matter, the further states deviate quantitatively and qualitatively from traditional obligations of non-participation and impartiality, the more the legal assessment of co-belligerency comes to the fore. Fundamentally, when a neutral state's violations of impartiality and abstention are of "such gravity as to justify the conclusion that the neutral state has become a party to the conflict," it is no longer a neutral state but a co-belligerent to the conflict.[57]

[50] *See* Wolff Heintschel von Heinegg, *"Benevolent" Third States in International Armed Conflicts: The Myth of the Irrelevance of the Law of Neutrality, in* International Law and Armed Conflict: Exploring the Fault Lines 543, 548–49 (Michael N. Schmitt & Jelena Pejic eds. 2007) (describing the rise of qualified neutrality); Alonso E. Illueca, *International Coalitions and the Non-Military Contributing Member States*, 49 Univ. of Miami Inter-Am. L.R. 1, 29 (2017) (arguing qualified neutrality "lacks factual basis").

[51] *See, e.g.*, Edwin Borchard, *The Attorney General's Opinion on the Exchange of Destroyers for Naval Bases*, 34 Am. J. Int'l L. 690, 696 (1940).

[52] Dr. von Heinegg, long a skeptic of qualified neutrality, has called Russia's invasion of Ukraine a "game changer" and claimed that now "there are good reasons to take a more nuanced position vis-à-vis 'qualified neutrality.'" *See* Wolff Heintschel von Heinegg, *Neutrality in the War Against Ukraine*, Articles of War (Mar. 1, 2022), https://lieber.westpoint.edu/neutrality-in-the-war-against-ukraine/.

[53] Such measures included "discriminatory embargoes or boycotts, as well as financial credits and furnishing of supplies and material, weapons and ships." *Id*. at 279.

[54] *See* Jürgen Rohwer, Axis Submarine Successes of World War Two: German, Italian, and Japanese Submarine Successes, 1939–1945, at 53–74 (1999).

[55] *See, e.g.*, Lassa F. L. Oppenheim, International Law: Disputes, War and Neutrality, §358, at 752 (Hersch Lauterpacht ed., 7th ed. 1952) ("Mere violation of neutrality must not be confused with the ending of neutrality.").

[56] *See, e.g.*, Michael Bothe, *The Law of Neutrality*, in The Handbook of Humanitarian Law in Armed Conflicts 485, 494 (Dieter Fleck ed. 1995). ("[I]t is not necessarily legal to attack a state violating the law of neutrality and to make it, by that attack, a party to the conflict."); *see also*, Kevin Jon Heller & Lena Trabucco, *The Legality of Weapons Transfers to Ukraine Under International Law*, Brill, Aug. 29, 2022.

[57] William H. Boothby & Wolff Heintschel Von Heinegg, The Law of War: A Detailed Assessment of the US Department of Defense Law of War Manual 377 (2018).

B. Co-belligerency

At the highest level of abstraction, co-belligerency "entails a sovereign State becoming a party to a conflict, either through formal or informal processes."[58] More practically, co-belligerency reflects the determination that a state's acts render it an "ally" to a party to the conflict, with legal and practical consequences for that state and its nationals.[59]

The test for co-belligerency, while unsettled, has become a crucial topic when considering limitations on the assistance that can be provided to Ukraine. According to press reports, the policies of Western states have largely been shaped by a legal assessment regarding how far the US can go before it becomes a co-belligerent in the conflict.[60] The European Parliament has stated that "providing military equipment and platforms" does not make the EU or member states co-belligerents.[61]

Few bright lines exist, but a neutral state most clearly becomes a co-belligerent when it "participates to a significant extent in hostilities," such as through the deployment of troops to the conflict.[62] Beyond direct deployments, however, the level of intervention required to trigger co-belligerency is legally unsettled. Co-belligerency is often said to attach when "direct support" is provided to a belligerent's military operations.[63] In an Office of Legal Counsel memorandum authored by Jack Goldsmith, co-belligerency status "turns on whether [the state's] participation" possesses a "direct nexus" to a belligerent's military objectives.[64] The International Criminal Tribunal for the former Yugoslavia's decision in *Blakić* suggested that determining co-belligerency turned on whether states "were allies and acted as such in conducting operations" in the conflict.[65]

Collectively, co-belligerency can generally be said to flow from a state's (1) direct participation (2) in hostile actions (3) intended to facilitate another belligerent's

[58] Christof Heyns (Special Rapporteur on Extrajudicial, Summary, or Arbitrary Executions), *Report of the Special Rapporteur on Extrajudicial, Summary or Arbitrary Executions*, UN Doc. A/68/382 (2013) at ¶ 60.

[59] The array of consequences is beyond the scope of this piece. Some have suggested that co-belligerency status renders the state's armed forces and military objects subject to attack "anytime, anywhere, and with any amount of force." Heller & Trabucco, *supra* note 56, at 263. *See also* Alexander Wentker, *At War: When Do States Supporting Ukraine or Russia Become Parties to the Conflict and What Would That Mean?* EJIL: Talk!, (Mar. 14, 2022) (impact on status determinations).

[60] *See* Ken Dilanian, Carol E. Lee, Courtney Kube & Dan De Luce, *Biden Admin Carefully Examining Legal Issues Around Providing Arms to Ukraine*, NBC News (Feb. 25, 2022), https://www.nbcnews.com/politics/national-security/biden-admin-carefully-examining-legal-issues-providing-arms-ukraine-rcna17758.

[61] Reply, Parliamentary question - E-001263/2022(ASW) (Sep. 23, 2022), https://www.europarl.europa.eu/doceo/document/E-9-2022-001263-ASW_EN.html.

[62] *Id*.

[63] *See* Christopher Greenwood, *Scope of Application of Humanitarian Law, in* The Handbook of Humanitarian Law in Armed Conflicts 39, 50 (Dieter Fleck ed. 1995).

[64] Thus, providing "general security" qualifies but "humanitarian support" does not. Office of Legal Counsel Memorandum Opinion for the Counsel to the President, "Protected Person" Status in Occupied Iraq Under the Fourth Geneva Convention, 45 (Mar. 18, 2004).

[65] Prosecutor v. Blakić, Case No. IT-95-14-T, Judgment, ¶ 137 (Int'l Crim. Trib. for the Former Yugoslavia, Mar. 3, 2000).

operational success. Each of these components has proven difficult to define with precision under conventional armed conflict and poses especially acute challenges in cyberspace.

4. NON-NEUTRALITY AND CO-BELLIGERENCY IN CYBER OPERATIONS

Government lawyers seeking to discern doctrinal lines regarding the kind of "support" a state could provide Ukraine appear to have differentiated acceptable activities in cyberspace. That divergence begs the question of whether the more permissive approach to cyber operations comports with our legal understanding of neutrality and co-belligerency and, if not, why the two genres of support are being approached differently.

A. Assessing Cyber Operations in Ukraine Under Traditional Neutrality

There can be little doubt that the extensive provision of weaponry to Ukraine by the United States and other NATO states violates the traditional neutral state obligations of non-participation and impartiality.[66] However, the neutrality analysis grows more complicated when considering the various strains of cyber operations support that Western states have allowed or directly provided.[67]

Under traditional neutrality rules, acts that would be prohibited if undertaken by the state are of no legal consequence if undertaken by private parties. For example, while a neutral state is barred from participating in acts of war by a party, that prohibition does not extend to precluding its citizens from entering into the conflict on behalf of one of the belligerents.[68] Similarly, states are prohibited from supplying war materials to belligerents, but only to the extent that the export of those arms is "controlled" by the state itself.[69]

Applied to the cyber warfare between Ukraine and Russia, the efforts undertaken by private corporations to thwart Russian cyber attacks, and even more assertive acts that facilitate counterattacks on Russian networks, would not offend traditional neutrality rules.[70] Likewise, Western corporations who have provided Ukraine with cybersecurity services, software, and equipment worth tens of millions of dollars similarly comport with the law of neutrality despite their potentially important role in defending Ukraine's military and civilian networks and infrastructure.

[66] *See, e.g.*, Bothe, *supra* note 56, at 485, 496.

[67] Distinct from neutrality rules, direct involvement in cyber operations may also constitute direct participation in hostilities.

[68] Bothe, *supra* note 56, at 498.

[69] *Id*. at 496–97.

[70] At least insofar as the tools provided remain beyond the scope of governmental regulation that would lend itself to a finding of "controlling" export.

In contrast, once the existence of the armed conflict is established, nearly all direct state cyber operations in support of Ukraine would violate the traditional neutral state obligations of non-participation and impartiality.[71] This would unquestionably include DCO or OCO as set out in *Joint Publication 3-12* and the provision of cyber "weapons," whether defensively or offensively oriented. Importantly, even the provision of intelligence regarding Ukrainian cyber vulnerabilities or impending Russian cyber attacks would likely be considered prohibited under traditional neutrality rules.[72]

B. Assessing Cyber Operations in Ukraine under Qualified Neutrality and Co-belligerency

While direct support of a state would violate traditional neutrality, that same support is consistent with a neutral state's obligations under qualified neutrality so long as the support is offered to a victim of aggression. Thus, the end of qualified neutrality and the beginning of co-belligerency converge on the same question: whether the state's acts transcend "support" to become an "act of war," thus shedding the rights of neutrality in favor of participation.

The range of acts that might constitute an "act of war" and thus create co-belligerency is "remarkably undertheorized."[73] At the heart of "acts of war" giving rise to co-belligerency are "hostile" acts designed to damage one belligerent's military capacity to the advantage of the other.[74] This formulation recognizes that once the existence of an IAC is established as a factual matter, a third-party state's acts do not need to independently establish an armed conflict to be considered a co-belligerent.[75] As such, acts that would not constitute an act of war outside of the armed conflict may reach the co-belligerency threshold once the armed conflict has begun.[76]

Recent decades have seen significant analytical efforts undertaken in identifying what constitutes direct participation in hostilities in the non-international armed conflict context, which are instructive here. The definition of "hostilities" relating to cyberspace is broader than that of an "attack" and includes acts intended to have

[71] Some have argued that US cyber operations do not violate traditional neutrality because they have likely been undertaken at Ukraine's request and thus can be considered collective self-defense lawful under Article 51 of the UN Charter. *See* Michael Schmitt, *Ukraine Symposium – U.S. Offensive Cyber Operations in Support of Ukraine*, Articles of War (Jun. 6, 2022), https://lieber.westpoint.edu/us-offensive-cyber-operations-support-ukraine/. The US could justify its cyber operations under this rubric, but contrary to past state practice, it has not made such claims to the UN Security Council as required under Article 51. This omission is presumably to avoid the co-belligerent status that such a reporting would spark.

[72] *See* Erik Castren, The Present Law of War and Neutrality 479–80 (1954) (neutral states prohibited from providing operational intelligence).

[73] Rebecca Ingber, *Untangling Belligerency from Neutrality in the Conflict with Al-Qaeda*, 47 Tex. Int'l L.J. 75, 90 (2011).

[74] *See* Tallinn Manual 2.0, *supra* note 10, at 429.

[75] *See* Alexander Wentker, *At War? Party Status and the War in Ukraine*, MPIL Research Paper Series 9 (Dec. 15, 2022).

[76] An example of this potentially includes pre-invasion US "hunt-forward" operations in which US forces coordinated closely with Ukraine to identify and neutralize Russian threats existing in Ukrainian networks.

the effect of "negatively affecting the adversary's military operations or capabilities" or causing physical harm or damage.[77] Hostilities include preparatory acts, "directly supporting" specific operations, and "identifying vulnerabilities in a targeted system."[78] Importantly, while engaging in a cyber attack constitutes participation in hostilities, simply providing a cyber weapon to a belligerent does not.[79]

Finally, the direct hostile acts must be accompanied by the requisite intent. Acts that would unmistakably constitute participation in hostilities, whether through conventional or cyber means, do not give rise to co-belligerency if they lack the relevant intent. The intent requirement is multi-fold. A state's hostile acts have to intend to diminish the military capacity of one belligerent in the conflict, with the goal of assisting another party to the conflict in the IAC itself (often referred to as "belligerent nexus"). Neither component is sufficient alone.

Within the examination of a belligerent nexus, certain peculiarities of cyber operations complicate a clean fulfillment of the co-belligerency standard.

1) The Problem of Persistent Conflict

It is difficult to ascribe an intent to influence a specific armed conflict to cyber operations that generally appear consistent with peacetime norms. And, unfortunately, cyber operations seeking to gather intelligence and identify and exploit vulnerabilities are a regular feature of the current global landscape.

Cybersecurity threats are targeting governments, businesses, and NGOs worldwide, at every level.[80] As these threats have proliferated, the powers granted to civilian and military authorities to counteract the threat have become increasingly expansive and institutionalized.[81] In turn, state-sponsored operations around the world seeking economic and military advantage have become pervasive. The result is an "actual and continuous, strategic competition in cyberspace that does not reach the level of armed conflict."[82]

That competition takes place through DCO and OCO described in *Joint Publication 3-12*. It includes "positioning of forces" and "gaining access to adversary, enemy, or intermediary links… to support future actions."[83] Doctrines such as "persistent

[77] Tallinn Manual 2.0, *supra* note 10, at 429.

[78] *Id*. at 431 (preparatory acts); *id*. at 430 (network vulnerabilities).

[79] At least without more particularized facts evidencing that the provision of the cyber weapon in question constitutes support of a specific operation. *See* Tallinn Manual 2.0, *supra* note 10, at 430.

[80] *See* Vasu Jakkal, *How Nation-State Attackers Like NOBELIUM Are Changing Cybersecurity*, Microsoft Security (Sep. 28, 2021), https://www.microsoft.com/en-us/security/blog/2021/09/28/how-nation-state-attackers-like-nobelium-are-changing-cybersecurity/.

[81] *See generally*, Myriam Dunn Cavelty, *The Militarisation of Cyberspace: Why Less May Be Better, in* 4th International Conference on Cyber Conflict (2012).

[82] Michael P. Fischerkeller & Richard J. Harknett, *Persistent Engagement, Agreed Competition, and Cyberspace Interaction Dynamics and Escalation*, 2019 Cyber Def. Rev. 267, 276 (2019).

[83] Joint Publication 3-12, *supra* note 10, at II-8.

engagement" seek to shift from a responsive posture to a more proactive one by encouraging "defending forward" operationally to force potential adversaries to expend resources defending national interests and away from attacking others.

Many of these activities would appear to qualify as "hostile" acts; however, given the pervasive nature of such cyber operations as a general matter, their connection to a specific armed conflict objective is less clear. Absent a sharp deviation in methodology or consequences, how can state cyber operations differentiate themselves from the ordinary course of business? In short, the persistency of "hostile" acts as a general matter makes ascribing the requisite intent more difficult during an armed conflict.

Chinese hacking at the outset of the Russian invasion demonstrates the quandary. According to news reports citing "Western intelligence officials," China allegedly engaged in hundreds of "cyber attacks" targeting Ukrainian government institutions on the eve of the Russian invasion.[84] Initial assessments of the activity indicated that China was engaged in cyber espionage that might assist its Russian partner. Curiously, however, the same Chinese cyber actors launched highly similar attacks against "government and military networks" in Russia and Belarus.[85] British analysts later described the Chinese cyber operations as "relatively routine" rather than demonstrative of collusion. In other words, given the high baseline level of similar cyber operations under normal conditions, it is impossible to conclude that the Chinese operations that had been identified were connected to the IAC between Russia and Ukraine.[86]

2) Attribution and Secrecy

Hostile acts can only lead to co-belligerency to the extent that they are identified and attributed to a third-party state. State preferences for using cyber operations over other alternatives often reflect the belief that such operations "offer low probability of detection."[87] The desire to avoid detection is likely especially acute when direct accountability might lend itself to a finding of co-belligerency. Western efforts in supporting Ukraine are instructive. The US and other NATO states have not hidden, but rather generally celebrated, the amount of lethal hardware provided to Ukrainian forces. However, in providing support, Western states have gone to great lengths to avoid a physical presence in the area. By contrast, the successes of the US Cyber Command have been articulated generally and without fanfare.

[84] Gordon Corera, *Mystery of Alleged Chinese Hack on Eve of Ukraine Invasion*, BBC News (Apr. 7, 2022), https://www.bbc.com/news/technology-60983346.

[85] *Id*.

[86] Further complicating the matter is that the underlying baseline and spikes in cyber operations activity themselves may reflect differences in monitoring rather than differences in reality. Prior to armed conflict, foreign state cyber operations may be under-surveilled and thus underestimated. Once an armed conflict has been initiated, identified cyber operations might be susceptible to misattribution in the opposite direction.

[87] Joint Publication 3-12, *supra* note 10, at IV-8.

The problem of attributing cyber operations to a state is magnified by Ukraine and Russia's encouragement of non-state actors to participate in the cyber conflict.[88] The addition of such actors offers an additional source for attribution and, if definitively tied to the acts in question, requires an assessment of state control prior to a finding of state responsibility.

Further, states rarely wish to publicize hostile acts that they are able to attribute to another party. Absent consequences dramatically affecting military operations or the civilian population, states appear to have little incentive to identify hostile cyber operations from third-party states.

5. CONCLUSION

Traditional principles of neutrality sought to avoid the spread of armed conflict. Russia's invasion of Ukraine, an act of naked aggression, rendered the impartiality and inaction of traditional neutrality unacceptable and brought questions of co-belligerency to the fore.

For years, government officials, military strategists, and academics have warned about the escalatory potential of cyber operations when attached to conventional armed conflict. One of the early storylines about the war in Ukraine was about Russia's failure to turn its cyber capabilities into battlefield gains. It increasingly seems that it is possible that the story that ultimately emerges, however, will be how third-party states found an emerging "long war" in cyberspace as the only dimension of the conflict in which they could directly engage with minimal fear of escalation.

[88] *See* Kate Conger & Adam Satariano, *Volunteer Hackers Converge on Ukraine Conflict With No One in Charge*, N.Y. Times (Mar. 4, 2022), https://www.nytimes.com/2022/03/04/technology/ukraine-russia-hackers.html.

2023 15th International Conference on Cyber Conflict
Meeting Reality
T. Jančárková, D. Giovannelli, K. Podiņš, I. Winther
2023 © NATO CCDCOE Publications, Tallinn

The Law of Neutrality and the Sharing of Cyber-Enabled Data During International Armed Conflict

Yann L. Schmuki
Advanced Studies Program in Cyber Security
Department of Computer Science
ETH Zurich, Switzerland

Abstract: The question of the extent to which neutral States are allowed to share (cyber-enabled) data during international armed conflict has rarely been addressed by governments and academia. There are two reasons for this gap: first, States are traditionally reluctant to publicly discuss or internationally regulate sharing of information with partners. Second, the law of neutrality has become a niche discipline in the past years when major international armed conflicts (IAC) were often considered to be passé. However, in today's digitalized societies, information has acquired a value similar to physical goods. Supporting a belligerent with data may therefore be just as problematic from a neutrality perspective as delivering weapons. This paper discusses the important implications of the law of neutrality for neutral States to share data obtained in cyberspace. After introducing a neutrality framework that takes contemporary State practice into account, I illustrate that the discussions on neutrality in the context of the Russia-Ukraine war are neither new nor unaddressed. A short case study will outline the inherent tensions between a neutral State's impartiality and its preventive obligations. Weighing these two factors in the context of an interconnected, cyber-driven security landscape, I argue that during an IAC, a neutral has the ability, but not the obligation, to share certain information with selected partners. However, this does not include militarily actionable data, as such sharing would violate the neutral State's fundamental impartiality obligations.

Keywords: *law of neutrality, cyber-enabled data, data sharing, impartiality, prevention*

1. INTRODUCTION

Two particular legal regimes apply to the ongoing international armed conflict (IAC) between the Russian Federation and Ukraine: international humanitarian law (IHL) and the law of neutrality (LoN).[1] While IHL in international and non-international armed conflicts enjoyed much attention in the past decades from most key actors, the law of neutrality has played a secondary role in public discourse. This has been steadily changing since Europe met reality in February 2022.

In the media and beyond, the question was raised as to the extent to which the law of neutrality, mainly based on 19th- and early 20th-century rules, is able to provide an appropriate framework for 21st-century IACs. At the same time, a number of States more or less directly referred to their 'neutrality' or the 'law of neutrality' when publicly explaining their (non-) support for one of the belligerent States in the war in Ukraine.[2] In fact, the secondary role the LoN played in the past has led to particular regulatory and research gaps in the area of cyberspace in general and data sharing in particular. The traditional rules of neutrality, notably the Hague Conventions, do not provide clear guidance in these areas, and their provisions can at best be operationalized by analogy. Furthermore, State practice is not sufficiently clear or public to make sound conclusions as to the applicable customary international law.

This paper addresses this gap by proposing a differential approach that ponders the neutral's impartiality and prevention obligations. It first expounds on how cyber means empower neutral States to potentially gather vast amounts of data that can be used for military or related purposes. A broad viewpoint is taken, and the concept of 'cyber-enabled' data is introduced. This notion describes any information a State gains from non-public sources by cyber means and it is subsequently subcategorized as 'actionable' and 'non-actionable' data.

[1] Certain authors consider the law of neutrality to form an inherent component of international humanitarian law. This paper treats the disciplines as distinct, due to their differing constitutive principles and the way they are applied during international armed conflict. However, both are part of the broader 'law of armed conflict'.

[2] Brazil: Al Jazeera, 'Russia-Ukraine war: What's behind Brazil's neutral position' (*Al Jazeera*, 22 April 2022) <www.aljazeera.com/news/2022/4/22/russia-ukraine-war-whats-behind-brazils-neutral-position> accessed 19 February 2023; India: 'PM Modi Explains Reasons for India's Neutrality in Russia-Ukraine War' (*Ani News*, 10 March 2022) <www.aninews.in/news/national/general-news/pm-modi-explains-reason-for-indias-neutrality-in-russia-ukraine-war20220310215913/> accessed 19 February 2023; Kazakhstan: 'Tokaev: Neutrality Corresponds with our National Interests' (*Informburo*, 27 September 2022) <https://informburo.kz/novosti/tokaev-nejtralitet-otvechaet-nashim-nacionalnym-interesam> accessed 19 February 2023; Kyrgyzstan: 'Sadyr Shoparov Pleads for Neutrality in the Current Situation Between Russia and Ukraine' (*Radio Azattyk*, 9 March 2022) <https://rus.azattyk.org/a/31744688.html> accessed 19 February 2023; Switzerland: 'Questions and Answers on Switzerland's Neutrality' (*Swiss MFA*, 9 September 2022) <www.eda.admin.ch/eda/en/fdfa/fdfa/aktuell/newsuebersicht/2022/03/neutralitaet.html>; Turkmenistan: 'Turkmenistan Will Continue Policy of Neutrality' (*Embassy of Turkmenistan in Kyiv*) <https://ukraine.tmembassy.gov.tm/en> accessed 19 February 2023.

In a second step, I introduce the concept of a ‘neutrality–belligerency continuum’ in reference to State practice and changing understandings of neutrality. This proposed contemporalisation of the LoN separates two core legal issues that arise in this context. The assessment of when a neutral becomes a belligerent is neatly distinguished from the question of whether a neutral violates its obligations under the LoN. This paper is limited to a discussion of the latter.

The central part of the paper addresses the law of neutrality and how it can be applied to a neutral State sharing cyber-enabled data during an IAC. I outline the existing legal framework and characterize the impartiality and prevention obligations of the neutral. Drawing analogies from State practice and a World War I case study, I argue that a neutral cannot share actionable data with belligerent or third States. However, it must be allowed to exchange non-actionable data, as this is essential to obtaining information necessary to perform its preventive obligations and to satisfy its own defensive needs.

2. THE PRACTICAL SIDE: THE NEUTRAL STATE AND ITS ACCESS TO DATA

Cyberspace: An Enabler for the Neutral and the Non-belligerent

In today’s digitalized international security environment, the capacities and challenges of neutral and non-belligerent States fundamentally differ from those of the past. For geographical and political reasons, a neutral State traditionally had very limited means to obtain reliable first-hand information on belligerents. Accessing the frontlines of conflict to gather first-hand information is a challenging endeavour, and the risk of getting drawn into conflict is high. Cyber means, however, can potentially provide (neutral) States with real-time remote access to data, notably on key military developments during an IAC.[3]

As the case study in the second part of this research will show, early telegraph or signal interception allowed neutral actors to obtain certain forms of wartime information. However, this access was very limited and primarily targeted only communication, not stored information. Contemporary bulk collection, computer network exploitation, or simple digital data transfer provide the neutral or the non-belligerent State with a whole new range of potential tools to gather and share information. At the same time, the neutral has become considerably more vulnerable to violations of its neutrality, notably by cyber means.

[3] Alexandr Galushkin, ‘On Cyberespionage and Cyberintelligence at the Present Stage’ (2014) 3 Bulletin of the Peoples’ Friendship University of Russia (RUDN) 43–44.

The Notion of 'Cyber-Enabled Data'

The 'data' referred to in this paper is addressed variously in academic literature and political discourse.[4] In the context of clandestinely obtained cyber-enabled data, the *Tallinn Manual 2.0* primarily uses the term '(cyber) espionage'.[5] However, States may obtain data in various ways, some of which do not involve any activity related to what is commonly referred to as 'espionage'. One can, for example, imagine a diplomat or military commander from belligerent State A deciding, for ideological or other reasons, to e-mail information to the Ministry of Foreign Affairs of neutral State B. In this case, the data is obtained without intelligence agencies or any form of active collection being involved. Only referring to cyber 'espionage' or 'intelligence' would, therefore, unnecessarily restrict the range of ways through which information can be obtained by neutrals during an IAC.[6]

Therefore, this paper uses, where possible, the term 'gathering of (cyber-enabled) data' instead of 'intelligence' or 'espionage'. The former is the broadest term, which leaves unspecified the method of acquisition and the stage of data processing, and refrains from normative qualifications. This broad understanding also reflects official statements of certain States. Poland, for example, refers in its public position paper on the applicability of international law to cyberspace to the 'theft of data' and not 'espionage'.[7]

'Cyber-enabled data' or simply 'data' as employed in this paper, is, therefore, any form of information that a State has gained from non-public sources by cyber means.

Actionable and Non-actionable Data

Literature on information and intelligence in the context of an armed conflict typically distinguishes between military, political, and economic forms. However, in the context of neutrality, this distinction is not useful. It is not the content or the method of acquisition of the data that is decisive from an LoN perspective, but rather its subsequent use (-ability). A binary qualification of whether data is actionable or not is thus more appropriate for the following analysis.

Sharing data that can provide a belligerent with a direct military advantage, allowing it to take kinetic or cyber action, is intuitively more problematic than providing data that only allows a better understanding of certain political or economic processes. Such a distinction, while not directly founded in legal documents, is justified by drawing

[4] Sulmasy and Yoo explained that international law does not provide for an '(...) internationally recognized and workable definition of "intelligence collection"'. Glenn Sulmasy and John Yoo, 'Counterintuitive: Intelligence Operations and International Law' (2007) 28 Michigan Journal of International Law 625, 637.

[5] Michael N Schmitt (ed), *Tallinn Manual 2.0 on the International Law Applicable to Cyber Operations* (2nd edn, Cambridge University Press 2017) 168ff.

[6] However, when only 'intelligence' is explicitly concerned, this term is employed.

[7] Council of Ministers of the Republic of Poland, 'Position of the Republic of regarding the Application of International Law in Cyberspace', point 2 (29 December 2022) <www.gov.pl/attachment/3203b18b-a83f-4b92-8da2-fa0e3b449131> accessed 19 February 2023.

an analogy between the material support of a belligerent and data sharing. While one-sidedly providing humanitarian goods to a belligerent in an IAC is generally not qualified as a violation of neutrality, the direct or indirect provision of weapons is broadly considered to be.[8]

It is, therefore, doubtful whether an absolutist prohibition or permission for a neutral's support for a belligerent, be it with goods or information, is a convincing approach under the current political and legal realities.

Pearson and Watson propose a definition of 'actionable intelligence' as 'intelligence that can be acted upon within a 12 to 72 hour period of time'.[9] For the sake of this paper, 'actionable data' is data that allows an actor to prepare and execute concrete military action, be it in kinetic form or in cyberspace. Data on tactical military developments or locations, for example, is actionable. Information concerning the nomination of a new political leader in an occupied territory can equally be directly linked to a subsequent airstrike and thus classified as actionable. On the other hand, data that leads to the imposition of sanctions, for example, is not considered actionable as it does not set the ground for immediate military action. As for other neutrality-related questions, a case-by-case assessment by the neutral is necessary to qualify data as actionable or non-actionable.

3. NEUTRALITY, NON-BELLIGERENCY, AND BELLIGERENCY

State Practice in the Area of Data Sharing During IAC

The IAC in Ukraine has led to an unprecedented amount of public State practice in the area of data sharing. Reportedly, Western States, without claiming to be neutral, have provided Ukraine with intelligence almost in real-time.[10] John Kirby, Press Secretary of the US Department of Defense, affirmed in May 2022 that: 'the United States provides battlefield intelligence to help Ukrainians… We do provide them useful, timely intelligence'.[11]

In its frankness, this statement implies that the information the US shares with the Ukrainian armed forces is militarily actionable, as it is concrete information that is

[8] Constantine Antonopoulos, *Non-Participation in Armed Conflict: Continuity and Modern Challenges to the Law of Neutrality* (Cambridge University Press 2022) 91ff.

[9] Stephen Pearson and Richard Watson, *Digital Triage Forensics: Processing the Digital Crime Scene* (Syngress, Elsevier Science [distributor] 2010) 9.

[10] Marko Milanovic, 'The United States and Allies Sharing Intelligence with Ukraine' (*EJIL:Talk!*, 9 May 2022) <www.ejiltalk.org/the-united-states-and-allies-sharing-intelligence-with-ukraine/> accessed 19 February 2023.

[11] 'Pentagon Press Secretary John F. Kirby Holds a Press Briefing' (*US Department of Defense*, 5 May 2022) <www.defense.gov/News/Transcripts/Transcript/Article/3022007/pentagon-press-secretary-john-f-kirby-holds-a-press-briefing/> accessed 19 February 2023.

used on the battlefield. According to US officials, the Russian flagship *Moskva* was notably sunk after the Ukrainian army struck it based on such intelligence assistance.[12]

Even if intelligence sharing is nothing new and is reportedly broadly practiced by States,[13] the way it has recently been confirmed by State officials is very progressive.[14] However, it seems impossible to establish a legally authoritative custom at the current stage, as neither practice nor *opinio iuris* are public and uniform. At the same time, even the United Nations (UN) has acknowledged the key role that information sharing may play in certain situations, notably in the realm of counterterrorism. UNSC Resolution 2396 (2017), for example, calls upon States 'to improve timely information sharing, through appropriate channels and arrangements'.[15] However, this reference was limited to the combatting of non-State actors and did not refer to States involved in an IAC.

Changing Understandings of Neutrality

In 1996, the International Court of Justice in its Advisory Opinion on Nuclear Weapons reaffirmed that the principle of neutrality 'is applicable… to all international armed conflict'.[16]

However, recent developments have challenged the strict applicability of the traditional law of neutrality to IACs. Besides extensive data sharing, weapons deliveries to belligerent States – be it Ukraine, Saudi Arabia, or Türkiye – are on the daily agenda of States that do not consider themselves parties to the conflict. More generally speaking, there are different visions on how the law of neutrality is still to be applied in the post-1945 UN system. These discussions have fulminantly reemerged in the context of the recent IAC between the Russian Federation and Ukraine.[17]

Today, it is difficult to maintain the traditional understanding of neutrality and belligerency as a binary function. State practice implies that States can actually

[12] Helene Cooper, Eric Schmitt, and Julian E Barnes, 'U.S. Intelligence Helped Ukraine Strike Russian Flagship, Officials Say' *New York Times* (5 May 2022) <www.nytimes.com/2022/05/05/us/politics/moskva-russia-ship-ukraine-us.html> accessed 19 February 2023.

[13] Kahana describes the cooperation between Mossad and the CIA in sharing (actionable) intelligence. See Ephraim Kahana, 'Mossad-CIA Cooperation' (2001) 14(3) International Journal of Intelligence and Counterintelligence 409.

[14] Julian Richards, 'Intelligence Sharing in Remote Warfare' *E-International Relations* (17 February 2021) <www.e-ir.info/2021/02/17/intelligence-sharing-in-remote-warfare/> accessed 19 February 2023.

[15] UNSC RES 2396 (21 December 2017) UN Doc S/RES/2396.

[16] Legality of the Threat or Use of Nuclear Weapons, Advisory Opinion, 1.C.J. Reports 1996, 226, 261, para 89.

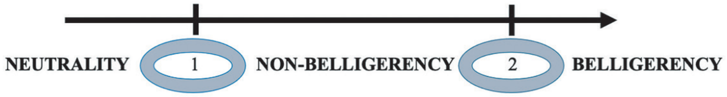
[17] Michael N Schmitt, 'Providing Arms and Material to Ukraine: Neutrality, Co-Belligerency, and the Use of Force' (*Lieber Institute*, 7 March 2022) <https://lieber.westpoint.edu/ukraine-neutrality-co-belligerency-use-of-force/> accessed 19 February 2023; Stefan Talmon, 'Waffenlieferungen an die Ukraine als Ausdruck eines wertebasierten Völkerrechts' (*Verfassungsblog*, 9 March 2022) <https://verfassungsblog.de/waffenlieferungen-an-die-ukraine-als-ausdruck-eines-wertebasierten-volkerrechts/> accessed 19 February 2023; Milanovic (n 10).

operate in a 'middle ground' between strict neutrality and belligerency.[18] In this context, Cordey and Kohler argue that 'according to modern State practice, the applicability of the law of neutrality depends on functional considerations, that often result in a differential or partial applicability of that body of law'.[19] At the same time, already Additional Protocol I to the Geneva Conventions and the Third 1949 Geneva Convention employed the notions of 'neutral or other State not party to the conflict'[20] or 'neutral or non-belligerent powers' respectively.[21]

The Neutrality–Belligerency Continuum

In the context of the ongoing IAC in Europe, different solutions have been proposed to reconcile State practice with traditional understandings of neutrality. Some authors have argued that neutrality has ceased to apply, while others have proclaimed 'the end of impartiality' or similar solutions.[22] It is outside the scope of this paper to holistically discuss the concrete actions that make States cease to be neutral. The working assumption of this paper is that the law of neutrality is best understood as a neutrality–belligerency continuum (see Figure 1).[23] In this understanding, States can adopt a neutral position that comes with certain duties, notably impartiality and prevention, and rights under the LoN.[24] At the same time, States can also decide to support one of the belligerents without directly participating in the conflict. It is general international law that applies in this second case.

FIGURE 1: IMPARTIALITY AND NON-PARTICIPATION THRESHOLDS



1: Impartiality threshold: Designates the border between the permitted behaviour of a neutral State and that of a non-belligerent State. Defining this threshold in the domain of data-sharing is the subject of this research.
2: Non-participation threshold: Designates the border between the permitted behaviour of a non-belligerent State and direct participation in an IAC. Establishing this threshold needs to be the subject of further research.

[18] This 'middle-ground-theory' has notably been described, and largely rejected, by Roscini. Marco Roscini, *Cyber Operations and the Use of Force in International Law* (Oxford University Press 2014) 267/268.

[19] Sean Cordey and Kevin Kohler, *The Law of Neutrality in Cyberspace* (ETH Zurich 2021) 7.

[20] Protocol Additional to the Geneva Conventions of 12 August 1949, and relating to the Protection of Victims of International Armed Conflicts, arts. 2(c), 30(3)(c), 31(1), 31(2), 31(4), and 47(f).

[21] Geneva Convention Relative to the Treatment of Prisoners of War (adopted 12 August 1949, entered into force 2 November 1950) 75 UNTS 135 (Geneva Convention) art 122.

[22] See Oona A Hathaway and Scott Shapiro, 'Supplying Arms to Ukraine Is Not an Act of War' (*Just Security*, 12 March 2022) <www.justsecurity.org/80661/supplying-arms-to-ukraine-is-not-an-act-of-war/> accessed 19 February 2023; Michael N Schmitt, 'Providing Arms and Material to Ukraine: Neutrality, Co-belligerency, and the Use of Force' (*Lieber Institute*, 7 March 2022) <https://lieber.westpoint.edu/ukraine-neutrality-co-belligerency-use-of-force/>; Talmon (n 17).

[23] Political scientists and State policies typically distinguish notions of 'integral neutrality', 'differential neutrality', 'active neutrality' or 'qualified neutrality'. These are political terms and will not be further addressed in the course of this research.

[24] This voluntarist understanding applies in the context of an IAC. It does not apply to States, like Switzerland or Turkmenistan, that have proclaimed and internationally been recognized as permanently neutral. The latter must be presumed to be neutral in any IAC.

4. THE LEGAL SIDE: DATA SHARING AND THE OBLIGATIONS OF THE NEUTRAL

The Tallinn Manual's Silence on Data Sharing Under the Law of Neutrality

The *Tallinn Manual 2.0* clearly states that 'the international Group of experts unanimously agreed that the law of neutrality applies to cyber operations'.[25] Several States have publicly adopted this position.[26] When it comes to 'cyber-espionage', the manual takes a circumstantial approach and argues that the 'lawfulness depends on whether the way in which the operation is carried out violates any international law obligations that bind the State'.[27] However, on the sharing of cyber-enabled data in the context of an IAC, the two Tallinn manuals have so far remained silent.

In IHL, explicit rules applying to 'reconnaissance' and 'espionage' do address the question of data gathering.[28] However, these rules are not helpful when it comes to the question of whether and to what extent available data can be shared with belligerents in the context of an IAC. Whether the data was obtained legally or through an internationally wrongful act has, as such, no implication for the neutrality-conformity of its sharing. The sharing State may be in violation of its neutrality obligations even if the data was obtained legally and vice versa.

The Law of Neutrality and the Principles Derived from It

The 1907 Hague Conventions V and VIII constitute the main black-letter foundation of the LoN. Influential States, or their predecessors, have ratified the two conventions, notably Brazil, China, France, Germany, the Netherlands, Japan, Switzerland, the Russian Federation, Ukraine, and the United States.[29] Furthermore, the conventions are considered to largely reflect customary international law.[30]

From the 1907 Hague Conventions, four constituting principles of neutrality have been derived: non-participation, prevention, impartiality, and acquiescence.[31] As argued above, the 'non-participation threshold' will not be discussed in this paper as it

[25] *Tallinn Manual 2.0* (n 5) 553.

[26] These States were France, Italy, the Netherlands, Romania, Switzerland, and the United States. Cordey and Kohler (n 19) 25.

[27] *Tallinn Manual 2.0* (n 5) r 32.

[28] See notably Protocol Additional to the Geneva Conventions of 12 August 1949, and relating to the Protection of Victims of International Armed Conflicts (adopted 8 June 1977, entered into force 7 December 1979) 1125 UNTS 3, art 46(1).

[29] International Committee of the Red Cross (ICRC) International Humanitarian Law Databases <https://ihl-databases.icrc.org/en/ihl-treaties/hague-conv-v-1907/state-parties?activeTab=undefined> and <https://ihl-databases.icrc.org/en/ihl-treaties/hague-conv-xiii-1907/state-parties?activeTab=undefined> accessed 19 February 2023.

[30] Wolff Heintschel von Heinegg, 'Neutrality in the War Against Ukraine' (*Lieber Institute*, 1 March 2022) https://lieber.westpoint.edu/neutrality-in-the-war-against-ukraine/ accessed 19 February 2023; Roscini (n 18) 247.

[31] Cordey and Kohler (n 19) 9.

is considered to typically pose less severe restrictions on the neutral than the qualified 'impartiality principle'.

When it comes to impartiality, Article 9 of the 1907 Hague Convention V on neutrality in case of warfare on land states that 'every measure of restriction or prohibition taken by a neutral Power… must be *impartially* applied by it to both belligerents'.[32] The duty of impartiality is not directly addressed in the Tallinn manuals' rules. However, in the commentary on Rule 151, the experts argue that restrictive measures by the neutral on its cyberspace 'must be impartially applied to all belligerents'.[33]

As for the preventive component of neutrality, Article 5 of the Hague Convention states that a 'neutral Power must not allow any of the acts referred to in Articles 2 and 4 to occur on its territory'.[34] Rule 152 of the *Tallinn Manual 2.0* transfers this approach into cyber law. Adopting the object and purpose of the Hague Convention, the experts argue in the manual that a 'neutral State may not knowingly allow the exercise of belligerent rights by the parties to the conflict from cyber-infrastructure located in its territory or under its exclusive control'.[35]

Applying the Law of Neutrality to Data Sharing in Times of IAC

Object and Purpose of Neutrality Conventions

The law of neutrality arose from a need for predictability in the relations between belligerents and non-belligerents.[36] Like IHL, the LoN accepts war as a reality and departs from the idea that, even in a conflict, all parties are better off when certain fixed rules of behaviour are respected. Neutrality is a concept that primarily applies to States, while also having repercussions for private actors, notably in a multi-stakeholder environment. However, this analysis focuses on the direct obligations of States, who may or may not delegate activities to or at least tolerate activities by third actors.

Legal Analogies

There is no clear rule, either in customary international law or in an international treaty, that directly addresses the connection between data sharing and neutrality.[37] Cordey and Kohler are among the few authors to address the issue in a cyber context, arguing that 'sharing *military* intelligence about a belligerent to another belligerent violates

[32] Convention (V) Respecting the Rights and Duties of Neutral Powers and Persons in Case of War on Land, The Hague, 18 October 1907, art 9 (emphasis added).

[33] *Tallinn Manual 2.0* (n 5) r 151, commentary 8, 557 (emphasis added).

[34] Hague Convention (V) art 5(1). Article 2 of the Convention prohibits the belligerents to move troops, munitions or supply through the territory of a neutral State. Article 4 states that no 'corps de combat' or 'recruiting agencies' can be opened on the neutral's territory to assist the belligerents.

[35] *Tallinn Manual 2.0* (n 5) r 152, 558.

[36] Antonopoulos (n 8) 222.

[37] *Oslo Manual on Select Topics of the Law of Armed Conflict* (Springer Nature 2021).

neutrality'.[38] Unfortunately, the authors do not elaborate upon why the intelligence shared must be military for the act to constitute a violation of neutrality.

They, however, make an important point when drawing an analogy between Article 47 of the 1923 Hague Rules of Aerial Warfare and data collection in the cyberspace of the neutral. Article 47 states that a 'neutral State is bound to take such steps as the means at its disposal permit to *prevent* within its jurisdiction aerial observation of the movements, operations or defences of one belligerent, with the intention of informing the other belligerent'.[39]

The authors convincingly argue that from this article, an obligation of the neutral State 'to conduct counterespionage… to prevent belligerents from exploring and observing neutral networks that would allow them to gain intelligence on the other belligerents' wartime action' can be derived.[40]

If this analogy is accepted and a neutral State is obliged to prevent and end wartime cyber-espionage in its territory, *a fortiori*, it cannot actively transfer such data and information to one of the belligerents. The *US Law of War Manual* makes a similar point when explaining the fact that a 'neutral State, if it so desires, may transmit messages by means of its communications facilities does not imply that the neutral State may use such facilities or permit their use to lend assistance to the belligerents on one side only'.[41]

However, these argumentations remain relatively constructed and notably rely on analogies and assumptions rather than black-letter or case law. One of the very few cases in which a court had to publicly pronounce itself on the neutrality-conformity of the sharing of data obtained by technical means is the so-called *Affaire des Colonels*.

Case Study: *L'Affaire des Colonels* (1915/16)

In the course of the first two years of World War I, two colonels of the Swiss Military Staff's intelligence section regularly shared classified intelligence briefings containing decoded Russian telegraphic correspondence with German and Austro-Hungarian military representatives.[42] Suspected of pro-German sentiments, the two colonels claimed that they acted in order to obtain essential military information from the Central Powers in exchange. At the same time, the representatives of the Entente in Switzerland expressed their strongest condemnation of the colonels' behaviour.[43]

[38] Cordey and Kohler (n 19) 56 (emphasis added).

[39] The Hague Rules of Aerial Warfare (1923) art 47 (emphasis added). The Rules are considered to be declarative of customary international law (Roscini [n 18] 249).

[40] Cordey and Kohler (n 19) 37.

[41] US Department of Defense, *Law of War Manual* (2015) 15.5.3.1.

[42] Sebastian Steiner, 'Oberstenaffäre', *Online International Encyclopedia of the First World War* (2016) <http://encyclopedia.1914-1918-online.net/article/oberstenaffare/2016-05-23> accessed 19 February 2023.

[43] Hand R Fuhrer, 'Die Gefahr aus dem Westen' *Neue Zürcher Zeitung* (13 January 2022) <www.nzz.ch/schweiz/die-gefahr-aus-dem-westen-ld.94548?reduced=true> accessed 19 February 2023.

As a first reaction, the Commander of the Swiss army General Ulrich Wille ordered members of the General Staff to immediately refrain from interacting with the military attachés of any belligerent nation, be it France, Russia, Germany, or Austria-Hungary. Under public pressure, the two colonels were subsequently charged by a Swiss military Tribunal with 'violation of neutrality', 'treason', and 'misconduct'.[44]

During the trial, the Chief of the Swiss General Staff Theophil Sprecher von Bernegg defended the accused. He noted in relation to information exchange with belligerents that if the information received from partners was of considerable value, the intelligence officers involved should be able to consider whether 'they want to offer something in return that is probably not in accordance with strict respect of the obligation of neutrality'.[45] All participants, including the defense and the defendants, did agree that the transfer of the daily briefing and decoded Russian correspondence constituted a violation of the obligations of impartiality. However, the defending side was seeking to justify the sharing based on the necessity to obtain strategically relevant information in exchange.

In its final judgment, the court found that there was an objective violation of neutrality, as the regular transfer of the daily intelligence briefings, containing information from the decoded Russian correspondence, included 'a certain, even if only formal and external, advantage for the concerned belligerent powers'.[46] Judge Major Emil Kirchhofer argued that 'the simple exploitation of the military affairs of others does not violate neutrality. In his opinion, 'the latter was only violated if in relation to the treatment of representatives of different groups of Powers, there is differentiation in proceeding'.[47] Acquitted from the accusation of treason, the two colonels were mildly punished by a release from their duties.

The Neutral's Obligation to Prevent Violations of Its Neutrality

As elaborated upon above, the LoN obliges a neutral State to be able to terminate and prevent violations of its neutrality, notably in cyberspace. However, the neutral can only do so by disposing of reliable information as to the intentions and capabilities of the belligerents. This concerns conventional threats and, probably even more strongly, cyber defense. Antonopoulos even argues that '… the maintenance and respect of one's neutrality in cyber warfare is not so much a matter of belligerent … but rather of neutral conduct'.[48] Knowing which software or cyber capacities a belligerent might employ to violate a neutral's neutrality is key to preventing it. In this context, one might think, for example, of a neutral State adopting technical measures to separate critical infrastructure from threatened systems.

[44] Hans R Fuhrer, 'Vor Hundert Jahren: Die Oberstenaffäre 1915/16', *Schweizer Soldat* (February 2016) <www.e-periodica.ch/digbib/view?pid=sol-004:2016:91::980> accessed 19 February 2023.

[45] Jürg Schoch, *Die Oberstenaffäre: eine innenpolitische Krise 1915/1916* (Lang 1972) 90.

[46] ibid 94.

[47] ibid 87.

[48] Antonopoulos (n 8) 211.

Already in the *Affaire des Colonels*, the defendants argued that information sharing was key to the neutral's capacity to defend itself. However, as Walsh explains, 'intelligence is a valuable commodity, and States bargain with one another to obtain the best possible return before agreeing to share it'.[49] Even if belligerents usually have an interest in preventing their 'enemy' from using the neutral's territory or cyber infrastructure to attack, there are many constellations in which it cannot be assumed that the neutral receives the required data, notably threat intelligence, 'for free'. In a digitally interconnected world, the neutral must therefore be permitted to share certain data with belligerents in exchange for data relevant to its preventive (cyber) obligations. This is still valid if one accepts that the duty of prevention in cyberspace is relative and that the neutral cannot be expected to prevent any violation of its neutrality.[50]

Which Data Are Neutral States Permitted to Share?

The neutral's impartiality and prevention obligations must be weighed against each other. If, as State practice suggests, a neutral is allowed to provide humanitarian aid to only one of the belligerent sides, it must equally be permitted to share non-actionable data to guarantee it receives the necessary information to assume its preventive obligations.

At the same time, the LoN can only be maintained as a relevant institution in contemporary international relations if it embraces at least a minimum level of military impartiality. The neutral must therefore exclude 'actionable data' from any sharing activities with only one belligerent side at the receiving end.

Roscini, and Cordey and Kohler, mention computer emergency response team (CERT) cooperation as a practical example of technical data sharing.[51] The question thus arises of whether CERT data can be considered actionable. This problem lacks a general answer, but if the shared data allows the belligerent to take subsequent military action in the form of targeted kinetic or cyber operations, the data initially shared must be considered 'actionable'. On the other hand, if the shared data 'only' implies the decision to take down the infected system (kill switch), it is non-actionable.

Are Neutrals Permitted to Transfer Data to Non-belligerents?

There is no problem with the neutral sharing non-actionable data with other non-belligerents, as the data, according to the argument made in this paper, may also be shared with belligerents. However, even when data-sharing agreements often contain a 'third-party rule' prohibiting the further transfer of information, it is almost impossible for a sharing neutral to ensure that (actionable) data is not passed on to

[49] James I Walsh, *The International Politics of Intelligence Sharing* (Columbia University Press 2010) 4.

[50] Antonopoulos (n 8) 218.

[51] Roscini (n 18) 25; Cordey and Kohler (n 19) 32.

a belligerent.[52] Therefore, the neutral is not permitted to share actionable data with other non-belligerents.

In this context, an analogy can be drawn with the delivery of weapons. In 2022, the Swiss Federal Council argued that due to the 'principle of equal treatment', it could not 'approve the transfer of Swiss war material by Germany and Denmark to Ukraine'.[53] In this logic, it would be equally problematic if a neutral allowed a third party to forward actionable data to a belligerent. However, while the transfer of weapons can be tracked by the neutral State, the latter completely loses control over the data it shares with partners. Therefore, actionable data must be excluded from any sharing from the very beginning.

What Measures Can a Belligerent Take to Bring a Neutral Back into Compliance?

If the neutral starts sharing actionable data with belligerents, it acts in violation of its neutrality obligations. This, as Schmitt and others have argued, does not render it a belligerent.[54] However, as the neutral has voluntarily chosen to submit itself to the special neutrality regime (be it ad-hoc or permanently), the 'harmed' belligerent may use the remedies proposed by that regime to respond to the violation.[55] In this situation, Article 153 of the *Tallinn Manual 2.0* can be applied and the 'aggrieved party to the conflict may take such steps, including by cyber-operations, as are necessary to counter that conduct'.[56] However, derived from the general law of State responsibility, subsequent action taken is required to be proportionate to the violation. As long as the sharing of the data by the neutral does not arise to the high benchmark of an armed attack, a retaliatory use of force by the 'aggrieved party', is not permitted.[57] The belligerent must first request the ending of the unlawful sharing activities from the neutral. If the latter does not follow up to this request, the belligerent may apply measures of self-help, notably in the cyberspace of the neutral, to make the data-gathering or its subsequent sharing stop.[58]

[52] Richards (n 14).

[53] The Swiss Federal Council, 'Ukraine: Federal Council Takes Decision on Various War Material Transactions' (3 June 2022) <www.admin.ch/gov/en/start/documentation/media-releases/media-releases-federal-council.msg-id-89141.html> accessed 25 January 2023.

[54] Schmitt (n 17). However, this is not the only view. Goldsmith and Bradly for example argue that 'one way that a state can become a co-belligerent is through systematic or significant violations of its duties under the law of neutrality' (Jack Goldsmith and Curtis Bradley, 'Congressional Authorization and the War on Terrorism' (2005) 118 Harvard Law Review 2047, 2112.

[55] Cordey and Kohler (n 19) 45.

[56] *Tallinn Manual 2.0* (n 5) art 153, 560 describes the 'Response by parties to the conflict to violations'.

[57] Roscini (n 18) 273: '… it is now the UN Charter that determines the legality of forcible reaction'.

[58] As explained in the Tallinn Manual, 'measures of self-help are subject to a requirement of prior notification that allows a reasonable time for the neutral State to address the violation'. *Tallinn Manual 2.0* (n 5) 561; Antonopoulos (n 8) 219.

5. CONCLUSION

This paper was written in the context of broader discussions on how neutral States may or may not support belligerents during an IAC. While the question of direct or indirect transfer of weapons is increasingly thematized, rules defining the neutral State's rights to share data are not in the spotlight. Departing from the observation that States apply the law of neutrality differentially, the proposed neutrality-belligerency continuum has allowed for a closer analysis of the latter problematic.

I have argued that a neutral must be able to share non-actionable data with belligerents (and third States) to ensure that it can prevent belligerents from operating on its territory or within its cyberspace. This is specifically true when on the one hand, neutrals may have an unprecedented amount of valuable data at their disposal, and on the other hand, are increasingly vulnerable to violations of their neutrality. Militarily actionable data, however, cannot be shared as this would violate the neutral's impartiality obligations.

In conclusion, I want to emphasize that the literature has rarely addressed this legal domain, which is (purposely) even less regulated by black-letter law. State practice largely remains obscure and non-consistent. As a consequence, legal positivism was intermingled with *lex ferenda* in this paper. Furthermore, additional research is necessary to expound on where the unaddressed 'non-participation threshold' lies and to discuss the role of non-State actors within the neutrality-belligerency continuum.

ACKNOWLEDGEMENTS

I would like to express my thankfulness to all those who contributed their valuable inputs in the course of this research. Furthermore, I want to say *merci* to the Swiss Study Foundation for supporting my academic formation in the past years.

2023 15th International Conference on Cyber Conflict
Meeting Reality
T. Jančárková, D. Giovannelli, K. Podiņš, I. Winther
2023 © NATO CCDCOE Publications, Tallinn

Obligations of Non-participating States When Hackers on Their Territory Engage in Armed Conflicts

Marie Thøgersen
PhD Fellow
iCourts, Faculty of Law, University of Copenhagen
Institute for Military Technology, Royal Danish Defence College
Copenhagen, Denmark
mth@jur.ku.dk

Abstract: One of the most striking aspects of cyberspace is the diffusion of power to the individual. Even a single person can, from the comfort of their own home, cause considerable harm to States on the other side of the globe. Since the Russian invasion of Ukraine, both belligerent States have successfully deployed novel techniques for the mobilization of individuals in cyberspace. The absence of geographical boundaries in cyberspace triggers important questions regarding the international legal implications for States whose territories are being used for such operations. To assess how the legal framework stands the test of reality, this article examines the possible international legal obligations of non-participating States hosting individuals conducting malicious cyber operations against Russia orchestrated by the IT Army of Ukraine. After a legal characterization of the activities of the IT Army, this article scrutinizes the legal norms conferring obligations on territorial States and accounts for the prevailing ambiguities surrounding their application. The principle of due diligence entails an obligation for States to not allow their territories to be used for cyber operations affecting the rights of, and producing serious adverse consequences for, other States. Special challenges surround the assessment in the context of an armed conflict; the status of a State as an aggressor entails important nuances to the *prima facie* rights of the State. Based on an analysis of how the legal framework applies to the activities of the IT Army of Ukraine, the article concludes that for non-participating States, the legality of refraining from exercising due diligence will often be contingent on contentious legal questions regarding countermeasures and self-defence.

Keywords: *cyberspace, non-State actors, countermeasures, self-defence, due diligence, obligations of non-participating States*

1. INTRODUCTION

Cyberspace makes it possible to conduct malicious operations against targets on the other side of the globe. Even a single individual can, from the comfort of their own home, involve themselves in distant armed conflicts. This provides a significant opportunity for belligerent States to utilize the skills of individuals in other States. The current conflict between Russia and Ukraine has provided several examples of the belligerent States' employment of individuals in other States to strengthen their cyber capacities. Most remarkable is perhaps the Ukrainian establishment of the IT Army of Ukraine (IT Army). Launched in a tweet by Ukraine two days after the Russian invasion, the IT Army quickly attracted thousands of Ukrainian as well as international volunteer hackers. Since then, the volunteers have been working in collaboration with officials from Ukraine's Ministry of Defence to target Russian infrastructure and websites.[1] In his thorough analysis of the IT Army, Soesanto argues that by allowing such participation, North Atlantic Treaty Organization (NATO) and European Union Member States might be setting unintended legal and ethical precedents that may create significant political blowback in the future.[2]

The ethical and political aspects aside, his concern about the legal precedents that States could be setting by allowing such participation points to important questions under international law. Particularly, it triggers the question of whether a State whose territory is being used for cyber operations in relation to a conflict to which the State is not a party is obliged to exercise due diligence. This article examines the possible international obligations of non-participating States hosting individuals conducting cyber operations against Russia orchestrated by the IT Army of Ukraine.

[1] 'Connect the Dots on State-Sponsored Cyber Incidents - Ukrainian IT Army' (*Council on Foreign Relations*) <www.cfr.org/cyber-operations/ukrainian-it-army> accessed 7 January 2023.

[2] Stefan Soesanto, *The IT Army of Ukraine – Structure, Tasking, and Ecosystem* (Center for Security Studies (CSS), ETH Zürich 2022) 18.

There is now broad agreement that international law applies in cyberspace.[3] As such, international norms imposing obligations on non-participating States apply equally to cyber activities.

To set the scene, Section 2 examines the legal characterization of the operations of the IT Army of Ukraine. Section 3 scrutinizes the international legal norms conferring obligations on non-participating States when non-State cyber activities are conducted from their territory in relation to an international armed conflict (IAC). Section 4 discusses the situation of a State neglecting an established obligation of due diligence. Against this background, Section 5 discusses the implications of the legal framework to current events as outlined in Section 2.

2. LEGAL QUALIFICATION OF THE OPERATIONS OF THE IT ARMY OF UKRAINE

The legal qualification of the operations of the IT Army influences the possible obligations of States from whose territories the operations are conducted. Therefore, before scrutinizing the possible obligations of territorial States, it is necessary to address two central questions, which must be assessed for every individual operation. *The first question* is whether the acts of the members of the IT Army are attributable to Ukraine. According to Article 8 of the International Law Commission (ILC) Articles of Responsibility of States for Internationally Wrongful Acts (ARSIWA), the conduct of a person or group of persons shall be considered an act of a State under international law if the person or group of persons is in fact acting on the instructions of, or under the direction or control of, that State in carrying out the conduct. As the operations of the IT Army are orchestrated via a Telegram portal controlled by Ukraine, the operations are presumably *instructed* by Ukraine in the sense of Article 8 of ARSIWA. However, an important feature of the IT Army is its successful employment of sub-organizations. The army claims:

> Quite a few channels… conduct DDoS [distributed denial of service] attacks on hostile services with us. Each community has a database of tutorials, as well as a sufficient number of involved participants. It is important to understand that each community is independent and chooses priority goals for itself. But we all communicate with each other and quite often they support us in attacks on our targets.[4]

Communication between distinct but ideologically connected communities may blur the fine lines between communication and coordination, between inspiration and

[3] GGE, 'Report of the Group of Governmental Experts on Advancing Responsible State Behaviour in Cyberspace in the Context of International Security' (United Nations 2021) A/76/135; OEWG, 'Final Substantive Report of the Open-Ended Working Group' (United Nations 2021).

[4] Soesanto (n 2) 18.

instruction. This networked feature inevitably complicates the attribution assessment of each individual operation, and thus, some operations conducted by members of the IT Army may be attributable to Ukraine, while others may not.

The second question is whether the operations of the IT Army constitute use of force contrary to the prohibition in Article 2(4) of the United Nations Charter (UNC). Alternatively, the operations could violate other international obligations such as the principle of non-intervention, dictating that a State may not intervene in the internal affairs of another State, or the principle of sovereignty, dictating that a State may not exercise its physical power in any form in the territory of another State.[5]

Recent operations include a series of DDoS attacks on specialized stores with the aim of preventing newly mobilized Russians from purchasing the required equipment;[6] a hack into the data of 650,000 members of the Russian platform Dobro, where they organized rallies in support of the war;[7] and a series of DDoS attacks against Russian banks, leading to hundreds of angry comments from bank customers.[8] Several of the most severe activities of the IT Army of Ukraine have been targeted at the banking sector. On 6 September 2022, massive DDoS attacks were conducted against Russia's third-largest bank, Gazprom, causing the shutdown of the bank's website, online banking, and call centre.[9] A few days after, the IT Army claimed to have breached the servers of the Central Bank of Russia, stealing thousands of internal documents. The files detailed the bank's operations, its security policies, and the personal data of some of its employees. It should be noted that the bank has denied the allegations.[10] So far, the operations have been non-forcible. However, they may constitute violations of the principle of non-intervention or the principle of sovereignty. For the scope of this paper, it suffices to conclude that some of the operations plausibly violate the said principles, and the possibility that future operations will reach the threshold of force cannot be excluded.

[5] Anders Henriksen, *International Law* (3rd edn, Oxford University Press 2021) 254ff <www.oxfordlawtrove.com/view/10.1093/he/9780198869399.001.0001/he-9780198869399> accessed 6 March 2023.

[6] IT Army of Ukraine, tweet, 30 November 2022 <https://twitter.com/ITArmyUKR/status/1598035202554892288?s=20&t=i_Rd-eWIKoaKf2StEARVng> accessed 13 April 2023.

[7] IT Army of Ukraine, tweet, 30 November 2022 <https://twitter.com/ITArmyUKR/status/1598038067625218048 accessed 13 April 2023.

[8] IT Army of Ukraine, tweet, 30 November 2022 <https://twitter.com/ITArmyUKR/status/1598039804847214593?s=20&t=i_Rd-eWIKoaKf2StEARVng> accessed 13 April 2023.

[9] IT Army of Ukraine, tweet, 6 September 2022 < https://twitter.com/ITArmyUKR/status/1567173639706972160> accessed 13 April 2023; Daryna Antoniuk, 'Ukrainian Hacktivists Claim to Leak Trove of Documents from Russia's Central Bank' (*Recorded Future*, 7 November 2022). <https://therecord.media/ukrainian-hacktivists-claim-to-leak-trove-of-documents-from-russias-central-bank/> accessed 13 April 2023.

[10] Antoniuk (n 9).

Based on those two questions, four possible scenarios can be identified:

(i) an operation attributable to Ukraine above the level of use of force;
(ii) an operation attributable to Ukraine below the level of use of force;
(iii) an operation *not* attributable to Ukraine above the level of use of force;
(iv) an operation *not* attributable to Ukraine below the level of use of force.

After the scrutiny of the legal framework in Sections 3 and 4, Section 5 returns to those scenarios to examine the legal implications of each scenario.

3. MAPPING THE LEGAL FRAMEWORK: OBLIGATIONS ON NON-PARTICIPATING STATES WHEN INDIVIDUALS CONDUCT CYBER OPERATIONS FROM THEIR TERRITORY

The legal relationship between States not participating in an IAC and the conflict parties has traditionally been governed by neutrality law.[11] However, the adoption of the UNC has arguably deprived this corpus of law much of its *raison d'etre*. Section A provides some introductory remarks on the status of neutrality law as of today. Based on those conclusions, Section B scrutinizes the rules potentially implicated when non-participating States' territories and cyber infrastructure are being used by individuals engaging in IACs.

A. Introductory Remarks: From Neutrality Law to Collective Security

Neutrality law is rooted in two Hague Conventions of 1907 – Convention (V) Respecting the Rights and Duties of Neutral Powers in Case of War on Land and Convention (XIII) Concerning the Rights and Duties of Neutral Powers in Naval War. It entails a series of mutual obligations between neutrals and belligerent States.[12] The rules are based on principles of non-participation in the conflict and impartiality among belligerent States. However, with the adoption of the UNC, the status of neutrality law has come into question.

At the outset, Article 2(5) of the UNC states that '[a]ll Members shall give the United Nations every assistance in any action it takes in accordance with the present Charter, and shall refrain from giving assistance to any state against which the [UN] is taking preventive or enforcement action'. This clause denotes that when measures

[11] Stephen P Mulligan, 'International Neutrality Law and U.S. Military Assistance to Ukraine' (Congretional Research Service 2022) LSB10735 3 <https://crsreports.congress.gov/product/pdf/LSB/LSB10735/3> accessed 8 March 2023.

[12] Wolff Heintschel von Heinegg, 'Territorial Sovereignty and Neutrality in Cyberspace' (2013) 89 International Law Studies 89 <https://digital-commons.usnwc.edu/ils/vol89/iss1/17>; Antonopoulos Constantine, *Non-Participation in Armed Conflict: Continuity and Modern Challenges to the Law of Neutrality* (Cambridge University Press 2022) 9.

of collective security are carried out by the UN in conformity with the UNC, Member States must help one side (the UN force) and refrain from aiding and abetting the other (the aggressor State).[13] In situations where the UN Security Council (UNSC) takes measures as proscribed in Chapter VII, non-participating States are obligated to make available to the UNSC their armed forces, assistance, and facilities, including right of passage. To say the least, the UNC obligations on non-participating States appear difficult to reconcile with the notion of neutrality in a traditional sense.[14] Therefore, a relevant question is whether the general prohibition of the use of force in Article 2(4) – and the UNSC as the guarantor of compliance with this prohibition – has deprived neutrality law of its scope of application.

Some scholars have argued that the failure of the UNSC to take action when Council deliberations reach a political impasse has meant that, in practice, neutrality law has remained relevant as a residual system of international law.[15] However, those arguments disregard the fact that the UNC also regulates the situation of an act of aggression not addressed by the UNSC; then, the inherent right to self-defence applies. Kelsen distinguishes between self-defence in a well-functioning system of collective security and self-defence when the system does not work. In the first case, self-defence is an exceptional and provisional interlude, and in the latter, 'it is not an inevitable measure taken within the framework of a working system of collective security, but is the replacement of this system, which is temporarily or definitely blocked, by the opposite principle of self-help'.[16] Self-defence when the system does not work (i.e. self-defence in accordance with Article 51 UNC) applies to all States due to the collective aspect of the provision. While collective self-defence under Article 51 is indeed a right, not a duty, and the exercise thereof implies a procedural risk, it still applies in all cases of an armed attack.[17] The UNSC not being able to identify the aggressor does not affect the rationale inherent in the system (i.e. that one party is always the aggressor) and that aggressor status affects the legal relationship with non-participating States as well.[18] Indeed, in the absence of UNSC action, there will be States who are non-participants in the hostilities, but they will not be neutral, as they will still be under an obligation under the UNC to help the UNSC to find a solution to the case.[19] Moreover, they will have a right to engage in collective self-defence under Article 51, subject to the procedural risk that such action entails.

[13] Yoram Dinstein, *War, Aggression and Self-Defence* (5th edn, Cambridge University Press 2011) 176 <https://www.cambridge.org/core/product/identifier/9780511920622/type/book> accessed 21 April 2023.

[14] ibid.

[15] Hitoshi Nasu, 'The Laws of Neutrality in the Interconnected World: Mapping the Future Scenarios' in *The Future Law of Armed Conflict* (2022).

[16] Hans Kelsen, 'Collective Security and Collective Self-Defense Under the Charter of the United Nations' in *The Use of Force in International Law* (1st edn, Routledge 2012) 785.

[17] Dinstein (n 13) 190; 236.

[18] Dinstein (n 13) 238.

[19] CG Fenwick, 'Is Neutrality Still a Term of Present Law?' (1969) 63 American Journal of International Law 100, 102.

On the basis of those observations, I hold the view that, at least for the scope of this paper, neutrality law has effectively been replaced by the system of collective security. As such, the relationship between non-participating States and belligerent States is regulated by the general rules of international law, as well as the rules of collective security.

B. Due Diligence Obligations on Non-participating States in Cyberspace

The legal norm regulating the obligations of States towards other States in relation to the harmful acts of individuals on their territory is the due diligence principle. Since the International Court of Justice (ICJ) ruled in the *Corfu Channel* case that every State has an obligation not to knowingly allow its territory to be used for acts contrary to the rights of other States, a general due diligence principle has been widely accepted in international law.[20] This section explores the due diligence obligations of non-participating States in relation to cyber operations conducted from their territory in the context of an IAC.

While the principle of due diligence has mainly played a role in environmental law, cyberspace constitutes a new field of potential application since hostile cyber operations often cross borders and often emanate from non-State actors.[21] The principle obliges States not to knowingly allow their territory to be used for internationally wrongful acts. The legally binding nature of the principle has been debated in the context of cyberspace since the 2015 Group of Governmental Experts caused confusion by holding that States *should* not knowingly allow their territory to be used for internationally wrongful acts, thereby indicating that such conduct is voluntary.[22] Some States have expressed similar views.[23] However, considering the general applicability of international law in cyberspace, and the fact that the due diligence principle is a general rule of international law, it is my view that a legally binding principle of due diligence applies in cyberspace independent of any custom specific to cyberspace.[24]

[20] *Corfu Channel case, Judgment of April 9th, 1949* (ICJ) 22.

[21] Michael N Schmitt, 'Below the Threshold Cyber Operations: The Countermeasures Response Option and International Law' (2013) 54 Virginia Journal of International Law 697, 706.

[22] GGE, 'Report of the Group of Governmental Experts on Developments in the Field of Information and Telecommunications in the Context of International Security' (United Nations 2015) A/70/174.

[23] New Zealand, 'The Application of International Law to State Activity in Cyberspace' (*Department of The Prime Minister and Cabinet*) <https://dpmc.govt.nz/publications/application-international-law-state-activity-cyberspace> accessed 5 January 2023; UK, 'Application of International Law to States' Conduct in Cyberspace: UK Statement' <www.gov.uk/government/publications/application-of-international-law-to-states-conduct-in-cyberspace-uk-statement/application-of-international-law-to-states-conduct-in-cyberspace-uk-statement> accessed 5 January 2023; GGE, 'Official Compendium of Voluntary National Contributions on the Subject of How International Law Applies to the Use of Information and Communications Technologies by States Submitted by Participating Governmental Experts in the Group of Governmental Experts on Advancing Responsible State Behaviour in Cyberspace in the Context of International Security Established Pursuant to General Assembly Resolution' (United Nations 2021) 141 (contribution of the United States).

[24] Dapo Akande, Antonio Coco, and Talita de Souza Dias, 'Drawing the Cyber Baseline: The Applicability of Existing International Law to the Governance of Information and Communication Technologies' (2022) 99 International Law Studies <https://digital-commons.usnwc.edu/ils/vol99/iss1/2> accessed 17 April 2023.

According to the *Tallinn Manual 2.0 on the International Law Applicable to Cyber Operations*, currently the most significant academic work on the international legal regulation of cyberspace, the due diligence obligation embraces all cyber operations that are 'contrary to the rights' of the affected State and have 'serious adverse consequences'.[25] These cumulative requirements are examined separately in the following section.

Serious Adverse Consequences

The term 'serious adverse consequences' implies that a certain threshold of harm must be reached to trigger a due diligence obligation. The precise threshold remains unsettled.[26] The *Tallinn Manual 2.0* holds that merely affecting the interests of the target State – for example, by causing inconvenience, minor disruption, or negligible expense – is insufficient.[27] States have taken different approaches to the question. Some speak of activities causing *harm* to other States or affecting them adversely, thereby setting a remarkably low threshold.[28] However, of these States, the majority do not consider the due diligence principle legally binding.[29] As such, only a few States support a legally binding principle applying at a low threshold.[30] A few States raise the threshold to 'significant harm'.[31] The most widespread standard, however, is that of 'serious adverse consequences', which is also applied in the *Tallinn Manual 2.0*.[32] Japan holds that damage to critical infrastructure would be an example of such serious adverse consequences.[33] The Netherlands holds that the consequences do not necessarily have to include physical damage.[34]

[25] Michael N Schmitt, *Tallinn Manual 2.0 on the International Law Applicable to Cyber Operations* (2nd edn, Cambridge University Press 2017) r 6.

[26] ibid.

[27] Schmitt (n 25) r 6, para 26.

[28] Richard Kadlčák, 'Statement by Czech Republic' (2nd substantive meeting of the OEWG, UN General Assembly, 11 February 2020); Roy Schondorf, 'Israel's Perspective on Key Legal and Practical Issues Concerning the Application of International Law to Cyber Operations' (2021) 97 International Law Studies <https://digital-commons.usnwc.edu/ils/vol97/iss1/21> accessed 17 April 2023; UK (n 23); GGE, 'Official Compendium' (n 23) 9 (contribution of Australia), 26 (contribution of Estonia), 141 (contribution of the United States).

[29] Schondorf (n 28); UK (n 23); GGE, 'Official Compendium' (n 23) 141 (contribution of the United States), 9 (contribution of Australia).

[30] Kadlčák (n 28); GGE, 'Official Compendium' (n 23) 26 (contribution of Estonia).

[31] Italy, 'Italian Position Paper on "International Law in Cyberspace"' <www.esteri.it/mae/resource/doc/2021/11/italian_position_paper_on_international_law_and_cyberspace.pdf>; Canada, 'International Law Applicable in Cyberspace' (GAC, 21 February 2017) <www.international.gc.ca/world-monde/issues_development-enjeux_developpement/peace_security-paix_securite/cyberspace_law-cyberespace_droit.aspx?lang=eng> accessed 5 January 2023.

[32] Government of the Kingdom of the Netherlands, 'Appendix: International Law in Cyberspace' <https://www.government.nl/binaries/government/documenten/parliamentary-documents/2019/09/26/letter-to-the-parliament-on-the-international-legal-order-in-cyberspace/international-law-in-the-cyberdomain-netherlands.pdf> accessed 13 April 2023; Ministry of Foreign Affairs of Japan, 'Basic Position of the Government of Japan on International Law Applicable to Cyber Operations'; GGE, 'Official Compendium' (n 23) 71–72 (contribution of Norway), 76 (contribution of Romania); Government Offices of Sweden, 'Position Paper on the Application of International Law in Cyberspace' <www.government.se/contentassets/3c2cb6febd0e4ab0bd542f653283b140/swedens-position-paper-on-the-application-of-international-law-in-cyberspace.pdf>accessed 13 April 2023.

[33] Ministry of Foreign Affairs of Japan (n 32).

[34] Government of the Kingdom of the Netherlands (n 32).

Applied to the activities of the IT Army of Ukraine, some of the outcomes plausibly constitute serious adverse consequences. For example, the operations against Gazprom allegedly prevented millions of users from carrying out any financial transactions.[35] Like all operations of the IT Army, those operations were conducted with the participation of individuals located in multiple States. Thus, a topical question is how to assess operations with cumulatively generating serious adverse consequences without the activities emanating from the individual State's territory reaching this threshold.[36] The Independent Group of Experts in the *Tallinn Manual 2.0* was divided on the question,[37] and no States have addressed it. Although the question therefore remains unsettled, I hold that the due diligence obligation is implicated even if the particular activity emanating from each State did not in itself generate sufficiently severe consequences. In that regard, it should be reiterated that the due diligence obligation is an obligation of conduct, not result; that the obligation requires knowledge; and that States are only obligated to do what is reasonable.

Contrary to the Rights of a State

The term 'contrary to the rights of a State' refers to cyber operations that breach an international obligation towards the target State.[38] Since international obligations are primarily imposed on States, operations by individuals will rarely breach them. Therefore, the determination of which cyber operations are contrary to the rights of a State requires a distinction between operations that are attributable to a State and operations of individuals.

Cyber operations attributable to a State and amounting to use of force as prohibited by Article 2(4) UNC are *prima facie* contrary to the rights of the target State. However, Article 51 UNC preserves a State's 'inherent right' to self-defence in the face of an armed attack, thus modifying the prohibition of the use of force.[39] A State using force in self-defence is not, even potentially, in breach of Article 2(4), and the act is not contrary to the rights of the target State.[40] Thus, it triggers no obligation of due diligence for the territorial State.

[35] IT Army of Ukraine, tweet, 6 September 2022, <https://twitter.com/ITArmyUKR/status/1567173639706972160> accessed 13 April 2023.

[36] Schmitt (n 25) r 6, para 29.

[37] ibid r 6, para 30.

[38] ibid r 6, para 15; Government of the Kingdom of the Netherlands (n 32); GGE, 'Official Compendium' (n 23) 71 (contribution of Norway), 76 (contribution of Romania); Italy (n 31); Government Offices of Sweden (n 32).

[39] ILC, 'Draft Articles on Responsibility of States for Internationally Wrongful Acts, with Commentaries' art 21.

[40] ILC (n 39) art 21; Nicholas Tsagourias, 'Self-Defence against Non-State Actors: The Interaction between Self-Defence as a Primary Rule and Self-Defence as a Secondary Rule' (2016) 29 Leiden Journal of International Law 801, 804.

A non-forcible cyber operation is contrary to the rights of a State if it violates, *inter alia*, the principle of non-intervention or the rule of sovereignty.[41] According to the mainstream approach, primary international law contains no exceptions to these principles. However, as the majority of cyber operations are non-forcible, the question arises of whether Article 51 may also constitute the legal basis for cyber operations contrary to international obligations other than the obligation not to use force. The question has remained relatively unexplored in the legal literature. In a recent article, Buchan claims that self-defence can be invoked to justify all measures necessary to repulse an armed attack, whether forcible or non-forcible. He bases his argument on an examination of the origins of the right of self-defence under customary law, the text of Article 51 UNC, the structure of the UNC, and State practice.[42] This broad approach to self-defence, he argues, enhances the effectiveness of the right of self-defence by broadening the available response options and helps prevent unnecessary escalations.[43] Indeed, in a system founded on a commitment to maintaining international peace[44] and a prohibition on the use of force, it appears illogical to allow the use of force in self-defence while prohibiting other less grave *prima facie* violations of international law. Such a distinction would also complicate the proportionality assessment when less intrusive measures are deemed sufficient.

However, there are still weighty reasons to challenge this broad concept of self-defence. Tsagourias argues that if self-defence as a circumstance precluding wrongfulness were to apply to any violation of international law, its scope would become so broad that it could potentially destabilize the international legal order.[45] In legal scholarship, self-defence has often been implicitly presumed to refer to forcible acts. This could be simply because force is often deemed necessary to counter an armed attack,[46] and consequently, the main importance of the concept of self-defence has been its character as an exception to that prohibition.[47] D.W. Bowett suggests that the mere fact that an act *prima facie* breaches an obligation other than the prohibition of the use of force does not mean that such measures ought to be denied the term 'self-defence'.[48] Thus, while acknowledging the validity of a broad concept of self-defence, theoretically also encompassing non-forcible measures, he holds that such measures will not normally be properly characterized as self-defence due to their invariably retaliatory character, since the whole purpose of self-defence must be the protection of the very rights that are endangered.[49]

[41] Notably, the nature of the principle of sovereignty has been subject to debate, see Kevin Jon Heller, 'In Defense of Pure Sovereignty in Cyberspace' (2021) 97 International Law Studies <https://digital-commons.usnwc.edu/ils/vol97/iss1/50> accessed 13 April 2023.

[42] Russell Buchan, 'Non-Forcible Measures and the Law of Self-Defence' (2023) 72 International and Comparative Law Quarterly 1, 7.

[43] ibid 2.

[44] Preamble of the United Nations Charter.

[45] Tsagourias (n 40) 820.

[46] MA Weightman, 'Self-Defense in International Law' (1951) 37 Virginia Law Review 1095, 1101.

[47] DW Bowett, *Self-Defence in International Law* (The Lawbook Exchange, Ltd 2009) 22.

[48] ibid 23.

[49] ibid.

However, it could also be that use of force is part of the very definition of self-defence. Robert Ago has pronounced that acting in self-defence means responding by force to forcible wrongful action carried out by another.[50] Kelsen has defined self-defence as 'the use of force by a person illegally attacked by another'.[51] A similar approach is reflected in the jurisprudence of the ICJ, albeit more subtly. By only considering self-defence in the context of the use of force, while resorting to the regime of countermeasures for non-forcible activities, the ICJ perhaps shows rather than tells that violations of rules other than the prohibition on the use of force fall outside the scope of Article 51.[52]

Another institution applying a narrow concept of self-defence is the ILC. In addition to the inherent right to self-defence in Article 51 UNC, ARSIWA Article 21 provides that the wrongfulness of an act of a State is precluded if the act constitutes a lawful measure of self-defence taken in conformity with the UNC. According to the ILC, the implication is that 'self-defence may justify non-performance of certain obligations other than that under (Article 2(4) UNC), provided that such non-performance is related to the breach of that provision'.[53] That would be the case, *inter alia*, with violations of the territorial sovereignty emanating from a forcible operation. Accordingly, the ILC adopts the view that the primary rule on self-defence in Article 51 only concerns forcible acts; violations of international law inevitably following from the use of force in self-defence are justifiable under Article 51, let alone violations of international law with no relation at all to the use of force, *inter alia*, malicious cyber operations.

As cyber operations often fall below the use of force threshold, the question of the scope of acts justifiable as self-defence is highly relevant and calls for further analysis beyond the scope of what is feasible in this brief paper. Rather than attempting to provide a definitive answer here, I will emphasize the significance of the question, particularly in the context of cyberspace. In the following, I will apply a narrow concept of self-defence, which, after all, appears to be the mainstream approach.

A narrow concept of self-defence inevitably results in the vast majority of the operations of the IT Army falling outside the scope of actions justifiable under Article 51. Instead, the general rules of State responsibility may preclude the wrongfulness – for example, if the operations are countermeasures. The ILC has stated that the underlying obligation is not thereby terminated or suspended.[54] Questions remain as to the legal consequences of the qualification of an operation as a countermeasure: Do countermeasures constitute *justified* conduct or merely *excused* conduct? In the

[50] United Nations, *Yearbook of the International Law Commission 1980, Vol. II, Part 1* (United Nations 1980) <www.un-ilibrary.org/content/books/9789213623381> accessed 7 January 2023.

[51] Kelsen (n 16) 784.

[52] *Case Concerning Military and Paramilitary Activities in and against Nicaragua* (ICJ, Judgment on the Merits) [201].

[53] ILC (n 39) art 21, para 2.

[54] ILC (n 39) art 22.

context of due diligence, that question determines whether an act contrary to the rights of a State – triggering a due diligence obligation for the territorial State – exists despite the operation constituting a countermeasure.[55]

Scholars have different views on the question.[56] According to one, a circumstance precluding wrongfulness *justifies* the conduct, thus providing permission to engage in the conduct. Conduct adopted in accordance with a justification is, therefore, lawful.[57] Applied to the scenario of a cyber operation attributable to Ukraine against Russia from the territory of a third State, the operation is justified if it constitutes a countermeasure. Thereby, the operation is not contrary to Russia's rights, and already for that reason, the third State has no due diligence obligation.

According to another view, a circumstance precluding wrongfulness merely *excuses* the conduct, meaning that the conduct remains illegal, but the consequences otherwise following from the illegality of the conduct are excluded.[58] This view implies that the function of a circumstance precluding wrongfulness is to relieve States from responsibility rather than to modify the substantial obligation.[59] Applying this view to the same scenario, the conduct remains illegal, and the third State carries a due diligence obligation towards Russia. As such, how one understands the effect of circumstances precluding wrongfulness determines the existence of a due diligence obligation.

In conclusion, while the complicated and unsettled question of the nature of countermeasures is beyond the scope of this paper, considering it in light of the question of due diligence adds a novel dimension to the discussion. The due diligence obligation is contingent upon the existence of an act contrary to the rights of a State. As a result, the nuance between justification and excuse is crucial in determining the existence of a due diligence obligation for States whose territories are used for acts constituting countermeasures. Therefore, to better comprehend the practical dimensions of the legal nature of countermeasures, future discussions on the subject could benefit from consulting the concept of acts 'contrary to the rights of a State' in the context of due diligence.

Cyber operations not attributable to a State are unlikely to breach an international obligation because international law imposes no such obligations on the individual. Instead, the due diligence implications depend on whether the operation would have

[55] Federica Paddeu, 'Clarifying the Concept of Circumstances Precluding Wrongfulness (Justifications) in International Law' in Lorand Bartels and Federica Paddeu (eds), *Exceptions in International Law* (Oxford University Press 2020) 205 <https://doi.org/10.1093/oso/9780198789321.003.0011> accessed 7 January 2023.

[56] Paddeu (n 55); Tsagourias (n 40); Buchan (n 42).

[57] Paddeu (n 55) 222.

[58] ibid 212.

[59] Tsagourias (n 40) 820.

been unlawful, had it been conducted by the territorial State.[60] Because of the absence of a special legal relationship between the territorial State and the target State, hostile activity is more likely to be unlawful in this scenario. Forcible operations may, in principle, constitute lawful collective self-defence. However, some monitoring from the territorial State is presumably necessary to ensure that the substantial requirements for self-defence are met. The procedural requirement of Article 51 regarding the immediate report of measures taken to the UNSC is also relevant to the assessment. The absence of such a report does not alone exclude an act from the scope of Article 51, but the ICJ has interpreted it to be one factor indicating whether the State in question was itself convinced that it was acting in self-defence.[61] Particularly in the absence of a UNSC report, an omission is unlikely to be accepted as collective self-defence.

In conclusion, the determination of whether an operation is contrary to the rights of the target State initially requires an assessment of the operation's attributability to another State. The legality of an operation attributable to a State must be assessed from the perspective of the responsible State. Operations constituting lawful self-defence trigger no due diligence obligation. The due diligence implications of countermeasures depend on whether they are considered *justified* or merely *excused*, which remains unsettled in international law. The legality of operations not attributable to a State must be assessed from the perspective of the territorial State. In principle, they may constitute lawful, collective self-defence. However, this requires the territorial State to ensure a certain level of monitoring and to report the activities to the UNSC.

4. CIRCUMSTANCES PRECLUDING THE WRONGFULNESS OF ABSTAINING FROM EXERCISING DUE DILIGENCE

Section 3 examined when cyber operations related to an IAC but conducted from the territory of a non-participating State trigger a due diligence obligation. This section concerns the situation where a State neglects an established due diligence obligation and the possibility of invoking a circumstance precluding wrongfulness as prescribed in ARSIWA Chapter V. As mentioned in Section 3, to the extent that an operation constitutes self-defence in accordance with Article 51, the operation is not even potentially unlawful. ARSIWA Article 21 further precludes the wrongfulness of an act of a State if the act constitutes a lawful measure of self-defence taken in conformity with the UNC. However, the provision only relates to violations of international law occurring in relation to the use of force in self-defence. Already for that reason, cyber operations below the level of use of force fall outside the scope of the provision.

[60] Schmitt (n 25) r 6, para 21.

[61] *Nicaragua v United States of America* (n 52) para 200.

Therefore, the relevant circumstance to potentially preclude the wrongfulness is Article 22 (countermeasures).

Article 22 provides that the wrongfulness of an act of a State not in conformity with an international obligation towards another State is precluded if it constitutes a countermeasure taken against the latter State in accordance with the specific rules and limitations governing the use of countermeasures. The regime of countermeasures is relevant for non-forcible response operations.[62] In contrast to self-defence, countermeasures may be taken in response to activities below the threshold of an armed attack. The relevant question here is whether the wrongfulness of neglecting a due diligence obligation may be precluded as a countermeasure taken in response to the target State's aggression against another State – in other words, whether to accept collective countermeasures. An earlier draft of ARSIWA accepting collective countermeasures in response to violations of *erga omnes* obligations was met with reluctance from States. Caught between the risk of abuse emphasized by reluctant States and the need for effective protection, the ILC ultimately decided not to decide.[63] Thus, the final draft left the question open in Article 54.[64]

The *Tallinn Manual 2.0* was divided on the question,[65] and only a few States have publicly addressed it in the context of cyberspace. Estonia was a pioneer in furthering the position, back in 2019, that States not directly injured may apply countermeasures to support the injured State.[66] In their analysis from 2020, Kjeldgaard-Pedersen and Schack argue that there may be State practice supporting a positive view on collective countermeasures.[67] Since then, more States have declared an openness to the proposition,[68] while other States remain sceptical of the idea.[69] The question remains contentious, and clarification will depend on further State practice.

[62] *Nicaragua v United States of America* (n 52) para 201.

[63] Christian J Tams, *Enforcing Obligations Erga Omnes in International Law* (Cambridge Studies in International and Comparative Law 44, Cambridge University Press 2005) 200.

[64] ibid.

[65] Schmitt (n 25) r 24, para 7.

[66] Kersti Kaljulaid, 'President of the Republic of Estonia at the Opening of CyCon 2019' (*Perma.cc*, 29 May 2019) <https://perma.cc/9F9M-EUYG?type=standard> accessed 13 April 2023.

[67] Marc Schack and Astrid Kjeldgaard-Pedersen, *Modforanstaltninger i cyberdomænet: Den folkeretlige ramme* (Københavns Universitet, Det Juridiske Fakultet 2020) <https://research.fak.dk/esploro/outputs/report/Modforanstaltninger-i-cyberdomnet-Den-folkeretlige-ramme/991815902603741> accessed 4 January 2023.

[68] New Zealand (n 23); Suella Braverman, 'International Law in Future Frontiers' (Chatham House, 19 May 2022) <https://chathamhouse.soutron.net/Portal/Public/en-GB/RecordView/Index/191224 > accessed 13 April 2023.

[69] Ministry of Defence of France, 'International Law Applied to Operations in Cyberspace' (UNODA) <https://documents.unoda.org/wp-content/uploads/2021/12/French-position-on-international-law-applied-to-cyberspace.pdf> accessed 17 April 2023; Canada (n 31).

5. RETURNING TO REALITY: POSSIBLE IMPLICATIONS FOR NON-PARTICIPATING STATES HOSTING MEMBERS OF THE IT ARMY OF UKRAINE

Thousands of individuals located in non-participating States have participated in hostile cyber activities against Russia orchestrated by the IT Army of Ukraine. Most States appear to prefer to refrain from taking action to prevent individuals on their territory from engaging in such operations, triggering the question of the possible obligations of the territorial States. In Section 2, four possible scenarios were identified. This final section concludes by returning to those four possible scenarios and examining how international law applies to each. It is assumed in the following that the operations have serious adverse consequences.

First, for forcible operations attributable to Ukraine, the acts may constitute lawful self-defence. Those acts are not internationally wrongful and are not contrary to the rights of Russia, and no due diligence obligation is implicated.

Second, for operations attributable to Ukraine that are non-forcible but violate other obligations, the operations may constitute countermeasures. In this scenario, the existence of a due diligence obligation depends on several contentious legal questions. The first question is whether the right to self-defence in Article 51 authorizes operations that are non-forcible but are *prima facie* in violation of other international obligations. I follow the mainstream approach in taking the view that non-forcible operations fall outside the scope of measures constituting lawful self-defence. However, I also emphasize that this important question requires further research. A non-forcible operation may, instead, constitute a countermeasure. Consequently, the second question is whether countermeasures are perceived as justifications or merely excuses. In the first view, the operations are not contrary to Russia's rights and trigger no due diligence obligation. In the latter view, a due diligence obligation is triggered, and refraining from exercising due diligence constitutes an internationally wrongful act. The wrongfulness of refraining from exercising due diligence may, then, be precluded if accepting the concept of collective countermeasures.

Third, for operations non-attributable to Ukraine, the wrongfulness of the acts must, instead, be assessed from the perspective of the territorial State. If the operation would be forcible if conducted by the territorial State, it could theoretically constitute lawful, collective self-defence. However, passively allowing individuals to use force against Russia without any monitoring and without any report to the UNSC is unlikely to be accepted as collective self-defence.

Fourth, for operations non-attributable to Ukraine that are non-forcible but would have violated other obligations if conducted by the territorial State, the lawfulness first depends on whether one applies a broad or a narrow concept of self-defence. However, while the likeliness of passively allowing individuals to use force against Russia to be accepted as self-defence is already low (third scenario), it is even more so in the case of non-forcible measures. Instead, the wrongfulness of refraining from exercising due diligence may be precluded as a countermeasure if accepting the contentious concept of collective countermeasures.

In conclusion, this brief article demonstrates that when international law is applied to cyber operations orchestrated by the IT Army of Ukraine, the obligations of non-participating States whose territories are being used depend on several highly contentious legal concepts. Further clarity ultimately requires that more States express their views, since States remain the main actors in shaping the cyber-specific content of international legal norms.

2023 15th International Conference on Cyber Conflict
Meeting Reality
T. Jančárková, D. Giovannelli, K. Podiņš, I. Winther
2023 © NATO CCDCOE Publications, Tallinn

Privatized Frontlines: Private-Sector Contributions in Armed Conflict

Tsvetelina J. van Benthem
University of Oxford
Oxford, United Kingdom
tsvetelina.vanbenthem@law.ox.ac.uk

Abstract: Technology companies have ramped up their support for Ukraine since Russia's full-scale invasion in February 2022. These companies interact with the Ukrainian authorities and infrastructure in a variety of ways, scanning for vulnerabilities in networks and issuing patches, ensuring internet access and providing threat intelligence. Through their contributions, these actors assist the military effort of a party to the conflict. What is the impact of these contributions on the status and protection of private sector employees and company infrastructure under the law of armed conflict? This article analyses the concept of direct participation in hostilities and the definition of military objectives and finds that some current contributions may come close to meeting the direct participation test for persons, and the definition of military objectives for objects. This, in turn, may expose persons and assets of technology firms to the risk of harm – a risk of which they may not be fully aware. Because of this risk, states are under an obligation to inform individuals under their jurisdiction of the legal qualification of their conduct, and of the legal implications of such qualification.

Keywords: *direct participation in hostilities, international human rights law, law of armed conflict, military objectives, private sector, right to information*

1. INTRODUCTION

An intricate web of contributions often underlies the military efforts of parties to a conflict. Political and military leaders take strategic decisions, arms producers supply means of warfare, and a complex machinery of organized entities and individuals creates, spreads and amplifies information campaigns. In today's highly digitized societies, the success of a military effort increasingly hinges on, among others, the security of networks supporting critical infrastructure and IT supply chains, the availability of essential digital services, and the provision of information on the location of military targets through apps and social media. The skillsets required for assisting parties to conflict on the digital front can often be found in the private sector, and the private sector finds itself increasingly drawn into situations of armed conflict.[1]

In the aftermath of Russia's full-scale invasion of Ukraine in February 2022, the Ukrainian Minister of Digital Transformation Mykhailo Fedorov used Twitter to appeal to Elon Musk to provide Ukraine with Starlink stations. Starlink is a system of satellites operated by SpaceX that provides off-grid high-bandwidth internet access.[2] Musk agreed to assist, and secured satellite internet service to the Ukrainian military and civilian authorities – a contribution without which, according to experts, 'the Ukrainian army would not have resisted the Russian onslaught, at least not as well'.[3] SpaceX was one of many private sector actors that heeded the call for contributions. Microsoft has been particularly vocal about its close collaboration with the Ukrainian authorities. According to Tom Burt, Microsoft's corporate vice president,

> Microsoft security teams have worked closely with Ukrainian government officials and cybersecurity staff at government organizations and private enterprises to identify and remediate threat activity against Ukrainian networks. [...] we established a secure line of communication with key cyber officials in Ukraine [...] This has included 24/7 sharing of threat intelligence and deployment of technical countermeasures to defeat the observed malware.[4]

[1] Nat Rubio-Licht et al., '*The war in Ukraine is putting tech – from companies to governments – to the test*' (*Protocol*, 1 March 2022) <https://www.protocol.com/policy/russia-ukraine-war-tech> accessed 17 April 2023.

[2] '*How Elon Musk's satellites have saved Ukraine and changed warfare*' (*The Economist*, 5 January 2023) <https://www.economist.com/briefing/2023/01/05/how-elon-musks-satellites-have-saved-ukraine-and-changed-warfare> accessed 17 April 2023.

[3] Elise Vincent, Alexandre Piquard and Cédric Pietralunga, '*Comment Starlink et les constellations de satellites d'Elon Musk changent la guerre*' (*Le Monde*, 15 December 2022) <https://www.lemonde.fr/economie/article/2022/12/15/starlink-et-les-constellations-de-satellites-nouvel-enjeu-militaire_6154463_3234.html> accessed 17 April 2023.

[4] Tom Burt, 'The hybrid war in Ukraine' (*Microsoft On The Issues*, 27 April 2022) <https://blogs.microsoft.com/on-the-issues/2022/04/27/hybrid-war-ukraine-russia-cyberattacks/> accessed 17 April 2023.

Google and Amazon have been active in providing Ukraine with cybersecurity support;[5] MDA, a Canadian intelligence firm, received approval to send satellite imagery of Russian troop movements to Ukrainian authorities;[6] and the Scotland-based Trustify secured Ukrainian government web domains.[7] The list goes on.[8] Private sector actors scan Ukrainian networks for vulnerabilities, issue security patches, share threat intelligence, and provide internet access. Through their contributions, they became instrumental in the war effort. Oleksii Vyskub, Ukraine's Deputy Minister of Digital Transformation, in thanking Trustify for its assistance, stated that he truly believes Trustify's support 'is an important contribution to our future victory over Russia, the victory of people of goodwill over evil'.[9]

This impulse to fight the good fight has been particularly prominent in the Russia-Ukraine conflict. Because the identification of 'victims' and 'aggressors', of 'right' and 'wrong' seems so straightforward, private sector technology companies see an opportunity to position themselves on the right side of history. Their assistance to Ukraine is not risk-free, however. That foreign support from both individuals and organizations has enhanced Ukraine's capacity to resist the aggression has not escaped the attention of the Russian authorities. In April 2022, the Russian Foreign Ministry published a statement condemning the wave of cyber operations against Russian websites mounted by 'international hackers'.[10] According to this statement, 'whoever sows the cyberwind will reap the cyberstorm'.[11] And more recently, Microsoft attributed the Prestige ransomware operation targeting Ukrainian and Polish transportation and logistics organizations to hackers with close ties to the Russian military.[12] The current conflict in Eastern Europe thus attests to both the varied nature of contributions by private sector actors and the tangible risks to which such contributions can give rise.

To what extent do private sector entities assisting the war effort of a party to conflict open themselves to forcible action by the other side? Put differently, which forms of

[5] Diya Li, 'On the Digital Front Lines: How Tech Companies are Supporting Ukraine' (*US Chamber of Commerce*, 29 March 2022) <https://www.uschamber.com/technology/on-the-digital-front-lines-how-tech-companies-are-supporting-ukraine> accessed 17 April 2023.

[6] Abishur Prakash, 'How Technology Companies Are Shaping the Ukraine Conflict' (*Scientific American*, 28 October 2022) <https://www.scientificamerican.com/article/how-technology-companies-are-shaping-the-ukraine-conflict/> accessed 17 April 2023.

[7] Ross Kelly, 'Edinburgh-headquartered Firm Praised for Ukraine Cyber Support' (*Digit News*, 24 October 2022) <https://www.digit.fyi/trustify-ukraine-support/> accessed 17 April 2023.

[8] Dina Temple-Raston, 'Rounding up a cyber posse for Ukraine' (*The Record*, 18 November 2022) <https://therecord.media/exclusive-rounding-up-a-cyber-posse-for-ukraine/> accessed 17 April 2023.

[9] Ross Kelly, 'Edinburgh-headquartered Firm Praised for Ukraine Cyber Support' (*Digit News*, 24 October 2022) <https://www.digit.fyi/trustify-ukraine-support/> accessed 17 April 2023.

[10] *Комментарий специального представителя Президента Российской Федерации по вопросам международного сотрудничества в области информационной безопасности, и.о. директора ДМИБ МИД России А.В.Крутских в связи с ростом числа хакерских нападений на Россию*, 14 April 2022, at: <https://embassylife.ru/post/5652> accessed 17 April 2023.

[11] ibid.

[12] Sean Lyngaas, 'Microsoft blames Russian military-linked hackers for ransomware attacks in Poland and Ukraine' (*CNN*, 14 November 2022) <https://edition.cnn.com/2022/11/10/politics/microsoft-russian-linked-hackers-poland-ukraine/index.html> accessed 17 April 2023. To be clear, this was a technical attribution to a particular group, not legal attribution towards the state.

contribution to the war effort provide a legal basis under the law of armed conflict for forcible action in response? The following section examines the legal consequences of private sector participation under the law of armed conflict *for individual employees* under the doctrine of direct participation in hostilities (DPH) and *for objects* under the definition of military objectives. It will be shown that the legal boundaries of DPH remain blurred and that the application of the definition of military objectives may pose difficulties in the context of private-sector digital providers. Following this analysis, the article will turn to the particular risks run by private sector actors through their assistance and the role of other rules of international law, such as the rules of the *jus ad bellum* regime, in constraining the possibility of forcible responses. Finally, the piece will analyse a positive obligation of states derived from international human rights law (IHRL) to inform those under their jurisdiction of the risks of engaging in conflict-related assistance. Because of the interpretative uncertainties around concepts such as DPH, and because of a lack of information campaigns on the legal implications of assistance, many actors may find themselves contributing to a war effort without a real understanding of the threat of harm that such contributions may entail. It may very well be that, on balance, the threat of harm will be deemed less significant than the incentive to participate. Be that as it may, the balancing exercise must be an informed one.

Three recommendations underlie the analysis in the article.

1. States must continue to clarify the relevant rules of international law through their national positions and statements in inter-governmental fora. While legal uncertainty may in some instances give a reason for pause and caution, it may equally be a significant driver of reduced restraint, and thereby of civilian harm.
2. The clarification exercise must be undertaken in a principled way, rather than on an *ad hoc* basis depending on the particular threat faced by a given state or group of states.
3. The private sector and the general population must be informed of the scope of rules relevant to their protection, and of the implications of loss of protection. Thus, states must be proactive in informing those under their jurisdiction of the risks inherent in conflict-related participation.

2. CHARACTERIZING TECHNOLOGY-SECTOR CONTRIBUTIONS UNDER THE LAW OF ARMED CONFLICT: DIRECT PARTICIPATION IN HOSTILITIES AND THE DEFINITION OF MILITARY OBJECTIVES

Under the law of armed conflict, determining the status of persons and objects is key. According to the principle of distinction, the basic rule of the conduct of hostilities regime is that 'the Parties to the conflict shall at all times distinguish between the civilian population and combatants and between civilian objects and military objectives and accordingly shall direct their operations only against military objectives'.[13] One's status determines the parameters of protection and exposure to risk.

Civilians receive a wide range of protections from military operations and their effects. Civilians must not be the object of attack.[14] Even when attacks are launched against military objectives, excessive incidental civilian harm can render such attacks unlawful.[15] And parties to conflict should take precautions in attack and against the effects of attacks to minimize harm to civilians.[16] These protections are enjoyed by civilians unless and for such time as they take a direct part in hostilities.[17] Like combatants, civilians who directly participate in hostilities can be lawfully targeted. Unlike combatants, they can only be targeted within the limited time frame of their direct participation. And, unlike combatants, they are not covered by combatant immunity and can be prosecuted for their acts of participation where domestic law criminalizes such participation.[18] Civilian objects also benefit from protection under the law of armed conflict.[19] Only military objectives can be the object of attack. Being clear on the boundaries of different categories is crucial to operationalizing the principle of distinction, which lies at the heart of the legal regime.

A. Direct Participation in Hostilities

In the past year, technology companies provided various forms of assistance to the Ukrainian authorities. This raises an important question about the impact of these acts of assistance on the protection of such persons from attack.

While it is perfectly possible for a state to incorporate employees or teams from the private sector into its armed forces, this is not the dynamic that emerged between persons working in technology companies and Ukraine. Similarly, they do not bear

[13] Protocol Additional to the Geneva Conventions of 12 August 1949, and relating to the Protection of Victims of International Armed Conflicts ('AP I'), 8 June 1977, art 48.

[14] ibid, art 51(2).

[15] ibid, art 51(5)(b).

[16] ibid, arts 57 and 58.

[17] ibid, art 51(3).

[18] Zhixiong Huang and Yaohui Ying, 'The application of the principle of distinction in the cyber context: A Chinese perspective' (2020) 102(913) IRRC 335, 350.

[19] AP I, art 52.

the hallmarks of groups belonging to a party to conflict.[20] The employees qualify as civilians under international humanitarian law (IHL).

Civilians enjoy protection from attack unless and for such time as they take a direct part in hostilities.[21] Thus, a civilian taking a direct part in hostilities loses their immunity from attack. The standard for loss of protection through DPH remains the same irrespective of whether one is examining the conduct of government employees, employees of private sector companies, or individuals acting on their own outside any institutional structure. The war effort is comprised of a multitude of contributions, and these contributions vary greatly in form and origin.[22] This is why, to fully ensure civilian protection, the standards for assessing contributions must be defined with sufficient clarity. Yet the contours of DPH remain blurred in ways that are particularly significant to today's forms of assistance. What is clear as a starting point, however, is that the doctrine applies to contributions carried out through information and communications technologies (ICTs).[23]

What, then, is the meaning of direct participation in hostilities? Clearly, lines must be drawn between forms of participation, yet the line-drawing exercise is riddled with challenges.[24] In the context of cyberspace, Brazil noted that one of the issues that deserve further reflection is 'when a civilian acting in the cyberspace might be considered as taking direct part in hostilities'.[25]

In drawing these lines, some states have resorted to the use of illustrative examples. The United Kingdom Manual of the Law of Armed Conflict thus posits that, while 'civilians manning an anti-aircraft gun or engaging in sabotage of military installations' are directly participating in hostilities, 'civilians working in military vehicle maintenance depots or munitions factories or driving military transport vehicles' are not.[26] Between these examples exists a wide spectrum of types of participation, the classification of which poses difficulties.

[20] See, in an analysis on private military and security companies, Nelleke van Amstel and Rain Liivoja, 'Private Military and Security Companies', in Rain Liivoja and Tim McCormack (eds.), *Routledge Handbook of the Law of Armed Conflict* (Routledge 2016), 629.

[21] AP I, art 51(3).

[22] Charles Garraway, 'The Changing Character of the Participants in War: Civilianization of Warfighting and the Concept of "Direct Participation in Hostilities"' (2011) 87 International Law Studies, 178.

[23] This has been made clear by states in their national contributions to inter-governmental fora. In addition, although the ICRC Interpretive Guidance did not take the use of such technologies as its primary focus, it acknowledges that electronic interference with military computer networks may qualify as DPH - Nils Melzer, *Interpretive Guidance on the Notion of Direct Participation in Hostilities under International Humanitarian Law* (ICRC 2009), 48.

[24] Kubo Mačák, 'Unblurring the lines: military cyber operations and international law' (2021) 6(3) Journal of Cyber Policy 411, 419.

[25] Official compendium of voluntary national contributions on the subject of how international law applies to the use of information and communications technologies by States submitted by participating governmental experts in the Group of Governmental Experts on Advancing Responsible State Behaviour in Cyberspace in the Context of International Security established pursuant to General Assembly resolution 73/266, Position of Brazil, p. 23.

[26] HM Government, *JSP 383: The Joint Service Manual of the Law of Armed Conflict* (London 2004), 5.3.3.

In 2009, the International Committee of the Red Cross (ICRC) issued its Interpretive Guidance on the Notion of Direct Participation in Hostilities under International Humanitarian Law. This Guidance offers an analytical toolkit for the examination of forms of participation. While *some* aspects of the Guidance have been met with *some* resistance from *some* quarters,[27] its recommendations on the constitutive elements of DPH were favourably received, with some controversies remaining around the scope of each element. According to the Interpretative Guidance, in order to qualify as DPH, a *specific act* must meet the following cumulative criteria:

> 1. the act must be likely to adversely affect the military operations or military capacity of a party to an armed conflict or, alternatively, to inflict death, injury, or destruction on persons or objects protected against direct attack (threshold of harm), and
> 2. there must be a direct causal link between the act and the harm likely to result either from that act or from a coordinated military operation of which that act constitutes an integral part (direct causation), and
> 3. the act must be specifically designed to directly cause the required threshold of harm in support of a party to the conflict and to the detriment of another (belligerent nexus).

Each of these three elements has given rise to interpretative controversies. Some of these controversies are of particular relevance to the types of contributions provided by technology companies today. For instance, under the Guidance, the threshold of harm element can be met in two ways – one, by 'causing harm of a specifically military nature', and two, 'by inflicting death, injury, or destruction on persons or objects protected against direct attack'.[28] The first way of reaching the threshold requires the unpacking of the ways in which conduct may *adversely affect* the military operations or military capacity of a party to the conflict.[29] Adverse effects on military operations or military capacity extend beyond killing, wounding, or damaging to sabotage, denial of the use of certain objects and equipment, and transmission of tactical targeting information, among others.[30] According to the ICRC, failing to *positively* affect a party to conflict is not to be equated with *adversely* affecting it. Some have considered that this construal of the threshold of harm is overly restrictive. Schmitt, for instance, favours an approach under which the threshold of harm criterion includes acts likely

[27] Michael N Schmitt, 'Deconstructing Direct Participation in Hostilities: The Constitutive Elements' (2010) 42 NYU Journal of International Law and Policy 697; Bill Boothby, 'And for Such Time as: The Time Dimension to Direct Participation in Hostilities' (2010) 42 NYU Journal of International Law and Policy 741; Ido Kilovaty, 'ICRC, NATO and the U.S. - Direct Participation in Hacktivities - Targeting Private Contractors and Civilians in Cyberspace under International Humanitarian Law' (2016 – 2017) 15 Duke Law and Technology Review 1, 10.

[28] ICRC Guidance, p. 47. The Guidance here goes beyond the *ICRC Commentary*, which speaks of 'acts of war which by their nature or purpose are likely to cause actual harm to the personnel and equipment of the enemy armed forces' - in Sandoz, Swinarski and Zimmermann (eds), *Commentary to the Additional Protocols of 8 June 1977 to the Geneva Conventions of 12 August 1949* (ICRC, Geneva 1987), para 1944.

[29] ibid.

[30] ICRC Guidance, p. 48.

to enhance a party's military operations or military capacity.[31] This was also the view of 'some members' of the International Group of Experts working on the Tallinn Manual.[32] This, of course, could open the door wide – indeed too wide. To illustrate, in the conflict in Ukraine technology companies have, without any doubt, enhanced Ukraine's military capacity by monitoring their networks, patching vulnerabilities, running scans for new vulnerabilities, and sharing their infrastructure and expertise. A broad interpretation that includes 'benefit' to a party would move beyond many states' understanding of DPH. Further, although Art. 51(3) of Additional Protocol I does not clarify the meaning of 'direct participation in hostilities', its restrictive wording ('unless and for such time') is worth emphasizing, as it implies that the concept is to be interpreted narrowly.

The second constitutive element, direct causation, does most of the heavy lifting in establishing whether a specific act amounts to DPH. To begin with, as emphasized in the ICRC Commentary to Additional Protocol I, 'there should be a clear distinction between direct participation in hostilities and participation in the war effort'.[33] If one were to construe any type of participation in sustaining the war effort as participation leading to loss of immunity from attack, then, given that the war effort is often a whole-of-population effort, this could turn virtually everyone into a target.[34] While a broad range of activities that are part of the general war effort or are war-sustaining in nature could eventually lead to the types of harm envisaged in the 'threshold of harm' criterion, the element of 'direct causation' is in place to delimit the acts that are sufficiently proximate to that harm from those that are too remote to qualify as DPH. Thus, not every type of involvement in or contribution to hostilities would satisfy the direct participation test.[35] The harm 'must be brought about in one causal step'.[36] Providing services to a party to conflict, general training, or scientific research would thus be seen as 'indirect', as they contribute to the building of or maintenance of capacity, rather than directly bringing about the requisite harm to the adversary.[37] Importantly for technology providers, however, when specific conduct is considered an integral part of a collective operation, such conduct may qualify as DPH even if it would not bring about the harm in and of itself. Not all forms of intelligence transmission would transform a contribution into direct participation, but the transmission of tactical intelligence on a target in the context of a concrete

[31] Michael N Schmitt, 'Deconstructing Direct Participation in Hostilities: The Constitutive Elements' (2010) 42 NYU Journal of International Law and Policy 697, 719.

[32] Michael N Schmitt (ed), *Tallinn manual 2.0 on the international law applicable to cyber operations* ('Tallinn Manual 2.0') (CUP 2017), Rule 97, para 5.

[33] ICRC Commentary, para 1945. *See also* Michelle Lesh, 'Direct Participation in Hostilities', in Rain Liivoja and Tim McCormack, *Routledge Handbook of the Law of Armed Conflict* (Routledge 2016).

[34] Dapo Akande, 'Clearing the Fog of War? The ICRC's Interpretive Guidance on Direct Participation in Hostilities' (2010) 59 ICLQ 180, 188.

[35] ICRC, Second Expert Meeting on the Notion of Direct Participation in Hostilities (Report, 25-26 October 2004), p. 10. *See also* ICTY, *Prosecutor v. Strugar*, Appeals Chamber Judgment, IT-0142-A, , 17 July 2008, paras 176-79.

[36] ICRC Guidance, p. 53.

[37] ibid.

military operation would.[38] Understanding the ways in which involvement in the war effort transitions from indirect to direct is becoming crucial, given the close threat intelligence-sharing links established between technology companies and parties to conflict in recent months.

And finally, the belligerent nexus element, while at first glance less complex in application, raises difficult questions in its relationship to subjective intent. According to the ICRC Guidance, 'belligerent nexus relates to the objective purpose of the act', rather than to an inquiry into the mind of the participant.[39] Admittedly, this understanding of the element avoids difficult questions of proof – and proving intent in those circumstances may indeed be both complex and operationally unrealistic. At the same time, following its logic would allow a party to conflict to consider persons as direct participants regardless of their mental capacity, age, and awareness of their contribution. The ICRC sought to open the door to limited exceptions for cases where 'civilians are totally unaware of their role' and where they are 'completely deprived of their physical freedom of action'.[40] The Tallinn Manual excludes from its scope 'unwitting persons' whose computers are used by someone else.[41] It is unclear, however, how those assessments are to be made, both legally and factually. A recent example exposes the challenges in applying this test. In May 2022, Ukraine's Security Service reported that Russia had developed a smartphone game seeking to attract Ukrainian children and utilize them as 'unwitting spies' in locating Ukrainian positions and infrastructure.[42] The children are not completely deprived of their freedom of action, and it is unclear how it could be determined whether they are 'totally unaware' of their role. There are good reasons to feel a degree of discomfort with the objective test and its implications.[43] This test has particular relevance for employees of technology companies who are, first, receiving assignments from the company's management, and, second, often working as part of large teams and may thus have little knowledge of how their work is being used.

To summarize, while the analytical test for assessing DPH is widely accepted, the steps of this test remain subject to interpretation and debate. It is of note that much of the contemporary discussions on DPH and the criticism of the ICRC Guidance originate from the experience of Western states fighting in asymmetric conflicts.[44] Attempts to

[38] ICRC Guidance, 54 – 55. *See also* position of Germany in Official compendium of voluntary national contributions, Group of Governmental Experts on Advancing Responsible State Behaviour in Cyberspace in the Context of International Security, p. 37.

[39] ICRC Guidance, 59.

[40] ibid, 60.

[41] Tallinn Manual 2.0, Rule 97, para 4.

[42] Tim McMillan, 'Russia Using Smartphone "Game" to Recruit Ukrainian Children as Unwitting Spies' (*The Debrief*, 25 May 2022) <https://thedebrief.org/russia-using-smartphone-game-to-recruit-ukrainian-children-as-unwitting-spies/> accessed 17 April 2023.

[43] Such a discomfort was also raised by Huang and Ying in the context of civilians whose computers are hacked by botnets – *supra* note 18, 354.

[44] Eric Talbot Jensen, 'Direct Participation in Hostilities: A Concept Broad Enough for Today's Targeting Decisions', in William Banks (ed.), *New Battlefields/Old Laws: Critical Debates on Asymmetric Warfare* (Columbia University Press 2011), 94.

broaden the standards and capture ever more remote forms of participation thus came from a particular perspective on threats and operational needs. Standards, however, should not be moulded with the belief that they will only be relevant *somewhere else* and to *someone else*. DPH is a standard under both treaty and customary law, and it applies equally to civilians involved in any conflict. Standards should be crafted in a principled way.

It is impossible to make a generalized statement on the qualification of specific acts of employees from technology companies as DPH. The need for careful case-by-case determination of DPH is in fact a part of the protective edge of the law of armed conflict regime.[45] If such employees shield networks from vulnerabilities and provide general cybersecurity training to officials, their involvement, even if crucial to capacity-building, will be only indirectly related to adverse effects on the adversary. If, however, such employees employ offensive capabilities, engage in acts of cyber sabotage, or provide concrete tactical intelligence on targets of attack, they will be participating directly and will thus lose protection from attack. Many contributions may be on a spectrum between these extremities. And, given the under-specification of the elements of the DPH test, it can be said that employees contributing to the war effort face risks of which they may not even be fully aware. Similar risks exist for infrastructure used to contribute to military action, and it is to the qualification of such infrastructure that the next section turns.

B. Defining Military Objectives

Tech sector infrastructure can play a crucial role in the military effort of a party to conflict. A perfect illustration of this role comes from SpaceX's Starlink satellites, which have enabled internet access in Ukraine since the February 2022 escalation of the conflict. So significant was this private sector contribution that other actors are now seeking to emulate the model of the satellite communications provider. With an eye to China, Taiwan recently initiated talks with investors to establish a similar project run by its own space agency.[46]

The satellites are infrastructure owned by a private actor, SpaceX, and they primarily support civilian internet access. Are these satellites civilian objects or military objectives?

As a basic premise, civilian objects are protected from attack, while military objectives are not. According to IHL, military objectives refer to those objects 'which by their nature, location, purpose or use make an effective contribution to military action and whose total or partial destruction, capture or neutralization, in the circumstances

[45] Michael N Schmitt, 'Humanitarian Law and Direct Participation in Hostilities by Private Contractors or Civilian Employees' (2005) 5 Chicago Journal of International Law 511, 534.

[46] Kathrin Hille, 'Taiwan plans domestic satellite champion to resist any China attack' (*Financial Times*, 6 January 2023 <https://www.ft.com/content/07c6e48b-5068-4231-8dcf-fe15cb3d0478> accessed 17 April 2023.

ruling at the time, offers a definite military advantage'.[47] Just as for persons, the status of an object is not static. A building designed for a civilian purpose, such as a school, can become a military objective if it is used in a way that effectively contributes to military action. A presumption of civilian status applies: according to the law of armed conflict, an object normally dedicated to civilian purposes must be presumed to remain a civilian object.[48] And the fact that infrastructure is *owned* by a private party does not mean that it cannot be a military objective if it satisfies the test set out in the law.

Two elements form the basis of the definition of a military objective. First, the nature, location, purpose, or use of the object must make an effective contribution to military action. In the case of Starlink satellites, this element is met without difficulty. There is no doubt that the provision of internet access has effectively contributed to Ukrainian military action. Second, the total or partial destruction, capture, or neutralization of the object, in the circumstances ruling at the time, must offer a definite military advantage. If a party to conflict is able to launch an operation that destroys a sufficient number of the satellite constellation to degrade its functionality,[49] then the element of definite military advantage will pose few difficulties, either. The drafting of Additional Protocol I does not accommodate an overly broad construal of the meaning of military objective, leaving indirect contributions and possible advantages out of its scope.[50]

Other scenarios are more complex. Consider, for instance, a high-rise building from which an IT security firm operates. It may be that only one floor of this building is used for the unit providing tactical intelligence to a party to conflict. Would the entire building constitute the military objective or only that floor? Here, the key question is determining what the 'object' is. According to some, structurally interdependent objects, such as the different floors of a building, are better qualified as one integral object.[51] Under that interpretation, the entire building would be a lawful object of attack.

This is not to say that an attack against this building would necessarily be lawful. For instance, it may be that civilians are present in the building, and that the harm they may foreseeably suffer would be excessive compared to the military advantage anticipated. Thus, the proportionality rule can play a role in constraining attacks against otherwise

[47] AP I, art 52(2).

[48] AP I, art 52. The United States rejects this presumption – Adil Ahmad Haque, 'Misdirected: Targeting and Attack Under the U.S. Department of Defense Law of War Manual' in Michael Newton (ed.), *The United States Department of Defense Law of War Manual: Commentary & Critique* (CUP 2019).

[49] Though note that an attack on a single satellite may cause insufficient disruption to count as a definite military advantage – see Tara Brown, 'Can Starlink satellites be lawfully targeted?' (*Articles of War*, 5 August 2022) <https://lieber.westpoint.edu/can-starlink-satellites-be-lawfully-targeted/> accessed 17 April 2023.

[50] Marco Sassòli, 'Legitimate Targets of Attack', (Informal High-Level Expert Meeting on the Reaffirmation and Development of International Humanitarian Law, Cambridge, January 27-29, 2003).

[51] Aurel Sari, 'Israeli Attacks on Gaza's Tower Blocks' (*Articles of War*, 17 May 2021) <https://lieber.westpoint.edu/israeli-attacks-gazas-tower-blocks/> accessed 17 April 2023.

lawful military objectives. It bears emphasis that the incidental harm envisioned by the proportionality rule only refers to loss of civilian life, and injury and damage to civilian objects. In this sense, the scenario of an attack against a high-rise building in use by corporate employees differs from that of attacking Starlink satellites. In the former case, the foreseeability of civilian death, injury, and damage is direct; in the latter, the disruption of the satellites will directly lead to loss of internet access, which may in turn lead to death, injury, or damage. Both direct and indirect (or reverberating) effects are encompassed within the proportionality rule,[52] though the indirect nature of the effects in the Starlink case merits a more careful analysis of the relationship between the satellite, internet access, and harm.[53]

3. DISTANCE AND RISK OF HARM

Through the emergence and proliferation of ICTs, geographical distance is no longer an obstacle to participation in conflict. In this sense, ICTs have brought about a true revolution in contemporary conflicts. Persons located thousands of miles away from the actual hostilities have the capacity to launch offensive cyber operations, engage in surveillance, or share intelligence. Thus, a person can pose a risk to a party to conflict even when they are geographically remote.[54]

On the level of legal standards, geographical proximity is not determinative in either the notion of DPH or in the definition of a military objective. In fact, the ICRC Guidance on DPH specifically notes that it is not geographical but *causal* proximity that matters for the purposes of the test.[55]

At the same time, matters of geographical location matter for the legality of targeting a particular person or object under other rules of international law. That is, the legality of targeting a person or object under international law is not determined solely on the basis of the concept of DPH and the definition of military objectives.[56] Other rules constrain the use of force. One example of such rules comes from the *jus ad bellum* regime. Under the *jus ad bellum*, a state shall not resort to force against another state except where it acts in self-defence or under Security Council authorization.[57] A state that has no relevant attribution links to persons engaged in forcible cyber operations

[52] International Law Association Study Group, *The Conduct of Hostilities and International Humanitarian Law: Challenges of 21st Century Warfare, Final Report* (ILA 2017), pp. 24-25.

[53] *See* Brown, *supra* note 49. Brown gives Tonga as an example, since the country does not have a backup internet system, and thus relies on Starlink for a range of functions, including humanitarian coordination efforts.

[54] François Delerue, 'Civilian Direct Participation in Cyber Hostilities' (2014) 19 Revista de Internet, Derecho y Política, 10, 13.

[55] ICRC Guidance, 55.

[56] Noam Lubell and Nathan Derejko, 'A Global Battlefield? Drones and the Geographical Scope of Armed Conflict' (2013) 11(1) Journal of International Criminal Justice 65.

[57] Charter of the United Nations, arts 2(4), 51 and 42.

does not open itself to forcible action by the state affected by these cyber operations, even if the forcible cyber operations would, if committed by a state, amount to an armed attack. The International Court of Justice in *Palestinian Wall* adopted an inter-state view of the right to self-defence, restricting it to responses to armed attacks coming from states.[58] While the discussions on the legality of self-defence against non-state actors are still ongoing,[59] the point worth emphasizing here is that the *jus ad bellum* provides safeguards against inter-state action even where persons satisfy the criteria for DPH and objects meet those for military objectives. These limitations on forcible action are particularly relevant to the contributions of technology companies in Ukraine. This is because most of these companies' employees and infrastructure are located in third states.

That international law provides constraints on forcible extraterritorial action does not necessarily imply that there is no risk of such action. After all, it has been reported that the Russia-based threat actor IRIDIUM launched a ransomware operation against the transportation and logistics sectors in Poland. The connections between Russia and this actor remain unclear, and the details of the ransomware operation do not seem to indicate that it can be qualified as an attack. At the same time, this incident does indicate that actors operating from states supporting one party to the conflict may find themselves increasingly vulnerable to operations seeking to curb this support. There is thus a distinct risk of harm associated with conflict-related contributions. This risk of harm does not necessarily imply an awareness of risk on the part of the persons and entities effectuating the contributions. It is then pertinent to ask how such persons and entities are to be protected, including through campaigns aimed at informing the public of the characterization and consequences of acts supporting parties to conflict.

4. PROTECTING FROM HARM THROUGH THE RIGHT TO INFORMATION

Contributions to the military effort of a party to conflict can expose persons and objects to danger. Military objectives and persons engaged in DPH can be attacked, subject to the applicable rules of international law. Persons and objects in the vicinity of military objectives or those directly participating in hostilities can similarly become vulnerable to the effects of forcible action. Further, civilians who are direct participants in hostilities do not enjoy immunity from domestic prosecution for lawful acts of war as combatants do. The legal consequences of loss of protection are invasive and potentially life-threatening. In light of this, the general population and the private

[58] *Legal Consequences of the Construction of a Wall in the Occupied Palestinian Territory* (Advisory Opinion) 2004 ICJ Rep 136, para 139.

[59] *See generally*, Dapo Akande, 'The Diversity of Rules on the Use of Force: Implications for the Evolution of the Law' (EJIL:Talk!, 11 November 2019) <https://www.ejiltalk.org/the-diversity-of-rules-on-the-use-of-force-implications-for-the-evolution-of-the-law/> accessed 17 April 2023. Ashley Deeks, 'The Geography of Cyber Conflict: Through a Glass Darkly' (2013) 89 International Law Studies 1.

sector must be fully informed of the qualification of different forms of contribution under the law of armed conflict and of the implications of such qualification. While IHL contains an obligation to disseminate the law,[60] the focus in this section is on IHRL.

States must safeguard the human rights of those under their jurisdiction. In addition to negative obligations, that is, obligations to abstain from interferences with rights, states are bound by a range of positive obligations, that is, obligations to take steps to protect rights. Consider the right to life. Under both international[61] and regional human rights instruments,[62] this right has been interpreted as extending to reasonably foreseeable threats and life-threatening situations that can result in loss of life. If a state reasonably foresees that the conduct of a person under their jurisdiction may expose them (or others) to harmful kinetic or cyber acts that can result in loss of life, then positive obligations would arise. Importantly, this obligation to protect against foreseeable threats would arise irrespective of whether the threatened harm would be lawful or unlawful, for example, whether it would be in contravention of the *jus ad bellum* regime. In fact, in many cases of positive obligations, the state is acting to protect individuals under its jurisdiction from the risk of *unlawful* violence.[63] The jurisprudence of the European Court of Human Rights supports the existence of a ‘public’s right to information’ as part of the positive obligation to take all appropriate steps to safeguard life.[64] This jurisprudence has developed in particular in relation to environmental risks under the right to life and the right to private life.[65]

The existence of positive obligations to protect against foreseeable risks does not imply that a state ought to prohibit direct participation. Rather, it would be required to inform those under its jurisdiction of the risks they run through their conduct, and perhaps, especially in the case of private sector actors operating digital infrastructure, seek to constrain the modalities of participation in the light of the large-scale harm that operations against their employees and infrastructure could entail. This positive obligation to inform can be discharged in a variety of ways, including through dedicated training for private sector companies and information campaigns designed for the general public. Ultimately, the goal is to make the relevant persons and entities aware of the legal significance of their conduct, and of the practical consequences to

[60] All four Geneva Conventions of 1949 contain obligations to disseminate IHL. The obligation requires parties to bring the text of the Conventions to the attention of the general public. *See* ICRC note, The Obligation to Disseminate International Humanitarian Law (2003).

[61] Human Rights Committee, General comment No. 36 on article 6: right to life, 2018, CCPR/C/GC/36, paras 7, 18.

[62] Alexandra Harrington, ‘Life as We Know It: The Expansion of the Right to Life Under the Jurisprudence of the Inter-American Court of Human Rights’ (2013) 35(2) Loyola of Los Angeles International and Comparative Law Review; African Commission on Human and Peoples’ Rights, General Comment No. 3 on the African Charter on Human and Peoples’ Rights: The Right to Life (Article 4) (2015).

[63] *Osman v UK* ECHR 1998–VIII 3124, paras 115-116.

[64] *Öneryıldız v. Turkey* (2005) 41 EHRR 20, paras 89-90.

[65] *Guerra and Others v Italy* (1998) 26 EHRR 357. *See* Council of Europe Factsheet, Environment and the Convention on Human Rights (October 2022).

which such conduct may expose them. Awareness of these consequences may in turn lead to either abstention from participation or adjustment of conduct. For instance, it was reported in February 2023 that SpaceX had taken steps to limit the use of Starlink in offensive operations carried out by the Ukrainian military. SpaceX COO Gwynne Shotwell stated that the company never intended its satellites to be 'weaponized', and that Ukraine had used them 'in ways that were unintentional and not part of any agreement'.[66] All relevant stakeholders must be provided with sufficient information on the content of the law to make informed decisions on their potential contributions.

5. CONCLUSION

This article examined the participation of technology companies in armed conflict through the experience accumulated in the Russia-Ukraine war. As ICTs remove geographical barriers, they enable remote operations and various forms of distanced participation in the military efforts of parties to conflict. In some ways, parties to conflict have come to rely on such participation, especially on the part of technology companies that can securitize networks by scanning for and patching vulnerabilities. Information-sharing has also become a pillar of state–private sector engagement.

Some of these forms of participation may qualify employees as direct participants in hostilities and objects as military objectives. There are tangible risks of harm associated with participation. These risks cannot be eliminated but must be managed. For one, states should clarify their positions on the elements of relevant rules of IHL, such as the notion of DPH. Factual uncertainty is already plaguing armed conflicts; if factual uncertainty is coupled with significant legal uncertainty, then the protection of the civilian population will find itself severely eroded. Importantly, this clarification of standards must be done in a principled and non-arbitrary way by advancing standards that states are content to accept being applied to their own populations and objects. And finally, the content of the relevant legal rules must be disseminated widely – not just as a matter of policy, but as a matter of legal obligation. Human rights law provides a basis for a right to information in circumstances of foreseeable threats to life. Information is the necessary condition for meaningful and balanced decision-making, and the private sector and the general public must be informed of the consequences of their contributions to conflict.

[66] Kate Duffy, 'SpaceX never intended Starlink internet to be "weaponized" in Ukraine, says COO' (*Business Insider*, 9 February 2023) <https://www.businessinsider.com/spacex-starlink-internet-never-intended-weaponized-ukraine-war-gwynne-shotwell-2023-2?r=US&IR=T> accessed 17 April 2023.

2023 15th International Conference on Cyber Conflict
Meeting Reality
T. Jančárková, D. Giovannelli, K. Podiņš, I. Winther
2023 © NATO CCDCOE Publications, Tallinn

Business@War: The IT Companies Helping to Defend Ukraine

Bilyana Lilly
Warsaw Security Forum

Kenneth Geers
Atlantic Council

Greg Rattray
Cyber Defense Assistance
Collaborative for Ukraine

Robert Koch
Universität der Bundeswehr

Abstract: During Russia's invasion of Ukraine, foreign private-sector information technology (IT) firms have provided hardware, software, and cyber intelligence to Kyiv. This assistance has helped Ukraine to stay online during the war by providing stronger network architecture and enhanced security. This paper examines the specific companies involved, the products and services they have offered, and the risks and opportunities associated with their assistance. The authors compile a list of lessons learned and offer actionable policy recommendations so that governments and IT firms are better able to navigate this crisis and similar crises in the future.

Keywords: *Ukraine, Russia, war, cybersecurity, public-private partnerships*

1. INTRODUCTION

In the Internet era, information technology (IT) companies will increasingly find themselves in the line of fire during international conflicts, whether they focus on basic architecture or specifically on security. During the 2014 Revolution of Dignity in Ukraine and Russia's subsequent invasion of Crimea and Donbas, there were many cyber operations against Ukraine's government, private sector, and civil society.[1]

[1] Kenneth Geers, ed., *Cyber War in Perspective: Russian Aggression against Ukraine* (Tallinn: NATO CCD COE Publications, 2015), https://ccdcoe.org/library/publications/cyber-war-in-perspective-russian-aggression-against-ukraine/.

This paper focuses on private-sector IT support to the Ukrainian government, civil society, or both during the ongoing war, which has helped Kyiv to defend against Russian cyber operations and cyberattacks, such as distributed denial of service (DDoS), data destruction, and critical infrastructure manipulation.

We consider contributions in three primary categories:

- Hardware
    - Computers
    - Mobile phones
    - Data centers
    - Critical technologies
    - Terminals
- Software
    - Operating systems, applications
    - Endpoint/network security, monitoring/security information and event management (SIEM) tools, vulnerability management
    - Remote data centers
    - Cloud architecture
- Cyber services
    - Training
    - Threat intelligence (advanced persistent threats (APTs)[2] / attack surface management (ASM) tools, vulnerability management)
    - Malware detection (indicators of compromise (IoCs), signatures)
    - Incident response
    - Security operations center (SOC) support

This analysis covers only publicly disclosed contributions of non-Ukrainian private-sector IT firms that have offered assistance to the Ukrainian government or civil society. It does not evaluate assistance provided by foreign governments, contributions supplied under non-disclosure agreements, or generic cyber intelligence analyses.[3]

2. PRIVATE-SECTOR IT FIRMS AND THEIR ASSISTANCE

Since February 24, 2022, foreign private-sector IT firms have provided the government of Ukraine with a wide range of hardware, software, and cyber services. This section lists many of these firms and their publicly disclosed contributions (see Table I on page 77).

[2] In this paper, "advanced persistent threat" is synonymous with a team working with or for a nation-state.

[3] The Russian government uses cyber operations and influence operations, which include disinformation, in combination to achieve maximum effect. In this paper, the authors examine only IT support that counters cyber operations. For analysis of the entire range of Russia's hybrid warfare activities, including cyber operations, disinformation, protests, coups, and assassinations, see Bilyana Lilly, *Russian Information Warfare: Assault on Democracies in the Cyber Wild West* (Annapolis, MD: Naval Institute Press, 2022).

A. Hardware

The week before the invasion, Amazon responded to a public call for help from Kyiv.[4] On February 24, Liam Maxwell, head of Government Digital Transformation at Amazon Web Services (AWS), met with Ukrainian Ambassador Vadym Prystaiko at the Ukrainian Embassy in London to compile a list of essential data from 27 Ukrainian ministries, 18 Ukrainian universities, and private-sector companies, including Ukraine's largest private financial institution, PrivatBank.[5] Experts set up a secure communication line with Ukrainian agencies and met to discuss the delivery of AWS Snowballs, data transfer devices each able to load 80 terabytes of encrypted data, to facilitate the transfer and storage of critical information infrastructure (CII) to the AWS cloud platform. Three days later, the first Snowballs arrived in Ukraine, and the migration of Ukrainian CII was underway. During the first four months of the war, AWS ingested more than 10 petabytes of data.[6]

On the first day of the war, a malicious firmware update reportedly rendered numerous Viasat KA-SAT modems unusable.[7] Similar to NotPetya in 2017, there was collateral damage across Europe. The U.S. Government attributed the Viasat hack to Russia.[8] Ukrainian vice prime minister sent a tweet to Elon Musk, asking for help,[9] and Musk approved the immediate delivery of his Starlink satellite Internet service to Ukraine. During the war, Starlink has been used for countless military and civilian communications: President Zelenskyy uses Starlink to stay connected with Allied leaders, and Ukrainian military commanders use it to call artillery strikes on the battlefield. Starlink's low-orbit system works in tandem with backpack-sized stations on the ground and offers high-speed, strongly encrypted, highly configurable service. Starlink has withstood increasingly sophisticated Russian hacks.[10] In July 2022, the Ukrainian army was using about 4,000 mobile terminals and requesting 6,700 more.[11] Ukraine's Deputy Prime Minister and Minister of Digital Transformation Mykhailo

[4] Kenneth R. Rosen, "The Man at the Center of the New Cyber World War," *Politico*, July 14, 2022, https://www.politico.com/news/magazine/2022/07/14/russia-cyberattacks-ukraine-cybersecurity-00045486.

[5] Russ Mitchell, "How Amazon Put Ukraine's 'Government in a Box'—and Saved Its Economy from Russia," *Los Angeles Times*, December 15, 2022, https://www.latimes.com/business/story/2022-12-15/amazon-ukraine-war-cloud-data.

[6] "Safeguarding Ukraine's Data to Preserve Its Present and Build Its Future," Amazon, June 9, 2022, https://www.aboutamazon.com/news/aws/safeguarding-ukraines-data-to-preserve-its-present-and-build-its-future.

[7] Matt Burgess, "A Mysterious Satellite Hack Has Victims Far Beyond Ukraine," *Wired*, March 23, 2022, https://www.wired.com/story/viasat-internet-hack-ukraine-russia/.

[8] Antony J. Blinken, "Attribution of Russia's Malicious Cyber Activity Against Ukraine," U.S. Department of State, May 10, 2022, https://www.state.gov/attribution-of-russias-malicious-cyber-activity-against-ukraine/.

[9] Hyunjoo Jin, "Musk Says Starlink Active in Ukraine as Russian Invasion Disrupts Internet," Reuters, February 26, 2022, https://www.reuters.com/technology/musk-says-starlink-active-ukraine-russian-invasion-disrupts-internet-2022-02-27/.

[10] Christopher Miller, Mark Scott, and Bryan Bender, "UkraineX: How Elon Musk's Space Satellites Changed the War on the Ground," *Politico*, June 8, 2022, https://www.politico.eu/article/elon-musk-ukraine-starlink/.

[11] Cade Metz, "Elon Musk Backtracks, Saying His Company Will Continue to Fund Internet Service in Ukraine," *New York Times*, October 15, 2022, https://www.nytimes.com/live/2022/10/15/world/russia-ukraine-war-news/musk-ukraine-internet-starlink.

Fedorov said that 150,000 people in Ukraine were using the service each day, which is paid for by a "range of stakeholders."[12] Ukraine's commander-in-chief, General Zaluzhnyy, praised the Starlink units' "exceptional utility."[13] SpaceX has withstood jamming signals and computer hacking.[14] However, SpaceX blocked control of Ukrainian military drones via Starlink in early February 2023, highlighting that it was "never meant to be weaponized."[15]

Some companies are already assisting with Ukraine's reconstruction efforts. In January 2023, at the World Economic Forum, Fedorov announced that Ukraine and Nokia had signed an agreement to help rebuild Ukraine's telecommunications infrastructure.[16]

B. Software

The Microsoft cloud has enabled Kyiv to provide critical services during the war. Before the Russian invasion, the Ukrainian government operated exclusively on servers located in government buildings inside Ukraine. These sites were vulnerable to missile attacks, and their physical destruction could paralyze the work of Ukraine's leadership. Recognizing this danger, on February 17, Ukraine decided to transfer existing local servers to the public cloud, in order to disburse its infrastructure and protect its data and digital services within European data centers outside of Ukraine. The Ukrainian government now administers its state data from Amazon and Microsoft cloud services, which have also helped to preserve Ukrainian education, banking, healthcare, and humanitarian services.[17]

At the start of the war, Cloudflare provided critical assistance. It offered its anti-DDoS tools, free of charge, to defend Ukraine's networks, which were heavily targeted by

[12] Michael Sheetz, "About 150,000 People in Ukraine Are Using SpaceX's Starlink Internet Service Daily, Government Official Says," CNBC, May 2, 2022, https://www.cnbc.com/2022/05/02/ukraine-official-150000-using-spacexs-starlink-daily.html.

[13] Alex Marquardt, "Exclusive: Musk's SpaceX Says It Can No Longer Pay for Critical Satellite Services in Ukraine, Asks Pentagon to Pick Up the Tab," CNN, October 14, 2022, https://www.cnn.com/2022/10/13/politics/elon-musk-spacex-starlink-ukraine/index.html.

[14] "How Elon Musk's Satellites Have Saved Ukraine and Changed Warfare," *Economist*, January 5, 2023, https://www.economist.com/briefing/2023/01/05/how-elon-musks-satellites-have-saved-ukraine-and-changed-warfare.

[15] Joey Roulette, "SpaceX Curbed Ukraine's Use of Starlink Internet for Drones -Company President," Reuters, February 9, 2023, https://www.reuters.com/business/aerospace-defense/spacex-curbed-ukraines-use-starlink-internet-drones-company-president-2023-02-09/.

[16] Mykhailo Fedorov, World Economic Forum Annual Meeting, January 17, 2023, https://www.weforum.org/events/world-economic-forum-annual-meeting-2023/sessions/press-conference-mykhailo-fedorov; Nokia, "Nokia's Statement on Ukraine," Nokia, March 3, 2022, https://www.nokia.com/about-us/newsroom/statements/nokia-statement-on-ukraine/.

[17] Brad Smith, "Extending Our Vital Technology Support for Ukraine," Microsoft, November 3, 2022, https://blogs.microsoft.com/on-the-issues/2022/11/03/our-tech-support-ukraine/; "Defending Ukraine: Early Lessons from the Cyber War," Microsoft, June 22, 2022, https://query.prod.cms.rt.microsoft.com/cms/api/am/binary/RE50KOK?utm_source=substack&utm_medium=email%27; Kaja Ciglic, Senior Director, Digital Diplomacy, Microsoft, correspondence with authors, January 27, 2023; Dan Black, "Russia's War in Ukraine: Examining the success of Ukrainian Cyber Defences," International Institute for Strategic Studies, March 28, 2023, https://www.iiss.org/research-paper//2023/03/russias-war-in-ukraine-examining-the-success-of-ukrainian-cyber-defences; Rosen, "The Man at the Center"; "How Amazon is Assisting in Ukraine," Amazon, December 1, 2022, https://www.aboutamazon.com/news/community/amazons-assistance-in-ukraine.

Russian hackers at that time. Following the invasion, Cloudflare helped to preserve the functionality of Ukrainian networks.[18] Google was also quick to respond, by offering expanded eligibility to Project Shield's DDoS protection for Ukrainian government and embassy websites worldwide. Google donated 50,000 Google Workspace licenses to the Ukrainian government, including a year of free access to Google Workspace solutions within its cloud-first, zero-trust security model. Currently, Google is working to scale Ukraine's national Diia digital education portal.[19]

Many smaller companies are also offering their software solutions. ESET provided its highest-grade service to Ukrainian critical infrastructure providers, and normal users received an automatic extension to expiring licenses.[20] Sophos offered Ukrainian organizations and consumers free access to its entire cybersecurity portfolio.[21] Sentinel One provided Ukrainian businesses with free access to its Singularity platform.[22] Vectra AI offered free cybersecurity software to organizations that may have been targeted as a result of the war.[23] Avast extended free licenses to users in Ukraine,[24] and a free decrypter for the HermeticRansom data wiper.[25] Outpost24 offered free real-time threat intelligence on Russian hacking groups.[26] Atlas VPN offered its software free of charge to Ukrainian journalists.[27]

C. Cyber Services

Before the invasion, Microsoft tracked six Russian APTs and eight malware families as they collected strategic intelligence and prepositioned destructive malware on nearly 50 Ukrainian agencies and enterprises.[28] In January, Microsoft alerted the Ukrainian government to a Russian "wiper" malware campaign, after which Microsoft established a 24/7 encrypted channel for communication with Ukrainian cybersecurity officials. By April, Microsoft had documented two or three "destructive" cyberattacks

[18] Interview with a cybersecurity expert, January 19, 2023.

[19] Kent Walker, "New Ways We're Supporting Ukraine," Google, December 1, 2022, https://www.blog.google/outreach-initiatives/public-policy/new-ways-were-supporting-ukraine.

[20] "UA Crisis—ESET Response Center," ESET, https://www.eset.com/int/ua-crisis/#eset-helps.

[21] "Ukraine Crisis Resource Center," Sophos, https://www.sophos.com/en-us/content/ukraine-crisis-resource-center.

[22] "A CISO's Guide to the Security Impact of the Attacks on Ukraine," Sentinel One, February 28, 2022, https://www.sentinelone.com/blog/a-cisos-guide-to-the-security-impact-of-the-attacks-on-ukraine/.

[23] "As the War in Ukraine Spirals, Vectra AI Announces Free Cybersecurity Services," Vectra, February 28, 2022, https://www.vectra.ai/news/as-the-war-in-ukraine-spirals-vectra-ai-announces-free-cybersecurity-services.

[24] Ondrej Vleck, "Avast's Response to the War in Ukraine," Avast, March 10, 2022, https://blog.avast.com/avast-response-to-war-in-ukraine.

[25] "Help for Ukraine: Free Decryptor for HermeticRansom Ransomware," Avast, March 3, 2022, https://decoded.avast.io/threatresearch/help-for-ukraine-free-decryptor-for-hermeticransom-ransomware/.

[26] "Staying Secure against Potential Cyber Attacks," Outpost24, accessed January 18, 2023, https://outpost24.com/cybersecurity-scan-offers.

[27] Edward G, "Atlas VPN Hands Out VPN Subscriptions to Support Journalists in Ukraine," Atlas VPN, February 24, 2022, https://atlasvpn.com/blog/atlas-vpn-hands-out-vpn-subscriptions-to-support-journalists-in-ukraine.

[28] Microsoft Digital Security Unit, "Special Report: Ukraine An overview of Russia's cyberattack activity in Ukraine," Microsoft, April 27, 2022, https://query.prod.cms.rt.microsoft.com/cms/api/am/binary/RE4Vwwd; Ciglic, correspondence, January 27, 2023.

per week (including one that targeted a nuclear power plant) that were believed to have been conducted by Russian military intelligence.[29] Microsoft has also provided real-time support to Ukrainian critical infrastructure.[30] By June, Microsoft had discovered Russian network intrusion attempts on 128 organizations in 42 countries outside Ukraine.[31] By July, Microsoft had committed $239 million worth of technological and financial assistance.[32] In November, Microsoft announced the extension of its technology support, free of charge, through 2023.[33]

Google helped the Ukrainian government to set up a system that sends rapid air raid alerts to mobile phones, and its Threat Analysis Group published threat intelligence on government-backed threat actors from Russia, Belarus, China, Iran, and North Korea who have targeted Ukrainian and Eastern European government and defense officials, military organizations, politicians, NGOs, and journalists.[34] Mandiant (now a part of Google Cloud) is providing cyber threat intelligence, monitoring, threat hunting, automated defense, malware detection, mitigation, incident response, and compromise assessments to the Ukrainian government.[35]

Many other companies have offered cyber services. The Cisco Talos threat intelligence team provided 24/7 security support to critical customers in Ukraine.[36] Recorded Future provided access to its threat intelligence portal,[37] as well as specialist engineers,[38] and plans to hire up to 100 personnel in Ukraine before 2025.[39] ESET provided threat intelligence and remediation services for critical infrastructure targets, and has collaborated with the Ukrainian Computer Emergency Response Team (CERT-UA).[40] Sentinel One provided the Ukrainian government with tailored threat intelligence.[41]

[29] David E. Sanger and Julian E. Barnes. “Many Russian Cyberattacks Failed in First Months of Ukraine War, Study Says,” *New York Times*, June 22, 2022, https://www.nytimes.com/2022/06/22/us/politics/russia-ukraine-cyberattacks.html; Microsoft Digital Security Unit, “Special Report.”

[30] Microsoft Digital Security Unit, “Special Report.”

[31] Brad Smith, “Defending Ukraine: Early Lessons from the Cyber War,” Microsoft, June 22, 2022, https://blogs.microsoft.com/on-the-issues/2022/06/22/defending-ukraine-early-lessons-from-the-cyber-war/; Ciglic, correspondence, January 27, 2023.

[32] Ciglic.

[33] Smith, “Extending Our Vital Technology Support for Ukraine.”

[34] Walker, “New Ways We’re Supporting Ukraine.”

[35] Walker, “New Ways We’re Supporting Ukraine”; Robert McMillan and Dustin Volz, “Google Sees Russia Coordinating With Hackers in Cyberattacks Tied to Ukraine War,” *Wall Street Journal*, September 26, 2022, https://www.wsj.com/articles/google-sees-russia-coordinating-with-hackers-in-cyberattacks-tied-to-ukraine-war-11663930801; “Ukraine Crisis Resource Center,” Mandiant, accessed January 18, 2023, https://www.mandiant.com/resources/insights/ukraine-crisis-resource-center.

[36] “The War in Ukraine: Supporting Our Customers, Partners, and Communities,” Cisco, June 23, 2022, https://www.cisco.com/c/m/en_us/crisissupport.html#~faqs.

[37] Kim Zetter, “Security Firms Aiding Ukraine During War Could Be Considered Participants in Conflict,” Zero Day, December 7, 2022, https://zetter.substack.com/p/security-firms-aiding-ukraine-during.

[38] “Ministry of Digital Transformation of Ukraine and Recorded Future sign Memorandum of Cooperation,” Recorded Future, December 6, 2022, https://www.recordedfuture.com/press-releases/120622.

[39] “Recorded Future Partners with Ukraine for Hiring Initiative,” Recorded Future, September 29, 2022, https://www.recordedfuture.com/press-releases/20220929.

[40] There are numerous reports available on the www.welivesecurity.com website.

[41] “Ukraine Crisis Resource Center,” Sentinel One, accessed March 4, 2023, https://www.sentinelone.com/lp/ukraine-response/.

Bitdefender offered threat intelligence and technical consulting.[42] Boldare created the web application UASOS to help Ukrainian refugees find transportation and accommodation outside of Ukraine.[43]

TABLE I: MAIN DISCLOSED IT SECTOR CONTRIBUTIONS TO UKRAINE SINCE FEBRUARY 2022

Company	IT Category	IT Category
Amazon	hardware, software, cyber services	Snowball devices, AWS cloud, software, educational devices to help children learn
Atlas VPN	software	VPN subscription
Avast	software	antivirus license
Bitdefender	cyber services	technical consulting, threat intelligence, cybersecurity technology
Boldare	software	app to find accommodation and transportation
Cisco	cyber services	threat intelligence, threat hunting, monitoring
Cloudflare	software	anti-DDoS tools
ESET	cyber services	threat intelligence, malware detection, remediation
Google	software, cyber services	technical infrastructure, digital skills, funding, training
Mandiant	cyber services	threat intelligence, malware detection, mitigation, incident response, compromise assessments
Microsoft	software, cyber services	data centers, cloud migration, storage, threat intelligence, malware detection, vulnerability discovery, patching
Nokia	hardware, software	software, telecommunications infrastructure
Outpost24	software, cyber services	vulnerability scans, threat intelligence
Recorded Future	software, cyber services	cyber threat intelligence, critical infrastructure protection
Sentinel One	software	endpoint protection
Sophos	software	endpoint protection, network security
Starlink	hardware, cyber services	satellite communication
Vectra AI	software	monitoring tools, incident response tools

[42] "Bitdefender & Romanian National Cyber Security Directorate (DNSC) Work Together in Support of Ukraine," Bitdefender, accessed January 18, 2023, https://www.bitdefender.com/ukraine/.

[43] Pawel Kansi, "Helping Ukraine: Boldare Strengthens Support for Tech to the Rescue," Boldare, March 4, 2022, https://www.boldare.com/blog/helping-ukraine-boldare-support-for-techtotherescue/; Natalia Zglinska, "Boldare Stands with Ukraine," Boldare, February 25, 2022, https://www.boldare.com/blog/boldare-stands-with-ukraine/.

D. Cyber Defense Assistance Collaborative for Ukraine

Numerous companies have joined forces in the Cyber Defense Assistance Collaborative (CDAC) for Ukraine, which was organized in March 2022 to facilitate the provision of cyber defense tools and services, intelligence support, security operations uplift, and strategic advice.[44] These include Avast, Cyber Threat Alliance, Looking Glass, Mandiant, Microsoft, Recorded Future, Sentinel One, Splunk, and Threat Quotient. CDAC collaborates closely with the Ukrainian National Security and Defense Council (NSDC) and Global Cyber Cooperation Center (GC3) to funnel requests for assistance from government and critical infrastructure organizations to the participants. In aggregate, during 2022, CDAC estimates that it has addressed 50+ requests, from 15+ recipients, for 1500+ hardware and software tools, and 550 training courses, with an estimated total value of over $10 million.[45]

3. RISKS AND OPPORTUNITIES OF DEFENDING A COUNTRY AT WAR

When private-sector firms are operating in a geopolitical conflict, they should prepare for a range of risks typically only associated with political, humanitarian, and military organizations. Their involvement, however, may also bring significant opportunities.

A. Risks

There are many traditional issues associated with IT deployment and maintenance abroad, including distance, language, cost, familiarity, and compatibility. International war only amplifies these operational considerations. Finding IT experts who can master geopolitics as well as technology is not easy. Large companies might be familiar with the challenge of integrating national security into their corporate calculus, but smaller or newer firms will likely face a considerable learning curve.

In war, private firms face risks that go far beyond typical corporate considerations. For example, all non-secure communications are subject to eavesdropping by the signals intelligence (SIGINT) infrastructure of numerous nation-states. The firms and their products may face threats to the confidentiality, integrity, and availability of communications. Team members may expect to be targeted by operational and even physical threats to their security. Further, the company's intellectual property is at a much higher risk of reverse engineering by adversaries.[46]

Russia—or any other nation—could consider private IT companies supporting Ukraine as participants in the war and treat them as "legitimate" targets for aggressive

[44] "CRDF Global becomes Platform for Cyber Defense Assistance Collaborative (CDAC) for Ukraine," PR Newswire, November 14, 2022, https://www.prnewswire.com/news-releases/crdf-global-becomes-platform-for-cyber-defense-assistance-collaborative-cdac-for-ukraine-301676373.html.

[45] Information provided by Greg Rattray, CDAC Executive Director, March 2, 2023.

[46] Interview with anonymous source, February 28, 2023.

operations, ranging from physical attacks on infrastructure and personnel to threats and sanctions. Russia placed some private-sector IT firm personnel on its sanctions list and blocked Microsoft from participating in the Open-Ended Working Group on Cybersecurity at the United Nations.[47] On October 27, a Russian foreign ministry official announced that Western commercial satellites, such as Starlink, could become targets in the war. White House spokesman John Kirby responded by saying that any Russian attack on U.S. infrastructure would not go unanswered.[48]

Even well-funded companies may struggle when operating in a conflict environment. On October 14, Elon Musk announced that SpaceX could not fund Internet service in Ukraine "indefinitely" and sent a letter to the Pentagon asking that the U.S. government take over funding the operation. He said the war had cost SpaceX $80 million and 20,000 Starlink terminals, and the company had suffered numerous cyberattacks. The government of Ukraine responded that Starlink had already become part of the nation's critical infrastructure, helping Ukraine survive the invasion, and that SpaceX must continue to provide a "stable connection" to the Internet.[49] One day later, Musk said his company would continue its service.[50]

The Starlink case study also illustrates one of the risks that governments in conflict face when relying on foreign private-sector IT assistance, namely, that governments may grow too dependent on a particular technology without guarantees for its continuous provision. After SpaceX's restrictions on Ukraine's use of Starlink for war operations, General James Dickinson, head of U.S. Space Command, cited it as a cautionary tale for the U.S. military's increasing dependence on commercial satellite providers.[51] Private firms, by definition, have the prerogative to choose their clients. As a result, Kyiv is reliant on the goodwill of foreign private-sector companies. There are no legally binding corporate rules of engagement for cyberspace, and there is no private-sector equivalent to the North Atlantic Treaty Organization's (NATO) Article 5, which would require companies to provide assistance to an allied government—even if that government is a victim of unprovoked aggression. Therefore, countries would be wise to adopt a policy of diversification of products and suppliers, when and where possible, to include the use of open-source solutions.

[47] Kaja Ciglic, Senior Director, Digital Diplomacy, Microsoft, correspondence with authors, March 6, 2023; Burhan Gafoor, Chair Open-Ended Working Group on security of and in the use of information and communications technologies 2021–2025, Permanent Mission of the Republic of Singapore, United Nations, New York, April 22, 2022.

[48] Steve Holland and Susan Heavey, "White House Vows Response if Russia Attacks U.S. Satellites," Reuters, October 27, 2022, https://www.msn.com/en-us/news/world/white-house-vows-response-if-russia-attacks-us-satellites/.

[49] Cassandra Vinograd and Helene Cooper, "Elon Musk Says SpaceX Can't Fund Internet Service in Ukraine 'Indefinitely,' Stirring Controversy," *New York Times*, October 14, 2022, https://www.nytimes.com/2022/10/14/world/europe/elon-musk-starlink-internet-ukraine.html.

[50] Metz, "Elon Musk Backtracks."

[51] Sandra Erwin, "Limits on Ukraine's Use of Starlink for War Operations Is a Lesson for U.S. Military," SpaceNews, March 9, 2023, https://spacenews.com/limits-on-ukraines-use-of-starlink-for-war-operations-is-a-lesson-for-u-s-military/.

B. Opportunities

In war, private-sector IT firms will have the opportunity to stress-test their hardware, software, and cyber services. They will have potential access to the vulnerabilities, exploits, attacks, malware, and tactics, techniques, and procedures (TTPs) of nation-state hackers, which companies can integrate into their products. They may also gain a positive reputational impact for their business.

As a direct result of its support for Ukraine, Microsoft confirmed two improvements in its cyber defense architecture. First, the artificial intelligence (AI) algorithm in its Defender for Endpoint cybersecurity tool detected, classified, and blocked Russian wiper malware, without any existing signatures or any human intervention. Second, Microsoft quickly distributed security software updates to Internet-connected endpoints and cloud-based assets.[52]

4. LESSONS LEARNED

Our research showed that much of the private-sector IT firm support to Ukraine has been more *ad hoc* than carefully planned. However, in either case, it appears critical that companies be able to understand and integrate national security concerns into their corporate calculus. At the same time, it is also apparent that most companies will, to a large degree, follow the lead of their respective governments, despite the inherent risks and opportunities.

In July, Yurii Shchyhol, the director of Ukraine's State Service for Special Communications and Information Protection (SSSCIP), said that, since the invasion began, there have been at least three difference-making cybersecurity gifts that the West has provided to Ukraine. First, Starlink was crucial in helping Ukraine to relaunch destroyed infrastructure. Second, servers and mobile data centers enabled Kyiv to create backup copies of entire institutions, allowing for the continuous operation of government. Third, powerful software, such as an Amazon private cloud, has allowed the Ukrainian government to securely manage its services.[53] Ukraine's Minister of Digital Transformation Mykhailo Fedorov stated: "Amazon AWS literally saved our digital infrastructure."[54]

This conflict has shown that countries could focus on acquiring the hardware, software, and cyber services required to strengthen and back up government services and civil infrastructure. The foundation for moving Ukrainian government and private-sector data to the cloud was a law passed by Ukraine's Parliament in February 2022,

[52] Smith, "Defending Ukraine."

[53] Rosen, "The Man at the Center."

[54] Beatrice Nolan, "Zelenskyy Awards Amazon the Ukraine Peace Prize after AWS Helped Save Its 'Digital Infrastructure,'" *Business Insider*, July 6, 2022, https://www.businessinsider.com/zelenskyy-amazon-ukraine-peace-prize-digital-war-support-aws-2022-7.

which repealed the previously valid requirements for data storage on the territory of Ukraine with regard to the increasing threat of war.[55] A cornerstone for the successful migration of Ukrainian data to the cloud was the Diia government smartphone app, which is "a single portal of public services for the population and business" and was already under development before the start of the war.[56]

However, there are limitations to applying the lessons learned from this war to any future war. The number of variables in international relations precludes drawing too many definitive conclusions. For example, in a China/Taiwan scenario, private-sector IT firms, as well as the respective national governments, may hesitate to act so decisively, because the Chinese economy is roughly ten times larger than Russia's and the potential economic ramifications are far greater.[57]

The war in Ukraine provides clearer lessons for countries closer to Russia's borders. Our research suggests that foreign private-sector IT firms could benefit from the development of four policy frameworks.

> **Voluntary Article 5:** NATO member states sign and agree to Article 5 of the NATO Treaty, the "musketeer clause," and recognize that an attack against one is an attack against all. The countries agree to assist Allies by taking action which those states deem necessary.[58] There is no equivalent to Article 5 in the private sector, which abides by a different legal framework. However, in the spirit of Article 5, private-sector IT firms could take similar action, as they deem appropriate. Nations often have domestic arrangements with private companies operating on their territory in matters of security and defense. However, this is strictly a sovereign affair, and private firms are less likely to support a coalition of nations without clear guidance from their government and a serious review of the risks and opportunities. In any case, private firms' support will be voluntary, based on principle, and non-binding.
>
> **Rules of engagement:** There is general agreement, at least in the European Union and NATO, that cyber operations are limited by the relevant permissions, prohibitions, and requirements of international humanitarian law (IHL). However, IHL's rules primarily regulate the behavior of states and organized armed groups. States and scholars have given limited attention to how IHL relates to private companies during armed conflicts.[59] Therefore, we should work to

[55] Tim Anderson, "'Russian Missiles Can't Destroy the Cloud': Ukraine Leader Describes Emergency Migration," *Register*, November 30, 2022, https://www.theregister.com/2022/11/30/ukraine_cloud_migration/.

[56] Dan Sabbagh, "Ukrainians Use Phone App to Spot Deadly Russian Drone Attacks," *Guardian*, October 29, 2022, https://www.theguardian.com/world/2022/oct/29/ukraine-phone-app-russia-drone-attacks-eppo.

[57] "The Reluctant Rise of the Diplomat CEO. Elon Musk Wants to Be a Statesman. Most Bosses Would Rather Not," *Economist*, October 27, 2022.

[58] NATO, "Collective Defence and Article 5," North Atlantic Treaty Organization, https://www.nato.int/cps/en/natohq/topics_110496.htm.

[59] Ciglic, correspondence with authors, January 27, 2023.

establish general guidelines that outline the recommendations for private-sector IT firms that seek to offer cyber defense assistance to a nation at war.

Multi-stakeholder incident response plan: During a crisis, threatened nations require the assistance of multiple stakeholders, including private-sector IT firms. The process through which this assistance is identified, solicited, and delivered can be accelerated if countries, through a public-private partnership, prepare for potential conflicts in advance. A multi-stakeholder incident response plan could help private-sector IT firms to prepare for geopolitical uncertainties. The plan should include potential crisis scenarios, the support nations will need, all relevant stakeholders, and current points of contact.

Coordination of contributions: Due to the challenging nature of the above recommendation, it would be wise to have multiple delegated entities that could coordinate the contributions of private-sector IT firms while actively liaising with the threatened government. Their goals could encompass understanding national needs, ensuring the rapid provision of assistance, avoiding duplication, and optimizing contributions. During this war, CDAC has served as a major coordinating body for such support to Ukraine, and it can serve as a model for future crises.

5. MOVING FORWARD: DEFENDING AND REBUILDING UKRAINE

Despite the existing risks associated with supporting a country at war, private-sector IT firms have continued to support Ukraine in its fight against Russia. Areas that various stakeholders are actively working on include defining rules of engagement in cyberspace for IT companies providing assistance during conflict, and ensuring that the coordination of assistance is improved through organizations like CDAC. Moving forward, the international community needs to create structures that can support advanced planning, as well as the effective and timely provision of assistance. The authors believe that two areas ripe for investment are collective defense and multi-stakeholder incident response.

Even with the ongoing war, Kyiv is focused not just on defending its networks and systems but also on rebuilding the infrastructure that Moscow has destroyed and on completing the digital transformation of Ukraine. In some de-occupied parts of Ukraine, this work has already begun. In January 2023, Ukraine signed a memorandum with Nokia to help rebuild Ukrainian telecommunications.[60] Kyiv's current vision is

[60] Mykhailo Fedorov, Vice-Prime Minister, Minister of Digital Transformation, Ministry of Digital Transformation of Ukraine, World Economic Forum Annual Meeting, January 17, 2023, https://www.weforum.org/events/world-economic-forum-annual-meeting-2023/sessions/press-conference-mykhailo-fedorov.

outlined on the website of the National Council for the Recovery of Ukraine from the Consequences of the War, in subsections such as "Restoration and development of infrastructure" and "Digitalization."[61] Foreign private-sector IT firms are already helping to defend Ukraine's systems and networks from Russian cyberattacks. Moving forward, we encourage IT companies to look beyond the horizon and start to plan for their participation in the reconstruction of Ukraine.

ACKNOWLEDGMENTS

We are grateful to Dan Black, Sean Costigan, Michael Daniel, Andrew Dwyer, Graham Price, and Zoe Spicer for their assistance and peer reviews.

[61] "National Council for the Recovery of Ukraine from the War," Government Portal, State Sites of Ukraine, https://www.kmu.gov.ua/en/national-council-recovery-ukraine-war; "Recovery and Development of Infrastructure," Government Portal, State Sites of Ukraine, https://www.kmu.gov.ua/storage/app/sites/1/recoveryrada/eng/recovery-and-development-of-infrastructure-eng.pdf; "Digitalization," Government Portal, State Sites of Ukraine, https://www.kmu.gov.ua/storage/app/sites/1/recoveryrada/eng/digitization-eng.pdf.

2023 15th International Conference on Cyber Conflict
Meeting Reality
T. Jančárková, D. Giovannelli, K. Podiņš, I. Winther
2023 © NATO CCDCOE Publications, Tallinn

Evaluating Assumptions About the Role of Cyberspace in Warfighting: Evidence from Ukraine*

Erica D. Lonergan
Assistant Professor
Army Cyber Institute
United States Military
Academy at West Point
West Point, NY, United States

Margaret W. Smith
Director, Cyber Project
Irregular Warfare Initiative
Washington, DC, United States

Grace B. Mueller
Postdoctoral Fellow
Army Cyber Institute
United States Military
Academy at West Point
West Point, NY, United States

Abstract: In the lead-up to Russia's February 2022 invasion of Ukraine, many experts offered predictions about how cyberspace would play a role in the conflict. Specifically, analysts expected Russia to launch a cyber "shock and awe" campaign against Ukraine, to integrate cyber operations into conventional military operations, and to launch significant cyber attacks against the West. We leverage an original dataset, as well as an analysis of several cyber incidents, to explore the extent to which these assumptions match up with reality. While the Ukraine conflict has witnessed a significant volume and diversity of cyber incidents, our research indicates that the cyber dimension of the war has not played out as analysts initially expected. Additionally, some of the more significant cyber incidents and cyber actors were not anticipated by experts, particularly the prominence of third-party non-state actors rather than more sophisticated nation-state actors, and the former's willingness to conduct cyber attacks beyond the theater of operations. We conclude by discussing the implications of these findings for future policymaking.

Keywords: *Ukraine, Russia, cyber conflict, offensive cyber operations, battlefield cyber operations*

* The views expressed by the authors are their own and do not reflect the policy or position of any U.S. government entity or organization with which they are affiliated.

1. INTRODUCTION

As it became increasingly clear that Russia was preparing for military action against Ukraine, cyber experts offered various predictions about what the cyber dimension of the conflict might look like. However, after one year of war, many of these assumptions have not been borne out. Therefore, we evaluate assumptions about the role of cyberspace in warfighting against the evidence, leveraging an original dataset of cyber incidents from the conflict developed through a partnership between the School of International and Public Affairs at Columbia University and the Army Cyber Institute at West Point. Overall, we find that many of the assumptions professed by experts are not supported by the data and, moreover, that the most impactful ways in which cyberspace has influenced the conflict have occurred through unexpected mechanisms. We conclude by exploring the generalizability of our findings beyond the Ukraine conflict, offering policy recommendations for the United States and Europe.

2. PREWAR ASSUMPTIONS

Three broad assumptions shaped the prewar cyber conversation.[1] First, many experts assumed that a Russian conventional assault on Ukraine would be preceded by or take place in conjunction with a major offensive cyber campaign. A related assumption was that Russia would conduct widespread and significant cyber attacks against Ukrainian critical infrastructure—or even do so in lieu of a conventional one.[2] For example, in February 2022, Jason Healey predicted, "A Russian invasion of Ukraine may redefine how we think about cyber conflict because it will be the first time a state with real capabilities is willing to take risks and put it all on the line."[3] This view was echoed by the former head of U.S. Army Europe, Ben Hodges, who declared, "We're not dealing with Boy Scouts here. These guys are absolutely ruthless at using cyber to wreck all the structures of a society."[4] Keir Giles hypothesized that Russia might even

[1] Erica D. Lonergan, "The Cyber-Escalation Fallacy: What the War in Ukraine Reveals About State-Based Hacking," *Foreign Affairs*, April 15, 2022; Maggie Smith, Erica D. Lonergan, and Nick Starck, "What Impact, if Any, Does Killnet Have?" Lawfare, October 21, 2022, https://www.lawfareblog.com/what-impact-if-any-does-killnet-have; "Defending Ukraine: Early Lessons from the Cyber War," Microsoft, June 22, 2022, https://query.prod.cms.rt.microsoft.com/cms/api/am/binary/RE50KOK; "Microsoft Digital Defense Report 2022: Illuminating the Threat Landscape and Empowering a Digital Defense," Microsoft, November 4, 2022; "Cyber Dimensions of the Armed Conflict in Ukraine: Quarterly Analysis Report – Q3 July to September 2022," Cyber Peace Institute, December 16, 2022; Maggie Miller, "NATO Prepares for Cyber War," *Politico*, December 3, 2022, https://www.politico.com/news/2022/12/03/nato-future-cyber-war-00072060; "NATO Secretary General Warns of Growing Cyber Threat," North Atlantic Treaty Organization, November 11, 2022, https://www.nato.int/cps/en/natohq/news_208889.htm?selectedLocale=en.

[2] For a critique of cyber coercion theory, see Erica D. Borghard and Shawn W. Lonergan, "The Logic of Coercion in Cyberspace," *Security Studies* 26, no. 3 (2017): 452–81.

[3] Joseph Marks and Aaron Schaffer, "Here's What Cyber Pros Are Watching in the Ukraine Conflict," *Washington Post*, February 24, 2022, https://www.washingtonpost.com/politics/2022/02/24/heres-what-cyber-pros-are-watching-ukraine-conflict/.

[4] "What are Putin's Intentions in Ukraine?" WTOP News, January 23, 2022, https://wtop.com/europe/2022/01/what-are-putins-intentions-in-ukraine/.

turn to cyber attacks to coerce Ukraine into capitulation, rather than invade: "Stand-off strikes using missiles, or potentially a destructive cyber onslaught, could target military command and control systems or civilian critical infrastructure and pressure Kyiv into concessions and its friends abroad into meeting Russia's demands."[5]

A second common assumption was that Russia would coordinate cyber operations with kinetic military operations on the battlefield. Jonathan Reiber, for instance, hypothesized, "This may end up being the first declared hostility where cyberspace operations are a part of an integrated offensive military invasion…. We could see a coordinated campaign of cyberspace operations targeting the Ukrainian government's senior leader communications, military critical infrastructure and communications, and aspects of Ukrainian national critical infrastructure."[6] Similarly, Keith Alexander, the first commander of U.S. Cyber Command, offered that "there can be little doubt that such a modern military campaign would almost certainly include an extensive cyber attack component."[7] In other words, experts expected Russia to act consistent with its own doctrine.[8]

Finally, many expected Russia to launch waves of sophisticated cyber attacks beyond the theater of operations in Ukraine, specifically targeting the U.S. and NATO to retaliate against Western actions, such as economic sanctions. For instance, in February 2022, the Cybersecurity and Infrastructure Security Agency (CISA) issued an advisory about Russian cyber threats to the United States, especially critical infrastructure, as part of its "Shields Up" campaign.[9] Around the same time, U.S. deputy attorney general Lisa Monaco warned, "Given the very high tensions that we are experiencing, companies of any size and of all sizes would be foolish not to be preparing right now as we speak."[10] Similarly, Britain's National Cyber Security Centre warned that Russia might conduct cyber operations targeting the United Kingdom, cajoling firms to "bolster their cyber security resilience in response to the malicious cyber incidents in and around Ukraine," with British leaders concerned about cyber spillover from

[5] Keir Giles, "Putin Does Not Need to Invade Ukraine to Get His Way," Chatham House, December 21, 2021, https://www.chathamhouse.org/2021/12/putin-does-not-need-invade-ukraine-get-his-way.

[6] Maggie Miller, "Russian Invasion of Ukraine Could Redefine Cyber Warfare," *Politico*, January 28, 2022, https://www.politico.com/news/2022/01/28/russia-cyber-army-ukraine-00003051.

[7] Keith Alexander, "Cyber Warfare in Ukraine Poses a Threat to the Global System," *Financial Times*, February 15, 2022, https://www.ft.com/content/8e1e8176-2279-4596-9c0f-98629b4db5a6.

[8] Gavin Wilde, "Cyber Operations in Ukraine: Russia's Unmet Expectations," Carnegie Endowment for International Peace, December 12, 2022, https://carnegieendowment.org/2022/12/12/cyber-operations-in-ukraine-russia-s-unmet-expectations-pub-88607.

[9] Garrett M. Graff, "The US Watches Warily for Russia–Ukraine Tensions to Spill Over," *Wired*, February 15, 2022, https://www.wired.com/story/russia-ukraine-cyberattacks-spillover/.

[10] Alexander Mallin and Luke Barr, "DOJ Official Warms Companies 'Foolish' Not to Share Up Cybersecurity Amid Russia Tensions," ABC News, February 17, 2022, https://abcnews.go.com/Politics/doj-official-warns-companies-foolish-shore-cybersecurity-amid/story?id=82959520.

the conflict.[11] Cyber security expert John Cofrancesco noted that Russia was likely to "make very strategic attacks against parts of our infrastructure that impact everyday Americans…. This is standard Russian operating procedure."[12]

3. WHAT DOES THE EVIDENCE DEMONSTRATE?

To evaluate how these assumptions have fared against reality, we collected original data on cyber incidents in the context of the Ukraine conflict.[13] Specifically, we focused on cyber activity from actors supporting Russia or Ukraine from March 2021 until the end of August 2022. We collected data dating back to nearly one year before Russia's invasion of Ukraine to identify evidence of Russian cyber behavior that might have represented preparation for conventional conflict, given the long time horizons that are often associated with offensive cyber operations, particularly against strategic targets. During this time period, we identified 131 cyber events, which we mapped to the phases of the conventional conflict to assess potential cyber-kinetic correlation.[14] Below, we evaluate trends based on the observations in the dataset and probe several cyber incidents in greater depth.[15]

A. Assumption 1: Cyber "Shock and Awe"

Despite a notable volume of cyber activity, there is little evidence of significant cyber attacks against Ukraine's critical infrastructure. Of the 131 recorded cyber events in our dataset, 61 (or 47%) took place in Ukraine, 29 (or 22%) took place in Russia, and the remaining incidents (31%) took place in other Western and/or European states.[16]

[11] Guy Faulconbridge, James Davey, and Kate Holton, "Brace for Russian Cyber-Attacks as Ukraine Crisis Deepens, Britain Says," *Reuters*, January 28, 2022, https://www.reuters.com/world/europe/brace-russian-cyber-attacks-over-ukraine-britain-says-2022-01-28/; Dan Sabbagh, "UK Firms Warned of Russian Cyberwar 'Spillover' from Ukraine," *Guardian*, February 23, 2022, https://www.theguardian.com/technology/2022/feb/23/uk-firms-warned-russia-cyberwar-spillover-ukraine-critical-infrastructure.

[12] Ken Dilanian and Courtney Kube, "Biden Has Been Presented with Options for Massive Cyberattacks Against Russia," NBC News, February 24, 2022, https://www.nbcnews.com/politics/national-security/biden-presented-options-massive-cyberattacks-russia-rcna17558.

[13] This is an ongoing data collection project. The current version contains information about incidents through August 2022.

[14] We rely on the U.S. military's definition of operational phases of a campaign. See "Joint Publication 3-0: Joint Operations," U.S. Department of Defense, January 17, 2017, incorporating Change 1, October 22, 2018.

[15] Microsoft, "Defending Ukraine"; James Andrew Lewis, "Cyber War and Ukraine," June 16, 2022, https://www.csis.org/analysis/cyber-war-and-ukraine; Jon Bateman, "Russia's Wartime Cyber Operations in Ukraine: Military Impacts, Influences, and Implications," Carnegie Endowment for International Peace, December 16, 2022, https://carnegieendowment.org/2022/12/16/russia-s-wartime-cyber-operations-in-ukraine-military-impacts-influences-and-implications-pub-88657.

[16] Twelve of NATO's 30 member states have also been targeted in relation to the conflict in Ukraine. See Table IV.

Concerning the cyber events that occurred in Ukraine and Russia, the overwhelming majority involved intelligence operations or cyber-enabled espionage, as depicted in Table I.[17, 18]

TABLE I: TYPES OF CYBER INCIDENTS SEEN IN UKRAINE AND RUSSIA

	Ukraine	Russia
Intelligence	**63.93% (39/61)**	**79.31% (23/29)**
Cyber-enabled IO	6.56% (4/61)	3.45% (1/29)
Cyber effects	29.51% (18/61)	17.24% (5/29)
Total cyber incidents	100% (61/61)	100% (29/29)

These incidents are characterized by probing, packet sniffing, and reconnaissance, all of which are intended to extract (or attempt to extract) private information from a target without affecting it. Most cyber intelligence operations are not intended to be uncovered by the target.[19] In general, we would not expect cyber intelligence operations to be associated with escalation, given the absence of effects coupled with the tacit understanding between states that cyber espionage is generally an acceptable form of statecraft.[20]

Concerning the targets of cyber activity, the overwhelming majority—for all states—are government entities, followed by private entities (see Table II). During the period under study, only 31.15% of all recorded cyber events in Ukraine targeted military entities (e.g., defense organizations, including cyber and non-cyber military or intelligence entities), and 0% of the cyber events in Russia targeted military actors. This is especially surprising given the kinetic warfighting recorded during this time.[21]

[17] Thirty-nine (or 63.93%) of the 61 total cyber events that took place in Ukraine, and 23 (or 79.31%) of the 29 total cyber events that took place in Russia were characterized by intelligence operations.

[18] In our data, following Valeriano and Maness' (2014) typology of three distinct types of cyber operations, we categorize cyber events into three different groups: 1) intelligence operations, 2) cyber-enabled information operations, and 3) cyber effects operations, which include disruption campaigns (e.g., DDoS), deny campaigns (e.g., website defacement and harassment), degrade campaigns (e.g., ransomware), and destroy campaigns (e.g., wiper malware).

[19] Joshua Rovner, "Cyber War as an Intelligence Contest," War On the Rocks, September 16, 2019, https://warontherocks.com/2019/09/cyber-war-as-an-intelligence-contest/.

[20] For an alternative perspective, see Ben Buchanan, *The Cybersecurity Dilemma*, 2016.

[21] There could be some bias in the data if there is a systematic disincentive for military organizations to publicly report information about cyber intrusions and attacks.

TABLE II: TARGETS OF CYBER ACTIVITY IN UKRAINE AND RUSSIA

	Private	Government	Military	Total incidents
Ukraine	45.90% (28/61)	**77.05% (47/61)**	31.15% (19/61)	61
Russia	**62.07% (18/29)**	41.38% (12/29)	0% (0/29)	29

Note: Multiple target types are possible for any given cyber event.

On a related note, when one looks at cyber activity in terms of critical infrastructure targeted, of the 61 recorded cyber incidents that took place in Ukraine, none is recorded as targeting the defense industrial base sector.[22]

Examining specific cyber incidents, the one that is most consistent with the assumption that Russia would conduct a paralyzing cyber strike at the outset of the conflict is the Viasat telecommunications hack, which took place on February 24, just one hour before the Russian army crossed into Ukraine. The evidence strongly suggests that the Viasat cyber attack was part of Russia's plan for the initial phase of the campaign. Reporting indicates that in the first 72 hours of the conflict, Russia sought to paralyze Ukrainian command and control (C2), likely as part of a decapitation strategy, through both kinetic means (such as missile and air strikes) and non-kinetic means (in addition to offensive cyber operations, this included jamming of Ukrainian communications systems, as well as direct messages to senior Ukrainian military officers not to resist the Russian invaders).[23]

Yet while this cyber incident might be considered an operational success, experts have concluded there is "no information that [the hack] worsened communications within Ukraine's military."[24] It does not appear that Russia was able to capitalize on the temporary disruption of Ukrainian C2 in the immediate aftermath of the Viasat attack. This could be attributed to the absence of advanced coordinated planning between

[22] We use a U.S. framework to define and categorize critical infrastructure. While there are limitations to this approach, we offer it as a general guide for evaluating different types of targets in Ukraine. According to this definition, the defense industrial base sector comprises companies and firms that contract with military organizations and military facilities. This is only one subset of the category of military targets.

[23] Mykhaylo Zabrodskyi, Jack Watling, Oleksandr V. Danylyuk, and Nick Reynolds, "Preliminary Lessons in Conventional Warfighting from Russia's Invasion of Ukraine: February–July 2022," RUSI, November 30, 2022, 8, 25, https://rusi.org/explore-our-research/publications/special-resources/preliminary-lessons-conventional-warfighting-russias-invasion-ukraine-february-july-2022; Michael Kofman and Jeffrey Edmonds, "Russia's Shock and Awe: Moscow's Use of Overwhelming Force Against Ukraine," *Foreign Affairs*, February 22, 2022, https://www.foreignaffairs.com/articles/ukraine/2022-02-21/russias-shock-and-awe; Helene Cooper, "Pentagon Gives a Grim Assessment of the First Stages of the Russian Invasion," *New York Times*, February 24, 2022, https://www.nytimes.com/2022/02/24/us/politics/pentagon-russia-ukraine.html; Congressional Research Service, "Russia's War in Ukraine: Military and Intelligence Aspects," Congressional Research Service, September 14, 2022, https://crsreports.congress.gov/product/pdf/R/R47068.

[24] Dustin Volz and Robert McMillan, "In Ukraine, a 'Full-Scale Cyberwar' Emerges," *Wall Street Journal*, April 12, 2022, https://www.wsj.com/articles/in-ukraine-a-full-scale-cyberwar-emerges-11649780203.

cyber operators and kinetic and maneuver forces, as well as the overall slow pace of Russian adaptation after it became clear that initial assumptions about how the war would unfold were wrong.[25] That Ukrainian forces were able to quickly shift to resilient and redundant capabilities—including communications capabilities provided by private sector actors—also likely blunted the attack's impact.

There is also evidence of Russian cyber attacks against Ukrainian critical infrastructure, including power, that suggests attempts at a more significant cyber campaign. However, these fell far short of the magnitude anticipated prior to the conflict. For example, a cyber incident took place on April 8, 2022, when Ukraine's power grid was targeted by Sandworm, the notorious Unit 74455 of Russia's Main Intelligence Directorate (GRU).[26] Although this incident occurred in April, the attackers had gained access to the power grid's system at the beginning of the war, highlighting the fact that cyber operations take time to plan and execute.[27] Had this malware been successful, it would have deprived roughly two million Ukrainians of power. But ultimately, Ukraine's national Computer Emergency Response Team and the Slovakia-headquartered cyber firm ESET were able to thwart the attack in progress.[28] Interestingly, the malware used in this incident was similar to that deployed in 2015, which succeeded in briefly knocking out power for 100,000 Ukrainians.

To launch another, more successful attack on the power grid, it would ostensibly take time for the GRU to develop new access and exploits, especially given that they would be operating in an environment in which network defenders would likely be even more vigilant about detecting malicious activity. Even though Ukrainian officials say this cyber incident was intended to support Russian military operations in eastern Ukraine, there is also no evidence of any follow-on military activity in the context of this particular cyber attack.[29] As the war progressed, the Russian military ultimately shifted its approach to more significant kinetic attacks against civilian critical infrastructure, including targeting the power grid, but there is no evidence that this had anything to do with the attempted cyber attack in April. Instead, it is more likely that this was in response to Russian setbacks on the battlefield.

What accounts for this lack of a more significant Russian cyber "shock and awe" campaign (especially in the early phases of the war), as well as the surprising

[25] Bateman, "Russia's Wartime Cyber Operations."

[26] Andy Greenberg, *Sandworm: A New Era of Cyberwar and the Hunt for the Kremlin's Most Dangerous Hackers*. (New York: Doubleday, 2019).

[27] Andy Greenberg, "Russia's Sandworm Hackers Attempted a Third Blackout in Ukraine," *WIRED*, April 12, 2022, https://www.wired.com/story/sandworm-russia-ukraine-blackout-gru/.

[28] Bateman, "Russia's Wartime Cyber Operations."

[29] Patrick Howell O'Neill, "Russian Hackers Tried to Bring Down Ukraine's Power Grid to Help the Invasion," April 12, 2022, https://www.technologyreview.com/2022/04/12/1049586/russian-hackers-tried-to-bring-down-ukraines-power-grid-to-help-the-invasion/.

ineffectiveness of Russian offensive cyber operations throughout the conflict?[30] Analysts have suggested a range of potential explanations. Gavin Wilde, for example, articulates three potential causes: the relative inexperience of Russia's information troops; bureaucratic rivalry and infighting over resources and prioritization, especially between the GRU and FSB (the Federal Security Service); and an inability to shape the war in its opening phase.[31] Others have attributed Russia's lack of strategic success in cyberspace to the effectiveness of Ukrainian cyber defenses, aided by collaboration with government and private sector partners.[32] This latter point sheds light on some of the cyber aspects of the conflict overlooked by experts, especially the role of private sector actors in conventional conflict.

B. Assumption 2: Cyber-Conventional Coordination

Similar to the first assumption, there is also minimal evidence of cyber-conventional coordination on the battlefield.[33] The U.S. military's dominant paradigm for joint operations is a planning construct consisting of six phases: 0) shape, 1) deter, 2) seize the initiative, 3) dominate, 4) stabilize, and 5) enable civil authority.[34] We use these operational phases for conventional military campaigns from the Russian perspective to categorize our cyber event data, as depicted in Table III.[35] Therefore, the coding for a cyber incident by campaign phase reflects the phase of the overall conventional campaign in which it occurred, rather than the phase of a cyber campaign. When one considers how the types of cyber operations have changed throughout the war for all cyber actors, including those in Ukraine, Russia, and beyond, some interesting patterns emerge.

[30] Lennart Maschmeyer and Nadiya Kostyuk, "There Is No Cyber 'Shock and Awe': Plausible Threats in the Ukraine Conflict," War on the Rocks, February 8, 2022, https://warontherocks.com/2022/02/there-is-no-cyber-shock-and-awe-plausible-threats-in-the-ukrainian-conflict/; Erica D. Lonergan, Shawn W. Lonergan, Brandon Valeriano, and Benjamin Jensen, "Putin's Invasion of Ukraine Didn't Rely on Cyberwarfare. Here's Why," *Washington Post*, March 7, 2022, https://www.washingtonpost.com/politics/2022/03/07/putins-invasion-ukraine-didnt-rely-cyber-warfare-heres-why/; Lennart Maschmeyer and Myriam Dunn Cavelty, "Goodbye Cyberwar: Ukraine as a Reality Check," *CSS ETH Zurich Policy Perspectives* 10, no. 3 (May 2022): 1–4; Bateman, "Russia's Wartime Cyber Operations."

[31] Wilde, "Cyber Operations in Ukraine."

[32] Sean Atkins, "A Web of Partnerships: Ukraine, Operational Collaboration, and Effective National Defense in Cyberspace," August 30, 2022, https://www.atlanticcouncil.org/content-series/airpower-after-ukraine/a-web-of-partnerships-ukraine-operational-collaboration-and-effective-national-defense-in-cyberspace/; Erica D. Lonergan and Brandon Valeriano, "What Ukraine Shows about Cyber Defense and Partnerships," *National Interest*, March 17, 2022, https://nationalinterest.org/blog/techland-when-great-power-competition-meets-digital-world/what-ukraine-shows-about-cyber.

[33] Nadiya Kostyuk and Yuri Zhukov, "Invisible Digital Front: Can Cyber Attacks Shape Battlefield Events?" *Journal of Conflict Resolution* 63, no. 2 (2019): 317–347. In particular, the authors find that "cyber operations have not created forms of harm and coercion that visibly affect their targets' actions" (p. 319).

[34] As described in the Department of Defense's "Joint Publication 3-0," the six operation phases begin with Phase 0 (Shape), as this precedes the operation order activation. See Figure V-7 on Chapter 5, p. 13. For additional information on Phase 0, see R. Bebber, "Information War and Rethinking Phase 0," *Journal of Information Warfare* 15, no. 2 (2016): 39052.

[35] We recognize that this is a U.S.-centric perspective and therefore use the phases as a general guide for organizing and categorizing the cyber event data.

TABLE III: TYPES OF CYBER INCIDENTS OVER TIME

	Phase 0: Shape	Phase 1: Deter	Phase 2: Seize the Initiative	Phase 3: Dominate
Intelligence	**100% (6/6)**	38.46% (5/13)	**59.09% (13/22)**	47.78% (43/90)
Cyber-enabled IO	0% (0/6)	15.38% (2/13)	9.09% (2/22)	1.11% (1/90)
Cyber effects	0% (0/6)	**46.15% (6/13)**	31.82% (7/22)	**51.11% (46/90)**
Total	100% (6/6)	100% (13/13)	100% (22/22)	100% (90/90)

During Phase 0, "shape," which took place between March 1, 2021, and January 31, 2022, all six of the recorded cyber incidents were characterized by intelligence operations. During Phase 1, "deter," which occurred in the first few months of 2022,[36] there were 13 recorded cyber events, the largest single category of which (46.15%) was cyber effects operations. In Phase 2, "seize the initiative," which took place between February 23, 2022, and April 7, 2022, we do not see an increase in the percentage of cyber effects operations, which we might expect in light of these cyber incidents occurring during the height of the initial invasion. Instead, we find that the majority (59.09%) was intelligence operations, as was the case in Phase 0. In Phase 3, "dominate," covering cyber activity between April 8, 2022, and August 18, 2022, there is a resurgence of cyber effects operations, which make just over half of the recorded 90 incidents that occurred during this time period. While it would be a step too far to make causal inferences from these data, the patterns of cyber operations mapped to the phases of the conventional campaign—moving from intelligence to effects in two waves—are consistent with the idea that cyber campaigns take time to develop, involve an initial investment in intelligence collection, and therefore may be difficult to synchronize with other aspects of a broader campaign.

A few cyber incidents stand out as plausible cyber-kinetic coordination, but there is little evidence of a direct causal link. For example, on August 15, 2022, the People's CyberArmy—a Russian hacktivist group—launched a distributed denial-of-service (DDoS) attack that targeted a Ukrainian state-owned nuclear power company, Energoatom. Even though hacktivists used 7.25 million bot accounts to flood Energoatom's website, the company was able to regain control of the website within three hours.[37] While it is hard to definitively say whether this brief cyber operation resulted in follow-on kinetic military action by Russia, Russian shelling did increase five days later, just 30 kilometers from the Pivdennoukrainsk nuclear power plant,

[36] In our dataset, Phase 1 occurred between December 1, 2021, and March 11, 2022.

[37] Daryna Antoniuk, "Ukraine's State-Owned Nuclear Power Operator Said Russian Hackers Attacked Website," August 17, 2022, https://therecord.media/ukraines-state-owned-nuclear-power-operator-said-russian-hackers-attacked-website/.

which is managed by Energoatom.[38] However, it is just as likely that these events cannot be correlated, particularly given the loose command-and-control relationship between the Russian government and Russian-aligned cyber hacktivist groups.

An additional potential example of coordination is another failed cyber attack on Ukraine's power grid, which occurred on July 1, 2022. In this incident, a pro-Russian hacker group, XakNet, targeted the DTEK power company. Moreover, the incident took place at the same time that Russian forces were carrying out missile attacks on DTEK's Kryvorizka thermal power plant in Kryvyi Rih,[39] leading one Ukrainian cyber official to point to this as evidence of kinetic-cyber coordination.[40] Nevertheless, according to Jon Bateman, kinetic attacks on Ukrainian power infrastructure have been a "routine feature of the war," and the cyber incident has not been specifically attributed to the Kryvorizka facility.[41] This suggests that coordination claims are not as strong as they appear. In fact, it has been reported that between September and December, DTEK energy facilities have been subjected to 21 conventional military attacks, compared to the single reported attempted cyber attack.[42] Moreover, related to the first assumption, the July cyber incident failed to destabilize DTEK infrastructure and therefore failed to disrupt the Ukrainian power system. This further corroborates the limitations of Russian-linked offensive cyber campaigns during the conflict.[43] Taken together, while we find little compelling evidence of cyber-conventional coordination, it is notable that many of the incidents in our dataset were perpetrated by pro-Russian hacktivist groups rather than threat actors more directly associated with the central government.

C. Assumption 3: Cyber Spillover

The assumption that has most closely stood up to empirical scrutiny is cyber spillover beyond the theater of operations in Ukraine. Indeed, our data reveal significant cyber activity targeting Western countries (as well as the Russian homeland). However, this activity has largely taken the form of disruptive, nuisance cyber attacks perpetrated by patriotic hackers and hacktivist groups, rather than major disruptive or destructive cyber campaigns ordered by the Russian government against Western critical infrastructure.

[38] Radio Free Europe, "Russian Shelling Hits Residential Area in Ukrainian Town Near Nuclear Plant," August 20, 2022, https://www.rferl.org/a/mykolayiv-ukraine-russian-shelling-nuclear-power-plant/31997199.html.

[39] DTEK Group (@dtek_en), July 1, 2022, https://twitter.com/dtek_en/status/1542884325015830528?s=20&t=o_Qz2y72EVvYBC8gAHDLMQ.

[40] Victor Zhora (@VZhora), July 1, 2022, https://twitter.com/VZhora/status/1542858906560512000.

[41] Bateman, "Russia's Wartime Cyber Operations."

[42] DTEK Energy, "Once Again One of DTEK Energy Facilities Got Hit as a Result of Russian Attack. It Stopped Generating Electricity," December 24, 2022, https://dtek.com/en/media-center/news/rosiya-znovu-obstrilyala-odin-z-energetichnikh-obektiv-dtek-vin-pripiniv-generatsiyu-elektroenergii/.

[43] DTEK Energy, "Enemy Launches Hacker Attacks on the Power System," July 1, 2022, https://dtek.com/en/media-center/news/vslid-za-raketnimi-udarami-po-tes-vorog-zavdae-khakerskikh-udariv-po-energosistemi/.

Although a large percentage (47%) of the malicious cyber activity related to the 2022 Russian invasion of Ukraine has taken place in Ukraine, more than half of the cyber incidents in our dataset have taken place beyond the theater of operations. As depicted in Table IV, 33 cyber incidents—making up 25% of the cyber events in our dataset—have targeted NATO member states during the time period we examined.

TABLE IV: TARGETS OF CYBER ACTIVITY IN NATO MEMBER STATES

	Private	Government	Military	Total incidents
Bulgaria	0% (0/1)	**100% (1/1)**	0% (0/1)	1
Canada	0% (0/1)	**100% (1/1)**	0% (0/1)	1
Czechia	**66.67% (2/3)**	**66.67% (2/3)**	33.34% (1/3)	3
Estonia	50% (1/2)	**100% (2/2)**	0% (0/2)	2
France	0% (0/1)	**100% (1/1)**	0% (0/1)	1
Germany	**100% (8/8)**	0% (0/8)	0% (0/8)	8
Italy	0% (0/4)	**100% (4/4)**	25% (1/4)	4
Latvia	40% (2/5)	**60% (3/5)**	0% (0/5)	5
Poland	0% (0/3)	**100% (3/3)**	0% (0/3)	3
Romania	100% (1/1)	100% (1/1)	100% (1/1)	1
United Kingdom	0% (0/1)	**100% (1/1)**	0% (0/1)	1
United States	**100% (0/3)**	0% (0/0)	0% (0/0)	3

Note: Multiple target types are possible for any given cyber event. NATO members without any recorded cyber events during the period under study include Albania, Belgium, Croatia, Denmark, Greece, Hungary, Iceland, Lithuania, Luxembourg, Montenegro, Netherlands, North Macedonia, Norway, Portugal, Slovakia, Slovenia, Spain, and Türkiye.

Germany, Latvia, and Italy have been the most frequently targeted NATO states, with Killnet carrying out the preponderance of these DDoS cyber events. The United States has also been on the receiving end of malicious cyber activity related to the Ukraine war, with three instances of cyber attacks against U.S. targets, one intelligence operation, and two cyber effects operations (none of which had significant impact).

Across these examples, given the threat actors involved (hacktivist groups rather than nation-state APTs), it is unlikely that this cyber activity represents deliberately ordered cyber attacks by the Russian government against the U.S. or NATO. Additionally, these attacks have had a negligible impact on the overall conflict. Despite affirming that

Article 5 applies to cyberspace in the context of the Ukraine conflict, NATO member states have largely focused on responding to cyber attacks with information-sharing and increasing defense and resilience measures.[44] In fact, while NATO members have escalated their involvement in the Ukraine conflict in terms of the quality and quantity of weapons provided (and associated training programs to instruct Ukrainians on their use), this has been a direct result of dynamics on the conventional battlefield rather than in cyberspace.

Cyber attacks against Russia have also been frequent during this war, again hinting at the fact that spillover in cyberspace extends even to the homeland of the initiator. Of the recorded 131 cyber incidents, 29 have targeted entities in Russia. The hacktivist group Anonymous was responsible for 22 of these incidents, followed by other anti-Russian (e.g., Network Battalion 65) and pro-Ukrainian (e.g., IT Army of Ukraine) actors.

4. WHAT DID THE EXPERTS MISS?

Our data also suggest important patterns that were overlooked by experts. In particular, prior to the conflict, there was little focus on the role of third-party actors.[45] Yet one of our most striking findings is the preponderance of cyber activity stemming from non-state actors that are not directly affiliated with governments, such as Anonymous, Killnet, and the IT Army of Ukraine. Specifically, we identified 25 different threat actors that have been responsible for the 61 incidents in our dataset that occurred within Ukraine, with some actors being more active than others (see Table V).

TABLE V: THREAT ACTORS TARGETING UKRAINE

Threat actor	Incidents
APT28	2
DEV-0586 APT	1
FreeCivilian	1
Gamaredon	2
GhostWriter	1
GRU	1
People's CyberArmy	4
Poss. Ghostwriter	1

[44] Miller, "NATO prepares for cyber war"; "Keynote Address by NATO Secretary General Jens Stoltenberg at the NATO Cyber Defence Pledge Conference in Italy," North Atlantic Treaty Organization, November 10, 2022, https://www.nato.int/cps/en/natohq/opinions_208925.htm.

[45] Another important oversight is the role of the private sector, such as Microsoft, Google, Starlink, and others, in enabling the defense and resilience of Ukraine in cyberspace. While we do not focus specifically on these actors in our dataset, this is another critical issue that experts did not anticipate.

Sandworm	1
UAC-0010	5
UAC-0020	1
UAC-0026	1
UAC-0035	1
UAC-0041	2
UAC-0056	12
UAC-0088	1
UAC-0094	1
UAC-0097	1
UAC-0098	1
UAC-0098, TrickBot/Conti	2
UAC-0101	1
UAC-0104	1
UNC1151	2
Unknown	9
Unknown Russian	6
Total incidents	61

The most active known threat actor in Ukraine during our time frame of analysis has been UAC-0056. It has launched 12 cyber espionage attacks against various entities in Ukraine, the majority of which have been government actors.[46] UAC-0056 phishing campaign lures are often based on important matters related to the war, with one email titled: "Help Ukraine."[47]

Additionally, our research indicates that threat actors appear to concentrate on certain targets and types of cyber campaigns. For example, the now-defunct ransomware group Conti[48] used a variety of tactics to break into victim networks, "including via spear-phishing campaigns, stolen Remote Desktop Protocol credentials, software vulnerabilities, and poisoned software."[49] Anonymous has a strong proclivity for launching cyber-enabled espionage operations, and it also targets private entities two-thirds of the time. All of the cyber events launched by Killnet, on the other hand, are cyber effects operations, and government entities are its main focus.[50] Additionally, Killnet relies on its low-skill, low-threat tactic of distributed denial-of-service attacks

[46] Georgia was also targeted by UAC-0056 in a COVID-19-themed campaign in March 2021.

[47] Malware Bytes, "Cobalt Strikes Again: UAC-0056 Continues to Target Ukraine in its Latest Campaign," July 13, 2022, https://www.malwarebytes.com/blog/threat-intelligence/2022/07/cobalt-strikes-again-uac-0056-continues-to-target-ukraine-in-its-latest-campaign.

[48] Conti members are still active despite the group's breakup.

[49] DarkReading, "Post-Breakup: Conti Ransomware Members Remain Dangerous," July 19, 2022, https://www.darkreading.com/attacks-breaches/breakup-conti-ransomware-members-dangerous.

[50] Of Killnet's 24 recorded cyber events, all 24 (100%) were disruption cyber operations.

to cause disruption, rallying supporters to contribute to the group's efforts to prevent the average user from accessing websites in countries that support Ukraine or oppose Russia. The majority of the threat actors recorded in our data, however, appear to prefer cyber intelligence operations over cyber effects operations.

5. POLICY IMPLICATIONS

An important policy implication of our analysis is that there are constraints on the "combat power" of cyber operations.[51] Much of the cyber activity seen in Ukraine has been intelligence operations or disruptive attacks—and the effects of the latter are limited and transient, with targets rapidly recovering and restoring functionality.[52] Other cyber operations seen in the conflict, such as destructive malware or ransomware, have not generated costs severe enough to lead to meaningful changes in the overall conflict, especially when compared to the violence wrought by conventional kinetic capabilities.[53]

This reality, coupled with the complexities associated with planning and implementing cyber campaigns, means that, especially in a wartime context, "offensive cyber will always be a fragile capability."[54] Cyber conflict in Ukraine has manifested itself in waves that correspond to cyber access and capability development times, potentially in ways that are mismatched to the actions on the battlefield or the overall conventional campaign plan. The fog of war complicates operational planning; the violence of kinetic conflict destroys key capabilities and forces personnel to focus on their own physical security; and belligerents no longer enjoy the luxury of long time horizons. These challenges do not necessarily indicate weakness or incompetence; they illuminate how difficult it is to integrate cyber capabilities into military planning. Therefore, policymakers should avoid overconfidence in drawing inferences about the (im)maturity of Russian cyber capabilities, as well as the feasibility of implementing their own operational concepts in a warfighting environment.

This also relates to the proliferation of non-state actors engaged in the digital battlespace. Specifically, the difficulties states experience conducting sophisticated, large-scale cyber operations synchronized with conventional campaigns could

[51] See Erik Gartzke, "The Myth of Cyberwar," *International Security* 38, no. 2 (Fall 2013): 41–73; and Thomas Rid, *Cyber War Will Not Take Place* (Oxford: Oxford University Press, 2013).

[52] Max Smeets and Herbert S. Lin, "Offensive Cyber Capabilities: To What Ends?" in *2018 10th International Conference on Cyber Conflict CyCon X: Maximising Effects*, eds. T. Minárik, R. Jakschis, L. Lindström (Tallinn: NATO CCD COE Publications, 2018), 55–71, https://ccdcoe.org/uploads/2018/10/CyCon_2018_Full_Book.pdf.

[53] Brandon Valeriano, Benjamin Jensen, and Ryan C. Maness, *Cyber Strategy: The Evolving Character of Power and Coercion* (New York: Oxford University Press, 2018); Max Smeets, "A Matter of Time: On the Transitory Nature of Cyberweapons," *Journal of Strategic Studies* 41, nos. 1–2 (2018): 6–32; Borghard and Lonergan, "The Logic of Coercion in Cyberspace."

[54] Defense Science Board, "Resilient Military Systems and the Advanced Cyber Threat," 2013, 49, https://nsarchive.gwu.edu/sites/default/files/documents/2700168/Document-81.pdf.

make it more appealing to permit, encourage, or condone third-party cyber activity. Policymakers should anticipate that future conflicts are likely to have a third-party cyber component. The fact that most of the cyber spillover in Ukraine has been conducted by these actors, rather than directly by a government, raises vexing questions about how to prepare for and defend against this kind of activity and ensure guardrails are in place to prevent these attacks from inadvertently escalating conflicts.

ACKNOWLEDGMENTS

We are grateful to Brandon Valeriano and Gavin Wilde for reviewing our codebook and data in the earlier stages of our project. We are also thankful for the expert input and feedback from the participants at the University of California San Diego workshop "Cyber Escalation in Conflict: Bridging Policy, Data, and Theory."

REFERENCES

Alexander, Keith. "Cyber Warfare in Ukraine Poses a Threat to the Global System." *Financial Times*, February 15, 2022. https://www.ft.com/content/8e1e8176-2279-4596-9c0f-98629b4db5a6.

Antoniuk, Daryna. "Ukraine's State-Owned Nuclear Power Operator Said Russian Hackers Attacked Website." *Record*, August 17, 2022. https://therecord.media/ukraines-state-owned-nuclear-power-operator-said-russian-hackers-attacked-website/.

Atkins, Sean. "A Web of Partnerships: Ukraine, Operational Collaboration, and Effective National Defense in Cyberspace." Atlantic Council, August 30, 2022. https://www.atlanticcouncil.org/content-series/airpower-after-ukraine/a-web-of-partnerships-ukraine-operational-collaboration-and-effective-national-defense-in-cyberspace/.

Bateman, Jon. "Russia's Wartime Cyber Operations in Ukraine: Military Impacts, Influences, and Implications." Carnegie Endowment for International Peace, December 16, 2022. https://carnegieendowment.org/2022/12/16/russia-s-wartime-cyber-operations-in-ukraine-military-impacts-influences-and-implications-pub-88657.

Bebber, R. "Information War and Rethinking Phase 0." *Journal of Information Warfare* 15, no. 2 (2016): 39-52.

Borghard, Erica D., and Shawn W. Lonergan. "The Logic of Coercion in Cyberspace." *Security Studies* 26, no. 3 (2017): 452–81.

Buchanan, Ben. *The Cybersecurity Dilemma: Hacking, Trust, and Fear Between Nations*. New York: Oxford University Press, 2016.

Congressional Research Service. "Russia's War in Ukraine: Military and Intelligence Aspects." September 14, 2022. https://crsreports.congress.gov/product/pdf/R/R47068.

Cooper, Helene. "Pentagon Gives a Grim Assessment of the First Stages of the Russian Invasion." *New York Times*, February 24, 2022. https://www.nytimes.com/2022/02/24/us/politics/pentagon-russia-ukraine.html.

Cyber Peace Institute. "Cyber Dimensions of the Armed Conflict in Ukraine: Quarterly Analysis Report—Q3 July to September 2022." December 16, 2022.

DarkReading. “Post-Breakup: Conti Ransomware Members Remain Dangerous.” July 19, 2022. https://www.darkreading.com/attacks-breaches/breakup-conti-ransomware-members-dangerous.

Department of Defense Defense Science Board. “Resilient Military Systems and the Advanced Cyber Threat.” 2013. https://nsarchive.gwu.edu/sites/default/files/documents/2700168/Document-81.pdf.

Dilanian, Ken, and Courtney Kube. “Biden Has Been Presented with Options for Massive Cyberattacks against Russia.” NBC News, February 24, 2022. https://www.nbcnews.com/politics/national-security/biden-presented-options-massive-cyberattacks-russia-rcna17558.

DTEK Energy. “Enemy Launches Hacker Attacks on the Power System.” July 1, 2022. https://dtek.com/en/media-center/news/vslid-za-raketnimi-udarami-po-tes-vorog-zavdae-khakerskikh-udariv-po-energosistemi/.

———. “Once Again One of DTEK Energy Facilities Got Hit as a Result of Russian Attack. It Stopped Generating Electricity.” December 24, 2022. https://dtek.com/en/media-center/news/rosiya-znovu-obstrilyala-odin-z-energetichnikh-obektiv-dtek-vin-pripiniv-generatsiyu-elektroenergii/.

DTEK Group (@dtek_en). “The Russian Federation has carried out a #cyber attack on #DTEKGroup’s #IT infrastructure. 1/4.” Twitter, July 1, 2022. https://twitter.com/dtek_en/status/1542884325015830528?s=20&t=o_Qz2y72EVvYBC8gAHDLMQ.

Faulconbridge, Guy, James Davey, and Kate Holton. “Brace for Russian Cyber-Attacks as Ukraine Crisis Deepens, Britain Says.” Reuters, January 28, 2022. https://www.reuters.com/world/europe/brace-russian-cyber-attacks-over-ukraine-britain-says-2022-01-28/.

Gartzke, Erik. “The Myth of Cyberwar.” *International Security* 38, no. 2 (Fall 2013): 41–73.

Giles, Keir. “Putin Does Not Need to Invade Ukraine to Get His Way.” Chatham House, December 21, 2021. https://www.chathamhouse.org/2021/12/putin-does-not-need-invade-ukraine-get-his-way.

Graff, Garrett M. “The US Watches Warily for Russia-Ukraine Tensions to Spill Over.” *Wired*, February 15, 2022. https://www.wired.com/story/russia-ukraine-cyberattacks-spillover/.

Greenberg, Andy. “Russia’s Sandworm Hackers Attempted a Third Blackout in Ukraine.” *Wired*, April 12, 2022. https://www.wired.com/story/sandworm-russia-ukraine-blackout-gru/.

———. *Sandworm: A New Era of Cyberwar and the Hunt for the Kremlin’s Most Dangerous Hackers*. New York: Doubleday, 2019.

Greig, Jonathan. “Pro-Russia Hackers Use Telegram, Github to Attack Czech Presidential Election.” *Record*, January 12, 2023. https://therecord.media/pro-russia-hackers-use-telegram-github-to-attack-czech-presidential-election/

Kofman, Michael, and Jeffrey Edmonds. “Russia’s Shock and Awe: Moscow’s Use of Overwhelming Force against Ukraine.” *Foreign Affairs*, February 22, 2022. https://www.foreignaffairs.com/articles/ukraine/2022-02-21/russias-shock-and-awe.

Kostyuk, Nadiya, and Yuri M. Zhukov. “Invisible Digital Front: Can Cyber Attacks Shape Battlefield Events?” *Journal of Conflict Resolution* 63, no. 2 (2019): 317–347.

Lewis, James A. “Cyber War and Ukraine.” CSIS, June 16, 2022. https://www.csis.org/analysis/cyber-war-and-ukraine.

Lonergan, Erica D. “The Cyber-Escalation Fallacy: What the War in Ukraine Reveals about State-Based Hacking.” *Foreign Affairs*, April 15, 2022. https://www.foreignaffairs.com/articles/russian-federation/2022-04-15/cyber-escalation-fallacy.

Lonergan, Erica D., Shawn W. Lonergan, Brandon Valeriano, and Benjamin Jensen. "Putin's Invasion of Ukraine Didn't Rely on Cyberwarfare. Here's Why." *Washington Post*, March 7, 2022. https://www.washingtonpost.com/politics/2022/03/07/putins-invasion-ukraine-didnt-rely-cyber-warfare-heres-why/.

Lonergan, Erica D., and Brandon Valeriano. "What Ukraine Shows about Cyber Defense and Partnerships." *National Interest*, March 17, 2022. https://nationalinterest.org/blog/techland-when-great-power-competition-meets-digital-world/what-ukraine-shows-about-cyber.

Mallin, Alexander, and Luke Barr, "DOJ Official Warms Companies 'Foolish' Not to Share Up Cybersecurity Amid Russia Tensions." *ABC News*, February 17, 2022. https://abcnews.go.com/Politics/doj-official-warns-companies-foolish-shore-cybersecurity-amid/story?id=82959520.

Malware Bytes. "Cobalt Strikes Again: UAC-0056 Continues to Target Ukraine in Its Latest Campaign." July 13, 2022. https://www.malwarebytes.com/blog/threat-intelligence/2022/07/cobalt-strikes-again-uac-0056-continues-to-target-ukraine-in-its-latest-campaign.

Marks, Joseph, and Aaron Schaffer. "Here's What Cyber Pros Are Watching in the Ukraine Conflict." *Washington Post*, February 24, 2022. https://www.washingtonpost.com/politics/2022/02/24/heres-what-cyber-pros-are-watching-ukraine-conflict/.

Maschmeyer, Lennart, and Myriam Dunn Cavelty. "Goodbye Cyberwar: Ukraine as a Reality Check." *CSS ETH Zurich Policy Perspectives* 10, no. 3 (May 2022): 1–4.

Maschmeyer, Lennart, and Nadiya Kostyuk. "There Is No Cyber 'Shock And Awe': Plausible Threats in the Ukraine Conflict." War on the Rocks, February 8, 2022. https://warontherocks.com/2022/02/there-is-no-cyber-shock-and-awe-plausible-threats-in-the-ukrainian-conflict/.

Microsoft. "Defending Ukraine: Early Lessons from the Cyber War." June 22, 2022. https://query.prod.cms.rt.microsoft.com/cms/api/am/binary/RE50KOK.

———. "Microsoft Digital Defense Report 2022: Illuminating the Threat Landscape and Empowering a Digital Defense." November 4, 2022. https://query.prod.cms.rt.microsoft.com/cms/api/am/binary/RE5bUvv?culture=en-us&country=us.

Miller, Maggie. "NATO Prepares for Cyber War." *Politico*, December 3, 2022. https://www.politico.com/news/2022/12/03/nato-future-cyber-war-00072060.

———. "Russian Invasion of Ukraine Could Redefine Cyber Warfare." *Politico*, January 28, 2022. https://www.politico.com/news/2022/01/28/russia-cyber-army-ukraine-00003051.

NATO. "Keynote Address by NATO Secretary General Jens Stoltenberg at the NATO Cyber Defence Pledge Conference in Italy." November 10, 2022. https://www.nato.int/cps/en/natohq/opinions_208925.htm.

———. "NATO Secretary General Warns of Growing Cyber Threat." November 11, 2022. https://www.nato.int/cps/en/natohq/news_208889.htm?selectedLocale=en.

O'Neill, Patrick H. "Russian Hackers Tried to Bring Down Ukraine's Power Grid to Help the Invasion." *MIT Technology Review*, April 12, 2022. https://www.technologyreview.com/2022/04/12/1049586/russian-hackers-tried-to-bring-down-ukraines-power-grid-to-help-the-invasion/.

Radio Free Europe. "Russian Shelling Hits Residential Area in Ukrainian Town Near Nuclear Plant." August 20, 2022. https://www.rferl.org/a/mykolayiv-ukraine-russian-shelling-nuclear-power-plant/31997199.html.

Rid, Thomas. *Cyber War Will Not Take Place*. Oxford: Oxford University Press, 2013.

Rovner, Joshua. "Cyber War as an Intelligence Contest." War on the Rocks, September 16, 2019. https://warontherocks.com/2019/09/cyber-war-as-an-intelligence-contest/.

Smeets, Max. "A Matter of Time: On the Transitory Nature of Cyberweapons." *Journal of Strategic Studies* 41, nos. 1–2 (2018): 6–32.

Smeets, Max, and Herbert S. Lin. "Offensive Cyber Capabilities: To What Ends?" In *2018 10th International Conference on Cyber Conflict CyCon X: Maximising Effects*, edited by T. Minárik, R. Jakschis, L. Lindström, 55–71, Tallinn: NATO CCD COE Publications, 2018.

Smith, Maggie, Erica D. Lonergan, and Nick Starck. "What Impact, if Any, Does Killnet Have?" Lawfare, October 21, 2022. https://www.lawfareblog.com/what-impact-if-any-does-killnet-have.

Sabbagh, Dan. "UK Firms Warned of Russian Cyberwar 'Spillover' from Ukraine." *Guardian*, February 23, 2022. https://www.theguardian.com/technology/2022/feb/23/uk-firms-warned-russia-cyberwar-spillover-ukraine-critical-infrastructure.

U.S. Department of Defense. "Joint Publication 3-0: Joint Operations." U.S. Department of Defense, January 17, 2017. https://irp.fas.org/doddir/dod/jp3_0.pdf.

Valeriano, Brandon, Benjamin Jensen, and Ryan C. Maness. *Cyber Strategy: The Evolving Character of Power and Coercion*. New York: Oxford University Press, 2018.

Valeriano, Brandon, and Ryan C. Maness. "The Dynamics of Cyber Conflict Between Rival Antagonists, 2001–11." *Journal of Peace Research* 51, no. 3 (2014): 347–360.

Volz, Dustin, and Robert McMillan. "In Ukraine, a 'Full-Scale Cyberwar' Emerges." *Wall Street Journal*, April 12, 2022. https://www.wsj.com/articles/in-ukraine-a-full-scale-cyberwar-emerges-11649780203.

Wilde, Gavin. "Cyber Operations in Ukraine: Russia's Unmet Expectations." Carnegie Endowment for International Peace, December 12, 2022. https://carnegieendowment.org/2022/12/12/cyber-operations-in-ukraine-russia-s-unmet-expectations-pub-88607.

WTOP News. "What Are Putin's Intentions in Ukraine?" January 23, 2022. https://wtop.com/europe/2022/01/what-are-putins-intentions-in-ukraine/.

Zabrodskyi, Mykhaylo, Jack Watling, Oleksandr V. Danylyuk, and Nick Reynolds. "Preliminary Lessons in Conventional Warfighting from Russia's Invasion of Ukraine: February–July 2022." RUSI, November 30, 2022. https://rusi.org/explore-our-research/publications/special-resources/preliminary-lessons-conventional-warfighting-russias-invasion-ukraine-february-july-2022.

Zhora, Victor (@VZhora). July 1, 2022. https://twitter.com/VZhora/status/1542858906560512000.

2023 15th International Conference on Cyber Conflict
Meeting Reality
T. Jančárková, D. Giovannelli, K. Podiņš, I. Winther
2023 © NATO CCDCOE Publications, Tallinn

The Irregulars: Third-Party Cyber Actors and Digital Resistance Movements in the Ukraine Conflict

Margaret W. Smith
Director, Cyber Project
Irregular Warfare Initiative
Washington, DC, United States
maggie.smith@irregularwarfare.org

Thomas Dean
Senior Engineer
Booz Allen Hamilton
Cambridge, MA, United States
dean_thomas@ne.bas.com

Abstract: The Russian invasion of Ukraine and the subsequent rise of the IT Army of Ukraine (IT Army) illustrate how an organized non-military, all volunteer, and multinational digital resistance movement can impact an ongoing conflict. For nation-states, third-party actors in cyberspace pose a different set of concerns than state-affiliated or state-sponsored cyber actors and are the latest incarnation of underground auxiliary forces that bring unconventional tactics to bear in a conventional conflict. To investigate the IT Army's role in the war in Ukraine, we created a dataset of the content collected from the "IT ARMY of Ukraine" public channel on the messaging app Telegram. The channel provides the most up-to-date information on proposed targets, why those targets are important, and if attacks are deemed successful, in both Ukrainian and English. Through the lens of nonviolent action strategy and theory, we assess the IT Army's effectiveness as a resistance movement, defined by Joint Publication 3-05, "Special Operations," as "an organized effort by some portion of the civil population of a country to resist the legal established government or an occupying power and to disrupt civil order and stability." The dataset enables us to develop a more complete picture of the IT Army's evolution as a digital resistance movement since its creation on February 26, 2022, to assess how it incorporates nonviolent action strategy and elements to manage the over 200,000 volunteers and concentrate resources and strength to disrupt Russian domestic targets in and through cyberspace.

Keywords: *digital resistance movements, Ukraine, third-party cyber actors, nonviolent action, cyberspace*

1. INTRODUCTION

Technology is ubiquitous in warfare, but one aspect not well understood is the impact of *modern* technology on nonviolent resistance movements that arise in response to, or during, a conventional conflict. Today, everyone with internet access or a smartphone is a potential resistance fighter and, with lower barriers to entry, a host of functions like intelligence reporting, fire coordination, and rapid information sharing can be carried out by anyone with a smartphone.[1] The Russian invasion of Ukraine on February 24, 2022, and the near-immediate rise of the IT Army of Ukraine—a group of volunteers taking nonviolent action against Russian targets in and through cyberspace—illustrate how an organized digital resistance movement can operate effectively during a conventional conflict to cause disruption to, and denial of, adversary domestic services through nonviolent action. For nation states, third-party actors in cyberspace pose a different set of concerns than state-affiliated or state-sponsored military cyber actors and are the latest instantiation of underground auxiliary forces that bring unconventional tactics to bear in a conventional conflict.

While the war in Ukraine is primarily a high-end conventional war, several cyber activities are being carried out on behalf of both parties to the conflict. These operations have ranged in sophistication from disorganized hacktivism and patriotic hacking (like the nuisance cyber operations carried out by Killnet, Anonymous, and others) to the more destructive cyber attacks conducted by Russian-government-linked advanced persistent threat actors targeting critical infrastructure (like the hacker group Sandworm).[2] Additionally, much of the third-party actor activity has occurred *external to* the theater of operations, with malicious cyber actors targeting NATO member states with disruptive, but low impact, cyber attacks.[3] War coverage has also largely concentrated on the kinetic aspects of the conflict, but analysis of cyber incidents has largely focused on the lack of a Russian cyber "shock and awe"

[1] Andrew Maher and Martijn Kitzen, "On Resistance: A Primer for Further Research," Modern War Institute, September 8, 2022, https://mwi.usma.edu/on-resistance-a-primer-for-further-research/.

[2] Maggie Smith et al., "What Impact, if Any, Does Killnet Have?" *Lawfare*, October 21, 2022, https://www.lawfareblog.com/what-impact-if-any-does-killnet-have; "Defending Ukraine: Early Lessons from the Cyber War," Microsoft, June 22, 2022, https://query.prod.cms.rt.microsoft.com/cms/api/am/binary/RE50KOK; "Microsoft Digital Defense Report 2022: Illuminating the Threat Landscape and Empowering a Digital Defense," Microsoft, November 4, 2022, https://query.prod.cms.rt.microsoft.com/cms/api/am/binary/RE5bUvv?culture=en-us&country=us; "Cyber Dimensions of the Armed Conflict in Ukraine: Quarterly Analysis Report – Q3 July to September 2022," Cyber Peace Institute, December 16, 2022.

[3] Maggie Miller, "NATO Prepares for Cyber War," *Politico*, December 3, 2022, https://www.politico.com/news/2022/12/03/nato-future-cyber-war-00072060; "NATO Secretary General warns of growing cyber threat," North Atlantic Treaty Organization, November 11, 2022, https://www.nato.int/cps/en/natohq/news_208889.htm?selectedLocale=en.

campaign in the early phases of the war, as well as the surprising ineffectiveness of Russian offensive cyber operations throughout the conflict.[4]

In this paper, we move beyond the current analysis on resistance movements, nonviolent action, and the cyber activity related to the war in Ukraine to investigate the IT Army of Ukraine (hereinafter "IT Army") and its role in the conflict. To that end, we evaluate the IT Army's activities using the key elements of nonviolent strategy and tactics, as developed by political scientist Gene Sharp,[5] to assess its effectiveness as a nonviolent resistance movement operating in and through cyberspace. We define the IT Army's activities as "disruptive," relying on the categories of resistance activities outlined by sociologist Zeynep Tufekci. We further use the US military doctrine—Joint Publication 3-05, *Special Operations*—to define a resistance movement as "an organized effort by some portion of the civil population of a country to resist the legally established government or an occupying power and to disrupt civil order and stability."[6] To investigate the IT Army, we downloaded the contents of the "IT ARMY of Ukraine" channel on the messaging app Telegram (i.e., the posts and all residual data, like attached images or files) to create a unique dataset.

Importantly, we extend the research on digital resistance groups and third-party cyberspace actors to an entity born of war. To date, most research on third-party cyber actors has occurred *after* a hack is discovered. The reason is that many groups, like Anonymous and REvil,[7] do not discuss operations in a public forum, preferring to keep their activities out of public view. Even a group like Killnet, which gloats about its activities and shares boastful videos on its public forums, does so after the given hack.[8] Additionally, research has largely focused on groups that are external to any

[4] Lennart Maschmeyer and Nadiya Kostyuk, "There Is No Cyber 'Shock and Awe': Plausible Threats in the Ukraine Conflict," *War on the Rocks*, February 8, 2022, https://warontherocks.com/2022/02/there-is-no-cyber-shock-and-awe-plausible-threats-in-the-ukrainian-conflict/; Erica D. Lonergan, Shawn W. Lonergan, Brandon Valeriano, and Benjamin Jensen, "Putin's Invasion of Ukraine Didn't Rely on Cyberwarfare. Here's Why," *Washington Post*, March 7, 2022, https://www.washingtonpost.com/politics/2022/03/07/putins-invasion-ukraine-didnt-rely-cyber-warfare-heres-why/; Lennart Maschmeyer and Myriam Dunn Cavelty, "Goodbye Cyberwar: Ukraine as a Reality Check," *CSS ETH Zurich Policy Perspectives* 10, no. 3 (May 2022): 1–4; Jon Bateman, "Russia's Wartime Cyber Operations in Ukraine: Military Impacts, Influences, and Implications," Carnegie Endowment for International Peace, December 16, 2022, https://carnegieendowment.org/2022/12/16/russia-s-wartime-cyber-operations-in-ukraine-military-impacts-influences-and-implications-pub-88657; Gavin Wilde, "Cyber Operations in Ukraine: Russia's Unmet Expectations," Carnegie Endowment for International Peace, December 12, 2022, https://carnegieendowment.org/2022/12/12/cyber-operations-in-ukraine-russia-s-unmet-expectations-pub-88607; Sean Atkins, "A Web of Partnerships: Ukraine, Operational Collaboration, and Effective National Defense in Cyberspace," Atlantic Council, August 30, 2022, https://www.atlanticcouncil.org/content-series/airpower-after-ukraine/a-web-of-partnerships-ukraine-operational-collaboration-and-effective-national-defense-in-cyberspace/; Erica D. Lonergan and Brandon Valeriano, "What Ukraine Shows about Cyber Defense and Partnerships," *National Interest*, March 17, 2022, https://nationalinterest.org/blog/techland-when-great-power-competition-meets-digital-world/what-ukraine-shows-about-cyber.

[5] Gene Sharp, *The Politics of Nonviolent Action*, 9th edition (Boston: Albert Einstein Institution, 2020).

[6] Joint Publication 3-05: Special Operations," U.S. Department of Defense, September 22, 2020, https://jdeis.js.mil/my.policy, GL-8.

[7] Jonathan Grieg, "Researchers Warn of REvil Return After January Arrests in Russia," *Record*, May 16, 2022, https://therecord.media/researchers-warn-of-revil-return-after-january-arrests-in-russia/.

[8] https://t.me/killnet_reservs/4759.

declared conflict, meaning that their formation is not linked to an attack on a nation's sovereignty or prompted by another state's aggressive actions. While many third-party actors have publicly declared support to one side or the other, the IT Army exists *because* Russia invaded. It is the first example of a nonviolent digital resistance movement formed out of war, acting as a third party to the conflict, carrying out its operations[9] wholly online, and having no intent of using its online activity to incite or coordinate physical violence on its behalf.

This paper first discusses resistance movements and nonviolent action to introduce Sharp's key elements for successful nonviolent action. Then we investigate the "IT ARMY of Ukraine" using data collected from the contents of its Telegram channel, with a specific focus on the patterns of behavior, themes, and group norms to assess organizational effectiveness and the impact assessments posted to assess the role of nonviolent digital resistance groups in modern warfare. We conclude by offering implications for policymakers to consider and areas of future research.

2. RESISTANCE MOVEMENTS AND NONVIOLENT ACTION

Modern history is replete with stories of resistance—where there are aggressors and occupiers, there is likely a resistance. Ukraine has actively fought Russia's aggression via several means, with the media placing heavy emphasis on the active or physical forms of resistance[10] like the subtle, but highly symbolic, act of the "sunflower lady"[11] who offered a Russian soldier seeds to put in his pocket, telling him that they would grow when he fell in battle. However, much of the literature on resistance movements deals with its physical and violent instances or when online activity risks escalation to physical violence, as is covered extensively in the terrorism-studies literature. Similarly, the phenomenon of digital social movements[12] receives attention while the potential, and realized, impact of a nonviolent digital resistance *during* periods of conflict or occupation remains largely underrepresented and understudied.

[9] Here we are referring to its cyber operations and activities, not the administrative or organizational processes that likely occur among a group of persons known to each other and potentially in offline settings.

[10] Danny Moriarty, "Pockets of Sunflower Seeds: Civil Resistance in Ukraine," Modern War Institute, June 13, 2022, https://mwi.usma.edu/pockets-of-sunflower-seeds-civil-resistance-in-ukraine/.

[11] "Ukrainian Woman Offers Seeds to Russian Soldiers So 'Sunflowers Grow When They Die,'" *Guardian*, February 25, 2022, https://www.theguardian.com/world/video/2022/feb/25/ukrainian-woman-sunflower-seeds-russian-soldiers-video.

[12] Catherine Powell, "The Promise of Digital Action—And Its Dangers," Council on Foreign Relations, March 21, 2022, https://www.cfr.org/blog/promise-digital-activism-and-its-dangers.

Nonviolent action comes in many forms, but sociologist Zeynep Tufekci divides civil resistance tactics into three broad categories: narrative, institutional, and disruptive.[13] Narrative forms of resistance are focused on persuading domestic and international audiences and include community-building efforts and awareness-raising. Institutional forms of resistance work within a society's legitimate political spaces and can include rallying citizens to vote against proposed legislation or running an opposition campaign for elected office. Lastly, Tufekci defines disruptive tactics as those that interrupt the basic functions of society, like a distributed denial-of-service (DDoS) attack on a government website that prevents citizens from making tax payments or a physical protest that blocks traffic at a major city intersection. Based on Tufekci's typology, we classify the IT Army as a group largely engaged in disruptive activities that target the goods and services relied upon by the Russian people.

Furthermore, we define the IT Army's disruptive tactics as nonviolent, adopting Thomas Rid's view of offensive cyber operations as disassociated from physical violence, harm, or damage.[14] For the most part, Rid's narrowly physical view of violence has shaped much of the research in the field of offensive cyber operations.[15] However, others opt for a wider view and suggest that we should consider even nonlethal or nonphysical offensive cyber operations as a type of violence. Florian J. Egloff and James Shires highlight how some researchers view cyber operations as violent, "especially those that intentionally cause harm to the affective life of individuals or community values and identities."[16] From this wider perspective on violence, it follows that any threat of violence or coercion is considered violent because it impacts on the affective life and community and may introduce limits to freedom of action. We adopt the narrower definition of violence, viewing the IT Army's activities as nonviolent and indirect and their effects as disruptive and unlikely to cause physical harm, as opposed to kinetic alternatives.

[13] Zeynep Tufekci, *Twitter and Tear Gas: The Power and Fragility of Networked Protest* (New Haven: Yale University Press, 2017); Rob Mobley, "Twitter and Tear Gas: The Power and Fragility of Networked Protest," *Antipode: A Radical Journal of Geography*, January 8, 2019, https://antipodeonline.org/2019/07/08/twitter-and-tear-gas-the-power-and-fragility-of-networked-protest/.

[14] Thomas Rid, *Cyber War Will Not Take Place* (Oxford: Oxford University Press, 2012), 9. Emphasis in original. For additional discussion on his bodily conception of violence and cyber operations, see Thomas Rid, "More Attacks, Less Violence," *Journal of Strategic Studies* 36 (2013), 139–42, https://doi.org/10.1080/01402390.2012.742012.

[15] Erica D. Borghard and Shawn W. Lonergan, "The Logic of Coercion in Cyberspace," *Security Studies* 26, no. 3 (2017), 452–81, https://dx.doi.org/10.1080/09636412.2017.1306396; Richard J. Harknett and Joseph S. Nye, "Is Deterrence Possible in Cyberspace?" *International Security* 42, no. 2 (Fall 2017), 196–99, https://doi.org/10.1162/ISEC_c_00290; Erik Gartzke and Jon R. Lindsay, "Coercion through Cyberspace: The Stability-Instability Paradox Revisited," in *Coercion: The Power to Hurt in International Politics*, eds. Kelly M. Greenhill and Peter Krause (New York: Oxford University Press, 2018), 179–203; Erik Gartzke and Jon R. Lindsay, eds., *Cross-Domain Deterrence: Strategy in an Era of Complexity* (New York: Oxford University Press, 2019); Brandon Valeriano, Ryan C. Maness, and Benjamin Jensen, "Cyberwarfare Has Taken a New Turn. Yes, It's Time to Worry," *Washington Post*, July 13, 2017, https://www.washingtonpost.com/news/monkey-cage/wp/2017/07/13/cyber-warfare-has-taken-a-new-turn-yes-its-time-to-worry/.

[16] Florian J. Egloff and James Shires, "Offensive Cyber Capabilities and State Violence: Three Logics of Integration," *Journal of Global Security Studies* 7, no. 1 (2021), 4, https://doi.org/10.1093/jogss/ogab028.

3. THE IT ARMY OF UKRAINE

Triggered by the Russian invasion, Ukraine's deputy prime minister and minister for digital transformation, Mykhailo Fedorov, announced the creation of the IT Army[17] on February 26, 2022,[18] in a post to his official Telegram channel.[19] The impact was almost immediate, and the initial list of 31[20] Kremlin-affiliated banks, corporations, and Russian government agencies that had been posted as targets were temporarily knocked offline via a series of DDoS attacks.[21] Shortly after Federov's call to action, Viktor Zhora, deputy chief of Ukraine's State Service of Special Communication and Information Protection, indicated that IT Army volunteers were also providing intelligence and attacking military systems.[22] Unlike the long-standing Estonian Cyber Defense League, the IT Army's genesis was chaotic and ad hoc, because its creation was in response to the Russian invasion. As such, the IT Army is not intended to be a surge capacity to incorporate civilians into a military structured response should the need arise but is instead a creative response to Russian aggression that skillfully leverages the global cyber community's talent, computing power, and will to defend Ukrainian sovereignty through nonviolent resistance. "All these actions are directed to make the aggressor weaker, to make him understand, to deliver truth to the Russian people," Zhora told reporters, and continued to explain that the IT Army volunteers are "doing everything possible to protect our land in cyberspace."[23]

It is important to note that as of the writing of this paper, the IT Army consists of two main parts that Stefan Soesanto describes as

(i) Public-facing: a continuous global call to action that mobilizes anyone willing to participate in coordinated DDoS attacks against designated—primarily civilian—Russian infrastructure targets.
(ii) In-house: a team likely consisting of Ukrainian defense and intelligence personnel who have been experimenting with and conducting increasingly complex cyber operations against specific Russian targets.[24]

[17] While being a tech-savvy user is helpful, the IT Army provides a library of resources and how-to guides for new technologists looking to assist their efforts, allowing anyone with a computer to contribute. Our research into the IT Army's Telegram channel also included the "IT Army of Ukraine Chat" channel, which is a place for supporters to share tips, access help, and troubleshoot issues.

[18] https://t.me/itarmyofukraine2022/1.

[19] Matt Burgess, "Ukraine's Volunteer 'IT Army' is Hacking in Uncharted Territory," *Wired*, February 27, 2022, https://www.wired.com/story/ukraine-it-army-russia-war-cyberattacks-ddos/.

[20] https://t.me/itarmyofukraine2022/4.

[21] Dan Goodin, "After Ukraine Recruits an 'IT Army,' Dozens of Russian Sites Go Dark," *arstechnica*, February 28, 2022, https://arstechnica.com/information-technology/2022/02/after-ukraine-recruits-an-it-army-dozens-of-russian-sites-go-dark/.

[22] Sam Schechner, "Ukraine's 'IT Army' has Hundreds of Thousands of Hackers, Kyiv Says," *Wall Street Journal*, March 4, 2022, https://www.wsj.com/livecoverage/russia-ukraine-latest-news-2022-03-04/card/ukraine-s-it-army-has-hundreds-of-thousands-of-hackers-kyiv-says-RfpGa5zmLtavrot27OWX.

[23] Schechner, "Ukraine's 'IT Army.'"

[24] Stefan Soesanto, "The IT Army of Ukraine: Structure, Tasking, and Ecosystem," *Cyberdefense Report* (June 2022): 4, https://doi.org/10.3929/ethz-b-000552293.

Both parts of the IT Army are characterized as offensive, with the more professional in-house team conducting a large amount of the development (e.g., creating and updating tools, generating training guides, troubleshooting tool issues, identifying targets, and likely fulfilling any intelligence support functions) needed to buttress Ukraine's offensive cyber efforts and to enable the nearly 200,000 followers of the IT Army's Telegram channel.[25] In this paper, we focus on the vast army of volunteers conducting DDoS and other attacks against Russian targets and directed by the instructions and posts shared via Telegram.

4. THE DATA

Telegram is a convenient platform for research because the application has built-in search and translation functions, and a channel's content can be downloaded as a .csv or .json file. We downloaded the IT Army's channel content as a .json file on November 1, 2022. Notably, on October 2, 2022, the IT Army stopped publishing its daily target lists to the channel, after learning that Russian actors were using them to automate defenses. Thereafter, the administrators changed tactics and posted target lists to the IT Army's tool suite and limited their posts to non-sensitive items. Despite the change, we can still develop a picture of the IT Army's evolution, activity, and role in the Ukrainian conflict through its daily themes, battle damage assessments (BDAs), and its organization through the lens of nonviolent action theory.

After downloading the channel's contents, we parsed the data to extract target lists, posted in the form of either hyperlinks or IP addresses and ports, and indexed it chronologically. We also indexed keywords related to recruiting, IT Army tools, and type of target. Lastly, we indexed posts that included images of attack results, or BDA screenshots of disabled websites. The compiled descriptive statistics in Table I present the scope of the IT Army's operations up to November 1, 2022. In total, the IT Army hyperlinked 9,547 domains, of which 3,896 are unique (41 percent). We manually searched for discussions about IT Army tools and found five instances that referenced a tool update. Additionally, because the IT Army relies on the time and resources of volunteers, we manually searched for discussions on recruiting and personnel and found nine posts in which the IT Army called for support. Also of interest are the 584 posts that include screenshots, often of a website's error code or health status report, for the purpose of reporting BDAs and impact. In all, from February 26, 2022, to November 1, 2022, the IT Army published 840 posts.

[25] Soesanto, "The IT Army of Ukraine," 4.

TABLE I: DESCRIPTIVE STATISTICS OF THE IT ARMY'S TELEGRAM CHANNEL CONTENT

Category	Total
Hyperlinked domains (total)[26]	9,547
Unique domains	3,896
IP addresses	3,973
Images	923
Uncategorized images[27]	322
Unrelated news or branding[28]	17 posts
Battle damage assessment or impact statement	584 posts
Recruitment	9 posts
Retooling or tool updates	2 posts
Data snatch and leak campaigns	2 events

5. ANALYSIS OF THE IT ARMY OF UKRAINE

We assessed the IT Army as a nonviolent resistance movement, along the seven elements of nonviolent strategy and tactics identified by Gene Sharp:

(i) the indirect approach to the opponent's power,
(ii) psychological elements,
(iii) geographic and physical elements,
(iv) timing,
(v) numbers and strength,
(vi) the issue and concentration of strength, and
(vii) the initiative.[29]

The list, as Sharp emphasizes, is incomplete—in part because strategic analysis of nonviolent struggles is difficult and relies on nebulous, difficult-to-measure factors like motivation and will to fight. Additionally, nonviolent tactics and strategy, unlike

[26] This number represents the total number of domains shared by the IT Army up to November 1, 2022, irrespective of whether the domain was posted more than once. Of 9,457 total domains shared, 3,896 are unique.

[27] "Uncategorized images" refers to images containing Ukrainian text that requires translation. For future iterations of research, we will have the image text translated.

[28] "Unrelated news or branding" refers to images that are unrelated to IT Army cyber operations (e.g., screenshots of a news event or other unrelated topic).

[29] Gene Sharp, *The Politics of Nonviolent Action*, 493–500.

military tactics, are much more dependent on individuals and intangible factors unique to each situation and need to be taken "within the context of the unique dynamics and mechanisms of nonviolent struggle."[30] Therefore, the list of guiding principles remains provisional.

Nevertheless, strategy and tactics *are* important in nonviolent struggle—especially because modern movements that quickly form on digital platforms are often victims of their own speed. Tufekci cautions that

> ...with this speed comes weakness, some of it unexpected. First, these new movements find it difficult to make tactical shifts because they lack both the culture and the infrastructure for making collective decisions.... Second, although their ability (as well as their desire) to operate without defined leadership protects them from co-optation or "decapitation," it also makes them unable to negotiate with adversaries or even inside the movement itself. Third, the ease with which current social movements form often fails to signal an organizing capacity powerful enough to threaten authority.[31]

However, as the next sections demonstrate, the IT Army appears to be avoiding some of these common pitfalls through consistency, transparency, responsiveness, and encouragement.

A. The Indirect Approach

To begin, the IT Army's approach is an example of Liddell Hart's "indirect approach" to military strategy, as opposed to a direct approach or the application of kinetic force by conventional forces on the battlefield.[32] The IT Army applies pressure on Russian domestic services and institutions with "such indirectness as to ensure the opponent's unreadiness to meet it," and its disruptions are intended to evoke a reaction from the Russian population and to generate pressure on the Russian government—the main target of the resistance.[33] Sharp refers to Liddell Hart's indirect approach as a "kind of political *jiu-jitsu*" in which the use of nonviolent action against an opponent using military means actually undermines the very sources of political power that enable the military action in the first place.[34] Therefore, the IT Army's disruptive tactics serve to confront Russia's military aggression indirectly and, in doing so, can cause Russia's aggression toward Ukraine to rebound against the government via domestic pressure.

The daily target lists illustrate how the IT Army avoids targeting known positions of strength that likely employ DDoS defensive tools—e.g., military, intelligence, or similar websites—and instead opts for targets of opportunity in sectors that

[30] Gene Sharp, *The Politics of Nonviolent Action*, 496.

[31] Tufekci, *Twitter and Tear Gas*, 71.

[32] B. H. Liddell Hart, *Strategy* (New York: Praeger, 1954), 25.

[33] Liddell Hart, *Strategy*, 25.

[34] Gene Sharp, *The Politics of Nonviolent Action*, 496.

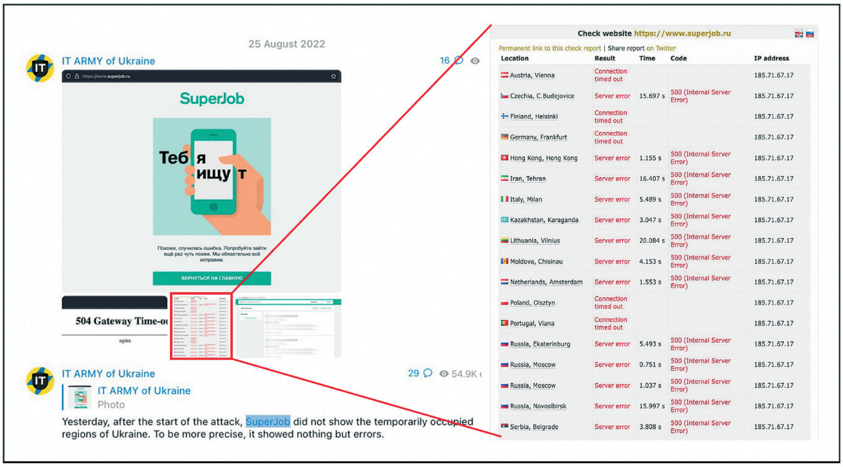
provide important domestic services. For example, the initial target list, posted on February 26, 2022, listed 31 domains of domestic importance, including media, oil and gas companies, and some government websites. On March 8, 2022, administrators encouraged supporters to "keep focusing on regional government websites"[35] and again on March 11, 2022, to "keep rolling with state companies… go for electronic signature services."[36] Over time, the targets diversified, ranging from lumber trading services[37] to the online payment system KoronaPay, but remain focused on Russian services that provide basic needs, thereby making the service either unavailable when the average Russian needs it or so unreliable that a task cannot be completed.[38] For example, the IT Army targeted Russian university admissions websites on June 20, 2022, the very day applications opened.[39]

B. The Psychological Elements

The second important aspect of nonviolent resistance is psychological. Sharp notes that morale is as important to war as it is to nonviolent action.[40] Intuitively, we can conclude that if group members feel valued and see that their actions are having an impact, they are more likely to feel motivated to continue their support.[41] The IT Army is beholden to the resources its volunteers bring to bear, and their posts reflect an organizational awareness of the role morale plays in group effectiveness. Particularly challenging for the in-house and administrative team is the IT Army's decentralized and largely anonymous composition. To foster community, the administrators actively communicate[42] attack outcomes[43] via screenshots, encourage[44] members to keep applying pressure on targets, and congratulate[45] successes. Figure 1 shows the BDA posted after a successful attack on a Russian job search website, SuperJob, which had listings for jobs in the four regions of Ukraine occupied by Russia in August 2022.[46]

[35] https://t.me/itarmyofukraine2022/147.
[36] https://t.me/itarmyofukraine2022/189.
[37] https://t.me/itarmyofukraine2022/246.
[38] https://t.me/itarmyofukraine2022/537.
[39] https://t.me/itarmyofukraine2022/446.
[40] Gene Sharp, *The Politics of Nonviolent Action*, 496.
[41] "APA Study Finds Feeling Valued at Work Linked to Well-Being and Performance," American Psychological Association, press release, March 8, 2012, https://www.apa.org/news/press/releases/2012/03/well-being.
[42] https://t.me/itarmyofukraine2022/255.
[43] https://t.me/itarmyofukraine2022/765.
[44] https://t.me/itarmyofukraine2022/775.
[45] https://t.me/itarmyofukraine2022/792.
[46] https://t.me/itarmyofukraine2022/542.

FIGURE 1: SCREENSHOT DISPLAYING THE HEALTH STATUS OF THE DOMAIN WWW.SUPERJOB.RU



25 August 2022

IT ARMY of Ukraine

16
https://www.superjob.ru
SuperJob
Теб я ищу т
Похоже, случилась ошибка. Попробуйте зайти ещё раз чуть позже. Мы обязательно всё исправим.
ВЕРНУТЬСЯ НА ГЛАВНУЮ
504 Gateway Time-ou
nginx

IT ARMY of Ukraine

IT ARMY of Ukraine
Photo

Yesterday, after the start of the attack, SuperJob did not show the temporarily occupied regions of Ukraine. To be more precise, it showed nothing but errors.

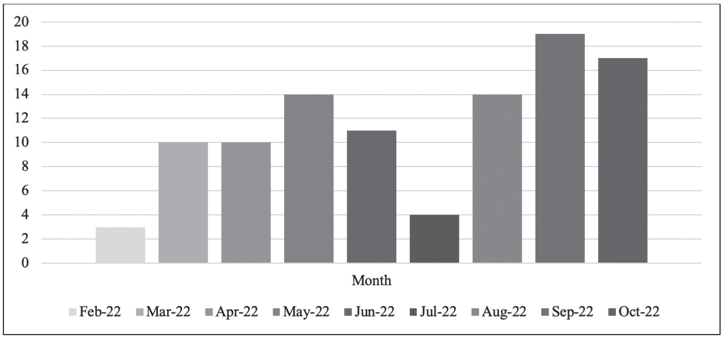
Check website https://www.superjob.ru

Permanent link to this check report | Share report on Twitter

Location	Result	Time	Code	IP address
Austria, Vienna	Connection timed out			185.71.67.17
Czechia, C.Budejovice	Server error	15.697 s	500 (Internal Server Error)	185.71.67.17
Finland, Helsinki	Connection timed out			185.71.67.17
Germany, Frankfurt	Connection timed out			185.71.67.17
Hong Kong, Hong Kong	Server error	1.155 s	500 (Internal Server Error)	185.71.67.17
Iran, Tehran	Server error	16.407 s	500 (Internal Server Error)	185.71.67.17
Italy, Milan	Server error	5.489 s	500 (Internal Server Error)	185.71.67.17
Kazakhstan, Karaganda	Server error	3.047 s	500 (Internal Server Error)	185.71.67.17
Lithuania, Vilnius	Server error	20.084 s	500 (Internal Server Error)	185.71.67.17
Moldova, Chisinau	Server error	4.153 s	500 (Internal Server Error)	185.71.67.17
Netherlands, Amsterdam	Server error	1.553 s	500 (Internal Server Error)	185.71.67.17
Poland, Olsztyn	Connection timed out			185.71.67.17
Portugal, Viana	Connection timed out			185.71.67.17
Russia, Ekaterinburg	Server error	5.493 s	500 (Internal Server Error)	185.71.67.17
Russia, Moscow	Server error	0.751 s	500 (Internal Server Error)	185.71.67.17
Russia, Moscow	Server error	1.037 s	500 (Internal Server Error)	185.71.67.17
Russia, Novosibirsk	Server error	15.997 s	500 (Internal Server Error)	185.71.67.17
Serbia, Belgrade	Server error	3.808 s	500 (Internal Server Error)	185.71.67.17

On average, the IT Army has posted 12 BDAs per month, with the highest volume occurring in September 2022, as seen in Figure 2. Posting BDA appears to boost morale and foster community bonding over the group's successes, but the administrators have also expressed frustration. On March 8, 2022, an exasperated administrator levied an ultimatum: "Time to figure out either you support the war and killing Ukrainians or go and fight against the criminal regime of Russian Federation. Nothing else on the table."[47] Yet the tone was quickly corrected, and positivity, encouragement, and snarkiness have followed. The administration candidly engages with the volunteer community, making the channel a fun place to be, as its posts are often humorous,[48] eliciting a sense of belonging and evoking a measure of humanity in an otherwise sterile online environment.

FIGURE 2: BATTLE DAMAGE ASSESSMENT (BDA) POSTS PER MONTH (FEBRUARY–OCTOBER 2022)



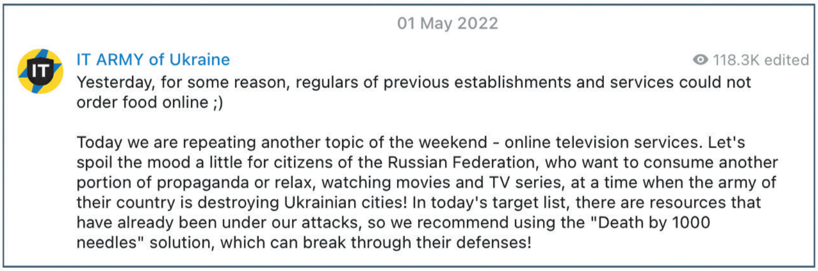
[47] https://t.me/itarmyofukraine2022/148.

[48] https://t.me/itarmyofukraine2022/763.

C. Geographic and Physical Elements

Sharp's third element of nonviolent action is related to geography and the physical elements of nonviolent action. Sharp explains that a "careful nonviolent strategist is likely to be attentive to the choice of the place at which given acts of opposition are to be undertaken."[49] Instead of considering the importance of a physical place, the IT Army often targets domains and IP addresses that have symbolic or temporal value to the Russian people. One example is the weekend themes (see Figure 3), which exhibit an understanding of what online services enable typical Russian leisure activities. On March 12, 2022, the IT Army encouraged its volunteers to target a list of food delivery services—the Russian equivalent of GrubHub—which Russians "so need" on the weekends.[50]

FIGURE 3: DISCUSSION OF WEEKEND THEMES



01 May 2022

IT ARMY of Ukraine

118.3K edited

Yesterday, for some reason, regulars of previous establishments and services could not order food online ;)

Today we are repeating another topic of the weekend - online television services. Let's spoil the mood a little for citizens of the Russian Federation, who want to consume another portion of propaganda or relax, watching movies and TV series, at a time when the army of their country is destroying Ukrainian cities! In today's target list, there are resources that have already been under our attacks, so we recommend using the "Death by 1000 needles" solution, which can break through their defenses!

Other popular weekend themes are television and streaming services[51] and the central website for purchasing movie tickets.[52] The targets are therefore symbolic as well as functional—the IT Army was effectively targeting the traditional outlets for Russian propaganda and state-run news.

D. Timing

Closely linked with Sharp's notion of geography is timing. Killnet, the pro-Russia hacking collective, is a group using reactive timing and tempo, having once "declared war" on 10 NATO member states for supporting the Ukraine war effort.[53] The IT Army takes a different approach and proactively selects targets of opportunity on important days or holidays to advertise the connection. For example, on May 9, 2022, Victory Day in Russia, the IT Army posted, "Let's switch to targets that are related to today's holiday in Russia," and included a long list of military domains and IP addresses.[54] In another example, on the Day of Knowledge, a holiday that celebrates

[49] Gene Sharp, *The Politics of Nonviolent Action*, 497.

[50] https://t.me/itarmyofukraine2022/197.

[51] https://t.me/itarmyofukraine2022/290.

[52] https://t.me/itarmyofukraine2022/477.

[53] Connor Jones, "Russian Hackers Declare War on 10 Countries after Failed Eurovision DDoS Attack," *ITPro*, May 16, 2022, https://www.itpro.co.uk/security/hacking/367685/russian-hackers-declare-war-on-10-countries-after-failed-eurovision-ddos.

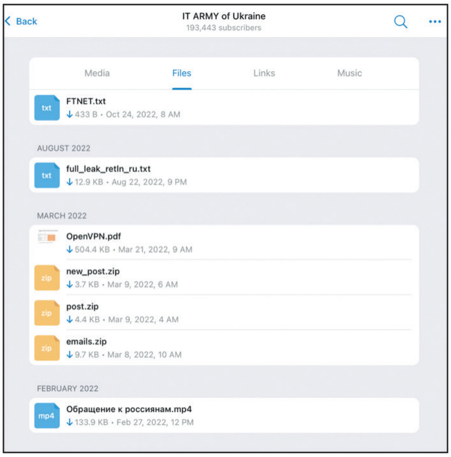
[54] https://t.me/itarmyofukraine2022/335.

the children of Ukraine and marks the official start of the academic year, the IT Army ensured that "everyone in occupied Crimea could see the greetings of the Ukrainian President. Simply because *Crimea is Ukraine*."[55] Both examples show how the IT Army leverages political and social context to assign meaning to its operations.

E. Numbers and Strength

The IT Army's narrow focus, friendly chat, easy-to-use tool set, and clear instructions (see Figure 4) have created an appealing way for thousands to contribute to the Ukraine war effort from anywhere and provide the mass required to disrupt Russian online targets.

FIGURE 4: RESOURCES AVAILABLE FOR VOLUNTEERS IN THE IT ARMY'S TELEGRAM CHANNEL



IT ARMY of Ukraine
193,443 subscribers

FTNET.txt
↓433 B • Oct 24, 2022, 8 AM

AUGUST 2022

full_leak_retln_ru.txt
↓12.9 KB • Aug 22, 2022, 9 PM

MARCH 2022

OpenVPN.pdf
↓504.4 KB • Mar 21, 2022, 9 AM

new_post.zip
↓3.7 KB • Mar 9, 2022, 6 AM

post.zip
↓4.4 KB • Mar 9, 2022, 4 AM

emails.zip
↓9.7 KB • Mar 8, 2022, 10 AM

FEBRUARY 2022

Обращение к россиянам.mp4
↓133.9 KB • Feb 27, 2022, 12 PM

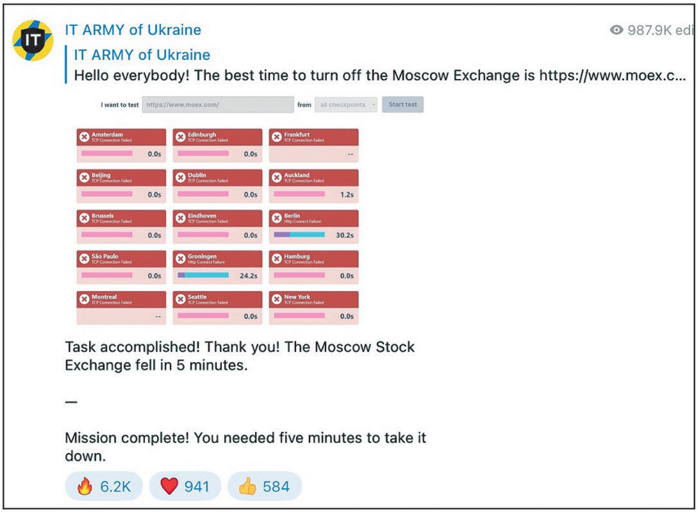
Since its inception, the IT Army has relied on its ability to mass resources against Russian targets to cause disruption. A dispersed resistance movement like the IT Army—whose membership has reached and now stabilized at roughly 200,000—can be effective only if it can leverage a critical mass of volunteers against a Russian target in a coordinated effort. Because the group is multinational and all-volunteer, the number of persons available for an operation at any given point in time is likely far less than 200,000, but the resource pooling potential remains significant. In many ways, the IT Army is unique among third-party actors in cyberspace because, despite its rapid growth, the organization has prevented mission creep and is committed to limiting its operations to the Russian-owned or controlled portions of the internet. It also sustains a largely professional discourse that is decidedly anti-Russia but more disciplined and inclusive than the language and content found, for example, in Killnet's forum.

[55] https://t.me/itarmyofukraine2022/577. Emphasis in original.

F. Concentration and Strength

However, while Sharp maintains that numbers and strength are critically important in both military and nonviolent action, he cautions that numbers alone may not be decisive.[56] For the IT Army, the element of nonviolent action that has proven essential is the "issue and concentration of strength."[57] A major challenge for the IT Army is sustaining the critical mass necessary to launch successful DDoS attacks, and in three periods—early March, mid-June, and the end of October 2022—they dedicated five posts to recruitment. During the group's first week, membership was growing fast, and, at the outset, the IT Army wanted anyone who showed interest in helping. It encouraged people to join[58] and asked them to contribute to the movement with comments like, "Hey IT Army! Internet Forces of Ukraine is a new impactful thing where everyone can contribute even having no IT background… Call everyone to join!"[59] The results were immediate, as shown in Figure 5.

FIGURE 5: BATTLE DAMAGE ASSESSMENT POST FROM FEBRUARY 28, 2022



IT ARMY of Ukraine

987.9K edi

IT ARMY of Ukraine

Hello everybody! The best time to turn off the Moscow Exchange is https://www.moex.c...

I want to test https://www.moex.com/ from all checkpoints Start test

☒ Amsterdam TCP Connection Failed 0.0s	☒ Edinburgh TCP Connection Failed 0.0s	☒ Frankfurt TCP Connection Failed --
☒ Beijing TCP Connection Failed 0.0s	☒ Dublin TCP Connection Failed 0.0s	☒ Auckland TCP Connection Failed 1.2s
☒ Brussels TCP Connection Failed 0.0s	☒ Eindhoven TCP Connection Failed 0.0s	☒ Berlin Http Connect Failure 30.2s
☒ São Paulo TCP Connection Failed 0.0s	☒ Groningen Http Connect Failure 24.2s	☒ Hamburg TCP Connection Failed 0.0s
☒ Montreal TCP Connection Failed --	☒ Seattle TCP Connection Failed 0.0s	☒ New York TCP Connection Failed 0.0s

Task accomplished! Thank you! The Moscow Stock Exchange fell in 5 minutes.

—

Mission complete! You needed five minutes to take it down.

🔥 6.2K ❤️ 941 👍 584

But after March 8, 2022, posts shifted to thanking volunteers for joining, and it was not until mid-June of that year that recruiting re-entered the conversation, giving the first indication that the IT Army's strength could be waning. So far, there has not been a clear pattern to the IT Army's recruitment efforts, but they appear to address the need as it arises.

[56] Gene Sharp, *The Politics of Nonviolent Action*, 498.

[57] Ibid., 499.

[58] https://t.me/itarmyofukraine2022/76.

[59] https://t.me/itarmyofukraine2022/124.

G. Initiative

Lastly, one of the more incredible feats of the IT Army is its ability to sustain the initiative. Sharp explains that "[t]he nonviolent leadership group needs to be able to control the situation and to demonstrate that it has control."[60] Moreover, Sharp emphasizes that, wherever possible, "the nonviolent group, not the opponent, will choose the time, issue and course of action and seek to maintain the initiative despite the opponent's repression."[61] While the IT Army has commented on events taking place on the battlefield, there are no signs that the ground war has disrupted or prevented the IT Army from launching an attack or inhibited the in-house leadership's ability to communicate with supporters. The result is a resistance movement that maintains its freedom of maneuver and operates largely independent of, or external to, conditions on the battlefield. Therefore, the IT Army can retain the initiative, apply persistent pressure on the Russian domestic space, and disrupt services for people who are otherwise physically removed from the conventional battlefield.

6. CONCLUSIONS AND FUTURE RESEARCH

The IT Army of Ukraine is a unique multinational, nonviolent resistance movement that leverages a creative structure to achieve operational impact in and through cyberspace and will likely inform cyber operational art in future conflicts. The employment of cyber operations in modern conflict is difficult, but the IT Army is using the internet's interconnectedness to achieve an asymmetric advantage over Russian domestic cybersecurity to disrupt services, websites, and corporations relied upon by the Russian people. Its public-facing Telegram channel coordinates roughly 200,000 volunteers and is a vehicle for the Ukrainian government to use to engage in persistent DDoS activities. As of November 1, 2022, the IT Army has targeted 3,896 unique domains and, while DDoS effects are temporary, their ability to target specific services means they can disrupt services at times when they are needed most and apply pressure on Russian domestic services and institutions. The in-house team likely remains directly connected to Ukrainian defense and intelligence agencies to facilitate targeting and to conduct the more boutique hacking operations, like the two instances of steal-and-leak operations we found evidence of in the Telegram data. In short, the IT Army has effectively employed disruptive tactics and nonviolent action on multiple occasions to interrupt the basic functions of Russian society and thus to be a nuisance to the Russian people. Their activities incorporate Gene Sharp's seven key elements of nonviolent action and, despite the difficulty of accurately assessing the IT Army's true impact, the 584 battle damage assessment posts and images show that they are causing disruptions at the times and places of their choosing.

[60] Gene Sharp, *The Politics of Nonviolent Action*, 500.

[61] Ibid., 500.

With media outlets sensationalizing the activities of malicious cyber actors like Killnet[62] and eager to attract readers, the story of the IT Army is a major human-interest story: everybody loves an underdog. But the IT Army's actions raise many questions about what self-defense and national defense mean in modern conflict.[63] Namely, the Ukrainian government has upended many assumptions and frameworks regarding the norms[64] for state behavior in cyberspace by enlisting the help of geographically dispersed volunteers. The IT Army challenges international legal frameworks too, as most volunteers are likely breaking laws in their home country by using their computer to hack in defense of Ukraine and largely at the request and direction of persons at least loosely tied with Ukrainian military and intelligence services. The IT Army's successes mean the international community cannot ignore their presence and role in the ongoing conflict in Ukraine, because doing so has implications for the future of international cyber norms, state behavior in cyberspace, and the national security landscape more broadly.

As an impactful nonviolent resistance movement, the IT Army of Ukraine raises several policy and legal questions. One unresolved question is what to do about the international volunteers. Are they a form of enemy combatant by virtue of their participation? Peter Pascucci and Kurt Sanger affirm that "any civilian seeking to impact the operations of a party to an armed conflict should be aware of the potential consequences of their participation."[65] But "participation" is hard to define, and even the International Committee of the Red Cross (ICRC) admits that the difference between "direct" and "indirect" participation "can be difficult to establish but is vital."[66] As Pascucci and Sanger note, the ICRC declares that "interfering electronically with military computer networks would qualify as directly participating," but beyond that, it is unclear whether a cyber operation carried out by a civilian would constitute direct participation. Even muddier is the reality that potentially thousands of supporters participate in creating a single effect. So who is really responsible—is it everyone involved? Another question not addressed in this paper is the IT Army's use of commercial infrastructure and platforms to host its administrative and operational resources. For example, the IT Army maintains GitHub repositories of its tools for volunteers to use to participate indirectly in the Ukrainian conflict. Does Microsoft, the owner of GitHub, have an acceptable use policy and, if so, what constitutes a violation? Other private companies continue to passively support the IT Army despite the same concerns about enabling participation in a declared conflict.

[62] Smith et al., "What Impact, if Any, Does Killnet Have?"

[63] Peter Pascucci and Kurt Sanger, "Cyber Norms in the Context of Armed Conflict," *Lawfare*, November 16, 2022, https://www.lawfareblog.com/cyber-norms-context-armed-conflict.

[64] Jay Healey and Olivia Grinberg, "Patriotic Hacking Is No Exception," Lawfare, September 27, 2022, https://www.lawfareblog.com/patriotic-hacking-no-exception.

[65] Pascucci and Sanger, "Cyber Norms in the Context of Armed Conflict."

[66] https://www.icrc.org/en/doc/resources/documents/faq/direct-participation-ihl-faq-020609.htm.

While the IT Army's activities are largely viewed as distinct from actions taken by the conventional Ukrainian military, additional research is needed to better understand how the IT Army augments or interferes with traditional military cyber operations. Because military cyber operations are highly classified, it is impossible to know at this time the extent to which the two may be coordinated, if at all. A related risk is that as the number of third-party actors increases, so too will the fog of war, making it even more difficult to distinguish between actors and their motivations. And with a capable and resourceful digital resistance group like the IT Army, there may be a risk of escalation if the group decides to become responsive to, or partake in, more conventional fighting activities. Ultimately, because the conflict in Ukraine was ongoing at the time of writing, future research on the IT Army needs to expand our initial dataset and incorporate the most recent months of data. This will help us further assess the IT Army's role in the war and build upon our knowledge of digital resistance movements and their relationship to conventional conflict.

2023 15th International Conference on Cyber Conflict
Meeting Reality
T. Jančárková, D. Giovannelli, K. Podiņš, I. Winther
2023 © NATO CCDCOE Publications, Tallinn

Analytical Review of the Resilience of Ukraine's Critical Energy Infrastructure to Cyber Threats in Times of War

Andrii Davydiuk
PhD Candidate
G. E. Pukhov Institute for Modelling in Energy Engineering NAS of Ukraine
Kyiv, Ukraine
andrey19941904@gmail.com

Vitalii Zubok
Doctor of Engineering
G. E. Pukhov Institute for Modelling in Energy Engineering NAS of Ukraine
Kyiv, Ukraine
vit@visti.net

Abstract: The Russia-Ukraine conflict has led to a significant increase in cyber attacks on critical infrastructure in Ukraine, with the energy sector being a primary target. The goal of these cyber attacks is to support military operations on the battlefield. Enhancing the resilience of the energy sector is a primary and urgent assignment for the security and defense sector of Ukraine.

This study aims to identify the cyber resilience factors of critical energy infrastructure and their possible dependencies and analyze the causes of their occurrence.

Accordingly, an analysis of the problems of the resilience of the critical energy infrastructure in Ukraine has been carried out. Based on this analysis, we have identified and studied some dependencies between cyber security for power energy infrastructure and other sectors, often referred to as cascade effects. By analyzing cause-and-effect relationships in power outages, the prerequisites for the emergence of negative factors affecting the resilience of critical infrastructure in the conditions of war have been determined.

Using the obtained information about cascade effects, procedures have been proposed to enhance resilience. These include implementing processes for collecting and processing big data on cyber statistics, optimizing public-private cooperation, and organizing cyber training.

The goal of these processes is to increase the level of cyber security for critical infrastructure. These processes are aimed at increasing the effectiveness of responding to cyber security crises in conditions of limited time and material resources.

The experience of Ukraine in conducting such research is unique. This can become the basis for the development of models and architectures for the resilience of electric power systems in other countries.

Keywords: *resilience, cyber attack, cyber security, Russia-Ukraine war, critical energy infrastructure*

1. OVERVIEW OF THE PROBLEM

The Russia-Ukraine war has a hybrid character. Cyber attacks pose a significant threat in this hybrid warfare [1]. In 2014, Russian forces invaded Ukraine [2], directly threatening the country’s territorial integrity and national security. Due to the Russian occupation of Crimea and parts of the Luhansk and Donetsk regions, the energy system of Ukraine has changed, as has the routing system of the Ukrainian segment of the internet [3]. These changes directly affect the resilience of both the energy system and the telecommunications (cyberspace) system and are a source of significant threats to these industries.

Ukrainian power engineering is unique in Europe due to the presence of a large transportation system with nodes up to 3000 MW and unique 750 kV transformers that are not found elsewhere in Europe. However, these nodes are the easiest targets for the enemy’s massive missile strikes. Greater generation of branching and localization (which the European Union strives for, by the way) increases the resilience of power engineering to physical impacts. However, large-scale generation requires intelligent digital management, which brings the problem of cyber threats to the forefront of the new energy sector.

Russia uses the potential of its special services [4] and countries that support Russia in this war [5] for cyber attacks. Russia poses a serious threat in cyberspace, as several Russian IT companies still have functioning computing capabilities worldwide [6]. A significant threat is the widespread use of Russian-made software [7]. Another problem is the involvement of Russian experts and employees in the Russian offices of international companies in building communication, cyber security, and energy systems in Ukraine before 2014. With the departure of such companies from the

Russian market, these employees were laid off, which serves as a motivation for collaboration with enemy hacker groups in this war [8].

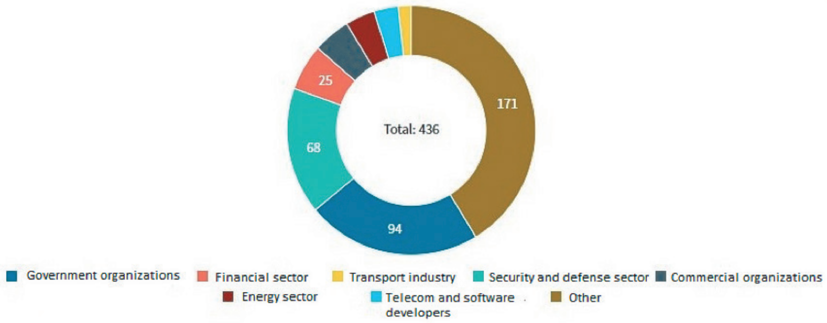
All these factors pose a cyber threat to critical infrastructure. Ukraine's energy sector has a direct impact on other sectors of the economy. The goals of cyber attacks on critical infrastructure facilities in Ukraine are to disrupt the functioning of electricity distribution systems, gather information, disrupt data exchange processes, and impact other dependent industries (critical infrastructure facilities).

The information obtained through cyber attacks on critical infrastructure helps the enemy plan missile strikes [9]. Disruption of information exchange processes is used as a distracting measure from the main intrusion in order to disable the system. System disruption is used to influence related industries. Russia's cyber attacks on energy systems also have political motives [10].

Cyber attacks on energy companies are more complex and difficult to detect. Companies that supply hardware and software to energy companies are also under constant threat. Therefore, supply chain attacks remain a growing source of threat [11].

Statistics on cyber attacks on the energy sector confirm the need for constant improvement of its resilience and security [12] (see Figure 1).

FIGURE 1: CERT-UA CYBER STATISTICS [13]



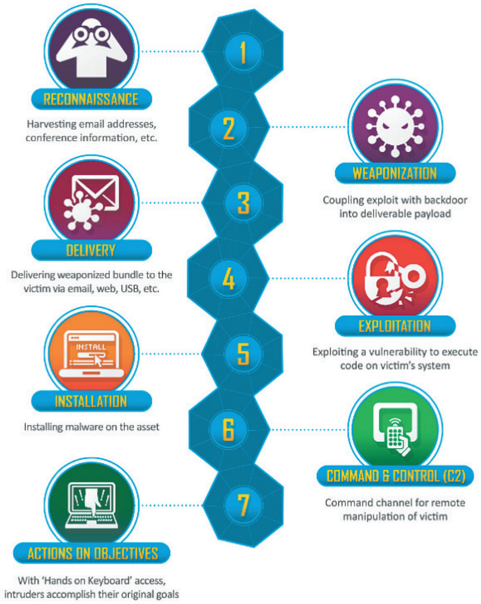
In view of the above, the aim of this study is to identify the resilience problems of the critical energy infrastructure in cyberspace, their potential consequences, and analyze the causes of their occurrence.

For identification, it is advisable to analyze the actions of the attacker. The attacker's actions depend on their ultimate goal, which may not only be to disrupt the energy object. Therefore, there is a need to analyze the dependency of the cyber security of critical energy infrastructure on other industries (cascade effects).

By utilizing the obtained information on cascade effects, it will be possible to propose procedures to improve resilience (compensatory measures).

Threat landscape analysis. Knowing the purpose of the attacker, we can trace the process of a cyberattack. For folding cyber attacks on critical infrastructure, the Cyber Kill Chain model [14] (Figure 2) is used.

FIGURE 2: CYBER KILL CHAIN MODEL [14]


1
RECONNAISSANCE
Harvesting email addresses, conference information, etc.
2
WEAPONIZATION
Coupling exploit with backdoor into deliverable payload
3
DELIVERY
Delivering weaponized bundle to the victim via email, web, USB, etc.
4
EXPLOITATION
Exploiting a vulnerability to execute code on victim's system
5
INSTALL
INSTALLATION
Installing malware on the asset
6
COMMAND & CONTROL (C2)
Command channel for remote manipulation of victim
7
ACTIONS ON OBJECTIVES
With 'Hands on Keyboard' access, intruders accomplish their original goals

The first step in this model is reconnaissance (information gathering) about the target. The cyber attack vector is determined by the volume of information about the target and the competence of the attacker. When collecting such information, the attacker aims to obtain the maximum amount of data using open source intelligence (OSINT) and human intelligence (HUMINT) means [15], without revealing themselves before the stage of carrying out malicious actions.

Having chosen the cyber attack vector, the attacker selects the tools for the cyber attack (e.g., exploit with a backdoor) [16]. An additional threat is the use by the adversary of tools (vulnerabilities) that Ukrainian hacktivists used against Russia without analyzing the possible risks of a reverse attack. After that, the tools are delivered to the target network. Upon entry into the target network, malicious code is exploited with the installation of malicious software on the computer in the target network. The installation of such software allows the attacker to gain control of the system. Having control, the attacker can perform destructive actions [17].

The landscape of cyber threats to critical infrastructure is described in Table I.

TABLE I: EXAMPLES OF THE EXISTING AND NEWLY ARISEN FACTORS OF CYBER THREATS TO ENERGY INFRASTRUCTURE

Existing threats	Deprecated versions of operating systems in operational technologies (OT)
	Many OT systems of worldwide brands designed and implemented in Ukraine by offices and personnel of de facto Russian companies [18]
	Wide usage of antivirus software developed by Russian software companies [19], [20]
	Accounting software developed by Russian software companies [19]
	Logistics software developed by Russian software companies [19]
	Possible Russian insiders among the top management of critical infrastructure facilities [20]
	Attacks on supply chains
Arisen	Seizure of equipment of state institutions and critical infrastructure facilities along with the occupation of territories by the troops of the Russian Federation [21]
	Obtaining forced access to critical systems in the occupied territories
	The enemy's use of means (vulnerabilities) used by Ukrainian hacktivists without analyzing such problems in the protection of Ukrainian critical infrastructure
	Creation of bot farms and bot networks (for DDoS attacks) [22]

The existing circumstances, outlined in Table I, testify to the significant influence of the Russian Federation. In the context of the Russia-Ukraine war, it is appropriate to define the concept of the country's cyberspace borders as the degree of dependence of one country on the IT solutions and information resources of another. In particular, dependence can be calculated as the share of IT solutions used in the state of one class of production of one country to the number of IT solutions of other countries. Given the above, the country's cyber security in 2014 was in a critical state and needed clear and systematic solutions. At the same time, the Russian Federation is significantly ahead of Ukraine in terms of technical support and has significant potential in the development of cyber security and information technologies.

At the same time, Russia is taking measures to increase the level of cyber security by switching to software of its own production. This measure is quite effective [23] and is also necessary for the critical infrastructure of Ukraine.

Another problem with the stability of the critical infrastructure of the energy industry is the use of typical cyber security solutions supplied by a limited number of vendors and system integrators to the Ukrainian market. These integrators are also a potentially less protected target for an adversary's cyberattacks to obtain data about their operations on critical infrastructure facilities. The use of cyber security configuration templates helps to increase the scale of attacks and, as a result, to increase damages.

The organizational and technical basis of cyber security in Ukraine. Within a limited period since 2014, Ukraine has started building its own organizational and technical model of cyber security [24], regulatory documents are being developed, and technical solutions are being implemented. However, it is worth noting that an objective assessment of the effectiveness of the steps taken in 2022 indicates insufficient improvement. At the same time, Ukraine is beginning to adapt to new processes in cyberspace:

1) a center for active countermeasures against Russian aggression in cyberspace was created [25];
2) cyber troops were formed [26];
3) the state authority for the security of critical infrastructure was determined [26];
4) a vulnerability detection system was created [27].

No doubt, in the conditions of war, the enemy has an advantage in Ukraine in terms of technological and time resources. The physical destruction of energy facilities by Russian missile strikes and shelling increases this advantage [28].

The risks of missile strikes affect a number of processes in critical infrastructure. These impacts include the death or disability of employees, disruption to emotional and psychological states, stress, turnover, and more. Where possible, organizations are sending employees to work from home, where there are additional cyber risks. The overall risks to critical infrastructure in the energy sector under conditions of systematic missile strikes include the following (see Table II).

TABLE II: EXAMPLE OF RISK FACTORS RELATED TO WORKPLACES

Working at the facility	Danger transportation to the facility and home
	The danger of staying at an object that is the target of a missile attack
Working remotely (at home)	Lack of resilience of electricity supply at home
	Unsecured electronic communications and remote access
	Compromised home computers
	Risks of mistakes due to inattention by roommates
Personnel changes	Lack of sufficient time for a detailed study of the infrastructure
	Insufficient experience and qualifications
	Insider risk

A cyber risk of a physical nature is the loss of information availability. The most common reasons for this loss are the lack of an internet connection (Internet Service Provider problem) and the lack of electricity at the user or the service provider.

A separate attack using power interruptions could be an example of the “voltage glitch” phenomenon [29], which can allow an attacker to access and modify chip programs. Also, such interruptions can cause other technical malfunctions.

Therefore, the critical infrastructure of the energy sector of Ukraine is at increased risk. The management of cyber risks and information security risks to energy-critical infrastructure is an element of ensuring its resilience.

2. LESSONS LEARNED

Ukraine is adapting, procuring, and widely installing alternative power sources (generators, batteries) [30], backing up data in cloud storage both in Ukraine and abroad [31], and training personnel [32]. However, it should be noted that Ukraine was not prepared for large-scale power outages. Prolonged adaptation, of course, affects cyber security operations.

Proposals for data analysis and forecasting. Ukraine’s experience indicates the need to develop models for analyzing the stability of the energy system. The basis for such models should be systems for collecting and analyzing big data. The main necessary data sets include:

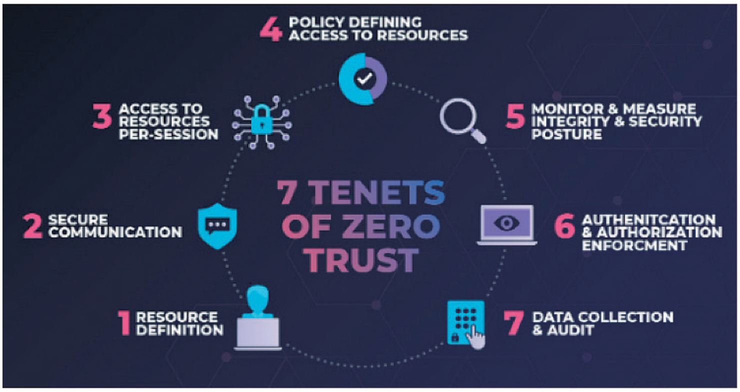
1) telemetry information from sensors on the perimeter of critical infrastructure information and communication systems [27];
2) data collected by public data collectors used for OSINT (e.g., Shodan, Censys, and ZoomEye) [33];
3) media information about the legal entity owner or operator of critical infrastructure, personnel, contractors, management decisions (especially personnel issues) [8];
4) the activities of hacker groups from hostile countries (Russia, Belarus, Iran);
5) cyber security actions on the enemy side;
6) information on procurement in IT and cyber security;
7) the connection between actions in cyberspace and active military actions against individual objects, including mass missile strikes on power plants.

This list is not exhaustive, but such sets of information are necessary components for modeling the information field of critical infrastructure protection objects. The modeling of such an information field consists of input/output information and how this is processed.

It can be argued that the resilience of critical infrastructure generally depends on the ability of an attacker to influence the content of input information or how this is processed. Therefore, modeling negative impacts using information from the information field of a critical infrastructure protection object with the aim of violating the properties of input/output information and how it is processed is a component of the resilience of critical infrastructure.

One effective approach to counteracting such impacts is the implementation of the Zero Trust model (see Figure 3).

FIGURE 3: ZERO TRUST MODEL [34]


4 POLICY DEFINING ACCESS TO RESOURCES
3 ACCESS TO RESOURCES PER-SESSION
5 MONITOR & MEASURE INTEGRITY & SECURITY POSTURE
7 TENETS OF ZERO TRUST
2 SECURE COMMUNICATION
6 AUTHENITCATION & AUTHORIZATION ENFORCMENT
1 RESOURCE DEFINITION
7 DATA COLLECTION & AUDIT

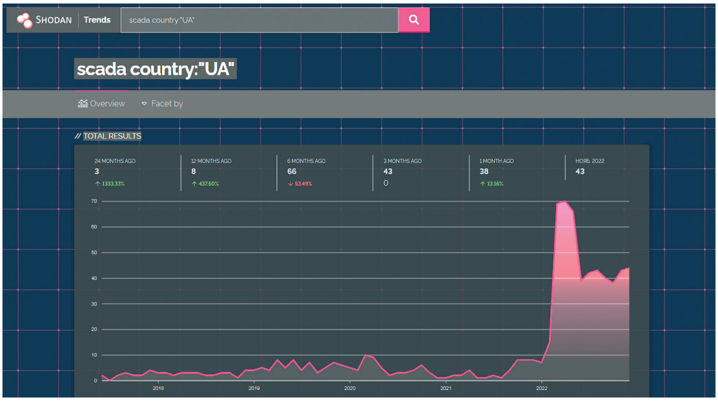
The results of such modeling can include measures to hide sensitive data and implement adequate backup measures (e.g., power supply lines and communications).

At the same time, this data forms the basis for risk analysis as an integral part of ensuring resilience. Therefore, resilience to cyber threats in the energy sector of Ukraine during wartime can be determined as a complex of measures for collecting and processing large amounts of data, risk management, adaptation measures, and analytical forecasting capabilities. The problems of energy supply in Ukraine have confirmed the need for resource and material support for developing effective plans to ensure the functioning of the energy system [21].

It is relevant to focus on what is possible through analytical forecasting, as these processes can significantly increase the effectiveness of implemented resilience measures. The task of analytical forecasting is to develop principles of correlation between the provided arrays of data. The result of such correlation is a sample of objects of critical infrastructure that are similar in their properties (e.g., integrator, vendor, equipment, and software). Such a sample makes it possible to quickly determine the possible scale of a cyber attack and take appropriate measures to localize the compromised environment. The reverse approach is also used by the adversary during cyber operations when they identify a vulnerable system and search for similar targets to increase the scale of the attack. Rapid localization contributes to a quality investigation of cyber incidents and reduces the consequences of cyber attacks.

Improving cooperation for resilience. The use of supervisory control and data acquisition (SCADA) systems is typical for most critical infrastructure. According to the Shodan system (shodan.io), which specializes in information searches in the field of the Internet of Everything, there has been an increase in the number of queries about SCADA in Ukraine (see Figure 4). This indicates the activity of hackers in the reconnaissance stages of cyber attacks [14]. Such statistical data is the basis for additional verifications of firewall logs and security information and event management (SIEM) systems. When anomalies are found, measures should be taken to block them to prevent them from recurring. Blocking the IP addresses of scanning services is also considered a best practice for security [33].

FIGURE 4: SEARCH STATISTICS IN THE SHODAN DATABASE [35]



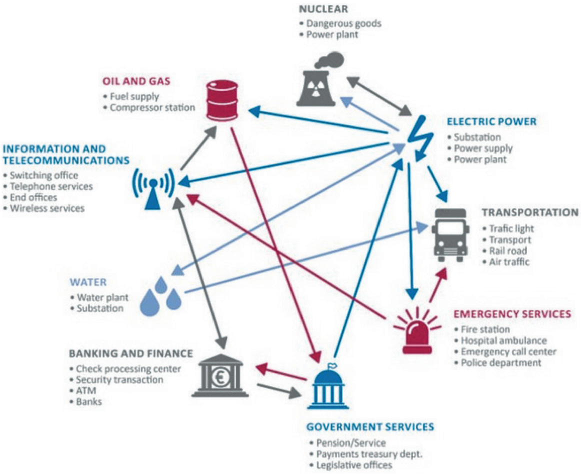
scada country:"UA"

// TOTAL RESULTS

24 MONTHS AGO 3 ↑ 1333.33%
12 MONTHS AGO 8 ↑ 437.50%
6 MONTHS AGO 66 ↓ 53.49%
3 MONTHS AGO 43 0
1 MONTH AGO 38 ↑ 13.16%
НОЯБ. 2022 43

At the same time, the problem of the centralized collection, analysis, and exchange of cyber security data (e.g., IP addresses of attackers, indicators of compromise, and data from phishing emails, etc.) between the government and private sector has not yet been solved. Partial implementation of such exchange among government organizations is only a part of the national cyber security system [36]. Therefore, one of the priorities for the cyber security of the critical infrastructure of the energy sector in Ukraine should be the implementation of automation for data exchange with the private sector in the interests of mutual security. Therefore, the main criteria for ensuring the resilience of critical infrastructure can include the observability of all participants and the availability of operational data exchange considering their interdependence (see Figure 5) [37].

FIGURE 5: INTERDEPENDENCIES OF CRITICAL INFRASTRUCTURE IN THE ENERGY SECTOR



Another example is context analysis, including information on the management decisions of international companies. Currently, many international companies are leaving the Russian market, and have been forced to lay off a large number of employees. On 9 January 2023, the Yale Chief Executive Leadership Institute published information that more than 1,000 international companies had shut down their operations in Russia. This increases the likelihood of the “disgruntled employee” threat. It is entirely possible that there may be consequences, such as cyberattacks on companies that use their products to harm the company.

Studying negative global experiences and implementing operational measures to prevent similar incidents is one of the tasks of a Chief Information Security Officer (CISO) in any organization. However, in the conditions of war and the use of the same IT technologies and products, the enemy can easily use the attack method used against them for their own purposes. The reverse task of analytical forecasting is to determine the priority of targets for attacks by the enemy, where the greatest results can be achieved with fewer resources.

Therefore, we can summarize that for resilience in cyberspace, in addition to resource provision, the development and implementation of enemy attack models using the “red card” and defense models using the “green card” are necessary.

To create such cards, the processing of large amounts of data is required, which requires the implementation of artificial intelligence and neural networks. At the same time, Ukraine is taking its first steps toward implementing artificial intelligence [38], which can significantly affect its defense potential in terms of increasing resilience to cyber threats. It is worth noting that data from Ukraine can greatly accelerate the training of these networks and make their use effective for the sake of overall security in the world. Data sources for such machine learning should also include information on the signature of the enemy’s tools, and the results of real cyber training of specialists within their own infrastructure. In particular, not every head of a cyber security unit can answer questions about what percentage of his unit’s specialists are interchangeable in terms of competencies and levels of access to the system. How long does it take for a specialist to access a hardware firewall terminal through a console port when its web interface is not available? Is this time critical for their system? Such and similar questions need to be worked throughout automation, which will significantly increase the resilience of critical infrastructure. Based on such training, it is advisable to develop regulations for crisis situations.

Improving awareness and cyber resilience skills. If we assume that the weakest point in any system is a human, then it is worth paying attention to cyber attacks related to humans. In the framework of this war, a unique situation has emerged, where the information space, in addition to hacker attacks, is filled with fakes and hostile propaganda. As a result, the number of web resource users and visits is actively growing. Counteracting propaganda by limiting access to enemy web resources at the provider level [39] has the opposite effect; in particular, it leads to an increase in the number of virtual private network (VPN) client installations, a significant part of which consists of Trojans that can have a significant negative impact on cyber security. In such an environment, phishing remains popular and quite effective. Potential targets of the enemy involve phishing on partner companies and suppliers. The enemy sees a person in the system as a point of entry into it. Reducing the threat of phishing will significantly reduce the number of cyber attacks and incidents. A possible solution for the cyber security of critical infrastructure could be the implementation of phishing protection infrastructure [40] and virtual traps infrastructure (honeypots).

Another direction of cyber security is reviewing the potential for conducting preventive cyber operations. The main goal of these operations is to reduce the resource capabilities of the adversary for the attack. This process is quite complex as it requires, in addition to technical knowledge, an assessment of the legal consequences. Ukraine needs the rapid development of a regulatory and legal framework to determine the legal status of cyber operations, which includes defining the concept of cyber warfare and the status of participants.

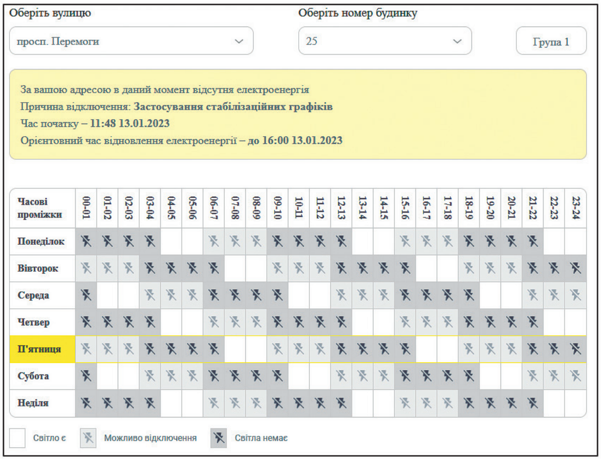
An additional problem is the issue of reintegrating the information and communication systems of temporarily occupied territories. This requires measures to audit information security and cyber defense, as well as the development of appropriate procedures where the Ukrainian experience in implementation is unique in the world. The complexity of such procedures lies in assessing the effectiveness of restoring destroyed infrastructure or creating a solution for building a new one. Implementing a new one may take more time and significant material resources, but it is a better approach to ensure business continuity and proper cyber defense than rebuilding old architecture. Adherence to principles of modularity and redundancy with the consideration of realized risks will significantly increase the level of resilience to such threats. It is also worthwhile considering the feasibility of reinstalling outdated software during restoration works. In this case, portable solutions (virtualization and emulation technologies) would be relevant, which could provide the rapid recovery of functionality, such as disaster recovery as a service (DRaaS), and could also be replaced by more complex and reliable software and hardware complexes as part of the implementation of new systems.

3. ENERGY CYBER RESILIENCE AND DIGITAL RESILIENCE

Ukraine was a leader in digital development prior to this massive Russian aggression. Mass rocket attacks on the power grid led to massive power outages that are difficult for energy grid operators to control. Enterprises, telecommunications operators, and the population are trying to adapt to the constant reduction in electricity supply.

The disruption of communication systems due to problems with the electricity supply occurs in several stages. One of the massive missile attacks led to an emergency shutdown of the entire power grid for 12 hours. At the end of November 2022, instead of planned power outages (for a few hours per day), scheduled power outages for a few hours per day were introduced, as shown in Figure 6. Further missile attacks and regular multi-day outages are expected. This is a real test for a modern digital society based on information technology and electronic communications. It is also important to realize the importance of personal electronic communication and the ability to be online. This concerns the open set of cyber social systems, the functioning of which ensures the stability of society.

FIGURE 6: ACTUAL SCHEDULE OF PLANNED AND EMERGENCY POWER OUTAGES FOR AN AVERAGE DISTRICT OF KYIV, PRESENTED BY THE ENERGY DISTRIBUTION COMPANY DTEK (http://dtek-em.com.ua)



Оберіть вулицю: просп. Перемоги

Оберіть номер будинку: 25

Група 1

За вашою адресою в даний момент відсутня електроенергія
Причина відключення: **Застосування стабілізаційних графіків**
Час початку – **11:48 13.01.2023**
Орієнтовний час відновлення електроенергії – **до 16:00 13.01.2023**

Часові проміжки	00-01	01-02	02-03	03-04	04-05	05-06	06-07	07-08	08-09	09-10	10-11	11-12	12-13	13-14	14-15	15-16	16-17	17-18	18-19	19-20	20-21	21-22	22-23	23-24
Понеділок	■	■	■	■			▨	▨	▨	■	■	■	■			▨	▨	▨	■	■	■	■		
Вівторок	▨	▨	▨	■	■	■	■			▨	▨	▨	■	■	■	■			▨	▨	▨	■	■	■
Середа	■			▨	▨	▨	■	■	■	■			▨	▨	▨	■	■	■	■			▨	▨	▨
Четвер	■	■	■	■			▨	▨	▨	■	■	■	■			▨	▨	▨	■	■	■	■		
П'ятниця	▨	▨	▨	■	■	■	■			▨	▨	▨	■	■	■	■			▨	▨	▨	■	■	■
Субота	■			▨	▨	▨	■	■	■	■			▨	▨	▨	■	■	■	■			▨	▨	▨
Неділя	■	■	■	■			▨	▨	▨	■	■	■	■			▨	▨	▨	■	■	■	■		

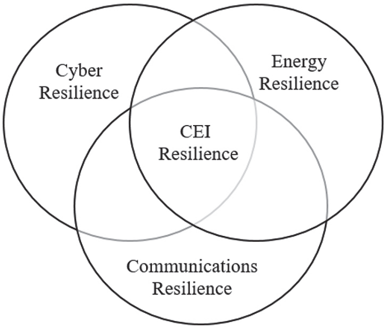
□ Світло є ▨ Можливо відключення ■ Світла немає

On weekdays in Figure 6, the darkest dots in the rows indicate scheduled power outages, the gray ones, possible power outages in the case of an overload in the power system, and the pairs of light dots show hours of guaranteed power supply. In reality, the schedule is rarely followed.

Therefore, Ukraine is experiencing cascading effects from failures in the power system. This affects the provision of banking and postal services, healthcare, public transportation, and others, which can be considered indirect economic losses. At the same time, these cascading effects also directly result from the lack of internet. Private electronic communications, closely tied to the internet, have been greatly impacted. Providers of electronic communication services (both national telecommunications operators and local internet providers) have begun to organize local power supply (e.g., large-capacity batteries and gasoline generators) for their equipment to ensure uninterrupted service provision, mostly at their own expense. These costs indirectly increased the cost of their services [41]. Therefore, a backup power supply for public services, taking into account cascading effects, is a component of resilience.

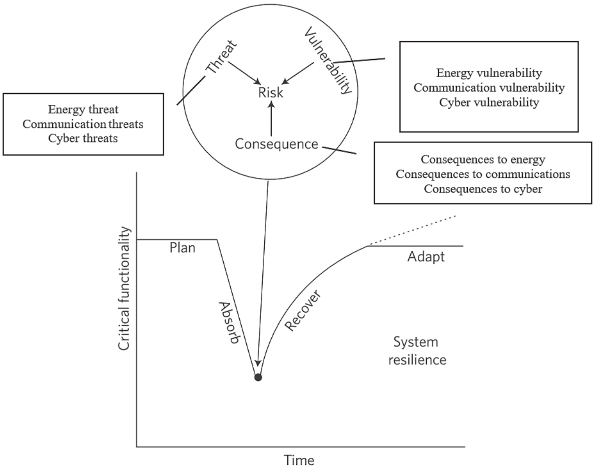
From this point on, it becomes clear that an important element of resilience is the independence of communication systems from vulnerable energy systems. Conversely, since information exchange between energy systems is carried out through communication systems, communication losses represent a risk for the energy system: disruption of such information exchange harms energy. Therefore, it is advisable to consider cyber attacks on telecommunications operators as another potential indirect impact on the energy system and to systematically analyze their dependencies. Let us expand our scope of activity to other non-communication IT systems and cyber-physical systems related to energy distribution, accounting and billing. They bring cyber threats aimed at challenging the resilience of critical energy infrastructure (CEI). Figure 7 shows such interdependencies at the macroscopic level.

FIGURE 7: RESILIENCE OF CEI AS A COMPONENT OF MANY INTERDEPENDENCIES



Taking into account the well-known structure of resilience management (Figure 8) presented in [42], we can update the components of risk, such as cyber risk, energy risk, and communication risk, and each of them can be decomposed into threats. From this point on, communication threats are part of energy threats and vice versa.

FIGURE 8: STRUCTURE OF THE STABILITY OF THE ENERGY SYSTEM REGARDING THE NATURE OF RISKS



Information availability and internet access are the most important factors for information security. Therefore, performing the following additional tasks will be a step toward enhancing digital resilience:

1) researching the most sensitive information needs of the population and businesses;
2) analyzing cyber threats that create disruptions in availability, and how cyber threats affect critical needs;
3) analyzing network architectures, types of communications and topologies, as well as determining which combinations can enhance the resilience of information communications involved in meeting the most sensitive information needs of the population and businesses.

The most effective strategies for adaptation and mitigation of the consequences of critical events are unlikely to have a significant effect if implemented by one person, a group of individuals, or a single business entity. Useful approaches, such as extending the service life of outdated telecommunications (e.g., wired telephones or ADSL), require support on a national scale. A scientific approach is needed to collect, analyze, and systematize existing experience (especially industry-specific) in order to reduce the cost of creating and operating resilient information and communication systems.

The war in Ukraine can be reasonably considered hybrid since, in addition to combat actions on the battlefield, attacks are also carried out in cyberspace. The existence of such a precedent indicates the possibility of similar situations in the future between other countries. Therefore, it is expedient to develop an index of the ability of countries to conduct cyber warfare, which cannot be considered a reverse index of cybersecurity [43], [44], since it should include an assessment of the possibilities for cyber attacks. Therefore, countries with a lower index should consider countries with a higher index as potential opponents or promising partners around the world.

In light of the above, it is necessary to create low-level and high-level frameworks with crisis management procedures in cyberspace during a state of war. These frameworks should include procedures for data exchange, systematic self-assessment and an evaluation of the percentage of the interchangeability of cybersecurity personnel within the organization.

Collaboration and partnership in the field of cybersecurity based on the principles of increasing the resilience of the energy sector to cyber threats is the key to security. This statement equates the concepts of resilience and security but does not make them interchangeable. Today, in the conditions of war, Ukraine demonstrates the implementation of the principle of openness to partnership for the sake of security [45], and perhaps this openness increases Ukraine's resilience in all aspects of the war.

Other necessary sources of stability for Ukraine, and for other countries, include assessing the criticality of dependencies on other countries in the field of cyber security and IT technologies, increasing human resources, and developing systems for collecting and analyzing big data.

The hybrid war in Ukraine has demonstrated the need to develop crisis response plans for cyber incidents and attacks [46], the need to create resilience (flexibility) models, and the need to implement processes for cyber operations and partnerships for security.

4. CONCLUSION

Modern wars are hybrid by nature, and the cyber domain is an important component of national security. A comprehensive analysis of the resilience problems of Ukraine's energy-critical infrastructure in cyberspace has made it possible to identify the dependencies of the cyber security of energy-critical infrastructure on other sectors (cascade effects). An analysis of the reasons for the emergence of factors affecting the resilience of critical infrastructure has been conducted, taking into account the specifics of waging war.

Energy systems play a key role in a digital civilization. The greater branching and localization of generation undoubtedly increases the resilience of the power system to strong physical actions. However, for large-scale generation, intelligent digital management is needed, which brings the problem of cyber threats to the forefront of the new energy industry. We have considered the cascade impact of energy on digital resilience and the reverse impact of electronic communications on the energy sector. Procedures have been proposed to increase the cyber resilience of energy, including the implementation of processes for collecting and processing big data, optimizing public-private interaction, organizing cyber training, and developing security frameworks for wartime. The basis for the proposed processes is the priority of protecting critical infrastructure by responding to crisis situations with limited resources.

Ukraine's experience in conducting such research is unique. This could become the basis for developing models and architectures of resilience for power systems in other countries.

ACKNOWLEDGMENTS

This article is partially funded by the US Army Engineering Research and Development Center. We are grateful to Drs Igor Linkov and Lance Fiondella and Mr Luke Hoguewood for their discussions and assistance in preparing the article.

REFERENCES

[1] "Cyberattacks in hybrid warfare: The case of Russia/Ukraine War." HeadMind Partners. Accessed: Mar. 5, 2023. [Online]. Available: https://www.headmind.com/en/cyberattacks-hybrid-warfare/

[2] "Full-scale Russian invasion of Ukraine: historical context." GOV.UA. Accessed: Mar. 5, 2023. [Online]. Available: https://uinp.gov.ua/aktualni-temy/povnomasshtabne-vtorgnennya-rosiyi-v-ukrayinu-istorychnyy-kontekst

[3] "How does the Internet work in the DPR – a letter from a reader from the occupied territories." Tokar.ua. Accessed: Mar. 5, 2023. [Online]. Available: https: //tokar.ua/read/44924

[4] "Russian state-sponsored and criminal cyber threats to critical infrastructure." Cybersecurity and Infrastructure Security Agency. Accessed: Mar. 5, 2023. [Online]. Available: https://www.cisa.gov/news-events/cyber security-advisories/aa22-110a

[5] G. Corera. "Iranian and Russian hackers targeting politicians and journalists, warn UK officials." BBC News. Accessed: Mar. 5, 2023. [Online]. Available: https://www.bbc.com/news/uk-64405220

[6] A. Soldatov and I. Borogan. "Russian cyberwarfare: Unpacking the Kremlin's capabilities." CEPA. Accessed: Mar. 5, 2023. [Online]. Available: https://cepa.org/comprehensive-reports/russian-cyberwarfare-unpacking-the-kremlins-capabilities/

[7] I. Levy. "Use of Russian technology products and services following the invasion of Ukraine." NCSC. Accessed: Mar. 5, 2023. [Online]. Available: https://www.ncsc.gov.uk/blog-post/use-of-russian-technology-products-services-following-invasion-ukraine

[8] "Ericsson divests its local customer support business in Russia." Ericsson. Accessed: Mar. 5, 2023. [Online]. Available: https://www.ericsson.com/en/press-releases/2022/12/ericsson-divests-its-local-customer-support-business-in-russia

[9] "Cyber, artillery, propaganda. general overview of the dimensions of Russian aggression." GOV.UA. Accessed: Mar. 5, 2023. [Online]. Available: https://cip.gov.ua/en/news/kiberataki-artileriya-propaganda-zagalnii-oglyad-vimiriv-rosiiskoyi-agresiyi

[10] “‘Ukrenergo’ under war conditions: Attacks increased threefold to block joining European power network.” GOV.UA. Accessed: Mar. 5, 2023. [Online]. Available: https://cip.gov.ua/en/news/ukrenergo-v-umovakh-viini- kiberataki-zrosli-vtrichi-shob-zupiniti-priyednannya-do-yevropeiskoyi-energomerezhi

[11] “CERT-UA has processed over 2,000 cyberattacks against Ukraine year to date.” GOV.UA. Accessed: Mar. 5, 2023. [Online]. Available: https://cip.gov.ua/en/news/cert-ua-vid-pochatku-roku-opracyuvala-bilshe-dvokh-tisyach-kiberatak-na-ukrayinu

[12] “Russian hackers attempted to cut electricity supply to many Ukrainians.” GOV.UA. Accessed: Mar. 5, 2023. [Online]. Available: https://cip.gov.ua/en/news/rosiiski-khakeri-namagalisya-pozbaviti-dostupu-do-elektroenergiyi-znachnu-kilkist-ukrayinciv

[13] “CERT-UA.” GOV.UA. Accessed: Mar. 5, 2023. [Online]. Available: https://cert.gov.ua/article/37121

[14] “Cyber Kill Chain®.” Lockheed Martin. Accessed: Mar. 5, 2023. [Online]. Available: https://www.lockheedmartin.com/en-us/capabilities/cyber/cyber-kill-chain.html

[15] C. Warner. “‘Attribution of advanced persistent threats’ notes.” Medium. Accessed: Mar. 5, 2023. [Online]. Available: https://warnerchad.medium.com/attribution-of-advanced-persistent-threats-notes-94008ea1f365

[16] “General recommendations for reducing the consequences of exposure to malicious software.” GOV.UA. Accessed: Mar. 5, 2023. [Online]. Available: https://cert.gov.ua/recommendation/2502

[17] “BlackEnergy APT Attacks in Ukraine employ spearphishing with Word documents.” Securelist | Kaspersky’s threat research and reports. Accessed: Mar. 5, 2023. [Online]. Available: https://securelist.com/blackenergy-apt-attacks-in-ukraine-employ-spearphishing-with-word-documents/73440/

[18] “Stop doing business with Russia.” #LeaveRussia: list of companies that have stopped or continue to work in Russia. Accessed: Jan. 2, 2023. [Online]. Available: https://leave-russia.org/uk?flt[108][eq][]=53953

[19] “Opendatabot and Netpeak Create List of Software of Russian Origin – Opendatabot.” Opendatabot. Accessed: Jan. 2, 2023. [Online]. Available: https://opendatabot.ua/analyt/russian-software

[20] A. D’Anieri. “Ukraine confronts the threat of Kremlin penetration into unreformed state bodies.” Atlantic Council. Accessed: Jan. 2, 2023. [Online]. Available: https://www.atlanticcouncil.org/blogs/ukrainealert/ukraine-confronts-kremlin-infiltration-threat-at-unreformed-state-bodies/

[21] M. Komarov, S. Honchar, and D. Dimitriieva, “Investigation of the problem of cyber survivability of objects of critical information infrastructure,” *Nuclear and Radiation Safety*, vol. 1(89), pp. 59–66, Mar. 2021, application date: Mar. 7, 2023, doi: 10.32918/nrs.2021.1(89).07.

[22] S. R. Team. “A Blog with No Name.” Team Cymru. Accessed: Mar. 5, 2023. [Online]. Available: https://www.team-cymru.com/post/a-blog-with-noname?utm_campaign=blog_noname_s2&amp;utm_medium=news&amp;utm_source=social+media

[23] “Official website of the unified register of Russian programs.” GOV.RU. Accessed: Mar. 5, 2023. [Online]. Available: https://reestr.digital.gov.ru/

[24] Ukraine, Cabinet of Ministers of Ukraine. (Dec. 29, 2021). *Resolution of the Cabinet of Ministers of Ukraine No. 1426, On approval of the Regulation on the organizational and technical model of cyber protection*. Application date: Jan. 2, 2023. [Online]. Available: https://zakon.rada.gov.ua/laws/show/1426-2021-п # Text

[25] Ukraine, Verkhovna Rada of Ukraine. (Feb. 23, 2006). *Law of Ukraine No. 3475-IV “On the State Service of Special Communication and Information Protection of Ukraine.”* Application date: Jan. 2, 2023. [Online]. Available: https://zakon.rada.gov.ua/laws/show/3475-15#Text

[26] Ukraine, President of Ukraine. (Aug. 26, 2021). *Decree of the President of Ukraine No. 447/2021 On the decision of the National Security and Defense Council of Ukraine dated May 14, 2021 “On the Cybersecurity Strategy of Ukraine.”* Application date: January 2, 2023. [Online]. Available: https://zakon.rada.gov.ua/laws/show/447/2021#Text

[27] Ukraine, Cabinet of Ministers of Ukraine. (Dec. 23, 2020). *Resolution of the Cabinet of Ministers of Ukraine No. 1295, Some issues of ensuring the functioning of the system for detecting vulnerabilities and responding to cyber incidents and cyberattacks*. Application date: Mar. 5, 2023. [Online]. Available: https://zakon.rada.gov.ua/laws/show/1295-2020-п#Text

[28] “Strikes on critical infrastructure of Ukraine during the Russian-Ukrainian war.” Wikipedia. Accessed: Jan. 2, 2023. [Online]. Available: https://uk.wikipedia.org/wiki/Strikes_on_critical_infrastructure_of_Ukraine_during_the_Russian-Ukrainian_war

[29] Invia. “How a voltage glitch attack could cripple your SoC or MCU – and how to securely protect it.” Design And Reuse. Accessed: Mar. 5, 2023. [Online]. Available: https://www.design-reuse.com/articles/48553/how-a-voltage-glitch-attack-could-cripple-your-soc-or-mcu.html

[30] O. Kadomska. “Energy survival: Alternative energy sources during power outages in Kharkiv.” OPINION. MEDIA. Accessed: Jan. 2, 2023. [Online]. Available: https://dumka.media/ukr/war/1669920291-temryavoyu-ne-zlyakati-yak-harkiv-yani-vchatsya-dolati-zatyazhni-blekauti

[31] *On cloud services: Law of Ukraine dated February 17, 2022 No. 2075-IX*. GOV.UA. Accessed: Mar. 5, 2023. [Online]. Available: https://zakon.rada.gov.ua/laws/show/2075-20#Text

[32] “Reforming the system of training professional personnel in the field of cyber security in Ukraine: the first six professional standards have been approved.” GOV.UA. Accessed: Mar. 5, 2023. [Online]. Available: https://cip.gov.ua/ua/news/reformuvannya-sistemi-pidgotovki-profesiinikh-kadriv-u-sferi-kiberbezpeki-v-ukrayini-zatverdzheni-pershi-shist-profesiinikh-standartiv

[33] “Cr0n. 20 search engines for pentesters. Average.” KR Labs. Accessed: Jan. 2, 2023. [Online]. Available: URL: https://blog.kr-labs.com.ua/20-poshukovyh-sistem-dlya-pentestera-28238b57db81?gi=b8bdd1009e94

[34] L. Yacono. “The 7 tenets of zero trust.” Cimcor. Accessed: Mar. 5, 2023. [Online]. Available: https://www.cimcor.com/blog/the-7-tenets-of-zero-trust

[35] “Shodan trends.” Shodan. Accessed: Jan. 2, 2023. [Online]. Available: https://trends.shodan.io/search?query=scada%20country: “UA” #facet/overview

[36] “Regarding the exchange of information on cyber threats.” GOV.UA. Accessed: Mar. 5, 2023. [Online]. Available: https://cert.gov.ua/article/39962

[37] “Enhancing the protection and cyber-resilience of critical information infrastructure.” Digital Regulation. Accessed: Mar. 5, 2023. [Online]. Available: https://digitalregulation.org/. https://digitalregulation.org/enhancing-the-protection-and-cyber-resilience-of-critical-information-infrastructure/

[38] Ukraine, Cabinet of Ministers of Ukraine. (Dec. 2, 2020). *Decree of the Cabinet of Ministers of Ukraine No. 1556-r, On the approval of the Concept of the development of artificial intelligence in Ukraine*. Application date: Jan. 2, 2023. [Online]. Available: https://zakon.rada.gov.ua/laws/show/1556-2020-p # Text

[39] Ukraine, President of Ukraine. (May 14, 2020). *Decree of the President of Ukraine No. 184/2020 On the decision of the National Security and Defense Council of Ukraine dated May 14, 2020 “On the application, cancellation and amendment of personal special economic and other restrictive measures (sanctions).”* Application date: Jan. 2, 2023. [Online]. Available: https://zakon.rada.gov.ua/laws/show/184/2020#Text

[40] A. Davidyuk, “Anti-phishing infrastructure as a means of countering threats to the functioning of critical purpose systems,” in *XIII International Scientific and Practical Conference of Young Scientists Information Technologies: Economy, Technologies, Education 2022, Kyiv, Ukraine, Oct. 26, 2022*, pp. 127–128.

[41] “Changes in the terms of providing tariffs and services.” Kyivstar. Accessed: Mar. 5, 2023. [Online]. Available: https://kyivstar.ua/uk/about/important_data/changes

[42] I. Linkov *et al.*, “Changing the resilience paradigm,” *Nature Climate Change*, vol. 4, pp. 407–409, 2014, doi: 10.1038/nclimate2227.

[43] “Global Cyber Security Index.” ITU. Accessed: Jan. 2, 2023. [Online]. Available: https://www.itu.int/en/ITU-D/Cybersecurity/Pages/global-cyber security-index.aspx

[44] M. M. Khudyntsev (gen. ed.), A. V. Zhilin, and A. V. Davidyuk, “World indices of cyber security: overview and methods of formation,” International University of Cyber Security, Institute of Modeling Problems in Energy, National Academy of Sciences of Ukraine, Kyiv, Global report / Catalog, Monograph, 2022. ISBN 978-966-136-887-2.

[45] “NCCC representatives discussed with the director of NATO’s Joint Center for Advanced Technologies on Cyber Defense the issue of deepening practical cooperation between Ukraine and the Center.” GOV.UA. Accessed: Jan. 2, 2023. [Online]. Available: https://www.rnbo.gov.ua/ua/Diialnist/5909.html

[46] Ukraine, President of Ukraine. (Feb. 1, 2020). *Decree of the President of Ukraine No. 37/2022 On the decision of the National Security and Defense Council of Ukraine dated December 30, 2021 “On the Implementation Plan of the Cybersecurity Strategy of Ukraine.”* Application date: Jan. 2, 2023. [Online]. Available: https://zakon.rada.gov.ua/laws/show/37/2022#Text

2023 15th International Conference on Cyber Conflict
Meeting Reality
T. Jančárková, D. Giovannelli, K. Podiņš, I. Winther
2023 © NATO CCDCOE Publications, Tallinn

Digital Supply Chain Dependency and Resilience

Lars Gjesvik
Senior Researcher
Norsk Utenrikpolitisk Institute
Oslo, Norway
larsg@nupi.no

Haakon Bryhni
Research Professor
Simula Metropolitan Centre for Digital Engineering
Oslo, Norway
haakonbryhni@simula.no

Niels Nagelhus Schia
Senior Researcher
Norsk Utenrikpolitisk Institute
Oslo, Norway
nns@nupi.no

Azan Latif Khanyari
PhD Candidate
Simula Metropolitan Centre for Digital Engineering
Oslo, Norway
azan@simula.no

Alfred Arouna
PhD Candidate
Simula Metropolitan Centre for Digital Engineering
Oslo, Norway
alfred@simula.no

Abstract: While a growing body of literature addresses how states increasingly aim to secure their digital domains and mitigate dependencies, less attention has been paid to how infrastructural and architectural configurations shape their ability to do so. This paper provides a novel approach to studying cyber security and digital dependencies, paying attention to how the everyday business decisions by private companies affect states' ability to ensure security. Every mobile application relies on a multitude of microservices, many of which are provided by independent vendors and service providers operating through various infrastructural configurations across borders in an a-territorial global network. In this paper, we unpack such digital supply chains to examine the technical cross-border services, infrastructural configurations, and locations of various microservices on which popular mobile applications depend. We argue that these dependencies have differing effects on the resilience of digital technologies at the national level but that addressing these dependencies requires different and sometimes contradictory interventions. To study this phenomenon, we develop a methodology for exploring this phenomenon empirically by tracing and examining the dispersed and frequently implicit dependencies in some of the most widely used mobile applications. To analyse these dependencies, we record raw traffic streams at a point in time seen across various mobile applications. Subsequently locating these microservices geographically and to privately owned networks, our study maps dependencies in the case studies of Oslo, Barcelona, Paris, Zagreb, Mexico City, and Dublin.

Keywords: *digital dependencies, digital supply chains, content delivery networks, resilience*

1. INTRODUCTION

The security of digital networks and technologies is an increasingly important issue for states. Early iterations of national security concerns in Western states primarily understood digital vulnerabilities from the lens of threats to critical infrastructures. In recent years, however, a richer understanding of cyber security has developed (Dunn Cavelty and Wenger 2020), drawing attention to infrastructural configurations (Musiani et al. 2016, 268), private interests (Srivastava 2021), and the unequal distribution of digital resources globally (Kwet 2019). Notably, scholars have increasingly started to dissect understandings of power and vulnerability rooted in complex interdependencies to depict how digitalization, geo-economics, and the security concerns of states intersect (Cartwright 2020; Nye 2020; Mügge 2023).

These concerns tie into an unease with growing dependencies on foreign actors and the potential risk that these dependencies could be utilized coercively (Farrell and Newman 2019). Additionally, they are related to a recasting of economic globalization in strategic terms (Leonard 2021; Walter 2021; Choer Moraes and Wigel 2022; Gertz and Evers 2020) that has been accelerated by COVID-19's exposure of globalization's fragilities (McNamara and Newman 2020). With economic dependencies increasingly understood also as strategic dependencies, questions about how to improve security and resilience in complex digital networks are both pressing and challenging.

Inspired by recent attention to and mapping of global supply chain dependencies,[1] this paper aims to unpack digital supply chains to examine the technical cross-border services, infrastructural configurations, and locations of various microservices on which popular mobile applications depend. With the growing popularity of software

[1] See, e.g., European Commission 2022.

ecosystems and accompanying digital supply chains, common applications and end-user software are only the final stage of a long list of distributed microservices and software tools (Decan et al. 2019; Cox 2019). Such software supply chains are a well-known security concern when it comes to human errors and malicious attacks (Ohm et al. 2020), but they also raise questions about dependencies, resilience and connectivity. Neither the physical location of these microservices nor ownership of the underlying infrastructure are equally distributed; instead, both reflect and reinforce the inequalities emerging from the uneven distribution of digital resources globally (de Goede 2020). Unpacking the distribution and dependencies of various microservices from the vantage point of different states can illustrate how states are differently positioned to address questions of digital dependency.

To study this phenomenon, we track the dependencies of seven globally popular consumer mobile applications across a selection of cases. The selected applications are five social media apps (Facebook, Instagram, Snapchat, Messenger, and TikTok) and two video conferencing apps (Google Meet and Zoom). Through a virtual private network (VPN),[2] we transport our experimental setup to different geographical locations, subsequently tracing dependencies at the packet level through their associated geographic locations and the corporate networks used by these services. By mapping the differences in digital dependency in Oslo, Barcelona, Paris, Zagreb, Mexico City, and Dublin, we highlight how digital dependency and resilience intersect with the economic choices of private companies.

For all the cases examined here, our study details how digital dependencies are simultaneously diverse and similar across countries, just as they are differently placed within global economic networks. When it comes to dependencies on private companies, all the cases are broadly similar in their dependence on a handful of key cloud infrastructure providers and content delivery networks (CDNs) based in the United States. In contrast, when it comes to dependencies on infrastructures and their territorial locations, there are substantial differences from one case to the other. While these differences largely reflect the size of the domestic market, Ireland and its ability to attract investments from global corporations remains an outlier. Crucially, addressing these different forms of dependency might require contradictory policies – especially for smaller and developing states – as public infrastructure investments are a costly alternative to attracting investments by global corporations. We argue that unpacking different forms of dependency – and how they affect the security and vulnerability of states' digital connectivity – is important for states to develop coherent strategies and interventions.

[2] https://nordvpn.com/

2. DIGITAL DEPENDENCY, RESILIENCE, AND INFRASTRUCTURES

Cyber security, insecurity, and resilience have become some of the key security concerns of the 21st century, sparking a maturing field that embraces a growing range of theories and methods (Dunn Cavelty and Wenger 2020; Stevens 2018). Yet, while a growing body of research examines the political implications of digital technologies, most work remains policy-centred and problem-solving (Stevens 2018). Thus, aspects of cyber insecurity are understudied or left at the margins. Key among such omissions is the lack of research examining the intersection between economic and political forces on the question of cyber security. This omission is especially important since the manufactured nature of cyberspace ties both security and insecurity to decisions of design and development primarily made by private companies (Dunn Cavelty and Wenger 2022, 2).

As a starting point for studies of cyber security, the importance of the decisions made by private companies has gained renewed relevance with the growing concerns over digital sovereignty and autonomy across a range of states previously championing a free and open internet (Autolitano and Pawlowska 2021; Christakis 2020; Couture and Toupin 2019). While the utilization of globally connected digital technologies has been widely perceived as a boon for modern societies, the inherent challenges to states' ability to govern and ensure national security have been a source of tensions. In recent years, such concerns have grown as a result of both coercive manipulation (Farrell and Newman 2019; Cartwright 2020; Ortiz Freuler 2022) and the growing criticality of digital technologies in modern societies.

To help fill this gap, our paper takes its cue from the literature on how technological developments and the consolidation of digital markets impact what cyber security and resilience mean, as well as from literature on the ability of states to provide security and address their concerns (Ilves and Osula 2020, 12). By doing so, we aim to couple the broadening idea of digital insecurity to a body of work that pays closer attention to how technological change and infrastructural configurations shape power over and through digital technologies (Musiani et al. 2016, 155–216).

We use the growth in software ecosystems and digital supply chains, and their connections with pre-existing digital infrastructures, as our prism for investigating these variations. Contemporary software and application development has become a complex and multifaceted affair. Rather than understanding development in terms of single software systems, an ecosystems approach can draw on software development and implement components that are both geographically and organizationally distributed (Decan et al. 2019). This shift to reusing software – creating novel dependencies in relation to code and software written and maintained by third parties – has been argued to have 'happened so quickly that we do not yet understand the best practices for choosing and using dependencies effectively' (Cox 2019). As a novel security concern, the multiple dependencies at the micro-level for most software and applications have been understood in terms of the lack of transparency and oversight by purchasing entities (Ellison et al. 2010). While, however, the risks inherent in these dependencies have received some attention, they have been investigated so far primarily in terms of vulnerability to mistakes (Cox 2019) or malicious attacks (Ohm et al. 2020; Harrand et al. 2021). We argue, however, that locating distributed digital supply chains also illustrates how technological developments and evolutions work in tandem with digital infrastructures to shape strategic dependencies.

There are two parts to this argument: Firstly, we propose that digital technologies are not static but mutate and interact with existing social and infrastructural configurations to have political effects. In proposing this, we consider how shifts in the provision of digital services and physical infrastructures can affect and shape politics. We also assess how everyday business decisions produce physical and virtual objects that have material consequences via 'both enabling and constraining effects' (Aradau 2010, 492). This draws our attention to the role of various infrastructures – physical ones as well as intangible services, standards, and protocols – and to how cyber security emerges as a consequence of these decisions as much as the actions of adversaries.

Secondly, we pay attention to how these effects are unevenly distributed and how control over and power through an ostensibly decentralized network is embedded in points of infrastructural control (Musiani et al. 2016, 4). We note that digital technologies ought to be considered through their embeddedness in physical digital infrastructures. Rather than the ephemeral and mythical conceptions of 'clouds' or other a-territorial metaphors (Amoore 2018), digitalization remains deeply rooted in physical objects and infrastructures to support everyday function. These physical infrastructures 'are necessary for the proper function of cyberspace, and most of those infrastructures are located on claimed territory' (Baezner and Robin 2018).

Crucially, the physicality of our digital world has a distributing effect as the 'effects on media industries, user experiences, and the politics of circulation occur unevenly around the world' (Parks and Starosielski 2015, 56). Thinking about digital insecurity from the bottom-up invites us to parse and dissect these asymmetries and their effect on political organizations such as states (Rosa and Hauge 2022).

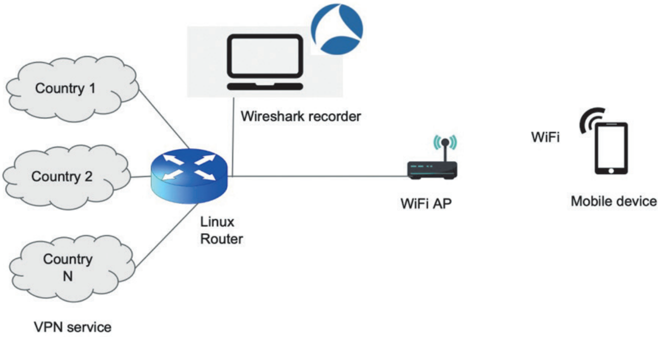
What such an approach to studying digital politics allows us to do is to consider how the increasing concern over digital dependency is impacted both by how software supply chains are ordered and their relation to the built physical world of data centres and fibre optic cables. We argue that studying how these effects play out requires micro-level studies of the complex dependencies that arise. This, in turn, requires an understanding of how a given state depends on services hosted outside its borders and how these dependencies alter the meaning of connectivity. We propose a novel approach to studying this phenomenon by unpacking and geolocating micro dependencies in commonly used applications.

3. TRACING MOBILE APPLICATION DEPENDENCIES

To track the dependencies of common software and applications, we have selected seven popular consumer mobile applications – five social media applications (Facebook, Messenger, Instagram, Snapchat, and TikTok) and two video conferencing applications (Google Meet and Zoom). These applications are globally available in mobile app stores, which is essential for performing measurements using a VPN.

To record the packet level traces, we deployed the measurement setup as shown in Figure 1, where the mobile device (an iPhone) connects to the internet via a Wi-Fi access point (AP). All background network traffic from the mobile device is recorded on a computer running a packet-capturing software called Wireshark.[3] A Linux router performs network address translation (NAT) for devices on the network and routes traffic over a VPN tunnel to the selected city.

FIGURE 1: LAB SETUP FOR MOBILE APPLICATIONS MEASUREMENTS



We start our measurements by factory resetting an iPhone. Next, we connect the iPhone to the Wi-Fi AP and record background network traffic for several hours to record IP addresses categorized as ‘noise’. This traffic is not the focus of our study, but we record this to isolate the application-specific network traffic from the seven apps we study.

[3] https://www.wireshark.org/

Subsequently, for each case country, we start with a factory reset iPhone and then install the seven applications one by one. After installing each application, we begin recording all network traffic exchanged by the app during its normal functioning. Note that we interact with the app during this recording to ensure all essential services are contacted. The Wireshark software eavesdrops on the Ethernet port that mirrors all traffic sent over the Wi-Fi AP. After we have application-specific traffic dumps for every app in each country, we process the dumps. This involves extracting all the IP addresses and subsequently finding the physical location of where they are geolocated –as described in further detail below.

After recording the traffic dumps for each of the selected mobile applications, we subsequently filter out all the previously identified noise-IPs. The app-specific IPs are thereafter traced to both corporate networks and their geographical location. The latter process represents a well-known research problem, and although commercial databases are available, they are not accurate in geolocating infrastructure IP addresses (Gharaibeh et al. 2017). Our first step to geolocating the IP addresses is to use the state-of-the-art tool called IPMap from RIPE NCC (Réseaux IP Européens Network Coordination Centre). IPMap uses a combination of active measurement latency-based methods, crowdsourced information, and reverse Domain Name System (rDNS) methods to geolocate IP addresses.[4] However, this tool does not provide broad coverage of the IP addresses from our measurements. For the remaining IP addresses, we rely on a manual geolocation method. Although this is a time-consuming method and not scalable for a large set of IPs, it is the only one available that can provide geolocations to a certain degree of accuracy. The first step of this method is to conduct multiple traceroute measurements to the IP addresses to be geolocated. The next step involves searching for rDNS hints in the subsequent hops in a traceroute. Operators often embed physical location hints in infrastructure IPs that help in geolocating them. With a combination of rDNS hints and the round-trip time associated with different country locations, we estimate the geolocation of IP addresses. There are, however, limitations to this method, as not all operators provide a reverse DNS name with an embedded location hint. There also exist challenges – like firewalls on some of the infrastructure IPs that block the ping and traceroute requests – which make it impossible to geolocate all IP addresses. However, in general, we see good geolocation coverage on the IP addresses we study.

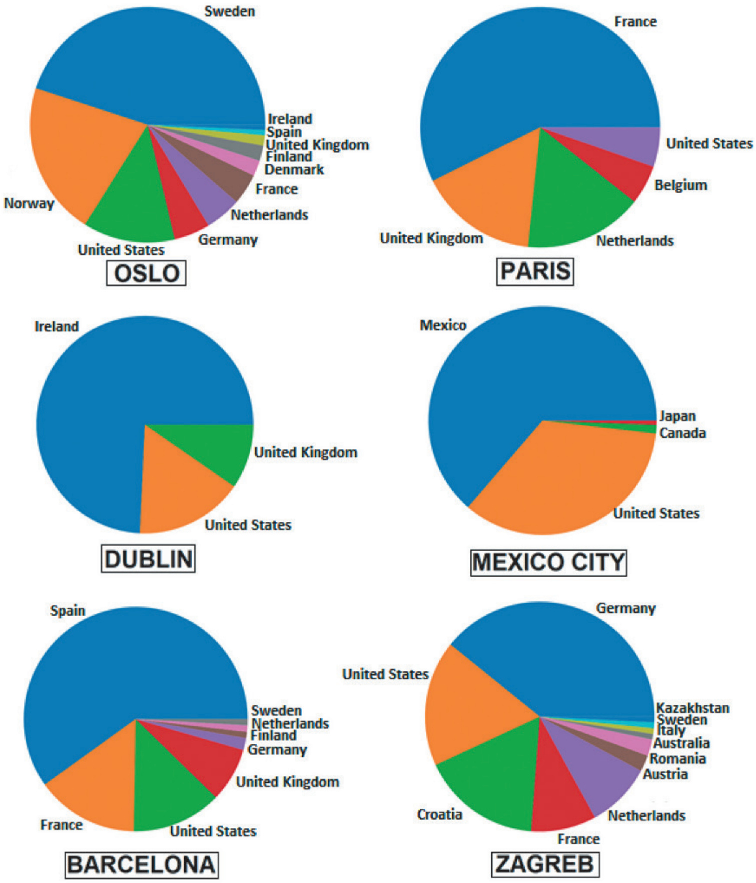
[4] https://ipmap.ripe.net/

While our findings ought to be of interest to governments, policymakers, and academics, we note that there are limitations in our approach that need to be considered. Geolocation was not possible for all the IP addresses we encountered, as many infrastructure IP addresses are firewalled and do not respond to traceroutes or pings by design. Conducting similar studies for other apps, including some that are of more obvious concern for national security, could highlight the strategic dimension of software supply chains to a greater extent. Moreover, expanding the cases to include more states beyond Europe and the West would better illuminate the unequal distribution of digital infrastructures and the political implications of that. Further, our study offers only a snapshot of the situation, and extending the project over time will likely unearth the extent to which service provision fluctuates. All these limitations were primarily a question of resources, and we hope that further studies can build on our findings.

4. GEOGRAPHIC AND CORPORATE DEPENDENCIES OF MOBILE APPLICATIONS

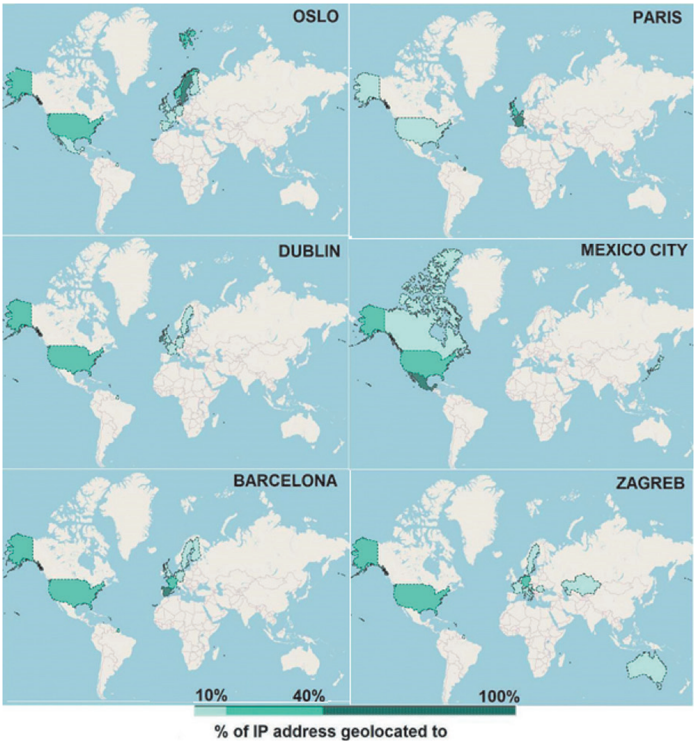
The geographic dispersion of dependencies illustrates how the different countries depend to varying degrees on services hosted outside of their borders. Using apps and services in Oslo and Zagreb, capitals in smaller countries of Norway and Croatia, requires significant global connectivity. For both, the majority of contacted infrastructure IP addresses are hosted abroad. For Oslo, there appears to be a high dependency on Sweden, while Zagreb is highly dependent on Germany. Both cities have 20% or less of the services hosted domestically. For our measurement, cities in larger states, as well as Dublin in Ireland, paint a different picture. All have domestic hosting for more than half of the microservices, with Dublin’s 68.7% representing the highest share of domestic hosting. Moreover, Oslo and Zagreb draw on more geographically distributed infrastructure dependencies than the other cases. Figure 2 shows the distribution of IP addresses that were geolocated to a specific country.

FIGURE 2: GEOGRAPHICAL DEPENDENCIES OF MOBILE APPLICATIONS



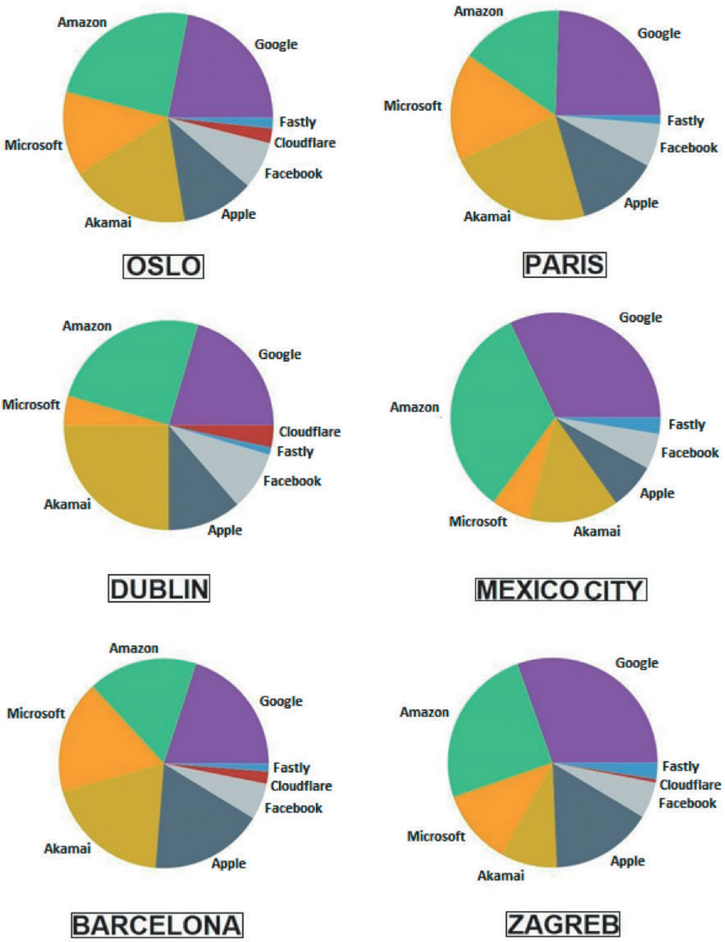
The geographical dependencies can also be illustrated by a map, as shown in Figure 3. The starkest difference is between Paris and Zagreb. The former draws largely on geographically proximate hosting infrastructures, as well as the United States, while Zagreb draws even on services hosted in Kazakhstan and Australia.

FIGURE 3: GEOGRAPHICAL DEPENDENCIES OF MOBILE APPLICATIONS



Turning to corporate centralization, the picture is slightly different. Most importantly, we see that the same five companies – Amazon, Google, and Microsoft (the three largest cloud infrastructure providers), Akamai (the largest CDN), and Apple – own the infrastructure IP addresses for all cases. Paris, Barcelona and Oslo depend on the same corporate networks to more or less the same degree, with Dublin having a similar if slightly skewed distribution. On the other hand, Zagreb and Mexico City are noteworthy for the outsized importance of Google and Amazon. We show the distribution of CDN providers in Figure 4.

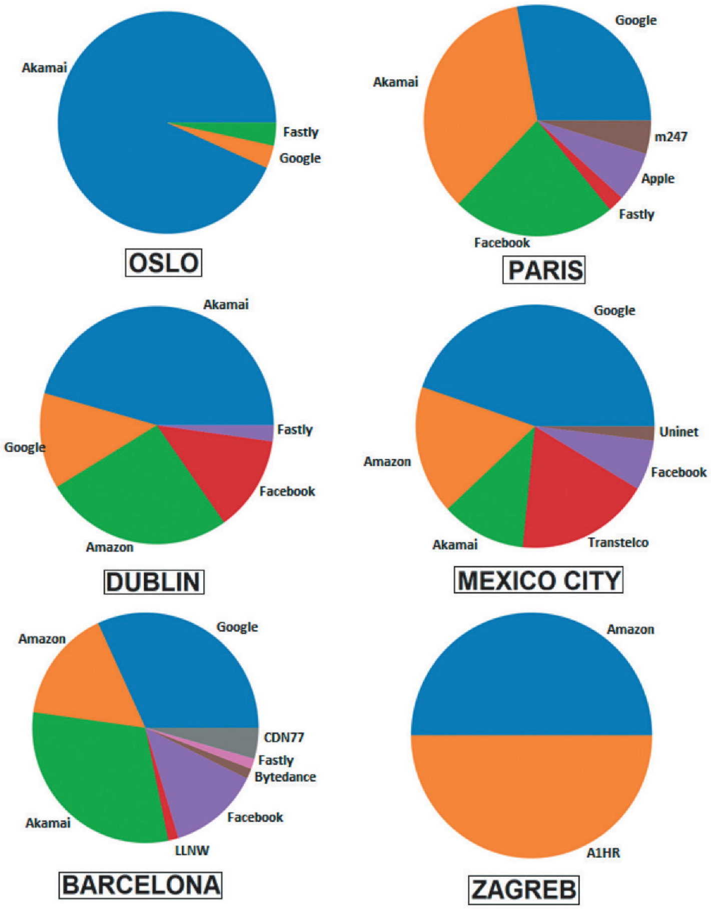
FIGURE 4: MAJOR CDN PROVIDERS FOR MOBILE APPLICATIONS



We further notice that the variation is even more pronounced when coupling the two measurements – as indicated by the heterogeneity of local IPs.

The cities where we measure a large share of foreign IPs also appear to depend locally on a single hosting provider. While 20.8% of IPs for Oslo are located in Norway, they overwhelmingly belong to a single CDN provider, Akamai. A similar pattern is apparent for Zagreb, where the local IP addresses belong to only two infrastructure providers, Amazon and A1HR. These are shown in Figure 5. While not conclusive, these findings indicate that cities at the margin of global networks have higher concentrations of dependence on both physical infrastructures and corporations. Contrasted with states that have more local hosting, an implication might be that smaller states ought to consider the necessity of attracting investments by infrastructure providers to enhance their resilience.

FIGURE 5: INTRA-COUNTRY GEO-DEPENDENCY CDN HETEROGENEITY FOR MOBILE APPLICATIONS



5. ECONOMIC DEPENDENCIES AND STRATEGIC CONSIDERATIONS

In recent years, states across the globe have started to re-evaluate their economic dependencies and rethink them as potential security concerns and vulnerabilities. For digital supply chains, the widespread use of software ecosystems and their embeddedness in consolidated physical infrastructure invites us to consider how and under what circumstances such dependencies could be problematic.

The selected apps are not necessarily problematic on their own. Instead, they are chosen as illustrative of the geographical and corporate distribution of digital infrastructure. While not perfect or exhaustive, these examples offer a preliminary step towards understanding cyber insecurity as the result of infrastructural configurations and economic decisions. First and foremost, this illuminates the extent to which dependence on assets beyond a state's boundaries intersects with connectivity infrastructures to create potential issues of national cyber resilience.

Partly, this is a story about the consolidation of key digital infrastructures in the hands of a few corporate entities. While the consolidation of digital markets has predominantly been identified as a strategic concern, it also poses potential resiliency issues as outages have global effects (Palmer 2021; Taylor 2021). Regardless of the coercive potential, the global provision of critical services by a handful of companies potentially poses a systemic risk – addressing which is seemingly beyond the reach of states.

With software relying on extensive ecosystems and supply chains and the distribution of microservices in various locations, the nature of connectivity infrastructure has changed, creating new vulnerabilities. Consequently, the states with significant dependencies beyond their borders also risk the lack of global connectivity and redundancy becoming recast as a security concern. Whether the concern is that submarine cables and other connectivity infrastructures can be targeted by other states (Brzozowksi 2020), disrupted by activists (Leicester 2022), harmed by commercial activity or accidents (Mauldin 2017), or simply severed by natural disasters (Schia et al. 2022), those states on the outskirts of communication networks are vulnerable to an extent to which others are not.

This intersection between mapping digital dependencies and the infrastructures that create connectivity generates complex topologies with unequal effects. Mapping the geographical distribution illuminates a notable difference between the six locations studied here. This dependency is partly shared, such as the extent to which all states depend on resources hosted in the US. Yet, it also involves ties with neighbouring countries or other regional hubs. For Oslo, this means a dependence on Sweden, while for Zagreb the lack of domestic options creates a dependence on Germany. These dependencies are also likely to endure, at least in the short to medium term, and perhaps even longer for those states without the means to replace global infrastructures and networks with domestic alternatives.

To some extent, the degree of dependency is a consequence of domestic market size, as larger markets are less dependent on infrastructures beyond their borders. Intuitively this makes sense, as global corporations are more likely to invest in comprehensive infrastructures to service larger markets. Yet this is by no means the entire story. Ireland is a telling example, as its economic strategy of attracting global corporations has seen a boon of investment in data centres and accompanying infrastructures by Big Tech. In a physical sense, how companies structure their global networks – where data centres are built and how networks of submarine and terrestrial fibre optic cables are constructed – matters for the dependencies that arise.

This latter point illustrates how smaller states can address dependency, in part, by playing into the hands of large digital companies. Our examination of corporate consolidation highlights how, for the selected apps, a limited number of companies are essential for all. Arguably, for a case like Ireland, the success in attracting investments from global corporations (Buckley and Ruane 2006) enhances the state's digital resiliency by limiting dependencies beyond its boundaries and enhancing resilience through private investment.

For those states on the outskirts of global networks, cyber security and resiliency might instead necessitate investments in better connectivity. In the case of Norway, concerns over limited redundancy have already triggered political interventions to diversify connectivity as a security mitigation (Norwegian Government 2019). From the perspective of developing states, such investments can be expensive, thus reinforcing digital divides as wealthy consumer markets are more likely to attract investments that bolster resiliency. Moreover, the need to connect to pre-existing hubs in the name of security and resiliency can trigger self-reinforcing effects, as central nodes in the global network attract more infrastructures that, in turn, enhance their attraction as locations for investment (Blum 2013). Furthermore, while attracting investments potentially addresses the issue of geographical dependencies, it only entrenches the dependence on companies and the possible systemic risks that come with that. Parsing the different types of dependencies, thus, highlights the extent to which they are in tension with each other.

6. CONCLUSIONS

This paper has examined the software supply chains of popular mobile phone applications or services by studying their geographic and corporate distribution. Considering renewed political attention to global supply chain fragilities and vulnerabilities, we identify the first steps to developing similar mappings of dependencies of individual states. The picture that emerges is mixed. On the one hand, our findings indicate significant variation from one location to the other. Users in Paris, Mexico City, Barcelona, and Dublin are seemingly far less reliant on international connectivity than those in Oslo or Zagreb. This possibly reflects how larger markets, as well as those able to attract investments in data centres by large digital corporations, are less dependent on outside infrastructures. This variety is mirrored in the local heterogeneity of infrastructure providers, as local hosting is often dominated by a single provider. However, this concentration at the corporate level overall primarily reflects the domination of five globally operating US-owned corporations.

We argue that these dependencies can have various effects on cyber resilience and security at the national level – whether that is exacerbating concerns over the impacts of regulatory changes in other states, heightening the criticality of global connectivity, or creating systemic risks of infrastructural failures by global companies. Crucially, addressing these different concerns might introduce contradictory solutions, in particular incentivizing the investments of large hosting providers. Unpacking how digital supply chains vary for individual states allows for greater granularity in our understanding of cyber resilience and dependencies, highlighting the intersection between political and economic forces in shaping different security contexts.

Our contribution is not to definitively prove that digital dependencies are always a security concern, that they always need to be addressed, or that any form of reliance on resources beyond borders are problematic. For the cases examined here, it is not given that the dependencies are problematic or worthy of addressing. However, with political attention turning, for a variety of reasons, to a re-examination of supply chains and dependencies, cyberspace as an entirely man-made domain poses a different set of questions for these trends. There is nothing given about where data resides or what states are central and not. The inequalities that arise emerge from the sum of decisions made by discrete entities – often private companies – and they affect states differently. With states re-examining their dependencies and vulnerabilities, taking stock of how this plays out for digital technologies is a complex affair and our contribution merely a first step.

REFERENCES

Aaronson, Susan Ariel. 2019. 'What Are We Talking about When We Talk about Digital Protectionism?' *World Trade Review* 18(4): 541–77. https://doi.org/10.1017/S1474745618000198.

Amoore, Louise. 2018. 'Cloud geographies'. *Progress in Human Geography* 42(1): 4–24. https://doi.org/10.1177/0309132516662147.

Aradau, Claudia. 2010. 'Security that Matters: Critical Infrastructure and Objects of Protection'. *Security Dialogue* 41(5): 491–514. https://doi.org/10.1177/0967010610382687.

Autolitano, Simona, and Agnieszka Pawlowska. 2021. 'Europe's Quest for Digital Sovereignty. GAIA-X as a Case Study'. Istituto Affari Internazionali. https://www.iai.it/sites/default/files/iaip2114.pdf.

Babić, Milan, Adam D. Dixon, and Imogen T. Liu, eds. 2022. *The Political Economy of Geoeconomics: Europe in a Changing World.* Cham: Springer International Publishing.

Baezner, Marie, and Patrice Robin. 2018. *Cyber Sovereignty and Data Sovereignty*. CSS Cyberdefense Trend Analyses. November 2018. https://doi.org/10.3929/ethz-b-000314613.

Blum, Andrew. 2013. *Tubes. A journey to the center of the Internet*. First Ecco paperback edition. New York: Ecco an imprint of HarperCollinsPublishers.

Bowker, Geoffrey C., Karen Baker, Florence Millerand, and David Ribes. 2010. 'Toward Information Infrastructure Studies: Ways of Knowing in a Networked Environment'. In *International Handbook of Internet Research*, edited by Jeremy Hunsinger, Lisbeth Klastrup and Matthew Allen. Dordrecht: Springer, 97–117.

Brzozowksi, Alexandra. 2020. 'NATO Seeks Ways of Protecting Undersea Cables from Russian Attacks'. EURACTIV. 23 October 2020. https://www.euractiv.com/section/defence-and-security/news/nato-seeks-ways-of-protecting-undersea-cables-from-russian-attacks/.

Buckley, Peter J., and Frances Ruane. 2006. 'Foreign Direct Investment in Ireland: Policy Implications for Emerging Economies'. *World Economy* 29(11): 1611–28. https://doi.org/10.1111/j.1467-9701.2006.00860.x.

Budnitsky, Stanislav, and Lianrui Jia. 2018. 'Branding Internet sovereignty: Digital media and the Chinese–Russian cyber alliance'. *European Journal of Cultural Studies* 21(5): 594–613. https://doi.org/10.1177/1367549417751151.

Cartwright, Madison. 2020. 'Internationalising State Power through the Internet: Google, Huawei and Geopolitical Struggle'. *Internet Policy Review* 9(3). https://doi.org/10.14763/2020.3.1494.

Choer Moraes, Henrique, and Mikael Wigel. 2022. 'Balancing Dependence: The Quest for Autonomy and the Rise of Corporate Geoeconomics'. In *The Political Economy of Geoeconomics: Europe in a Changing World*, edited by Milan Babić, Adam D. Dixon, and Imogen T. Liu. Cham: Springer International Publishing, 29–55.

Christakis, Theodore. 2020. '"European Digital Sovereignty": Successfully Navigating Between the "Brussels Effect" and Europe's Quest for Strategic Autonomy'. *SSRN Journal*. https://doi.org/10.2139/ssrn.3748098.

Couture, Stephane, and Sophie Toupin. 2019. 'What Does the Notion of "Sovereignty" Mean When Referring to the Digital?' *New Media & Society* 21(10): 2305–22. https://doi.org/10.1177/1461444819865984.

Cox, Russ. 2019. 'Surviving Software Dependencies'. *Communications of the ACM* 62(9): 36–43. https://doi.org/10.1145/3347446.

Decan, Alexandre, Tom Mens, and Philippe Grosjean. 2019. 'An Empirical Comparison of Dependency Network Evolution in Seven Software Packaging Ecosystems'. *Empirical Software Engineering* 24(1): 381– 416. https://doi.org/10.1007/s10664-017-9589-y.

Dunn Cavelty, Myriam. 2013. 'From Cyber-Bombs to Political Fallout: Threat Representations with an Impact in the Cyber-Security Discourse'. *International Studies Review* 15(1): 105–22. https://doi.org/10.1111/misr.12023.

Dunn Cavelty, Myriam, and Andreas Wenger. 2020. 'Cyber Security Meets Security Politics: Complex Technology, Fragmented Politics, and Networked Science'. *Contemporary Security Policy* 41(1): 5–32. https://doi.org/10.1080/13523260.2019.1678855.

Dunn Cavelty, Myriam and Andreas Wenger, eds. 2022. *Cyber Security Politics. Socio-Technological Transformations and Political Fragmentation*, 1st ed. Milton Park, UK; New York, NY: Routledge.

Ellison, Robert J., John B. Goodenough, Charles B. Weinstock, and Carol Woody. 2010. 'Evaluating and Mitigating Software Supply Chain Security Risks'. Software Engineering Institute. https://apps.dtic.mil/sti/pdfs/ADA522538.pdf.

Epifanova, Alena. 2020. 'Digital Sovereignty on Paper. Russia's Ambitious Laws Conflict with Its Tech Dependence'. Kennan Institute. The Russia File. https://web.archive.org/web/20220105204833/https://www.wilsoncenter.org/blog-post/digitalsovereignty-paper-russias-ambitious-laws-conflict-its-tech-dependence.

European Commission. 2022. *EU Strategic Dependencies and Capacities. Second Stage of In-depth Reviews*. Commission Staff Working Document. Brussels: European Union. Brussels. https://ec.europa.eu/docsroom/documents/48878.

Farrell, Henry, and Abraham L. Newman. 2019. 'Weaponized Interdependence: How Global Economic Networks Shape State Coercion'. *International Security* 44(1): 42–79. https://doi.org/10.1162/isec_a_00351.

Flensburg, Sofie, and Signe Sophus Lai. 2020. 'Networks of Power. Analysing the Evolution of the Danish Internet Infrastructure'. *Internet Histories* 5(2): 79–100. https://doi.org/10.1080/24701475.2020.1759010.

Gertz, Geoffrey, and Miles M. Evers. 2020. 'Geoeconomic Competition: Will State Capitalism Win?' *Washington Quarterly* 43(2): 117–36. https://doi.org/10.1080/0163660X.2020.1770962.

Gharaibeh, Manaf, Anant Shah, Bradley Huffaker, Han Zhang, Roya Ensafi, and Christos Papadopoulos. 2017. 'A Look at Router Geolocation in Public and Commercial Databases'. In *Proceedings of the 2017 Internet Measurement Conference*, 463–69. New York, NY: Association for Computing Machinery.

Gjesvik, Lars. 2022. 'Private infrastructure in weaponized interdependence'. *Review of International Political Economy* 3(2): 1–25. https://doi.org/10.1080/09692290.2022.2069145.

Goede, Marieke de. 2020. 'Finance/Security Infrastructures'. *Review of International Political Economy* 28(2): 1–18. https://doi.org/10.1080/09692290.2020.1830832.

Goede, Marieke de, and Carola Westermeier. 2022. 'Infrastructural Geopolitics'. *International Studies Quarterly* 66(3). https://doi.org/10.1093/isq/sqac033.

Goldsmith, Jack L., and Tim Wu. 2006. *Who Controls the Internet? Illusions of a Borderless World*. New York: Oxford University Press.

Harrand, Nicolas, Thomas Durieux, David Broman, and Benoit Baudry. 2021. 'Automatic Diversity in the Software Supply Chain'. https://doi.org/10.48550/arXiv.2111.03154.

Hong, Yu, G. Thomas Goodnight. 2020. 'How to Think about Cyber Sovereignty: The Case of China'. *Chinese Journal of Communication* 13 (1): 8–26. https://doi.org/10.1080/17544750.2019.1687536.

Hughes, Thomas P. 1994. ‘Technological Momentum’. In *Does Technology Drive History? The Dilemma of Technological Determinism*, edited by Merritt Roe Smith, Leo Marx. Cambridge, MA: MIT Press.

Hummel, Patrik, Matthias Braun, Max Tretter, and Peter Dabrock. 2021. ‘Data Sovereignty: A Review’. *Big Data & Society* 8(1). https://doi.org/10.1177/2053951720982012.

Ilves, Luukas, and Anna-Maria Osula. 2020. ‘The Technological Sovereignty Dilemma and How New Technology Can Offer a Way Out’. *European Cybersecurity Journal* 6(1).

Inkster, Nigel. 2021. *The Great Decoupling. China, America and the Struggle for Technological Supremacy*. London: Hurst Publishers. https://ebookcentral.proquest.com/lib/kxp/detail.action?docID=6665367.

Irion, Kristina. 2012. ‘Government Cloud Computing and National Data Sovereignty’. *Policy & Internet* 4(3–4): 40–71. https://doi.org/10.1002/poi3.10.

Irion, Kristina, Mira Burri, Ans Kolk, and Stefania Milan. 2021. ‘Governing “European Values” Inside Data Flows: Interdisciplinary Perspectives’. *Internet Policy Review* 10(3). https://doi.org/10.14763/2021.3.1582.

Kwet, Michael. 2019. ‘Digital Colonialism: US Empire and the New Imperialism in the Global South’. *Race & Class* 60(4): 3–26. https://doi.org/10.1177/0306396818823172.

Lambach, Daniel, and Kai Oppermann. 2022. ‘Narratives of Digital Sovereignty in German Political Discourse’. *Governance*. https://doi.org/10.1111/gove.12690.

Leicester, John. 2022. ‘French Police Probe Multiple Cuts of Major Internet Cables’. AP News. 21 October 2022. https://apnews.com/article/technology-europe-france-marseille-business-49d27ccc0195f1c48b33a5634232031f.

Leonard, Mark. 2021. *The Age of Unpeace. How Connectivity Causes Conflict*. London: Transworld Publishers Ltd.

Mauldin, Alan. 2017. ‘Cable Breakage. When and How Cables Go Down’. TeleGeography. 3 May 2017. https://blog.telegeography.com/what-happens-when-submarine-cables-break.

McNamara, Kathleen R., and Abraham L. Newman. 2020. ‘The Big Reveal: COVID-19 and Globalization’s Great Transformations’. *International Organization* 74(S1): E59–E77. https://doi.org/10.1017/S0020818320000387.

Monsees, Linda, and Daniel Lambach. 2022. ‘Digital Sovereignty, Geopolitical Imaginaries, and the Reproduction of European Identity’. *European Security* 31(3): 377–94. https://doi.org/10.1080/09662839.2022.2101883.

Mueller, Milton. 2010. *Networks and States: The Global Politics of Internet Governance*. Cambridge, MA: MIT Press.

Mueller, Milton L. 2020. ‘Against Sovereignty in Cyberspace’. *International Studies Review* 22(4): 779–801. https://doi.org/10.1093/isr/viz044.

Mügge, Daniel. 2023. ‘The Securitization of the EU’s Digital Tech Regulation’. *Journal of European Public Policy*. https://doi.org/10.1080/13501763.2023.2171090.

Musiani, Francesca, Derrick L. Cogburn, Laura DeNardis, and Nanette S. Levinson. 2016. *The Turn to Infrastructure in Internet Governance*. New York, NY: Palgrave Macmillan.

Nocetti, Julien. 2015. ‘Contest and Conquest. Russia and Global Internet Governance’. *International Affairs* 91(1): 111–30. https://doi.org/10.1111/1468-2346.12189.

Norwegian Government. 2019. *Tiltaksoversikt til nasjonal strategi for digital sikkerhet*. Accessed 10 March 2019. https://www.regjeringen.no/contentassets/c57a0733652f47688294934ffd93fc53/tiltaksoversikt---nasjonal-strategi-for-digital-sikkerhet.pdf.

Nye, Joseph S. 2020. 'Power and Interdependence with China'. *Washington Quarterly* 43(1): 7–21. https://doi.org/10.1080/0163660X.2020.1734303.

Ohm, Marc, Henrik Plate, Arnold Sykosch, and Michael Meier. 2020. 'Backstabber's Knife Collection: A Review of Open-Source Software Supply Chain Attacks'. In *Detection of Intrusions and Malware, and Vulnerability Assessment*, Lecture Notes in Computer Science, edited by Clémentine Maurice, Leyla Bilge, Gianluca Stringhini, and Nuno Neves. Cham: Springer International Publishing, 23–43.

Ortiz Freuler, Juan. 2022. 'The Weaponization of Private Corporate Infrastructure: Internet Fragmentation and Coercive Diplomacy in the 21st Century'. *Global Media and China* 8(1): 6–23. https://doi.org/10.1177/20594364221139729.

Palmer, Annie. 2021. 'Dead Roombas, Stranded Packages and Delayed Exams. How the AWS Outage Wreaked Havoc Across the U.S.'. CNBC. 9 December 2021. https://www.cnbc.com/2021/12/09/how-the-aws-outage-wreaked-havoc-across-the-us.html.

Parks, Lisa D., and Nicole Starosielski, eds. 2015. *Signal traffic. Critical studies of media infrastructures*. Urbana, IL: University of Illinois Press.

Pohle, Julia, and Thorsten Thiel. 2020. 'Digital Sovereignty'. *Internet Policy Review* 9(4). https://doi.org/10.14763/2020.4.1532.

Privacy Company. 2021. *Google Workspace DPIA for Dutch DPA*. Commissioned by Dutch Ministry of Justice and Security. https://www.rijksoverheid.nl/documenten/publicaties/2021/02/12/google-workspace-dpia-fordutch-dpa.

Rhodes, R. A. W. 1994. 'The Hollowing Out of the State. The Changing Nature of the Public Service in Britain'. *Political Quarterly* 65(2): 138–51. https://doi.org/10.1111/j.1467-923X.1994.tb00441.x.

Rosa, Fernanda R., and Janice A. Hauge. 2022. 'GAFA's Information Infrastructure Distribution: Interconnection Dynamics in the Global North versus Global South'. *Policy & Internet* 14(2): 424–49. https://doi.org/10.1002/poi3.278.

Sassen, Saskia. 1996. *Losing Control? Sovereignty in the Age of Globalization*. Leonard Hastings Schoff Lectures. New York: Columbia University Press. http://gbv.eblib.com/patron/FullRecord.aspx?p=1273980.

Schia, Niels Nagelhus, Lars Gjesvik, and Ida Rødningen. 2022. 'Loss of Tonga's Telecommunication. What Happened, How Was It Managed and What Were the Consequences?' NUPI. https://www.nupi.no/en/publications/cristin-pub/loss-of-tonga-s-telecommunication-what-happened-how-was-it-managed-and-what-were-the-consequences.

Srivastava, Swati. 2021. 'Algorithmic Governance and the International Politics of Big Tech'. *Perspective on Politics*, 1–12. https://doi.org/10.1017/S1537592721003145.

Stevens, Tim. 2018. 'Global Cybersecurity: New Directions in Theory and Methods'. *Politics and Governance* 6(2): 1–4. https://doi.org/10.17645/pag.v6i2.1569.

Taylor, Josh. 2021. 'Facebook Outage. What Went Wrong and Why Did It Take So Long to Fix After Social Platform Went Down?' *Guardian*, 5 October 2021. https://www.theguardian.com/technology/2021/oct/05/facebook-outage-what-went-wrong-and-why-did-it-take-so-long-to-fix.

Walter, Stefanie. 2021. 'The Backlash Against Globalization'. *Annual Review of Political Science* 24(1): 421–42. https://doi.org/10.1146/annurev-polisci-041719-102405.

Winseck, Dwayne. 2019. 'Internet Infrastructure and the Persistent Myth of U.S. Hegemony'. In *Information, Technology and Control in a Changing World*, edited by Blayne Haggart, Kathryn Henne, and Natasha Tusikov. Cham: Springer International Publishing, 93–120.

2023 15th International Conference on Cyber Conflict
Meeting Reality
T. Jančárková, D. Giovannelli, K. Podiņš, I. Winther
2023 © NATO CCDCOE Publications, Tallinn

Modeling 5G Threat Scenarios for Critical Infrastructure Protection

Gerrit Holtrup
Kudelski IoT Security
Cheseaux-sur-Lausanne, Switzerland
gerrit.holtrup@nagra.com

William Blonay
NATO CCDCOE
Tallinn, Estonia
william.blonay@ccdcoe.org

Martin Strohmeier
armasuisse Science and Technology
Thun, Switzerland
martin.strohmeier@ar.admin.ch

Alain Mermoud
armasuisse Science and Technology
Thun, Switzerland
alain.mermoud@ar.admin.ch

Jean-Pascal Chavanne
Federal Department of Justice and Police
Bern, Switzerland
jean-pascal.chavanne@
isc-ejpd.admin.ch

Vincent Lenders
armasuisse Science and Technology
Thun, Switzerland
vincent.lenders@ar.admin.ch

Abstract: Fifth-generation cellular networks (5G) are currently being deployed by mobile operators around the globe. 5G is an enabler for many use cases and improves security and privacy over 4G and previous network generations. However, as recent security research has revealed, the 5G standard still has technical security weaknesses for attackers to exploit. In addition, the migration from 4G to 5G systems takes place by first deploying 5G solutions in a non-standalone (NSA) manner, where the first step of the 5G deployment is restricted to the new radio aspects of 5G. At the same time, the control of user equipment is still based on 4G protocols; that is, the core network is still the legacy 4G evolved packet core (EPC) network. As a result, many security vulnerabilities of 4G networks are still present in current 5G deployments. To stimulate the discussion about the security risks in current 5G networks, particularly regarding critical infrastructures, we model possible threats according to the STRIDE threat classification model. We derive a risk matrix based on the likelihood and impact of eleven threat scenarios (TS) that affect the radio access and the network core. We estimate that malware or software vulnerabilities on the 5G base station constitute the most impactful threat scenario, though not the most probable. In contrast, a scenario where compromised cryptographic keys threaten communications between network functions is both highly probable and highly impactful. To improve the 5G security posture, we discuss possible mitigations and security controls. Our analysis is generalizable and does not depend on the specifics of any particular 5G network vendor or operator.

Keywords: *5G, next-generation networks, threat scenarios, critical infrastructures, cyber defense, security*

1. INTRODUCTION

The arrival of the fifth generation of cellular networks (5G) enables new use cases compared to previous mobile telecommunications standards. Examples range from the support of stationary devices in the Internet of Things (IoT) to highly mobile settings in vehicular networks. Power, latency, and data rate requirements vary widely across these different device classes. The introduction of the network slice and network function virtualization concepts in 5G are expected to address these differences in functional requirements.

Currently, the migration from 4G to 5G systems is taking place by first deploying 5G solutions in a non-standalone (NSA) manner, where the first step in 5G deployment is restricted to the new radio aspects of 5G (5G-NR). At the same time, the control of user equipment is still based on 4G protocols; that is, the core network is still the legacy 4G network.

Previously unsolved privacy concerns in 4G are addressed in the 5G standard. Contrary to the previous generation, the analysis of the security of the 5G system, as defined in [1], was already an active concern of researchers before the wide deployment of the standard [2]. A formal analysis of the security procedures by Basin et al. [3] has revealed weaknesses that may potentially still be fixed before 5G standalone systems are deployed.

While previous work focuses on the radio interface, this paper analyzes a full standalone system, including the 5G core network (5GC) architecture [4]. However, given the reality that immediate deployments of 5G in the field is NSA deployments, these will also be covered where appropriate.

We build our security analysis on existing literature focusing on the use of 5G in critical infrastructures [5]–[8], including recent research papers published by the CCDCOE

[9]. We first present the STRIDE methodology to achieve this. Then, various threat scenarios (TS) are analyzed in more detail, as well as the associated security controls to address them. Our work lays the foundation for risk analysis of 5G networks in critical infrastructure protection.

2. BACKGROUND

A. STRIDE Methodology

Our threat analysis follows the STRIDE (spoofing, tampering, repudiation, information disclosure, denial of service, and the elevation of privileges) classification [10], [11] of threats developed by Microsoft, which requires data flows between different components to be formalized. The threat assessment methodology is illustrated by six steps in Figure 1. Each component, process, data flow, external entity, and data store is exposed to a subset of threat categories, as described in Table I.

FIGURE 1: STRIDE THREAT ASSESSMENT METHODOLOGY

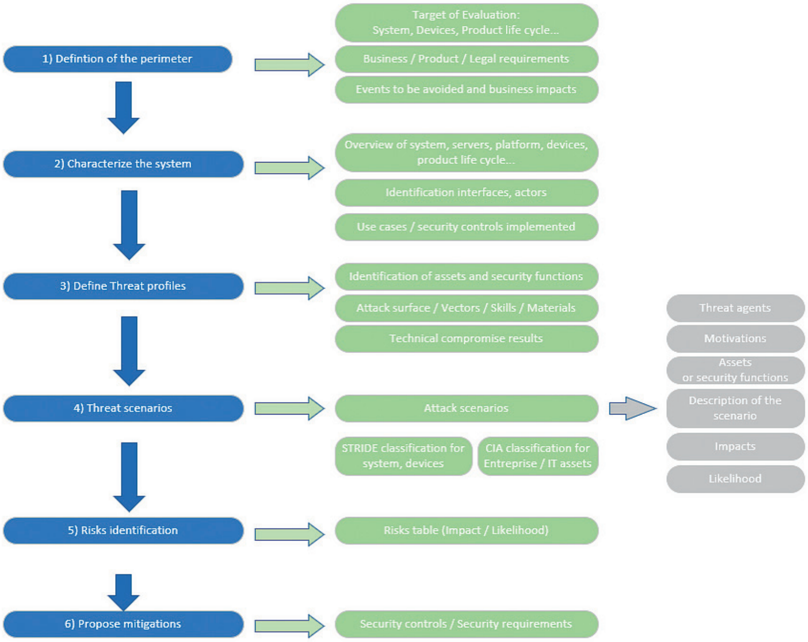

1) Defintion of the perimeter
Target of Evaluation: System, Devices, Product life cycle...
Business / Product / Legal requirements
Events to be avoided and business impacts
2) Characterize the system
Overview of system, servers, platform, devices, product life cycle...
Identification interfaces, actors
Use cases / security controls implemented
3) Define Threat profiles
Identification of assets and security functions
Attack surface / Vectors / Skills / Materials
Technical compromise results
4) Threat scenarios
Attack scenarios
STRIDE classification for system, devices
CIA classification for Entreprise / IT assets
Threat agents
Motivations
Assets or security functions
Description of the scenario
Impacts
Likelihood
5) Risks identification
Risks table (Impact / Likelihood)
6) Propose mitigations
Security controls / Security requirements

TABLE I: THREATS AFFECTING COMPONENTS WITH STRIDE CLASSIFICATION

Components	Spoofing	Tampering	Repudiation	Information disclosure	Denial of service	Elevation of privileges	STRIDE
External entity or interactors	X		X				SR
Process	X	X	X	X	X	X	STRIDE
Data / Keys storage		X		X	X		TID
Data flow		X		X	X		TID
Devices	X	X		X	X	X	STRIDE

1) 5G System Overview

There are several foundational changes in the 5G architecture compared to 4G. First, the 5G system extends to new frequency spectra, which increase data rates and are well suited for massive MIMO (multiple-input multiple-output) applications and micro-cells. Indeed, transmitters for frequencies in the mm-wave range have intrinsically high directivity, thereby also providing spatial multiplexing capabilities with more ease than at lower frequencies. However, power generation within these frequency ranges is still difficult, and absorption rates by the atmosphere tend to be high. They are therefore unsuitable for macro-cells, which are expected to continue to use frequency bands previously allocated to 3G and 4G cellular networks.

2) 5G New Radio (5G-NR)

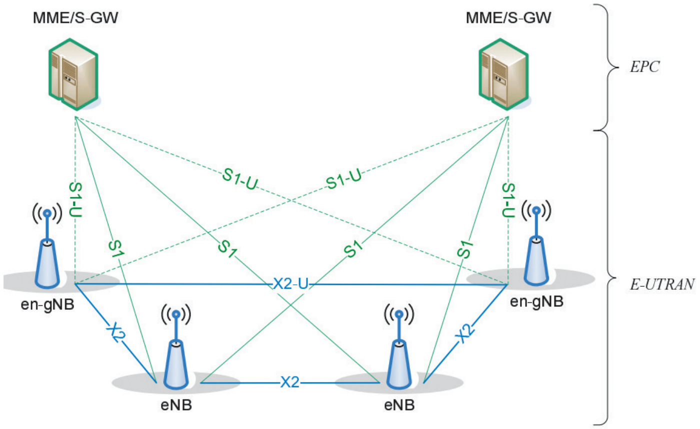
The 5G radio interface uses the same frequency ranges as 4G plus additional frequency bands. This includes frequencies in the sub-6GHz band, particularly the newly attributed frequencies around 3.5 GHz and frequencies around 24–26 GHz. The frequency bands above 6 GHz offer inherently higher bandwidth but present higher absorption rates and thus limit the size of a single cell. Furthermore, at these frequencies it is getting more complicated to use antennas with wide beamwidth as the antenna-to-wavelength ratio has the tendency to result in more directive antennas than at lower frequencies. To adapt to the higher frequency bands and ensure adequate coverage while meeting the increasing demands for end-user performance in uplink and downlink, mobile network operators deploy advanced antenna array systems with beamforming and MIMO capabilities. The frequency bands below 1 GHz still offer the means of achieving coverage with a minimum number of cells (thus achieving coverage in rural areas where the high-density deployment of nano-cells would be too costly).

3) 5G Non-standalone

The first stages in 5G deployment focus on the integration of 5G-NR base stations (known as gNodeBs or gNBs) into the existing 4G system in the context of a multi-radio dual connectivity implementation (see Figure 2). This is done by adhering to standard TS 37.340 [12]. The core network is still the 4G evolved packet core (EPC), and the master nodes for dual connectivity are 4G base stations (eNBs). The 5G base station is integrated as an en-gNB into the system and acts as a secondary node. It only exchanges user plane data with the core network. All control data is exchanged with the eNB over the X2 link. From a user equipment (UE) perspective, the control plane is located in the eNB, while user plane data are transmitted over the gNB. This dual connectivity system also implies that the UEs that support this mode have to integrate concurrent 4G and 5G radio interface support. The increased power consumption might be unsuitable for low-power applications in the IoT context.

Finally, UEs supporting this mode of operation must use the standard 4G network attach procedures, which implies sending their unique international mobile subscriber identity (IMSI) clear to the network during the first attach. This means that the identity concealment feature introduced for 5G is not usable in non-standalone deployments, and IMSI catching is still possible without any increased difficulty.

FIGURE 2: NSA 5G NETWORK ACCORDING TO [12], DEPICTING THE INTERACTION BETWEEN THE 4G EPC USING THE MOBILITY MANAGEMENT ENTITY / SERVING GATEWAY (MME/S-GW) AND THE 5G EVOLVED UNIVERSAL TERRESTRIAL RADIO ACCESS NETWORK (E-UTRAN)

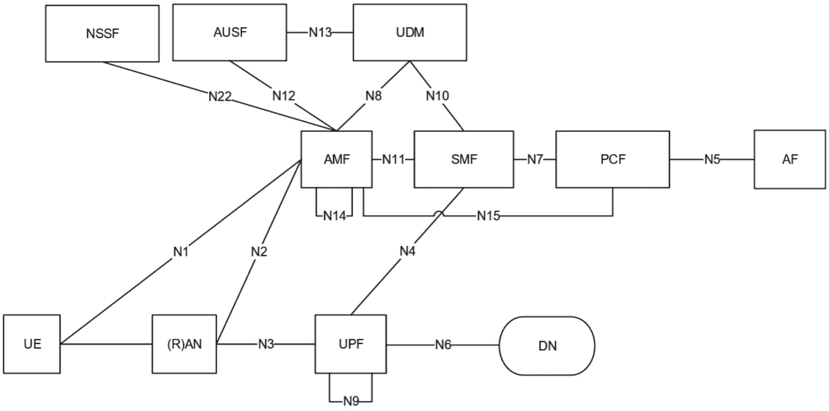


4) 5G Standalone

In the case of a standalone 5G deployment (or of a dual connectivity deployment using a 5G core network), the radio interface and core network differ from 4G. In 5G, the architecture has been designed to achieve a cleaner separation of control and user planes. The core network has been redesigned using a service-based architecture, which makes the virtualization of some network functions easier. Once virtualized, the network functions can be implemented as cloud instances. To guarantee the security of virtualized network functions, the operator of the 5G system has to pay attention to the isolation mechanisms between the virtual machines. The implicit level of trust in a serving network has also been reduced, and some new security features have been implemented. Authentication and access management functions are now in two different building blocks of the system.

Figure 3 shows the various reference points in the 5G system architecture if no roaming is involved, that is if the serving network corresponds to the home network (roaming is out of the scope of this paper due to space constraints). The access and mobility management function (AMF) is clearly separated from the session management function (SMF). The unified data management (UDM) of the home network and the Universal Subscriber Identity Module (USIM) of the UE contain the same long-term keys used for further key derivation during the authentication process. The authentication server function (AUSF) is located in the home network of the device and performs its authentication. It also provides high-level keys to the AMF that initiated the authentication session.

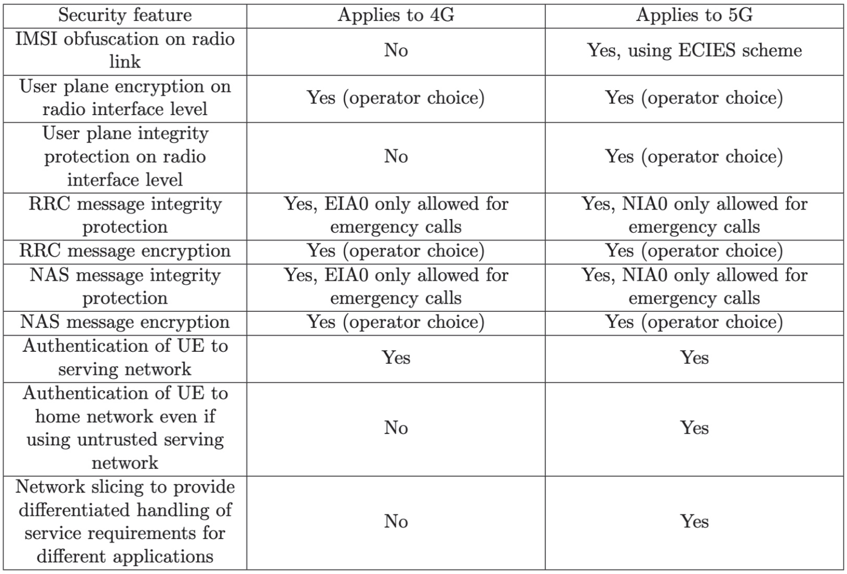
FIGURE 3: REFERENCE ARCHITECTURE IN THE 5G SYSTEM IN A NON-ROAMING CONTEXT – FOR DETAILED EXPLANATIONS AND MEANINGS OF ABBREVIATIONS, PLEASE REFER TO [2]



B. Novel 5G Security Features

In 5G standalone implementations, some new security features mitigate previously identified vulnerabilities. Contrary to previous versions of the 3rd Generation Partnership Project (3GPP) standards, the universal integrated circuit card (UICC) of the UE now contains an asymmetric key element, the public key of the home network for use in elliptic curve algorithms. The (limited) use of asymmetric cryptographic algorithms allows the transmission of protected information to the core network without previous key negotiation with this network. This mechanism avoids IMSI catcher attacks that track mobile phones as the unprotected IMSI in the initial attach request has been replaced by an obfuscated subscription concealed identifier (SUCI) in the initial registration request. Further differences between 4G and 5G security features are summarized in Table II.

TABLE II: COMPARISON OF 4G AND 5G SECURITY FEATURES



Security feature	Applies to 4G	Applies to 5G
IMSI obfuscation on radio link	No	Yes, using ECIES scheme
User plane encryption on radio interface level	Yes (operator choice)	Yes (operator choice)
User plane integrity protection on radio interface level	No	Yes (operator choice)
RRC message integrity protection	Yes, EIA0 only allowed for emergency calls	Yes, NIA0 only allowed for emergency calls
RRC message encryption	Yes (operator choice)	Yes (operator choice)
NAS message integrity protection	Yes, EIA0 only allowed for emergency calls	Yes, NIA0 only allowed for emergency calls
NAS message encryption	Yes (operator choice)	Yes (operator choice)
Authentication of UE to serving network	Yes	Yes
Authentication of UE to home network even if using untrusted serving network	No	Yes
Network slicing to provide differentiated handling of service requirements for different applications	No	Yes

C. Protection Goals

We will now discuss the assets that need to be protected in the 5G ecosystem.

1) User Identity and Location

The first assets are user identity and location. The novel concept of transmitting a concealed SUCI instead of the IMSI in an initial registration/attach procedure provides some level of privacy protection. The visiting network is not supposed to be aware of the unconcealed subscription permanent identifier (SUPI) until the end of the authentication procedure. At this point in time, the home network has effectively authenticated the serving network to be trusted. Even when the SUPI is transmitted to the AMF of the visiting network, the identity is still not provided to the gNB.

However, in some cases (e.g., emergency procedures), the UE will still directly communicate its globally unique SUPI. Other temporally persistent identifiers are also still visible during the registration procedures, such as the global unique temporary identifier (GUTI). The core network can request the device's unique international mobile equipment identity (IMEI), which may allow the correlation of a connection with a specific user (particularly if the user connects to both 4G and 5G networks).

If an attacker is capable of correlating the 5G-GUTI with the SUPI or IMEI of a user, it is still possible to track the position of the UE. Indeed, all initial requests in the case of the change of the serving cell will still reveal the 5G-GUTI.

2) Service Availability

The impact of denying a device connectivity varies from small annoyance because a phone call cannot be placed to endangering human life if even emergency calls are no longer possible. For machine-to-machine communications, the systems are expected to be robust in the absence of reliable communications even though the consequences might be anything up to a "graceful" standby of the system.

3) Data Integrity

It is important that the data sink can trust that the incoming data stream is coming from an authentic source. If it is possible to also inject fake data, these pieces of data may not only result in wrong decisions on the receiving end, but the level of trust in any authentic data is also decreased. This can lead either to false-alarm-type situations or to a genuine alarm being disregarded by the system.

4) Data Confidentiality

In all communication contexts, the data transmitted over the radio link is the main asset of this link. Depending on the use case, the data may be sensitive, and its confidentiality has to be protected.

The keys involved in protecting the data both in terms of confidentiality and integrity are secondary assets that must be protected. Indeed, leakage of a device's keys allows an attacker to directly leverage this knowledge to decrypt confidential data and impersonate the device.

5) Network Performance

For safety-critical functions, the general availability of the network service might be insufficient but additionally requires a communications channel that fulfills certain boundary conditions. Such services rely, for example, on low latency or a minimum data rate (quality of service). If the network performance is downgraded below a given threshold either in terms of latency or data rate, then for these devices, this situation can be equivalent to a complete denial of service condition.

3. THREAT SCENARIOS

In this chapter, we identify the threat scenarios for 5G. Table III lists the scenarios and their contexts according to the STRIDE methodology, which we discuss in detail in the following.

TABLE III: LIST OF POTENTIAL THREAT SCENARIOS

STRIDE	Threat scenarios	Context and potential security controls
STRIDE	TS 01: A disgruntled employee with access to the database of all device keys makes a copy of the keys and sells them to a criminal organization	The UDM manages all keys used inside the network. Security control: strict access control and use of a hardware security module (HSM) to protect the keys, update mechanism of keys stored in the operator's UICCs
STRIDE	TS 02: Key extraction through hardware attacks on the UICC element. First the attacker extracts the keys from the UICC of a valid device. The keys are then used to create clones and attack the network or, if an attack is invasive/ destructive to spy on communications of the legitimate user	Difficulty depends on the robustness of the UICC
STRIDE	TS 03: Malware on the mobile equipment (ME) with sufficient privilege dumps the current security context of a device. The dumped keys can then be used to impersonate the device to the network and to decrypt all previous communications of the device	Keys derived in the context of a registration procedure are held outside the UICC in the context of the ME security control: Regular renewal of the device security context by the network
D	TS 04: Physical or logical jamming of devices through fake gNB	- Impact per jammer limited to its coverage - Except for protocol-based jamming during the attach procedure of a device, the duration of impact is only as long as the jammer is active - Security control: Blacklist of fake gNB broadcast in nominal network
I	TS 05: Partial SUCI and permanent equipment identifier (PEI) catcher through interception of radio link	Security control: encryption of signaling messages both on radio and non-access stratum (NAS) level to protect PEI
D	TS 06: Physical or Logical jamming of gNB	- Impact per jammer limited to one gNB - Impact only as long as the jammer is active - Security control: beam forming networks (BFN) to eliminate the jammer's radio signal
TRIDE	TS 07: Exploit software vulnerability in a gNB (or malicious firmware update) to install backdoors to data buffers and extract signaling information in clear or might result in attacker-managed DoS	- Tampered gNB might share handled data - Might provide access to gNB level key-vulnerabilities might be built in unintentionally or by malicious supplier and actions triggered through radio interface - Security control: External audit of gNB code and secure coding rules Authentication of firmware

TRID	TS 08: Exploit software vulnerability in a network function (or malicious firmware update) can lead to misconfiguration of UEs, data leakage and bypass of security controls; in a virtualized network function this can include data leakage through side-channel attacks between virtual machines using the same physical resources	Tampered network function (e.g., AMF) might disclose current security context of a device or not implement all optional security features
TI	TS 09: Extraction of keys used to establish IPSec connection from link node memory. - If a link node (gNB, AMF, etc.) uses software implementation of IPSec, keys might be exposed through heartbleed-style attacks - In gNB, they might not be stored in secure storage and extracted through local physical access	- Software vulnerabilities - Software implementation of cryptographic suites - Security control: Use of robust hardware module for handling of root keys used for secure channel establishment
D	TS 10: Stealing or modifying the physical configuration of a gNB - Disrupting access to the backhaul - Removal of gNB or its antennas in insufficiently secured physical location	Mitigations: - Physical security for gNB access - Overlap in the cell coverage
D	TS 11: Overloading traffic in high priority slice at the cost of lower priority slices (or slices associated with another public land mobile network (PLMN) in the radio access network (RAN) sharing case)	Mitigations: - Proper implementation of service level agreements and resource management function in gNBs

A. TS 01: Operator UDM Database Theft

The keys contained in the UDM database are also stored in the UICC elements of the UEs. Having control of this database allows an attacker to fully impersonate the network. As it is difficult to update the long-term keys in the UICCs (in particular in embedded systems), it is very costly to respond to this attack and a root key update may be the better option.

The main mitigation is strict physical access control to the UDM. Using a hardware security module (HSM) also forces the attacker to make time-consuming attacks once in possession of the HSM to extract the data. This time window might be sufficient for the operator to be aware of the loss of the device and to deploy new keys in the UICCs in their network.

Threat agents: malicious/compromised employee with access to the UDM storage. Given the amount of confidential information being disclosed through one attack, the motivation for a criminal organization or hostile nation can be considered high.

B. TS 02: Device Long-Term Key Extraction Through Hardware Attacks on the UICC Element

The UICC contains the keys used at the root of the key derivation and agreement process between the UE and the network. If an attacker can extract the key material from a legitimate UICC, the attacker can generate clones of the device, eavesdrop on the communication and inject fake data. One attack vector would be to extract the keys before the initial use of the UICC in a UE, but tampering detection is also difficult later on in certain machine-to-machine contexts. As a mitigation, the network should only authorize one active security context at any given time, and thus the cloned (and legitimate) devices cannot function in parallel.

Threat agents: security researchers to check the robustness of products and test their technical capabilities, criminal organizations, and foreign government agencies.

C. TS 03: Non-permanent Key Extraction from Mobile Equipment

Most keys inside the UE are handled inside the ME and not the USIM. While the security requirements are clearly specified for the USIM [1], the requirements are less clear for the ME. While the baseband and application space inside normal UEs are often separate subsystems, both might be handled in the same processor, particularly for low-cost components.

This opens up the possibility that a malicious application running inside the ME has knowledge of the current security context and allows attackers to eavesdrop and inject messages nominally from the UE to the network. Unless the network triggers the renewal of the security context, these keys will remain valid. For a stationary IoT device, the network might want to limit the amount of exchanged data and thus only renew the security context within long intervals.

Depending on the security mechanisms used by the ME to protect against the installation of malware, this attack can be much easier to perform than TS 02, with a nearly comparable result. Even if more complex ME architectures are used, it is expected that the extraction of a security context from the ME is much less costly than extracting secrets from the UICC. The extraction of the security context can, however, only be achieved once the device is operational.

Threat agents: opportunistic hackers, criminals, and security researchers.

D. TS 04: Physical or Logical Jamming of Devices

The basic physical jamming of devices will only affect the UE if the jammer is active. Depending on the covered frequency bands and the beamforming capabilities of the device, the device might even be capable of blocking the angle of arrival of the jammer. In the case of a logical jammer, however, equivalents to known 4G attacks [2] are possible, and their impact persists until the device has undergone a power cycle. Indeed, if a UE tries to switch to this rogue gNB following the cell selection and reselection mechanism described in [13], then the rogue gNB can trigger a new registration procedure followed by transmitting an unprotected REGISTRATION REJECT non-access stratum (NAS) message. As stated in [14, section 4.4.4.2], this message must be processed before a valid security context is established between the UE and the network.

In the case of stationary devices, the cell reselection criteria might be difficult to achieve by the rogue gNB as long as the current cell on which the device is camped remains powerful enough. For mobile devices, the rogue gNB only needs to provide a slightly better signal than other candidate cells in the attacked network. Given that some rejection causes require the device to either follow a power cycle or to have its USIM reinserted, this can have a near-permanent effect on some types of devices. For example, a drone being controlled through 5G would naturally either have to disregard the 5G specifications or go into a safe return mode, as there would be no means of a human manually triggering a power cycle while flying.

The cost of the rogue gNB can be estimated to be lower than a high-end physical jammer.

Threat agents: criminal and terrorist organizations.

E. TS 05: Location Tracking Through Standard Radio Link Interception

Depending on the choice of the network operator, signaling messages can only be integrity protected. Even though the SUPI will only be transmitted in its concealed form, an attacker can still gather the same amount of information through the home network identifier transmitted in the context of the authentication procedure, and the PEI transmitted inside the SECURITY MODE COMPLETE message.

If the network operator chooses to use encryption for signaling messages, an attacker can only capture the SUCI and the associated home network identifier. This may be of interest if the target user's home network is more uniquely identifiable (e.g., a visit of a foreign delegation).

If the attacker possesses a network of (potentially low-cost) radio sensors with sufficient density, it is possible to match and continuously track the location of a given set of UEs. Importantly, with knowledge of the target's location at the beginning of the tracking session, it might be possible to track the target without physically following it after this initial matching phase.

Threat agents: In the absence of the encryption of signaling data, location tracking might interest criminals, terrorist organizations, or foreign government agencies. If only the home network identifier could be intercepted, foreign government agencies might remain motivated to implement this attack. If the tracking is based on a sensor network, then it is likely that only government agencies have the resources to install this type of network.

F. TS 06: Jamming of a gNB

The effect of a physical jammer on a gNB will disappear as soon as it is no longer active. From a protocol point of view, it should, however, be quite easy to obtain a modified rogue UE that continuously jams the random access channels of a gNB. Such a logical jammer would deny new UEs from requesting access to the cell. The gNB would be severely impacted in its operations, and network performance for this cell would decrease drastically.

Suppose the gNB detects the presence of this logical jammer and is capable of locating its position. In that case, the gNB might configure its beam forming networks (BFN) to suppress the jammer signal's arrival direction. However, this suppression capability will depend on the size of its antenna array (and indirectly on cell center frequency). Standard external anti-jamming detection and mitigation by providers or authorities can also mitigate this attack.

This attack would only impact a single gNB.

Threat agents: criminals.

G. TS 07: Malware or Software Vulnerabilities on a gNB

The software stacks inside a gNB and the network functions of the 5G core are complex. The manufacturers of the equipment might also not be willing to share the code even with the network operators, as the scheduling function might contain highly proprietary optimizations. The software of a gNB is expected to be updatable.

Vulnerabilities may be present because of backdoors mandated by the government of the equipment manufacturer, due to coding errors, or after the replacement of the original firmware with malicious firmware. Consequences include threats to all availability, integrity and confidentiality. If the vulnerability is already present in the official firmware, it might be exploitable through the radio network. In this case, all gNBs with the same vulnerability would be at risk, and the result could be catastrophic for the infrastructure of a network operator or even a country.

The modified gNB could also be used as an entry point to attack core network functions through the existing link between the gNB and the core network (particularly the user plane function and the AMF). However, the feasibility of this attack depends on the absence of any load balancer in front of the 5G core.

Thanks to virtualization concepts, the non-time critical sections of the gNB central unit can be located in the cloud, which may handle more than one physical radio access network (RAN). In this case, a successful attack on the cloud instance (e.g., physical access to the data center hosting the cloud VM) directly impacts more than a single physical gNB instance.

Threat agents: disgruntled member of the development team for malicious inclusion of a backdoor in the firmware code base, member of the development team unintentionally inserting exploitable vulnerability into the firmware, security researcher analyzing the firmware and detecting a vulnerability, government agency mandating the inclusion of a backdoor in code provided to foreign operators that the mandating government agency can activate at will.

H. TS 08: Malware or Software Vulnerability in 5G Core Network Functions

Similar to the gNB, an attacker might be able to exploit a vulnerability in a network function such as the AMF. Given the key derivation schemes used in 5G, knowledge of lower-level keys does not provide knowledge of higher-level keys. However, this reasoning does not apply in the other direction. A misconfigured SMF could also instruct the gNB to configure the data bearers as not being confidentiality protected.

As the network functions do not require being distributed to cover the territory of the operator, they can be located in physically secure locations. This makes a local attack on network functions less likely.

If virtualization is used, they can also be operated from the cloud and thus be physically hosted in the data centers of cloud service providers. Besides the potential legal consequences, this may enable micro-architectural attacks or open up vulnerabilities in the hypervisor managing the virtual machines.

Threat agents: opportunistic hackers if the control interface of the network function is exposed on the public internet; criminal organizations for blackmailing the network operators; government agencies for espionage and control of foreign infrastructure.

I. TS 09: Stealing Keys Used for Link Protection Between Network Functions

If the network equipment is physically accessible, an attacker might also use physical attacks to extract the network keys. However, the network operator should not rely on physical security alone to protect the data in transit between different network functions. Alternatively, an attacker can extract the keys securing the link through a zero-day exploit against the software running inside the network function. It is also possible to attack cloud solutions via side-channel leakages [16] to other functions executed on the same hardware.

Threat agents: criminals, hackers, and security researchers.

J. TS 10: Theft or Physical Misconfiguration of a gNB

Depending on the type of gNB (stationary or mobile) and its location (e.g., a dedicated building or a shared space), physical access to its antenna may be difficult to protect. The connection between the gNB and backbone is likely even more difficult to protect. Given the skepticism related to 5G radio transmissions in parts of the population, it is possible to imagine that a small community of hacktivists disregards planning or court decisions and actively removes or destroys the antennas of 5G base stations whenever easily accessible.

Threat agents: hacktivists.

K. TS 11: Exploiting Bad Resource Management in Slice Resource Allocation

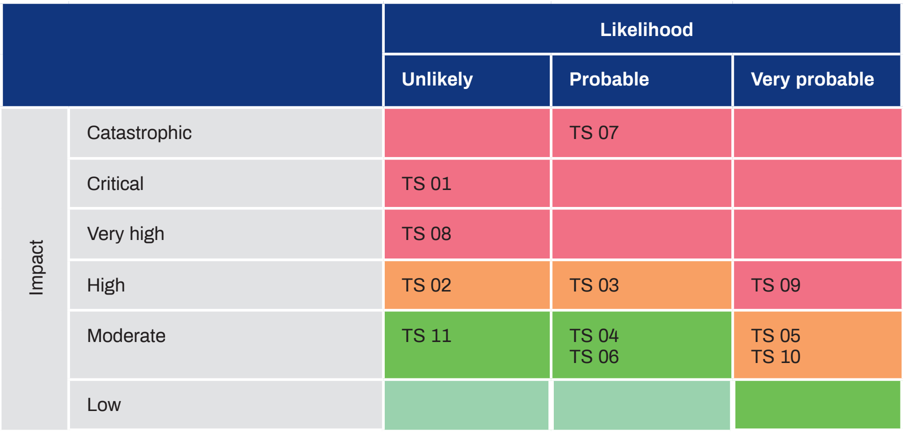
The sharing of the RAN between operators and, to a lesser extent, slice management by a single operator, opens up the issue of proper resource management under high loads. In the case of RAN sharing, the primary owner of the radio resource might privilege its radio resource requirements and no longer guarantee sufficient bandwidth to the sharing operator in the case of network overload. Apart from the generic network overload aspect, this attack will, however, heavily depend on implementation choices made by the network operator.

Threat agents: criminals, terrorists.

4. RISK ASSESSMENT

Figure 4 shows the summarized risk matrix for all identified threat scenarios, classified by impact and probability of occurrence. The dangerousness of the scenarios decreases from red to light green. The likelihood of an attack is related to various factors, such as a remote or local attack, logical or partial hardware attack, the time required to implement, the cost of equipment, and the expertise required for an attack.

FIGURE 4: RISKS MATRIX OF THREAT SCENARIOS



		Likelihood		
		Unlikely	Probable	Very probable
Impact	Catastrophic		TS 07	
	Critical	TS 01		
	Very high	TS 08		
	High	TS 02	TS 03	TS 09
	Moderate	TS 11	TS 04 TS 06	TS 05 TS 10
	Low			

5. MITIGATIONS AND SECURITY CONTROLS

Several threat scenarios are only possible due to the under-specification of the 5G standard. Indeed, if an operator implements all optional security and follows the recommendations inside the specifications, then some scenarios are impossible to exploit.

Other threat scenarios rely on an insufficient level of protection by the security features. Indeed, security is not achieved by merely activating a feature but by activating it in a robust manner that withstands attacks against its bypass or deactivation. Concerning TS 01 (lifting of the key database of the subscribers), if it is possible to update the keys in the UICCs used by the network operator and if the used HSM is sufficiently robust, then it might be possible to mitigate the attack before the attacker has been able to extract the keys of the lifted database. However, the robustness of the protection mechanism of the database in the UDM is highly dependent on its logical and hardware implementation. It might even be possible that the operator is dependent on the physical security of their cloud service provider.

Extracting the keys of a single subscriber through an attack on the associated UICC (TS 02) might be made more difficult by using hardware elements with additional countermeasures against both passive and active attacks. External certification of the UICC might provide an increased level of confidence in its robustness.

Attacks that are based on potential vulnerabilities or non-compliances inside the ME of the user equipment (TS 03 and TS 06) can only be mitigated by the network operator inside the core network. Indeed, only the network operator has control over the UICC inside the terminal. Knowing that the trust in the security of the ME is limited, the network operator should force a renewal of the security context on a regular basis (TS 04) in order to limit the duration of a security breach and be able to suppress some directions of arrival to filter out logical and physical jammer signals (TS 06).

In the current version of the specifications, a compliant device has no means of mitigating logical jamming attacks of some REGISTER REJECT causes sent by the rogue network (TS 04). Indeed, this message can be sent before the establishment of a security context, and the network currently has no means of authenticating itself before the security context has been configured between the network and the device. A potential mitigation of this situation could be as follows: All current global reject causes should be limited to a single network. The network would identify itself by broadcasting a network pre-security context authentication public key (e.g., in one of the system information blocks) and signing the reject message using the associated private key. Therefore, an attacker without knowledge of the real network private key cannot fully impersonate this network.

A rogue gNB could naturally broadcast its own public key and reject the registration of any UE. However, the UE would still be authorized to try to re-register to another network broadcasting a different public key. Note that currently the impact of a fake gNB is potentially much higher for an IoT device (and particularly a moving IoT device) than for a normal mobile phone. In principle, an IoT device is more vulnerable than a mobile phone, since there are fewer optional security features implemented. If the real network is made aware of the presence of a rogue gNB in one of its cells, it can also blacklist this rogue gNB in the system information broadcast by the surrounding legitimate gNBs.

The disclosure of the PEI described in TS 05 is only possible if the network operator chooses not to apply NAS and radio-level encryption for control plane messages. The exploiting of vulnerabilities that allow the extraction of key material or tampering with the firmware in the gNB or other network elements (TS 07 to TS 09) depends on the robustness of the authentication functions at boot time (but not only) and the presence of vulnerabilities inside the software. As these functions are essential for the correct operation of the network, the network operator should be aware of their importance and implement procedures that make it possible to increase trust in the correct and robust implementation of these functions. This applies to both the equipment manufacturer and other service providers (e.g., cloud operators). The operator should also evaluate the design features used to protect the authenticity of executed functions and the confidentiality of secrets.

Concerning threat scenario TS 10, increasing the acceptance of 5G systems by open discussions with the public should at least reduce the risk of the destruction of base stations by hacktivists. To avoid a network disruption by criminals or terrorists, the physical security of access to the base stations and redundancy in cell coverage are the only means to maintain network operations at all times and in all places.

For TS 11, appropriate resource management between slices taking into account their criticality and general QoS requirements should mitigate this threat.

6. DISCUSSION

Outside the context of UEs in a limited service state, exchanges with the gNB at radio resource control (RRC) level and with the 5GC at NAS level are expected to be integrity protected from a certain state onwards. However, it is unclear to which level UEs implement this part of the specifications and discard messages that are not protected using at least level NIA1. UEs that reply to unprotected Security Mode Commands will still expose their IMEI to a rogue network and thus indirectly disclose the identity of the subscriber. Verification of the adherence of a UE to the standard could be achieved by modifying a fully functional standalone Software-Defined Radio (SDR) implementation of a 5G network that allows deactivating the integrity protection for selected messages and using test SIM cards under the control of the researcher.

For data confidentiality, the activation of data encryption at the radio level and at NAS level is entirely under the network operator's control. To which extent operators activate RRC, NAS, and user plane encryption needs to be verified. Suppose in the control plane, an operator only relies on integrity protection. In that case, the IMEI/PEI and the associated 5G-GUTI of the device can still leak and allow tracking of the user even if the user plane data is encrypted. Using a fully instrumented test UE that provides access to this level of information would verify the protection level used by operators in the field.

On the network side, it is unclear to which extent operators implement IPSec between all network functions. If an operator relies on the physical security of the network links, then this might allow interception of confidential data (including key material) between the network entities. Without physically forcing access to the operator's network, IPSec can only be verified by auditing the network operators.

In the latest 5G releases, 3GPP has added new services such as edge computing or proximity services with their related network functions that increase the complexity of the operator's networks. These new services and procedures may bring some additional risks or vulnerabilities that will have to be carefully analyzed and assessed. Furthermore, roaming architectures and procedures have been devised for 5G, not all of which have been fully specified by the GSM Association [15], and the use of these intermediate actors significantly increases the attack surface.

7. CONCLUSION

Our comprehensive analysis shows that 5G networks are still exposed to many threats previously identified in 4G implementations. This remains even more true in NSA deployments where the network is 5G in name only (or, to be more precise, only 5G for some aspects of the radio channels). Due to performance constraints in some 5G devices, the network operator might be tempted not to use all possible security controls (e.g., user plane encryption and integrity protection) for the communications of these device classes. The virtualization concepts create additional challenges for the operators, as they potentially create new trust relationships between the operator and third parties, such as cloud service providers.

ACKNOWLEDGMENTS

A special thank you goes to Max Duparc for the proofreading of this article, as well as contributors and reviewers from Kudelski SA for their insightful observations: Alain Paschoud, Nicolas Mutschler, and Benoît Gerhard.

REFERENCES

[1] "TS 33.501. Security architecture and procedures for 5G systems, V17.7.0." 3GPP. Sep. 2022. [Online]. Available: https://portal.3gpp.org/desktopmodules/Specifications/SpecificationDetails.aspx?specificationId=3169

[2] R. Piqueras Jover and V. Marojevic, "Security and protocol exploit analysis of the 5G specifications," *IEEE Access*, vol. 7, pp. 24956–24963, 2019, doi: 10.1109/ACCESS.2019.2899254.

[3] D. Basin, J. Dreier, L. Hirschi, S. Radomirovic, R. Sasse, and V. Stettler, "A formal analysis of 5G authentication," in *Proceedings of the 2018 ACM SIGSAC Conference on Computer and Communications Security*, New York, NY, USA, Oct. 2018, pp. 1383–1396. doi: 10.1145/3243734.3243846.

[4] "TS 23.501. System architecture for the 5G system (5GS), V17.6.0." 3GPP. Sep. 2022. [Online]. Available: https://portal.3gpp.org/desktopmodules/Specifications/SpecificationDetails.aspx?specificationId=3144

[5] J. Śliwa and M. Suchański, "Security threats and countermeasures in military 5G systems," in *2022 24th International Microwave and Radar Conference (MIKON)*, Sep. 2022, pp. 1–6. doi: 10.23919/MIKON54314.2022.9924818.

[6] E. Yocam, A. Gawanmeh, A. Alomari, and W. Mansoor, "5G mobile networks: reviewing security control correctness for mischievous activity," *SN Applied Sciences*, vol. 4, no. 11, p. 304, Oct. 2022, doi: 10.1007/s42452-022-05193-8.

[7] J. P. Mohan, N. Sugunaraj, and P. Ranganathan, "Cyber security threats for 5G networks," in *2022 IEEE International Conference on Electro Information Technology (eIT)*, May 2022, pp. 446–454. doi: 10.1109/eIT53891.2022.9813965.

[8] T. Yang et al., "Formal Analysis of 5G Authentication and Key Management for Applications (AKMA)," *Journal of System Architecture*, vol. 126, p. 102478, May 2022, doi: 10.1016/j.sysarc.2022.102478.

[9] V. Oeselg *et al*., "Research Report: Military Movement Risks From 5G Networks," CCDCOE, Tallinn, Estonia, 2022.

[10] B. Potter, "Microsoft SDL threat modelling tool," *Network Security*, vol. 2009, no. 1, pp. 15–18, Jan. 2009, doi: 10.1016/S1353-4858(09)70008-X.

[11] L. Kohnfelder and P. Garg, "The threats to our products," *Microsoft Interface, Microsoft Corporation*, vol. 33, 1999. https://adam.shostack.org/microsoft/The-Threats-To-Our-Products.docx

[12] "TS 37.340 Multi connectivity, overall description, stage-2, V17.1.0." 3GPP. Jul. 2022. [Online]. Available: https://portal.3gpp.org/desktopmodules/Specifications/SpecificationDetails.aspx?specificationId=3198

[13] "TS 38.304. User equipment (UE) procedures in idle mode and in RRC inactive state, V17.2.0." 3GPP. Oct. 2022. [Online]. Available: https://portal.3gpp.org/desktopmodules/Specifications/SpecificationDetails.aspx?specificationId=3192

[14] "TS 24.501. Non-access-stratum (NAS) protocol for 5G system (5GS); stage 3, V17.8.0." 3GPP. Sep. 2022. [Online]. Available: https://portal.3gpp.org/desktopmodules/Specifications/SpecificationDetails.aspx?specificationId=3370

[15] GSMA, "NG.132. Report 5G Mobile Roaming Revisited (5GMRR) Phase 1, Version 2.0," Apr. 2022.

[16] Y. Zhang, A. Juels, M. K. Reiter, and T. Ristenpart, "Cross-VM side channels and their use to extract private keys," in *Proceedings of the 2012 ACM Cconference on Computer and Communications Security*, Oct. 2012, pp. 305–316.

2023 15th International Conference on Cyber Conflict
Meeting Reality
T. Jančárková, D. Giovannelli, K. Podiņš, I. Winther
2023 © NATO CCDCOE Publications, Tallinn

Toward Mission-Critical AI: Interpretable, Actionable, and Resilient AI

Igor Linkov
Senior Scientific Technical Manager
U.S. Army Corps of Engineers
Concord, MA, United States
Igor.Linkov@usace.army.mil

Andrew Strelzoff
Principal Data Scientist
U.S. Army Corps of Engineers
Vicksburg, MS, United States
Andrew.Strelzoff@erdc.dren.mil

Jeffrey Keisler
Professor, Management Science and Information Systems
University of Massachusetts Boston
Dorchester, MA, United States
Jeff.Keisler@umb.edu

Alexander Kott
Senior Research Scientist
Army Research Laboratory
Adelphi, MD, United States
alexander.kott1.civ@mail.mil

Petar Tsankov
Co-founder and CEO
LatticeFlow
Zurich, Switzerland
petar.tsankov@latticeflow.ai

Kelsey Stoddard
Network Scientist
U.S. Army Corps of Engineers
Concord, MA, United States
Kelsey.S.Stoddard@usace.army.mil

S.E. Galaitsi
Research Environmental Scientist
U.S. Army Corps of Engineers
Concord, MA, United States
Stephanie.E.Galaitsi@usace.army.mil

Benjamin D. Trump
Senior Research Social Scientist
U.S. Army Corps of Engineers
Concord, MA, United States
Benjamin.D.Trump@usace.army.mil

Pavol Bielik
Co-founder and CTO
LatticeFlow
Zurich, Switzerland
pavol.bielik@latticeflow.ai

Abstract: Artificial intelligence (AI) is widely used in science and practice. However, its use in mission-critical contexts is limited due to the lack of appropriate methods for establishing confidence and trust in AI's decisions. To bridge this gap, we argue that instead of aiming to achieve Explainable AI, we need to develop Interpretable, Actionable, and Resilient AI (AI3). Our position is that aiming to provide military commanders and decision-makers with an understanding of how AI models make decisions risks constraining AI capabilities to only those reconcilable with human cognition. Instead, complex systems should be designed with features that build trust by bringing decision-analytic perspectives and formal tools into the AI development and application process. AI3 incorporates explicit quantifications and visualizations of user confidence in AI decisions. In doing so, it makes examining and testing of AI predictions possible in order to establish a basis for trust in the systems' decision-making and ensure broad benefits from deploying and advancing its computational capabilities. This presentation provides a methodological frame and practical examples of integrating AI into mission-critical use cases and decision-analytical tools.

Keywords: *artificial intelligence, trust, mission-critical AI*

1. INTRODUCTION

"Can I trust the recommendation of an AI agent?" This question is difficult to answer, especially if the decision at stake is complex and may heighten existing or introduce new risks to humans. Yet such high-stakes decision-making has become routine within systems incorporating artificial intelligence (AI), such as controls for chemical plants, defense systems, and health insurance rate determinations. Stakeholders must be not only able to configure AI and its enabling technologies for a given industry or task but must also have the tools and methodologies to examine and address failures, limitations, and needs for quality control at various stages of the AI development and application process.

Trust in social situations grows based on performance over time [1], and trust in AI can be developed the same way. But both objectives and situations are liable to change in time and space, and a static decision made under specific circumstances may have limited utility in divergent futures. The contemporary world changes quickly and sometimes dramatically, and AI decisions must be contextualized within a changing and uncertain threat space.

The ultimate goal of AI is to provide users with actionable recommendations that meet both the implicit and explicit goals of decision-makers and stakeholders. Recommendations generated from AI-based approaches hold advantages over human decision-makers through their ability to analyze vast bodies of information quickly in an objective and logic-centered fashion, as long as they are trained to do so. In many situations, these benefits are clear and already implemented in practice, such as machine learning systems for detecting phishing attempts [2]. AI applications are also capable of providing multistep and adaptable strategies, as demonstrated by programs that play chess or Go, as well as AI-based cybersecurity systems [3], [4].

However, it is important to note that AI recommendations may not account for decision-makers' values or specific mission needs. For example, following a cyber attack, an AI-generated decision engine may recommend disabling an application on the compromised computer system. This action may neutralize the threat posed by the compromised system but could simultaneously endanger a mission, negatively impact a user's ability to perform critical tasks, or enable the adversary to extend the cyber attack's duration or scope.

Because the broader-scale impacts of the recommended path forward may not have been incorporated into the AI's design or scope, the AI decision processes may omit critical conditions that a human operator would implicitly account for. Such incomplete scoping of AI-driven analysis is especially problematic when unspoken, unacknowledged, or subjective variables influence or shape what a successful outcome looks like to a human manager. The AI solves the problem it is given, but it is the human's responsibility to ensure the recommendation's suitability in context. Similarly, the human users making this judgment will benefit from understanding the factors that led to the AI's decision, as this can help them see the value of factors they might have overlooked.

2. HUMAN-MACHINE TEAMING FOR DECISION-MAKING

Although AI-driven analysis can greatly enhance our decision-making ability, providing insight into AI's shifts in its analysis of needs, expectations, and mission requirements will ensure the relevance and credibility of its decisions and make its expectations for the future explicit. If AI's analytical outputs do not account for these and other broader and potentially subjective concerns, an overly myopic focus on a tactical decision can derail strategic mission requirements. As such, more effective deployment of AI in decision-making must resolve the black box concerns of AI –

in that it is unclear how to explain, interpret, and act upon AI's conclusions as its underlying algorithm and parameters are hard to decipher.

We can expect AI recommendations to differ from the choices an operator would make alone. For yes/no decisions, there are three possibilities: 1) the AI is more risk-averse than the human, 2) the AI is more risk-tolerant than the human, or 3) the AI and the human agree.

Assuming that the AI is correct more often than a human under the same time and resource constraints (underscoring the utility of AI applications), the human who disagrees with the AI should still follow the AI's recommendations. The challenge of that moment of discordance, then, is to convince humans to trust the AI's output despite their own opposing judgment.

There are already situations in which trust between AI and human users is fragile: the term "techlash" refers to the growing animus toward technology, especially information technology. Techlash is a distrust that technologies have the users' best interests at heart, given some questionable behavior from the organizations that build and/or promote them [5]. If the benefits of superior AI decision-making are to be realized and further developed, it is essential to establish a foundation for users to build trust in the AI's decisions [6].

To create such a foundation, it is crucial to consider four dimensions: Explainable AI, Interpretable AI, Actionable AI, and Resilient AI.

3. EXPLAINABLE AI

Explainability refers to the extent that a system's internal mechanics can be explained in terms that are salient to human cognition. The inability of AI algorithms to articulate the reasons for specific predictions and recommendations arises from the complexity of the underlying deep learning algorithms and the training data provided. To address this challenge, there are various initiatives that aim to produce AI models that are more easily understood without overly sacrificing the accuracy of the AI's predictions. For example, DARPA's explainable AI (XAI) program aims to develop new algorithms with "the ability to explain their rationale, characterize their strengths and weaknesses, and convey an understanding of how they will behave in the future." This will enable users to better understand and improve trust in AI's decisions, as well as appreciate their value added for specific applications.

There are several limitations of Explainable AI in its current framing. First, while there may be situations in which AI can be explained, some processes, in spite of their practical benefits, are too complex for human cognition. However, it is still possible to render some relationships more transparent: in image processing, saliency methods use digital neural networks to provide maps according to pixel relevance within the image. The true fidelity of various methods can be difficult to measure [7], but their application promotes the idea that some relationships between inputs and outputs can be held consistent for both human and AI cognition. This, however, may leave much of the algorithmic processes unexplained.

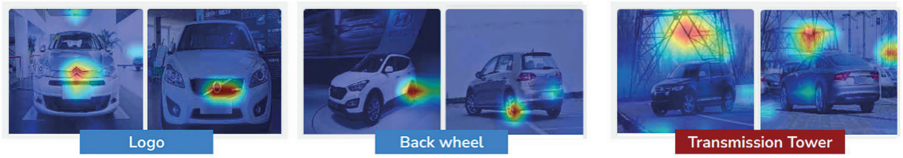
Second, a key advantage of AI may well be its ability to avoid human-like behavior when that behavior is not actually optimal: AI can provide innovative strategies for achieving objectives deduced from the framed problem. In some cases, an AI game-playing agent triumphs over human rivals not because it improves upon known human strategies but precisely because it deviates from those strategies. However, mandating that AI explain something that is counterintuitive to human operators may not help in trust-building. AI arrives at decisions through convoluted and complex algorithms (the black box) that are generally shrouded from or impenetrable to human operators. Inviting humans into the box may jeopardize AI's true power by forcing it to conform to human recognition.

Yet human understanding (and, typically, acceptance) is predicated on AI conformity to recognizable cognition processes. Just because humans do not see the reason for a process does not make it inconsequential. AI may arrive at decisions by avenues that are unfamiliar or obtuse compared to those upon which humans have historically relied. Truly benefiting from AI may entail excusing it from the onus to explain itself to humans because such a demand constrains AI to the same values and limitations that have always underpinned human decisions.

Case Study

To illustrate the usefulness and limitation of AI explainability, consider the task of object type identification. We use the Comprehensive Cars dataset [8], which consists of 136,726 images of cars annotated with 163 car makers. Given an AI model trained on this dataset, we use saliency maps [9] to produce visual explanations for the AI model's predictions, as shown in Figure 1. Here, in addition to the images, we overlay saliency maps where red color highlights regions identified as important for the AI model prediction. The explanation can be useful to confirm that the model uses features relevant to the task, such as the logo, and to discover spurious features, in this case, the presence of the transmission towers behind the car.

FIGURE 1: ILLUSTRATION OF VISUAL EXPLANATIONS OF THE AI MODEL PREDICTION IN THE APPLICATION OF CAR MODEL PREDICTION



By visually inspecting the saliency maps, the human can review the visual cues that support the AI's car model prediction, such as brand logo (Figure 1, left), back wheels (Figure 1, middle), spare wheels (Figure 2, left), and the transmission tower behind the car (Figure 1, right). The use of brand logos is clearly intuitive and salient for human decision-makers. While features such as back wheels are less intuitive, they may indeed be used as fine-grained visual cues to differentiate between similar car makers. However, the AI also uses features in the images that are clearly spurious to the exercise, such as the presence of a transmission tower behind the car.

In Figure 2, the AI model uses a spare wheel as a strong signal for predicting Jeep class (left images). However, the same model focuses on unrelated parts of the image (right images), which may seem unintuitive. One reason for this behavior, however, could be that the AI model uses the "lack" of a spare wheel, which is difficult to visualize using saliency maps alone.

FIGURE 2: AN EXAMPLE HIGHLIGHTING THE LIMITATIONS OF SALIENCY MAPS FOR EXPLAINING AI DECISIONS

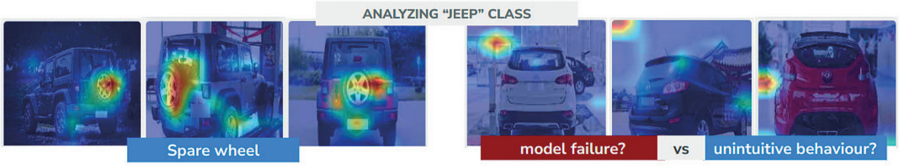


Figure 2 (right) provides seemingly non-intuitive explanations. Here, the attribution suggests that when detecting the Jeep class, the AI model focuses on an unrelated image region without even looking at the car. Despite this, however, the model still produces the correct prediction. When trying to align the AI model with human-like behavior, one explanation is that for a model to predict a particular class, one of two cases are possible: 1) the AI model detects the presence of a feature that is highly predictive of a class, such as the spare wheel, or 2) the AI model interprets the absence of a predictive feature as a negative signal. That is, even though the AI model's decision may seem unintuitive based on the saliency maps, the AI model decision may still be grounded.

Determining whether the decision is valid or the AI model learned a spurious feature is non-trivial, and using AI explainability alone is not sufficient to answer this question. Next, we turn our attention to AI interpretability, which will help address this question from a different perspective.

4. INTERPRETABLE AI

The transition from explainability to interpretability means moving from providing a reason for a decision to assessing meaning in the context of a specific decision or mission. Like Explainable AI, Interpretable AI recognizes the tradeoff between transparency and accuracy enabled by computational power. Rather than seeking to optimize both, Interpretable AI emphasizes understanding cause and effect within the AI system [10]. Users can examine the sensitivities of the output recommendations to changes in the parameter inputs without needing to understand the complex internal computations of the algorithms. Interpretable AI should allow users to toggle the parameters that are most uncertain in order to study the impacts of their changes, as well as to test the AI's reaction to changes against the users' own beliefs about underlying relationships between inputs and outputs.

For example, in determining which car a person should purchase, income should be an important factor. Within an Interpretable AI system, income relevance and effect could be verified by varying the income input within the model and viewing the subsequent changes in model output. If an AI system exhibited extreme sensitivity to income and little sensitivity to the difference between a three- and four-person family, the user could conclude that the AI system reflects at least some of the factors that the user deems most important in car selection.

Instead of explaining the AI results to humans, Interpretable AI models allow users to place AI recommendations in the context of the decision problem. Interpretation does not imply that operators must understand the process driving the AI recommendations. To this end, forging meaning, more than explanation, allows the AI to build functionality around accuracy and complexity while ensuring that humans can find sufficient meaning in the outcomes to implement them.

Case Study

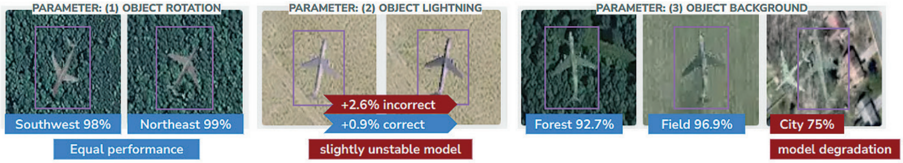
We illustrate the interpretability of an AI model that detects airplanes from satellite images. Concretely, we use a publicly available dataset [11] containing 3,999 objects, all labeled as "plane" class. We train a state-of-the-art object detection model, Yolov5 [12], which is able to identify 97.7% of the airplanes in the test dataset (recall), with predictions that were 66.2% correct (precision).[1]

[1] By making the model more conservative, we can also obtain 92.2% recall and 87.8% precision.

To better interpret how the AI model detects airplanes, we generate a new set of images that differ in some parameters while keeping the remaining parameters the same. As we will see, the analysis will allow us to not only better understand how the AI model makes decisions but also uncover hidden blind spots (i.e., cases where the model systematically underperforms). We note that while varying some parameters can be trivial when working with structured data, or natural language text, this requires sophisticated computer vision approaches when applied to images to preserve realism.

In Figure 3, we illustrate the effect of changing three parameters: 1) the plane orientation, 2) the plane lighting, and 3) the ground under the plane. Changing the plane's orientation – for example, from west to east – reveals that the AI model has learned to work well across various orientations. We notice that the AI model's predictions are, however, less stable when we vary the plane's lightning: after increasing the lightning, the AI model misses 0.9% of planes that were previously detected correctly and fails to detect 2.6% of planes that were previously identified correctly. Finally, we illustrate a serious model degradation when changing the ground parameter: The AI model's ability to detect planes above urban environments drops to 75%, which is significantly worse compared to other ground types such as forests, fields, plains, and water. Based on these insights, the user can make an informed decision about whether such instabilities can be tolerated or if they need to be explicitly addressed.

FIGURE 3: ILLUSTRATION[2] OF INTERPRETING AI MODELS FOR AERIAL OBJECT DETECTION BY VARYING SELECTED IMAGE ATTRIBUTES: (1) ROTATION, (2) LIGHTNING, AND (3) BACKGROUND. FOR EACH ATTRIBUTE, WE EVALUATE THE AI MODEL'S RECALL, WHICH REVEALS THAT: (1) THE MODEL IS NOT AFFECTED BY OBJECT ORIENTATION, (2) IT BECOMES SLIGHTLY UNSTABLE WHEN THE LIGHTNING CHANGES, AND (3) DETECTING PLANES ABOVE URBAN ENVIRONMENTS IS DEGRADED.



This case study highlights the usefulness of AI interpretability in understanding and improving the performance of AI models. However, AI interpretability alone is not sufficient as it only focuses on the AI model's decisions while ignoring the important mission-critical context. To address this limitation, in the next section, we turn to Actionable AI.

[2] Note that here we show only a small crop of the full-resolution image containing the objects to be detected. Further, the bounding box around the plane is not part of the image and is included only as a visual cue.

5. ACTIONABLE AI

Ultimately, the foundations of decision-maker actions are grounded upon evidence-based data (including AI recommendations) as well as strategic and tactical considerations, which include factors such as decision context, mission needs, and available resources, and so on. AI systems should provide a level of confidence and sufficient information so that the decision-maker can trust a suggested course of action [13]. Limitations in AI recommendations may be unclear to the user until the results are applied and evaluated, since limitations can potentially cause costly mistakes. Most AI systems are designed for specific contexts, but users, for lack of other options, may apply them to circumstances outside the design capabilities. This is especially important in situations where systems perform under widescale threats; therefore, it is necessary to change both the operational and decision environments outside of the AI system performance range.

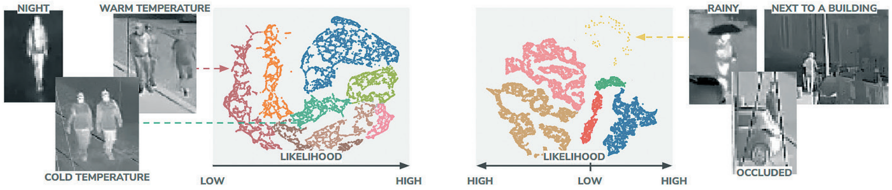
Decision theory in general, and multi-criteria decision analysis (MCDA) [14] in particular, provides a template for visualizing tradeoffs between choices and quantifying the relative level of confidence that an AI system is placing on its recommendations. MCDA approaches typically require input scores across several dimensions associated with different management alternatives and outcomes that reflect objective evidence-based data and subjective weights related to tradeoffs across these dimensions relevant to the mission and values. A basic but typical approach is to calculate the total value score for an alternative as a weighted sum of its scores across several criteria. These scores can be translated into utility functions or other metrics relevant to confidence in courses of action recommended by AI systems. Linking MCDA with Scenario Analysis [15] allows the integration of the movable threat space to ensure AI decisions are applied in ways that will be most beneficial given the uncertainty of the future. The implication for AI is that there may be value in a layer that maps the content of generic explanations into the specific terms a rational human decision-maker would use to infer that a course of action is appropriate – for example, in terms of the criteria such a decision-maker would use in the absence of AI.

Case Study

Let us consider the application of detecting people using thermal cameras. We use a dataset collected over eight months and containing more than one million images across a wide range of environmental conditions, including fog, occlusions, night, dew point, and wind speed [16]. Rather than evaluating an AI model using aggregate statistics, we aim to explicitly model the tradeoff across dimensions relevant to the mission. To this end, we start by defining a set of scenarios curated either manually or, as in our case, semi-automatically by first clustering the data along relevant dimensions.

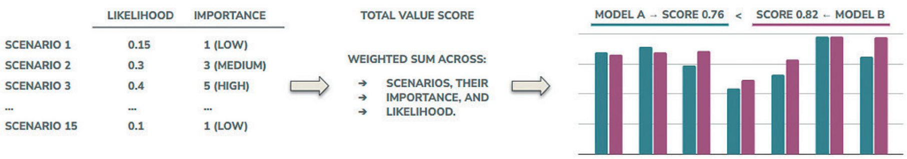
We illustrate the scenario definition results in Figure 4, containing 13 different scenarios visualized with different colors and using two different dataset representations. Here, each point corresponds to an image from the dataset, and the two dataset representations were created by projecting selected metadata attributes, as well as image context, into a two-dimensional space. The representation in Figure 4 (left) includes attributes of temperature and humidity, while the representation in Figure 4 (right) includes precipitation and sunlight intensity. Depending on the weather, the infrared camera produces high-contrast images during the night or low-contrast images when the temperature is high (left images). Other scenarios include persons that are occluded, next to a building, or holding additional objects, such as umbrellas when it rains (right images).

FIGURE 4: ILLUSTRATION OF A WIDE RANGE OF ENVIRONMENTAL CONDITIONS AND CONTEXTS IN WHICH AN AI MODEL NEEDS TO OPERATE AND THEIR FORMALIZATION AS A SET OF SCENARIOS, VISUALIZED AS POINT CLUSTERS OF THE SAME COLOR



Next, for each scenario, we define its likelihood and importance with respect to the mission, as shown in Figure 5. Here, importance is defined by a human decision-maker, while the likelihood can be estimated directly from the data. These are then used to compute the total value score for each AI model. This is useful for measuring the model uncertainty for each scenario, and to allow fine selection criteria for determining which model to use.

FIGURE 5: AN EXAMPLE OF EXPLORING TRADEOFFS ACROSS DIMENSIONS RELEVANT TO THE MISSION BY ASSIGNING LIKELIHOOD AND IMPORTANCE TO EACH SCENARIO FROM FIGURE 4



	LIKELIHOOD	IMPORTANCE
SCENARIO 1	0.15	1 (LOW)
SCENARIO 2	0.3	3 (MEDIUM)
SCENARIO 3	0.4	5 (HIGH)
...	...	...
SCENARIO 15	0.1	1 (LOW)

TOTAL VALUE SCORE

WEIGHTED SUM ACROSS:

- SCENARIOS, THEIR IMPORTANCE, AND LIKELIHOOD.

In this case study, we highlighted how Actionable AI can help to account for uncertainty in complex environments by breaking down the operation domain into individual scenarios. These scenarios can then be consulted by the human decision-maker to make an informed decision on whether the system was designed for a given context, together with the expected performance. However, such evaluation still assumes no external disruptions are affecting the system, a challenge addressed in the next section.

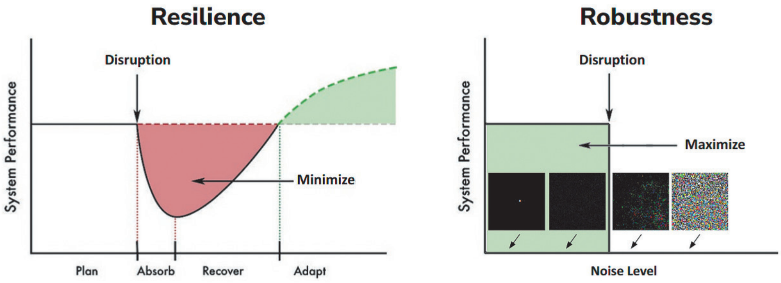
6. RESILIENT AI

AI models are typically deployed as part of a larger system that performs a more complex task than that solved by the AI model alone. In this system, the AI model may interact with other AI models as well as non-AI-based components. As a result, there is an inherent gap between evaluating AI models in isolation and the AI-based system as a whole. For example, an AI model trained to detect objects in an image is evaluated in terms of its accuracy – how many objects are predicted correctly (true positives), how many objects were missed (false negatives), and how many spurious objects were detected (false positives). Yet, when this model is deployed in a system for automated threat detection, the system's overall performance is evaluated in terms of its ability to continuously identify targets, assess their threat level, and track them over time. Similarly, when such a model is used to identify traffic signs for autonomous driving, the object detection model is only one of the components providing inputs to a control system whose ultimate goal is to drive safely and avoid collisions. Evaluating such AI systems requires explicitly modeling the system as a whole, including its individual components, and defining its critical function.

Beyond modeling the overall system and its critical function, an important property of deployed systems is their resilience (i.e., the system's ability to recover from disruption) [17]. This is crucial, as the deployed system is inevitably subject to failures, either due to an active adversary or the inherent difficulty of the task at hand. After a failure, resilience quantifies both the negative impact of the disruption on the system's critical function and the time it takes for the system to recover (Figure 6, left). We note that while there has been enormous progress in training AI models robust to different types of failures, such as adversarial or natural noise [18], [19], this body of work treats AI models in isolation, overlooking the overall system and the recovery of the system's critical function whenever a failure happens (Figure 6, right). In resilience, the goal is to minimize system degradation after a disruption occurs. In robustness, the goal is to maximize the area and the amount of noise that can be applied to the AI model, before a disruption occurs.

To bridge this gap, the design and evaluation of AI-based systems can be improved in two ways. First, the system's critical function should be explicitly modeled and used as a criterion for AI model selection. This is because a more accurate and robust AI model does not imply better performance for the overall AI-based system. For example, consider an AI model that detects obstacles to ensure an autonomous car does not crash. An AI model that is more accurate on average, but less accurate on obstacles right in front of the car, can lead to worse performance (i.e., more car crashes). Note, here we are not interested in understanding that this is the AI model's behavior, which is where Explainable and Interpretable AI is used, but rather training AI models to have this behavior. Second, the AI model robustness and system-level resilience should be optimized jointly, to account for objects that have different importance for the system's critical function.

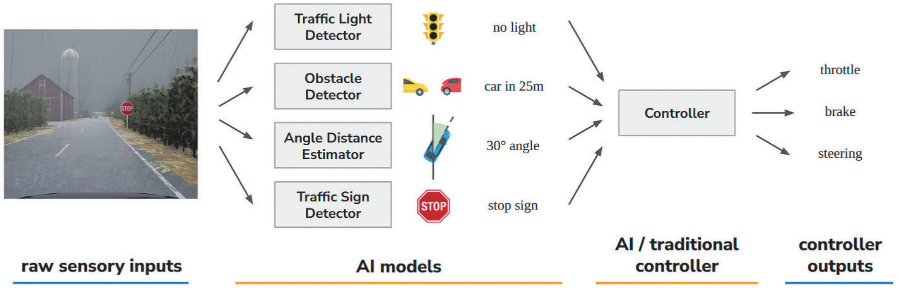
FIGURE 6: THE KEY DIFFERENCE BETWEEN (LEFT) RESILIENCE (IMAGE CREDITS [17]) AND (RIGHT) ROBUSTNESS



Case Study

Consider an autonomous driving application where the system takes multiple sensory inputs with the goal of predicting three continuous control actions at each step – throttle, steering angle, and brake, as shown in Figure 7. The system is split into two parts: (1) a set of base AI models trained to take raw sensory inputs and predict high-level affordances, including traffic light status, obstacles, and their distance, angle of the car with respect to the road center line, and a traffic sign detector, and (2) a controller, which takes the predicted affordances as inputs and predicts control actions for throttle, steering, and braking. Even though the base AI models have been trained to solve their individual tasks, such as detecting traffic lights, to assess the resilience of the overall system, we need to explicitly consider all AI models together and the system's critical function.

FIGURE 7: AN EXAMPLE OF A COMPLEX SYSTEM THAT INPUTS RAW SENSORY INFORMATION AND USES A SET OF AI MODELS TO OBTAIN HIGH-LEVEL AFFORDANCES THAT DESCRIBE THE SCENE, WHICH ARE THEN PROCESSED BY A CONTROLLER TO COMPUTE THROTTLE, BRAKE AND STEERING ACTIONS



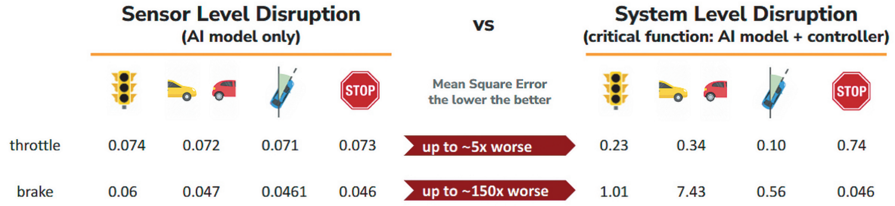
In this case, the critical functions can be defined either (1) externally, such as in passenger safety or the safety of other vulnerable road users, or (2) internally in terms of the controller's outputs (i.e., whether the system is braking when it should be). To illustrate the gap between AI robustness and resilience, consider the evaluation of such an autonomous driving system in Figure 8, trained using data from the CARLA simulator [20]. Here, we compare two different types of disruptions. The first disruption affects each sensor in isolation, which in our case corresponds to adding noise to the raw sensory data with the goal of disrupting AI operation.[3] Even though we are disrupting a single sensor at a time, we are interested in evaluating the overall system's performance as measured by the (internal) critical functions. In this case, compared to the operation without disruptions,[4] the throttle and brake critical function performance decreases[5] by +4% and +27%, respectively. While these results look very promising, they are biased by the gap of evaluating AI models in isolation, compared to evaluating the system as a whole. Second, we perform the same disruptions, but this time to the whole system, including the controller. This allows us to obtain a reliable assessment of the system's limitations, which reveals worse performance of up to 5 and 150 times for throttle and brake, respectively.

[3] We follow the approach of "Croce and Hein [22]" instantiated with ε = 1/255.

[4] The mean square error for throttle and brake in normal system operation is 0.071 and 0.046, respectively.

[5] The performance decrease is measured as the magnitude of the error margin (i.e., mean square error) with respect to the optimal controller action. Ideally, resilience would be measured with respect to the external critical functions, such as passenger safety, and would translate into the number of accidents and their severity.

FIGURE 8: CRITICAL FUNCTION PERFORMANCE UNDER SENSOR AND SYSTEM-LEVEL DISRUPTIONS



	Sensor Level Disruption (AI model only)				vs	System Level Disruption (critical function: AI model + controller)			
	🚦	🚗 🚗	🚗	STOP	Mean Square Error the lower the better	🚦	🚗 🚗	🚗	STOP
throttle	0.074	0.072	0.071	0.073	up to ~5x worse	0.23	0.34	0.10	0.74
brake	0.06	0.047	0.0461	0.046	up to ~150x worse	1.01	7.43	0.56	0.046

7. CONCLUSION

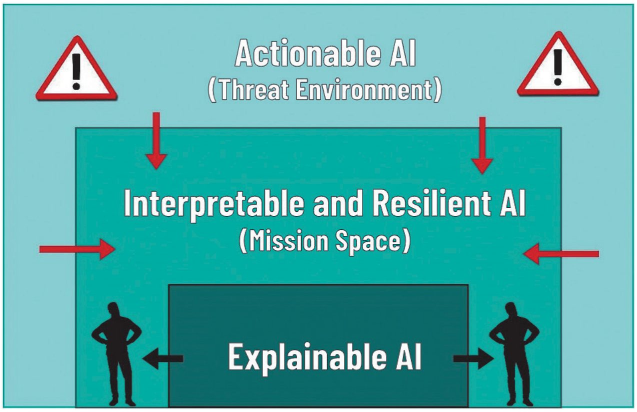
The validity or trustworthiness of decisions is predicated upon AI's analysis, and the uncertainty surrounding AI's decision-making algorithms makes it difficult to understand which parameters were used to arrive at a conclusion or how those parameters were weighted for importance relative to one another. While such parameters will typically not be analyzed by the end users, addressing these concerns within AI's earliest stages of research and development, together with the knowledge that they exist, is important for gaining trust and incorporating AI to complement their operations and decision-making needs. Stakeholders who use AI may encounter AI-driven guidance that is antithetical to their core values or mission requirements. This can cause users to reject AI's analysis in favor of human decision-making abilities alone, or possibly to adopt the AI-driven conclusions to their own detriment for the longer-term future, as measured by impacts according to their values. Neither outcome is desirable.

However, this human-AI tension can be avoided by creating a more effective, ethical, and transparent process that combines the decision-making needs and abilities of both actors. To benefit society and ensure applicability, AI systems need a mechanism to build operator confidence in AI recommendations for increasingly complex decision-making processes. To this end, we propose that many of the goals of Explainable AI can be realized with Actionable, Interpretable, and Resilient AI (AI3) without penalties in terms of AI computational power and accuracy.

Figure 9 illustrates that building trust in AI systems requires transferring meaning and relationships from one coherent system of understanding to another, from AI to human cognition. Explainable AI may be possible in some circumstances but is inevitably couched within the context (threat environment) and the objectives and predetermined notions of the user (mission space). By rendering these more visible within the AI interface structure, the user can better access aspects of understanding even if the black box itself, meaning the actual decision-making computations, cannot be fully explained within human cognition constraints.

AI systems should be able to capture the values of the decision-maker in selecting courses of action but should also provide a level of confidence and sufficient information such that the decision-maker can critically evaluate its recommendation. AI3 requires that operators understand enough about the decisions and their assumptions to anticipate how well-suited AI recommendations will be to the given problem. Therefore, in addition to communicating the reasoning processes, AI must communicate important contextualizing factors to its users. Decision output should include projections of performance to various changes or challenges that may arise, according to the user's objectives. AI output could also anticipate how those objectives might change, at least in framing, within different futures.

FIGURE 9: EXPLAINABLE, INTERPRETABLE, ACTIONABLE, RESILIENT AI AND THEIR INTERACTIONS (ADAPTED FROM [21])



Ultimately, a near-term requirement to enhance AI includes deepening the contextualized interactions between AI and its users to build human trust in AI outputs. Interpretable AI allows users to toggle parameter inputs to study the effects on the decisions, Actionable AI provides insights into the value of AI decisions in different and uncertain futures, and Resilient AI explicitly accounts for the system's critical function and its recovery upon inevitable failures. Together they enhance Explainable AI as Actionable, Interpretable, and Resilient AI.

ACKNOWLEDGMENTS

Research was sponsored in parts by the US Army Corps of Engineers (FLEX) and Army Research Office and was accomplished under Grant Number W911NF-20-1-0317. The views and conclusions contained in this document are those of the authors and should not be interpreted as representing the official policies, either expressed or implied, of the Army Research Office or the U.S. Government.

REFERENCES

[1] R. J. Lewicki and C. Wiethoff, "Trust, trust development, and trust repair," *The Handbook of Conflict Resolution: Theory and Practice*, vol. 1(1), pp. 86–107, 2015.

[2] M. Khonji, Y. Iraqi, and A. Jones, "Phishing detection: a literature survey," *IEEE Communications Surveys & Tutorials*, vol. 15(4), pp. 2091–2121, 2013.

[3] E. Al-Shaer, J. Wei, K. W. Hamlen, and C. Wang, "Towards intelligent cyber deception systems," in *Autonomous Cyber Deception: Reasoning, Adaptive Planning, and Evaluation of Honeythings*, New York, NY: Springer, 2019.

[4] A. Kott *et al*., "Autonomous Intelligent Cyber-defense Agent (AICA) Reference Architecture, Release 2.0," US Army Research Laboratory, Adelphi, MD, 2019.

[5] M. Finn and Q. DuPont, "From closed world discourse to digital utopianism: the changing face of responsible computing at Computer Professionals for Social Responsibility (1981–1992)," *Internet Histories*, vol. 4(1), pp. 6–31, 2020.

[6] K. Siau and W. Wang, "Building trust in artificial intelligence, machine learning, and robotics," *Cutter Business Technology Journal*, vol. 31(2), pp. 47–53, 2018.

[7] R. Tomsett, D. Harborne, S. Chakraborty, P- Gurram, and A. Preece, "Sanity checks for saliency metrics," 2019, *arXiv: 1912.01451*.

[8] L. Yang et al., "A large-scale car dataset for fine-grained categorization and verification", in *Proceedings of the IEEE conference on computer vision and pattern recognition*, 2015, pp. 3973–3981.

[9] R. R. Selvaraju, M. Cogswell, A. Das, R. Vedantam, D. Parikh, and D. Batra, "Grad-CAM: Visual Explanations from Deep Networks via Gradient-Based Localization," *IEEE International Conference on Computer Vision (ICCV)*, 2017, pp. 618–626.

[10] C. Benzaid and T. Taleb, "AI-driven zero touch network and service management in 5G and beyond: Challenges and research directions", *Ieee Network*, vol. 34, no. 2, pp. 186–194, 2020.

[11] Kaggle, 2019. "CGI Planes in Satellite Imagery." Kaggle. [Online]. Available: https://www.kaggle.com/datasets/aceofspades914/cgi-planes-in-satellite-imagery-w-bboxes

[12] G. Jocher, "YOLOv5 SOTA Realtime Instance Segmentation (v7.0).", 2022[Online]. Available https://github.com/ultralytics/yolov5/discussions/10258.

[13] A. Preece, D. Braines, F. Cerutti, and T. Pham, "Explainable AI for intelligence augmentation in multi-domain operations," 2019, *arXiv: 1910.07563*.

[14] I. Linkov, E. Moberg, B. Trump, B. Yatsalo, and J. Keisler, *Multi-Criteria Decision Analysis: Case Studies in Engineering and the Environment*. CRC Press, 2020.

[15] Y. Tourki, J. Keisler, and I. Linkov, "Scenario analysis: a review of methods and applications for engineering and environmental systems," *Environment Systems & Decisions*, vol. 33(3), pp. 3–20, 2013, doi: 10.1007/s10669-013-9437-6.

[16] I. Nikolov et al., "Seasons in Drift: A Long Term Thermal Imaging Dataset for Studying Concept Drift," in *35th Conference on Neural Information Processing Systems (NeurIPS 2021) Track on Datasets and Benchmarks*, 2021.

[17] I. Linkov and B. D. Trump, *The Science and Practice of Resilience*. Cham: Springer, 2019.

[18] T. Bai, J. Luo, J. Zhao, B. Wen, and Q. Wang, "Recent Advances in Adversarial Training for Adversarial Robustness," in *Proceedings of the 13th International Joint Conference on Artificial Intelligence*, 2021, pp. 4312–4321, doi: 10.24963/ijcai.2021/591.

[19] N. Akhtar and A. Mian, "Threat of Adversarial Attacks on Deep Learning in Computer Vision: A Survey," *IEEE Access*, vol. 6, pp. 14410–14430, 2018.

[20] A. Dosovitskiy, G. Ros, F. Codevilla, A. Lopez, and V. Koltun, “CARLA: An Open Urban Driving Simulator,” in *Proceedings of the 1st Annual Conference on Robot Learning*, 2017, pp. 1–16.

[21] I. Linkov, S. Galaitsi, B. Trump, J. Keisler, and A. Kott, “Cybertrust: From Explainable to Actionable and Interpretable Artificial Intelligence,” *Computer*, vol. 53, pp. 91–96, 2020.

[22] F. Croce and M. Hein, “Reliable Evaluation of Adversarial Robustness with an Ensemble of Diverse Parameter-free Attacks,” *ICML*, pp. 2206–2216, 2020.

2023 15th International Conference on Cyber Conflict
Meeting Reality
T. Jančárková, D. Giovannelli, K. Podiņš, I. Winther
2023 © NATO CCDCOE Publications, Tallinn

Zero-Day Operational Cyber Readiness

Barış Egemen Özkan
İstanbul, Türkiye

İhsan B. Tolga
Binalyze
Tallinn, Estonia

Abstract: As we move all our business practices into cyber terrain, the unique characteristics of cyberspace assets and threats require a different perspective to define and implement the concept of cyberspace readiness. The connected and dependent nature of functional and core services in and through cyberspace has created a nondeterministic security environment with unpredictable, ubiquitous and ambiguous threat perceptions. Building, increasing and sustaining cyber readiness requires producing, training, equipping, deploying and sustaining cyber warriors with competent capabilities against a continuously mutating threat landscape in a timely manner. Traditional military readiness approaches geared for kinetic services do not suit the unique requirements of cyber warfare readiness. A unit at "60 days notice to move" has 60 days to get ready to act. If the average time to detect a cyber attack is 200 days, cyber defenders must be ready for cyber attacks on average 200 days before they start. Hence, we propose the term "zero-day readiness" to describe agile and vigilant cyber readiness. In this paper, we offer a novel cyberspace readiness model based on principles, resources, activities, capabilities and benefits. While resource-demanding to build, improve and sustain, the proposed Zero-Day Readiness model has the potential to significantly increase the assessment and visibility of gaps as well as support judgment on the allocation of limited resources. The added value of this research is in developing a more revisionist readiness perspective for cyberspace operational readiness than the traditional kinetic operational domains, particularly for organizational and military cyber defense perspectives.

Keywords: *cyber security architecture, cyber capability building, zero-day readiness, cyberspace operations*

1. INTRODUCTION

Cyberspace is expanding at an increasing pace, covering more elements of business practice and daily activities, both in the professional and personal lives of people (Tabansky 2011). In addition, the Internet-of-Things has already created another vast domain with countless moving parts working in reasonable harmony.

In the modern world, communication, energy, life-support, healthcare, finance, transportation, the military, population registries, education and agriculture are just some of the sectors that almost completely operate on digital terrain. Due to its asymmetrical, quasi-anonymous and dual-use features, cyberspace challenges our traditional understanding of key concepts such as security, borders, human rights, privacy and sovereignty (Slack 2016). The sheer number of nodes that are interconnected and dependent on each other, as well as the limited amount of control at each party's disposal on the cyber landscape, make it impossible to have complete coverage of all its operations. Therefore, cyberspace and its affiliates currently possess a nondeterministic nature, for which only reactive measures are employed for any task.

Now, it is quite fair to estimate that in the near future, cyberspace and the countless multilateral connections within it will grow at an even faster rate. With the introduction of artificial intelligence (AI) and autonomous systems, the offset will become even greater between the nominal complexity of cases in cyber security and the laws/regulations aiming to govern them.

The interconnected and interdependent nature of systems, as well as the rapid migration of information networks to the cloud, constitutes the main driving factor for the increasing trend in cyber attacks (Forums 2023; CCDCOE 2022). For legitimate reasons (remote working, interconnection requirements, procurement and operational costs, automation, etc.), isolated corporate and organizational networks behind air gaps are now a thing of the past, and the separation between different domains happens on an abstract logical level. Along with its numerous advantages, the new cloud computing paradigm and virtual networks bring a new array of threats (Kushwaha, Roguski and Watson 2020), usually rooted in configuration errors, embedded design flaws, the utilization of different mediums for data-in-transit, vulnerabilities in supply-chain-dependent services (CCDCOE 2020; ENISA 2023) and the human factor (Das et al. 2018). In other words, as we are making things more convenient and smarter through digitization, we are also creating more vulnerable terrain for malign actors to exploit.

Organizations are continuously generating a huge amount of data and associated logs, and it is very likely that a threat actor with privileged access inside their information networks can go undetected. And the trust relationships across different organizations and parties, which can easily be exploited, increase the area accessible with malign intentions exponentially. In addition, there are various avenues threat actors can exploit to gain unauthorized access, and it is virtually impossible to track the source location, assuming the source is inside the same or cooperating jurisdiction, which is usually not the case (Kushwaha, Roguski and Watson 2020).

The current complexity of cyberspace and the lack of proportionality between this complexity and control and security mechanisms renders cyberspace a nondeterministic security environment with unpredictable, ubiquitous and ambiguous threat perceptions. Analogous to the distinction between climate and daily weather forecasts, while reports depicting aggregations of millions of cyber incidents provide overall trends regarding cyber attacks and their characteristics, it is virtually impossible to predict where and when the next cyber attack will come. Hence, we have to be ready for all kinds of attacks at all times.

The term "cyber defense structure" is used here to encompass both cyber security frameworks and cyber security architectures, but more aligned with the cyber defense concept. While cyber defense is the adopted strategy to protect the designated domain and assets in cyberspace, cyber security is the core component, encompassing the actual activities in this regard.

The current cyber defense frameworks of organizations, including but not limited to military organizations, usually orient their activities around forthcoming frameworks or architectures, such as the NIST Cyber Security Framework (NIST 2018), ISO/IEC 27000 Family (ISO 2023), Cloud-Native Application Protection Platform (CNAPP) (CheckPoint 2023), Cybersecurity Mesh Architecture (Fortinet 2023), AICPA Service Organization Controls 2 (AICPA 2023), MITRE ATT&CK (MITRE 2023), ISACA Control Objectives for Information Technology (ISACA 2023) and Center for Internet Security (CIS) Controls (Security 2023).

Similar to conventional military and defense doctrines, in its greatest common denominator set, a robust cyber defense structure is based on the assets it aims to protect as its bedrock in a prioritized fashion. At this point, it is also important to note that there is a unique relationship between kinetic and cyber warfare/attacks, the complications of which usually blend into each other's habitual domains (Libicki 2020). The structure is then reinforced by the vision and mission statements, and mapped with the capabilities in current and projected timeframes, to form a detailed set of steps, scenarios and processes (EDA 2023). At the final stage, the aim is to assist in standardizing operations and facilitating readiness by establishing common ways of accomplishing relevant tasks.

While we have a number of cyber security frameworks built upon risk management, due to the unique nature of cyber attacks as mentioned above, the term "readiness" requires a shift of focus toward measures and activities in the stages that precede attacks. Analogous to conventional levels of military readiness, the cyber defense frameworks/architectures need to be customized in a proactive fashion and layered with respect to cyber defense maturity and the perceived effects of the threats if and when they take place.

2. IMPLEMENTATION OF ZERO-DAY READINESS IN THE MILITARY CONTEXT

Readiness is a standard military concept that refers to the ability to react in the intended fashion within a given timeframe. In other words, the definition of military readiness includes preparedness in the event of uncertainty. The preparation is required to cover both the physical and strategic components of missions, and it helps to ensure a greater chance of success when faced with challenges (Institute for Defense & Business 2021). Although the traditional armed forces maintain a persistent readiness posture at all times, the readiness concept outlining those activities is necessary to gain and maintain the capability for deployment and action within a given timeframe. These activities mainly consist of capabilities for transportation, provisions, storage, weapons, medical operations, training, communication and leadership.

A leading example of military readiness is the 60 Days' Notice to Deploy/Move concept, which is also referred to in the United Nations Manual for the Generation and Deployment of Military and Formed Police Units to Peace Operations (United Nations 2021). Considering the major outbreaks in history, the escalation of tensions or certain developments related to military operations or threats usually serve as the trigger point for the X Days' Notice to Move concept.

Cyberspace does not possess the same characteristics as the landscape of conventional defense operations, hence continuous resilience is still a considerable challenge (IBM Institue 2023). In conventional wars and military operations, effects and realizations are instantaneous. However, according to recent reports, the average detection time for data breaches is 287 days (Blumira 2022). The sum of detection and response time makes the duration even greater, resulting in organizations suffering the greatest of losses.

While there is considerable research into minimizing detection times for cyber attacks, particularly through applying new concepts such as AI, machine learning (Anastopoulos 2022) and quantum technologies (Libicki and Gompert 2021) to existing problems, the cyber defense postures of organizations cannot afford any delay in the response stage. Therefore, the concept we named Zero-Day Readiness emerges as a promising model to enable organizations to mitigate the offset between detection and response to cyber attacks.

While industry has led the momentum of technological transformation in recent decades, the readiness concept is one of the rare military doctrines that the civilian domain has benefited from. As discussed above, readiness at a designated state (i.e., 60 days to move) in the military context enable elements of force to be ready to act within a given timeframe. The readiness state can be depicted as a duration to act or a member of a nominal set among which each represent a set of predefined activities and/or the operational mode of a given system. Normally, threat alarms and readiness states go hand in hand. Threat alarm states are driven by perceived threats and depict how imminent an attack is, while the readiness state is determined based on the threat alarm state in order to respond to possible threat activity in a timely manner.

In a typical military readiness approach, one would define a hierarchy of alarm levels and corresponding readiness states to list a predefined set of actions, as given in Figure 1.

FIGURE 1: EXAMPLES OF TYPICAL CYBER THREAT ALARM AND READINESS STATES

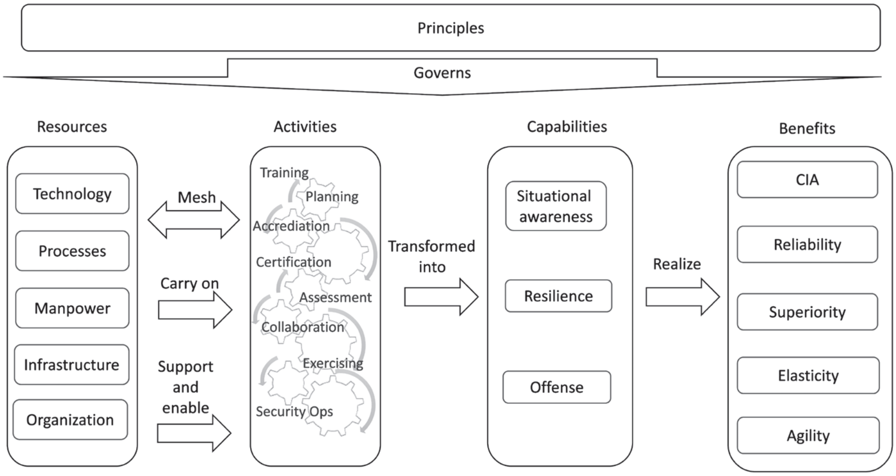
Threat Alarm State		Cyber Readiness State	
RED	Attack has started or is imminent	I	• Prohibit cross-boundary solutions • Set all perimeter security devices in aggressive mode • Prohibit all mobile devices • Warm up the disaster recovery facilities • …
YELLOW	Attack is likely	II	• Disable remote working • Limit cross-boundary solutions • Exercise alternative/contingency/emergency solutions • 7/24 SOC operations • …
GREEN	Attack is not likely	III	• Regular business operations • 5/8 SOC operations • Set all perimeter security devices in moderate mode • …

Due to the unique nature of cyber attacks and the dynamic nature of cyberspace, the conventional military readiness approach, as given above, cannot be transformed for the cyber domain. Unlike kinetic domains of land, sea and air, the indication and warnings (I&W) of malign cyber intentions are time delayed. Hence, the I&Ws of malicious cyber activities do not indicate an ongoing or imminent attack but rather a delayed indicator of a compromise which is already set forth. In light of these fundamental disparities, the typical military readiness approach cannot be used, as viable cyberspace readiness calls for a practical but novel approach to address the unique properties pertaining to this domain.

3. ZERO-DAY OPERATIONAL READINESS MODEL

In order to streamline the various tasks, zero-day readiness[1] is based on a model that acts as a guide and a roadmap for organizations to avoid additional complexity and duplicate actions. In its attempts to provide an agile and vigilant cyber readiness posture, the Zero-Day Readiness model in Figure 2 is composed of principles, resources, activities, capabilities and benefits.

FIGURE 2: THE ZERO-DAY READINESS MODEL



Theory: Under the governance of the Principles, organizations utilize the Resources to carry out Activities, whose outputs are then transferred into Capabilities, which will eventually be realized as Benefits.

Lemma 1: Resources carry on, enable and support the Activities.

Lemma 2: Organizations carry out Activities to perform day-to-day cyber security operations and baseline activities.

[1] The term "zero-day readiness" here refers to the ability to react at the same moment a designated act takes place.

Lemma 3: As part of the defense planning process, organizations invest in Resources to build Capabilities.

Lemma 4: Capabilities realize Benefits as a measurement of the success of Zero-Day Readiness.

In the following sections, we will not give an in-depth definition of each element of the readiness model. We expect readers to have a reasonable degree of cyber literacy to capture an overview of each element. We will present the significance of each element of the model for cyberspace readiness.

A. Principles

Principles are one of the fundamental pillars of the Zero-Day Readiness model and set the guiding rules and good practices in and through cyberspace. The principles in Figure 3 are considered the core values of the readiness model that must always exist and sets the operational boundaries of the solution space. Organizations with zero-day readiness must relentlessly apply those principles as they build capacity, define and improve processes, integrate and configure cyber security technologies and carry out security activities. While all of those principles are well known to the cyber defense communities, some of them are defined in detail while others are left at the conceptual level for users to scope and tailor as needed.

- Minimum attack surface: Minimum attack surface will increase the efficiency of cyber resilience capabilities by minimizing attack feasibility area and simplifying the security design.
- Zero-trust: Zero-trust will improve the availability of services and data, while hardening the authentication and authorization capabilities by verifying all requests with no exception, regardless of the previous grants.
- Risk-based decision-making: As cyber security cannot be provided 100%, risk-based decision-making increases the adaptability of the cyber security architecture to emerging threats. It also balances mission and security equities by not sacrificing the security to the mission and vice versa.
- Defense in depth: Defense in depth creates a resilient cyber security architecture through segmentation and layered defense, including supply chain security.
- Persistent situational awareness: Persistent situational awareness increases resilience and threat awareness via a recognized cyber picture supported by cyber intelligence, surveillance and reconnaissance activities providing awareness of threat actors (red picture), data, network and services (blue picture) as well as missions.
- Data centric defense: Data has become an immense source of power for

all organizations as emerging technologies such as artificial intelligence are built upon it. One of the five warfare imperatives of NATO's Warfighting Capstone Concept (NATO 2022) is cognitive superiority, and data is the atomic element of this imperative. Therefore, zero-day readiness is built by positioning and keeping the data at the center of all initiatives and capability development activities.

- Proactive defense: Sustainable readiness at zero-day forces cyber defenders to be proactive. The asymmetric nature of cyber defense and offense requires defenders to be more proactive to understand the adversary's capabilities and intentions as well as develop defense strategies in advance with the aim to minimize the risk and increase the resilience posture.

FIGURE 3: PRINCIPLES OF THE ZERO-DAY READINESS MODEL

Minimum Attack Surface	Zero-Trust	Risk-Based Decision-Making	Defense In Depth	Persistent Situational Awareness	Data-Centric Defense	Proactive Defense
- Decrease attack feasibility area - Simplify the security design - Increase efficiency of defense, detection and response	- Improve availability of services - Increase efficiency of defense, detection and response - Increase flexibility and visibility	- Increase flexibility - Improve adaptibility	- Layered defense - Segmentation - Managing supply-chain risks	- Real-time recognized cyber picture - Cyber intelligence	- Cognitive superiority - Optimizing emerging and disruptive technologies	- Holistic Approach - Threat modeling

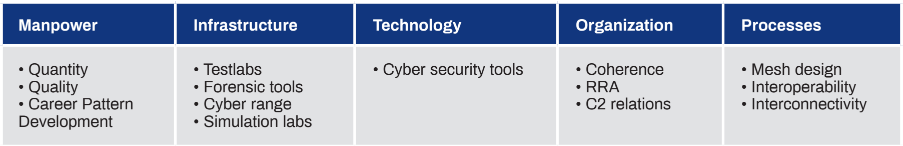
B. Resources

In order to establish zero-day readiness, organizations must build and secure the necessary resources given in Figure 4. The resources act as the decision variables of the proposed model by which an organization can tune its level of readiness. All of the resource elements except the processes represent the nodes of the solution space, while the processes act as arcs of the model to mesh the mentioned resources.

- **Manpower:** Proficient manpower with sufficient quantity and quality is fundamental to zero-day readiness and requires organizations to develop career patterns and continuous individual training programs for personnel to carry on cyber security activities.
- **Infrastructure:** Zero-day readiness requires necessary infrastructure for Security and Network Operations Centers (SOC and NOC), data forensics capabilities, secure communication lines, testing, training and exercise facilities, broadband data channels, data storage, virtualization, network sophistication and cryptographic infrastructure capabilities.

- **Technology:** Zero-day readiness requires the purchase, integration and configuration of necessary technologies within the security infrastructure. It is crucial to note that technology is neither a capability nor an objective, but a means to achieve objectives and build capabilities.
- **Organization:** Organizations must have coherent structures where the command and control (C2) relations, roles, responsibilities and authorities (RRA) are clearly defined and deconflicted. As part of continuous assessment, it is crucial to backward deconflict RRAs as new elements are created.
- **Integrated processes:** The technologies, infrastructure, organization and manpower must be integrated into each other in a meshed structure by means of well-defined and continuously improved processes in order to get the most out of them.

FIGURE 4: RESOURCES OF THE ZERO-DAY READINESS MODEL



Manpower	Infrastructure	Technology	Organization	Processes
• Quantity • Quality • Career Pattern Development	• Testlabs • Forensic tools • Cyber range • Simulation labs	• Cyber security tools	• Coherence • RRA • C2 relations	• Mesh design • Interoperability • Interconnectivity

C. Activities

Resources perform the activities given in Figure 5 to transform the outputs of those activities into capabilities. In order to establish zero-day readiness, organizations must carry out those activities in a consistent and persistent manner. Continuity and persistency in all of the given activities below are key to successfully implementing a sustainable high state of zero-day readiness.

- **Planning:** As Dwight D. Eisenhower once said, "Plans are useless, but planning is indispensable." Organizations with zero-day readiness must have the capability to conduct at minimum business impact analysis, plan for business continuity and disaster recovery using the Primary, Alternative, Contingency, Emergency (PACE) approach (Williams 2021). Since asset management, mapping assets to missions and prioritization are intrinsic parts of planning activities, these enable situational awareness and cyber resilience.
- **Certification and Accreditation:** The Zero-Day Readiness model requires continuous risk-based, threat-informed assessment, certification and accreditation. Accreditation decisions must be based on mission assurance as systems go through compliance checks to ensure critical outputs and support for mission objectives. Stagnant readiness based on standards

on dusty shelves will fail the next Patch Tuesday since the threat vectors continuously change their TTPs and find new ways to get in.

- **Assessment:** Cyber security requirements must be reviewed, designed into systems, and continuously updated to reflect the emerging threat perception supported by the technology and threat forecast. Organizations with higher readiness level must continuously evaluate the capabilities and skillset of their cyber security architectures against current and forecasted cyber threat vectors.
- **Training and Exercise:** Organizations seeking higher readiness must convert the day-to-day operations into individual and collective training opportunities to build capacity, measure and improve resources and capabilities.
- **Collaboration:** Since cyberspace and cyber threats have no boundaries as we traditionally conceive them, sharing threat intelligence, best and worst practices, threat and technology forecasts, personnel exchange programs to the maximum extent are key to establishing a credible readiness posture. In addition, collaboration invigorates the required momentum to facilitate resource-demanding persistence in zero-day readiness.
- **Security Operations:** Zero-day readiness requires a recognized cyber picture to be established on red, blue and white cyber terrain, prioritizing critical assets to defend (perimeter, network, endpoint, data, mobile, cloud, user, privileged access), threat-informed detection, responding to incidents and recovering from cyber attacks while maintaining delivery of security services under contested environments.

FIGURE 5: ACTIVITIES OF THE ZERO-DAY READINESS MODEL

Planning	Certification and Accreditation	Assessment	Training and Exercise	Collaboration	Security Operations
- Business impact analysis - Business continuity plan - Disaster recovery plan	- Security targets - Target of evaluation - Common criteria - Compliance - Risk management	- Review - Threat and technology forecast - Standardization	- Training need analysis - Measurement and improvement	- Sharing - Lessons learned	- Situational awareness - Defend - Detect - Incident response - Recovery

D. Capabilities

NIST defines five fundamental functions within the cyber resilience framework to better endure against malicious cyber activities: identify, defend, detect, respond and recover (NIST 2018). Those functions are mainly focused on the defensive portion of cyber security. However, as pointed out by the International Institute for Strategic Studies in their military cyber maturity study (Blessing and Austin 2022), offensive capabilities are an indispensable portion of the cyber readiness and deterrence posture. Without testing against credible offensive capabilities, the evaluation of the maturity of defensive capabilities will fall short. In addition, the cyber deterrence will be crippled as deterrence-by-punishment will not exist (Libicki 2009). Therefore, we add offensive capability to our model depicted in Figure 6 to achieve complete readiness.

1) Resilience

Cyber resilience is the most effective way to achieve cyber deterrence-by-denial through being able to deliver business outputs even in cyber-contested environments. It encompasses understanding and appreciating the cyber terrain, protecting assets, detecting malicious activities and responding to them, and finally recovering from attacks by replacing the lost data and services. A higher level of resilience maturity increases the cost of a successful attack, which is believed to be a deterrent factor when the cost exceeds the expected benefits. Hence, deterrence-by-denial through resilience is an enabler of zero-day readiness.

Competent defensive capabilities include defined, applied, monitored and improved security policies, sustainable skill development, AI-powered cutting-edge technologies, and configuring the defensive capabilities properly. Detection of attacks is the main driver of the Zero-Day Readiness model; hence, the length of average detection time is one of the major parameters of cyber readiness. Risk and threat-informed detection is best implemented in order to keep up with dynamically changing threat landscapes. Once the cyber events are detected, containing them, executing a response plan,[2] assessing and remediating the damage, communicating to stakeholders, and collecting forensic data are essential parts of the response capability. Individual and collective training as well as automatization via emerging technologies to respond timely even under stressful times are key to achieving a high readiness on response capability.

2) Offense[3]

Offensive cyber capabilities are not necessarily needed for all organizations, but those who have functions to develop credible cyber deterrence by retaliation and operationalize cyberspace as a new military domain need to have tools to use a certain degree of offensive capabilities or have processes in place to demand cyber targeting in support of their objectives. Offensive capabilities are crucial to achieving complete freedom of maneuvering and superiority in and through cyberspace (USA 2018). Without superiority, achieving a credible cyber readiness at zero day will always be contested by the adversaries. While the norms and legal aspects of utilizing offensive capabilities are still maturing, authorized organizations will be better prepared if they develop and improve targeting capabilities to use if and when necessary.

[2] Considering the broad spectrum of cyber attack types and associated response scenarios, the Response Plan refers to tailored steps of reaction with respect to the actual nature of the incident, instead of a rigid or one-for-all response scenario.

[3] The challenges of attribution in the context of retaliation, legal aspects and rules of engagement are factors to consider when planning to use offensive capabilities.

Even in those organizations with no intention of creating offensive cyber effects, a certain level of offensive capabilities is still required to achieve zero-day readiness for the purpose of red-teaming and penetration testing activities to measure and improve defensive capabilities.

While most of the defensive activities are carried out by automated tools, some of the attacks, namely sophisticated ones, need a special response to understand the details of the TTP and damage. Threat hunting is another crucial capability for zero-day readiness to minimize the dwell time of attackers in and through the cyber terrain.

3) Situational Awareness

Readiness requires a thorough understanding of the adversary's capabilities, intention and activities to make better intel-driven decisions as it sheds light on the unknown. Without prior information on the enemy's attack types and behaviors, proactive defense cannot be achieved. Hence, zero-day readiness requires timely collected and analyzed tactical, operational and strategic level cyber threat intelligence for the red picture.

Like any competent commander understands the battlefield with all threats and opportunities, the cyber security posture starts with a thorough understanding of the cyber terrain, where all the services and data reside. The identification of key cyber terrain to defend is a continuous activity. The development of indicators and warnings, security policies, identifying assumptions and constraints are crucial functions of this capability.

FIGURE 6: CAPABILITIES OF THE ZERO-DAY READINESS MODEL

Situational Awareness	Resilience	Offense
- Defensive intelligence - Blue cyber-picture - Red cyber-picture - Mission awareness	- Protection - Detection - Response - Recovery	- Offensive intelligence - Threat hunting - Attribution - Targeting - Rules of engagement - Legal aspects

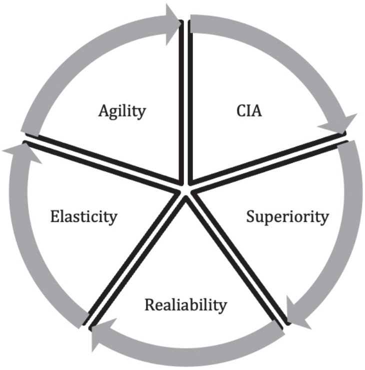
E. Benefits

The objective of the Zero-Day Readiness model is to ensure that the capabilities also realize benefits. Those benefits given in Figure 7 are confidentiality, integrity and availability (CIA), freedom of maneuvering and superiority in cyberspace, and the reliability, elasticity and agility of the services supporting or enabling the objectives of the organization. Those benefits are not a thorough list and may be amended in accordance with the strategic objective of the organization.

Availability is the probability of a service's availability when needed for use at a given time (Hagenimana, Lixin and Kandege 2016). The reliability of a service, on the other hand, is the probability of a service's state of being operational during the time of operation (US DoD 2009). Elasticity is the ability to adapt and change as necessary. As the threat landscape continuously changes, the security services must adapt to the continuum of change in a timely manner. The agility of an organization is its ability to be responsive to a mutating threat landscape and technological developments.

Superiority in cyberspace is best explained using the analogy of Anti-Access, Area Denial (A2AD) in the military domain. While A2 aims to prevent a hostile from entering the operational area, AD limits the hostile's freedom of maneuvering in that operational area by preventing it from using the available capacity and resources. With the same token, superiority in cyberspace can best be achieved by means of cyber A2AD by preventing malicious cyber actors from penetrating and operating in the cyberspace area of responsibility.

FIGURE 7: BENEFITS OF THE ZERO-DAY READINESS MODEL



4. DISCUSSION

We built the proposed model upon generally accepted cyber security frameworks and concepts. However, it exhibited a novel viewpoint on the military readiness concept in and through cyberspace, which we referred to as "zero-day cyberspace readiness." While this sounds resource-demanding, with enough political will, if the organizations can transform the readiness activities into day-to-day baseline operations, transforming the outcomes of those activities into capabilities will compensate for the initial investment in resources.

In order to be ready for all sorts of threat vectors at any time, organizations must maintain persistent readiness. One can challenge the proposed Zero-Day Readiness model for its perceived need for the sustainability of a persistent zero-day alert state. We note that performing those functions in a continuous manner might prove resource-demanding, assuming that the majority of the organizations perform those functions once or very intermittently during the course of a system's life cycle due to limited resources. However, our premise is that extending those functions listed in Figure 5 to be performed persistently over time will eventually result in lower costs by removing the higher fixed setup costs of intermittent routines.

We acknowledge that the term "zero-day readiness" may cause slight confusion among readers as it may give the impression that the proposed model is to enhance defense posture against zero-day attacks. However, the name is solely used to emphasize a persistent approach in increasing and sustaining readiness against all types of attacks. The main purpose of this paper is to serve as an abstract framework for policymakers in designing and implementing cyber security architecture at a zero-day readiness state. Nonetheless, each organization will have to scope and tailor the proposed model before validating it through a real-life implementation. While there are multiple cyber security frameworks with varying similarities, the added value of the proposed model is the consolidation of those disparate approaches around the readiness concept within a novel persistency approach.

Due to the connected and interdependent nature of cyberspace, military operations are becoming more and more dependent on national critical infrastructures owned and operated by non-military entities. Therefore, in an environment calling for all-inclusive end-to-end defense efforts, it is crucial to establish a comprehensive cyber readiness posture by incorporating the readiness levels of enabling/supporting infrastructures governed by non-military organizations.

5. CONCLUSION

Conventional military readiness doctrine sets a time or a selection of readiness states with corresponding activities for existing kinetic domains. However, due to the nature of cyber attacks and the average dwell time of attackers in blue networks for 200 days, we need a different approach to cyberspace readiness to effectively detect and defend against cyber attacks. The proposed model for cyberspace zero-day readiness is simple yet novel and comprehensive in capturing all the necessary aspects of readiness in cyberspace.

Due to the unique nature of cyber terrain and cyber attacks, we have to change the way we train and equip our cyber forces for them to be ready to act when required. Contemporary cyber threats have an ultimate advantage over defenders due to digital transformation. In order to compensate for the existential handicap of the average of 200 days to detect an ongoing attack, organizations must notionally be in a readiness state at least 200 days ago to detect and respond to the attack attempt. In other words, organizations must be in a zero-day readiness state today in order to detect and respond to an attack that will likely happen in 200 days. Rolling backward over the time domain, the proposed Zero-Day Readiness model ensures improved capabilities with competent resources carrying out activities under the governing principles.

While maturing the readiness state is an evolutionary journey as the threat landscape continuously changes, the proposed cyberspace readiness level at zero-day model can be used as a benchmark for organizations to analyze and assess the gap and to achieve the desired readiness in due course.

One of the major contributions of the proposed readiness model to underlying industry cyber security frameworks is the inclusion of offensive capabilities in the readiness concept. While the norms are still developing for offensive capabilities, it is crucial to note the importance of offensive capabilities for readiness in order to establish freedom of maneuvering and action in and through cyberspace.

The Zero-Day Readiness model enables and supports superiority in and through cyberspace by utilizing the resources in security activities in order to generate capabilities which will eventually be realized in benefits.

REFERENCES

AICPA. 2023. “SOC for Service Organizations: Trust Services Criteria.” https://us.aicpa.org/interestareas/frc/assuranceadvisoryservices/aicpasoc2report.

Anastopoulos, Vasileios, and Davide Giovannelli. 2022. “Automated / Autonomous Incident Response.” https://ccdcoe.org/uploads/2022/05/Automated-Autonomous-Davide-Giovannelli.pdf.

Blessing, Jason, and Greg Austin. 2022. “Assessing Military Cyber Maturity : Strategy, Institutions and Capability.” IISS. February 3, 2022. https://www.iiss.org/blogs/research-paper/2022/02/assessing-military-cyber-maturity.

Blumira. 2022. “Blumira’s State of Detection & Response.” https://www.blumira.com/wp-content/uploads/2022/05/State-of-Detection-and-Response-1.pdf.

CCDCOE. 2020. “Recent Cyber Events and Possible Implications for Armed Forces.” https://www.ccdcoe.org/uploads/2020/10/Recent-Cyber-Events-and-Possible-Implications-for-Armed-Forces-6-October-2020_Final.pdf.

CCDCOE. 2022. “Recent Cyber Events: Considerations for Military and National Security Decision Makers.” https://www.ccdcoe.org/uploads/2022/02/Report_Reflections_on_2021_A4.pdf.

CheckPoint. 2023. “What Is a Cloud-Native Application Protection Platform (CNAPP).” https://www.checkpoint.com/cyber-hub/cloud-security/what-is-a-cloud-native-application-protection-platform-cnapp/.

Das, Pankaz, Rezoan A. Shuvro, Mahshid Rahnamay-naeini, Nasir Ghani and Majeed M. Hayat. 2018. “Efficient Interconnectivity Among Networks Under Security Constraint.” In *MILCOM 2018 – 2018 IEEE Military Communications Conference (MILCOM)*, 88–93, Los Angeles: IEEE. doi: 10.1109/MILCOM.2018.8599813.

EDA. 2023. “Capability Development: Cyber.” https://eda.europa.eu/what-we-do/capability-development/cyber.

ENISA. 2023. “The ENISA Threat Landscape (ETL) Report.” https://www.enisa.europa.eu/topics/cyber-threats/threats-and-trends.

Fortinet. 2023. “What Is Cybersecurity Mesh?” https://www.fortinet.com/resources/cyberglossary/what-is-cybersecurity-mesh.

Forums, GRC World. 2023. “Cloud Migration and Remote Working Attracting New Breed of Cyber-Threat.” https://www.grcworldforums.com/cloud-migration-/cloud-migration-and-remote-working-attracting-new-breed-of-cyber-threat/3987.article.

Hagenimana, Emmanuel, Song Lixin and Patrick Kandege. 2016. “Computation of Instant System Availability and Its Applications.” *SpringerPlus* 5, no. 1: 954. https://doi.org/10.1186/s40064-016-2590-x.

IBM Institute. 2023. “5 Trends for 2023.” https://www.ibm.com/downloads/cas/JLKJK1ZP.

Institute for Defense & Business. 2021. “What Is Military Readiness?” https://www.idb.org/what-is-military-readiness/.

ISACA. 2023. “COBIT ISACA Framework.” https://www.isaca.org/resources/cobit.

ISO. 2023. “ISO/IEC Information Security Management.” https://www.iso.org/isoiec-27001-information-security.html.

Kushwaha, Neal, Przemysław Roguski and Bruce W. Watson. 2020. “Up in the Air: Ensuring Government Data Sovereignty in the Cloud.” In *2020 12th International Conference on Cyber Conflict 20/20 Vision: The Next Decade*, edited by T. Jančárková, L. Lindström, M. Signoretti, I. Tolga, G. Visky, 43–61. Tallinn: NATO CCDCOE Publications. https://ccdcoe.org/uploads/2020/05/CyCon_2020_book.pdf.

Libicki, Martin C. 2020. “Correlations Between Cyberspace Attacks and Kinetic Attacks.” In *2020 12th International Conference on Cyber Conflict 20/20 Vision: The Next Decade*, edited by T. Jančárková, L. Lindström, M. Signoretti, I. Tolga, G. Visky, 199–213. Tallinn: NATO CCDCOE Publications. https://ccdcoe.org/uploads/2020/05/CyCon_2020_book.pdf.

Libicki, Martin C., and David Gompert. 2021. “Quantum Communication for Post-Pandemic Cybersecurity.” In *2021 13th International Conference on Cyber Conflict: Going Viral*, edited by T. Jančárková, L. Lindström, G. Visky, P. Zotz, 371–386. Tallinn: NATO CCDCOE Publications. https://ccdcoe.org/uploads/2021/05/CyCon_2021_book_Small.pdf.

Libicki, Martin. 2009. *Cyber Deterrence and Cyberwar*. Santa Monica: RAND Corporation. https://www.rand.org/content/dam/rand/pubs/monographs/2009/RAND_MG877.pdf.

MITRE. 2023. “MITRE ATTACK.” https://attack.mitre.org.

NATO. 2022. “NATO Warfighting Capstone Concept.” https://www.act.nato.int/nwcc.

NIST. 2018. “Cybersecurity Framework Version 1.1.” https://www.nist.gov/cyberframework/framework-documents.

Security, CIS Center For Internet. 2023. “CIS Critical Security Controls.” https://www.cisecurity.org/controls.

Slack, Chalsey. 2016. “Wired yet Disconnected: The Governance of International Cyber Relations.” *Global Policy* 7, no. 1: 69–78.

Tabansky, Lior. 2011. “Basic Concepts in Cyber Warfare.” *Military and Strategic Affairs* 3, no.1: 75–92.

United Nations. 2021. *United Nations Manual for the Generation and Deployment of Military and Formed Police Units to Peace Operations May 2021*. https://pcrs.un.org/Lists/Resources/04-%20Force%20and%20Police%20Generation%20Process/Force%20Generation%20Documents%20(Military)/2021.05%20Manual%20for%20Generation%20and%20Deployment%20of%20MIL%20and%20FPU.pdf.

US DoD. 2009. “Reliability, Availability, Maintainability, and Cost Rationale Report Manual.” https://www.dau.edu/tools/Lists/DAUTools/Attachments/133/DoD-RAM-C-Manual.pdf.

USA. 2018. “National Cyber Strategy of United States of America.” https://trumpwhitehouse.archives.gov/wp-content/uploads/2018/09/National-Cyber-Strategy.pdf.

Williams, Tim. 2021. “Applying PACE Methodology across the Full Spectrum of Modern Day Operational Deployments.” QINETIC. February 1, 2023. https://www.qinetiq.com/en/blogs/applying-pace-methodology-across-the-full-spectrum-of-modern-day-operational-deployments.

2023 15th International Conference on Cyber Conflict
Meeting Reality
T. Jančárková, D. Giovannelli, K. Podiņš, I. Winther
2023 © NATO CCDCOE Publications, Tallinn

AI-assisted Cyber Security Exercise Content Generation: Modeling a Cyber Conflict

Alexandros Zacharis
European Union Agency
for Cyber Security
alexandros.zacharis@enisa.europa.eu

Razvan Gavrila
European Union Agency
for Cyber Security
razvan.gavrila@enisa.europa.eu

Constantinos Patsakis
Department of Informatics
University of Piraeus
kpatsak@unipi.gr

Demosthenes Ikonomou
European Union Agency
for Cyber Security
demosthenes.ikonomou
@enisa.europa.eu

Abstract: A cyber conflict can be defined as a cyberattack or a series of attacks that target the critical functions of a country. Such attacks can potentially wreak havoc on government and civilian infrastructure and disrupt critical systems, resulting in damage to the state and even loss of life. National bodies are usually expected to run cyber crisis exercises to prevent such attacks and prepare for their impact. Developing risk scenarios that are both relevant and up to date with the current threat landscape is a critical element in the success of any cyber exercise, especially a cyber conflict scenario.

Our work explores the results of applying machine learning to unstructured information sources to generate structured cyber exercise content in preparation for or during a destructive cyber conflict. We collected a dataset of publicly available cyber security articles and used them to assess future threats and as a skeleton for new exercise scenarios. We utilize named-entity recognition to structure the information based on a novel ontology. With the help of graph comparison methodologies, we match the generated scenarios to known threat actors' tactics, techniques, and procedures and enrich the final scenario accordingly, with the help of synthetic text generators following our novel artificial-intelligence-assisted cyber exercise framework (AiCEF). Our framework has been evaluated on its efficiency and speed and can produce structured cyber exercise scenarios in real time, provided with incident descriptions in raw text format or a set of keywords. By deep diving into a pool of pre-tagged incidents, AiCEF can build exercise content from scratch, assisting inexperienced exercise planners in generating a scenario quicker and achieving a level of quality similar to an experienced planner or subject matter expert.

We have assessed our methodology for relevance and preparedness by applying it to a real cyber conflict use case to model two categories of crisis management exercise scenarios: pre-conflict and post-conflict initiation. Thus, we assess whether the generated scenarios match the attack trends and the news feeds that were not used in training the AiCEF and prove that we can provide targeted and customized awareness of upcoming incidents.

Keywords: *cyber conflict, cyber awareness, cyber exercises scenario, artificial intelligence, machine learning, named-entity recognition*

1. INTRODUCTION

Cyber security exercises (CSE) are increasingly becoming an integral part of the cybersecurity training landscape [1], providing hands-on experience to personnel of both public and private organizations worldwide. The ISO Guidelines for Exercises [2] define a CSE as "a process to train for, assess, practice, and improve performance in an organisation." Similarly, ENISA defines a CSE as "a planned event during which an organisation simulates cyber-attacks or information security incidents or other types of disruptions to test the organisation's cyber capabilities, from being able to detect a security incident to the ability to respond appropriately and minimise any related impact." [3]

The success of such exercises can only be measured by their outcomes and ability to address real-world needs. As a result, each exercise must prioritize the prompt identification of training objectives, topic-specific preparation, and mirroring of real-world scenarios. To achieve these aims, exercise planners (EPs) must invest substantial effort, particularly when creating material miming cyber operations.

The Current Problem

Cyber operations are best used in combination with electronic warfare, disinformation campaigns, anti-satellite attacks, and precision-guided munitions. The objective is to degrade informational advantage and intangible assets (e.g., data) for operational advantage. Cyber operations can also be used for political effect by disrupting finance, energy, transportation, or government services [4]. This inherited complexity makes creating realistic scenarios difficult, especially for inexperienced cyber exercise planners or those lacking subject matter expertise. The planner must make a plethora of choices when creating a scenario to not only meet the learning objectives but also make it timely and appealing to the target audience. The research question is whether the above process can be automated. While previous work [5] has shown that the planning process can be modeled [6], [7] and automated to a great extent, the relevance and timeliness aspects have not yet been explored. Therefore, the core problem that this work studies is the relevance of the generated exercises and their usefulness in a real-world scenario based on public information sources.

Proposed Solution and Contribution

Based on the above, this paper illustrates how an EP can generate structured and realistic CSE scenarios based on a pool of pre-tagged incident information despite having little prior experience in scenario development. We also demonstrate how using machine learning (ML) and our artificial-intelligence-assisted cyber exercise framework (AiCEF) [5], such scenarios can be generated from the ground up, in particular the threats, attacks, assets, vulnerabilities, and attack vectors used by sophisticated threat actors, as well as the potential impact and severity of the resulting cyber incidents.

Thus, the main contributions reported in this article include but are not limited to:

1) mapping real-world threat information using our cyber exercise scenario ontology (CESO);
2) using this mapping to generate AI-enriched scenarios to improve their realism;
3) identifying the best strategy for using known, relevant incidents to predict and train on incidents that will materialize in the future.

Overview of This Paper

The rest of the paper is divided into four sections. First, we summarize previous literature on cybersecurity exercises and cyber drills. Then, we provide an overview of the AiCEF framework and how it uses ML to generate cybersecurity exercise scenarios. We focus on the cyber exercise scenario ontology (CESO) and the scenario augmented model (SAM) used, and on how EPs can use advanced persistent threat (APT) enhancers to simulate the activity of sophisticated threat actors more precisely. Next, we present the methodology to generate the test scenarios step by step. We have created simulated incident datasets for the public administration, energy, and information and communications technology (ICT) sectors for this work. In Section 5, we compare our datasets to observed cybersecurity attacks to provide an overview of our main use cases and findings by assessing whether the synthetic scenarios match observed real-world cyber trends. Our main focus has been the cyber incidents that public sources linked to Ukraine before and after the early phases of the 2022 aggression. We made this choice based on the duration of the conflict and the abundance of resources covering it, providing the perfect testbed for the development of future exercise scenarios in an attempt to cover future events.

2. RELATED WORK

Previous research in the field of cybersecurity exercises [8]–[10] has highlighted the importance of cyber drills [11] in aiding teams in designing, implementing, managing, and defending a computer network [3], [12]–[17]. Patriciu and Furtuna [18], [19] propose various processes and criteria for developing a cybersecurity exercise [20].

Green et al. [21] and Rursch et al. [22] have conducted additional research on cyber defense competitions [23]; [24] examines the best-suited architecture, tools, and strategies to create an active learning experience [25]–[27], whereas [28] takes a different approach to live cyber drills, offering lessons learned and making recommendations to assist organizations in conducting their own exercises. Other studies have examined how to run a cybersecurity exercise using service providers [29], [30].

In the literature, most researchers discuss serious games and how they can be used to train future practitioners, focusing on various aspects of capture-the-flag (CTF) challenges. The core concept is that the gamification of a thought process can significantly improve the learning process [31], [32] and seems rather effective in cybersecurity [33]–[36].

Nevertheless, despite the issues that CTF challenges and other gamification approaches might have, the crucial issues are the identification of the learning objectives, the positioning, and the timeliness of the content [37], [38]. Undoubtedly, for core cybersecurity skills, these aspects are not that thorny. Yet they become a dominant problem when the target training group is not a novice or when multiple skills have to be trained across diverse groups. As a result, these aspects require a lot of manual effort, good knowledge of the field, and an overview of how different sectors operate [39].

Moreover, cybersecurity exercises can be used to generate scientifically useful datasets for future security research [40], [41] and to identify hidden risks from ineffective security policies and/or processes [42], [43].

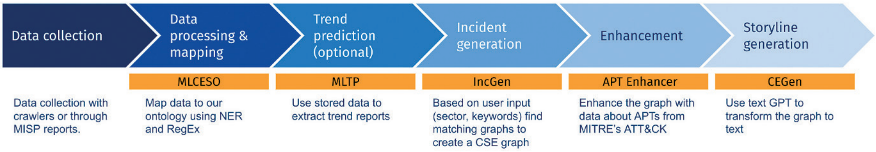
3. CESO AND THE AiCEF FRAMEWORK

AiCEF [5] is an ML-powered cyber exercise generation framework developed in Python consisting of various tools to help an EP create a timely and targeted CSE scenario regardless of their experience level. Its main components relevant to the work presented in this paper are the following:

1) CESO: the cyber exercise scenario ontology used to describe the various components of a CSE;
2) AiCEF: the cyber exercise framework used to model CSEs based on CESO using ML;
3) MLCESO: the ML models trained to parse text and extract objects based on CESO;
4) IncGen: the incident generation module that models a CSE incident from the MLCESO-extracted objects based on CESO;
5) CEGen: the cyber exercise generation module that models a CSE from the MLCESO-extracted objects based on CESO;
6) KDb: a knowledge pool of incidents stored in a database. Extracted objects and other characteristics, including the STIX 2.1 blob, are stored in the database;
7) APT Enhancer: a module that helps enhance any STIX 2.1 incident graph with objects of the modeled activity of known APTs [44].

Figure 1 shows these components in a timeline diagram to help the reader get a quick grasp of the role of each component in the flow and navigate through the rest of the sections understanding how these pieces fit in the bigger picture.

FIGURE 1: PROCESS FLOW AND THE CORRESPONDING MODULES OF AICEF [5]


Data collection
Data processing & mapping
Trend prediction (optional)
Incident generation
Enhancement
Storyline generation
MLCESO
MLTP
IncGen
APT Enhancer
CEGen
Data collection with crawlers or through MISP reports.
Map data to our ontology using NER and RegEx
Use stored data to extract trend reports
Based on user input (sector, keywords) find matching graphs to create a CSE graph
Enhance the graph with data about APTs from MITRE's ATT&CK
Use text GPT to transform the graph to text

We briefly describe each of these components in the following paragraphs, but the interested reader can refer to [5] for more detail.

Cyber Exercise Scenario Ontology (CESO)

A CSE is a collection of simulated incidents provided to players in an orchestrated way to achieve the exercise's objectives. CESO [5], the ontology developed to support AiCEF, is incident-centric, focusing on using a bottom-up approach that allows us to identify and describe incidents first so we can group them into events and then cover the full generation of CSE scenarios that fit the high-level objectives set.

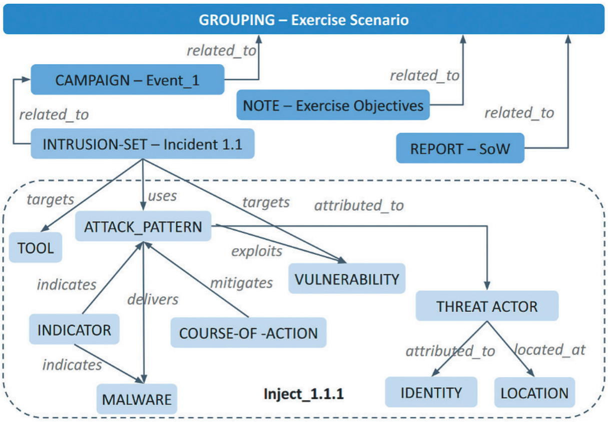
We have explored a set of existing ontologies, taxonomies, frameworks, standards, and formats relevant to cyber security and focus on the representations of the critical elements of CSEs, considering that their building blocks are the very incidents to be simulated. Our research concluded that a combination of ISO 22398 [2], MITRE ATT&CK [45] and Cyber Kill Chain [29], MITRE CVE [17], and STIX 2.1 [22] would provide the necessary means.

We chose STIX 2.1 as the basis for our ontology, which defines a taxonomy of cyber threat intelligence to be extended to cover our need to describe a CSE scenario.

Scenario Augmented Model

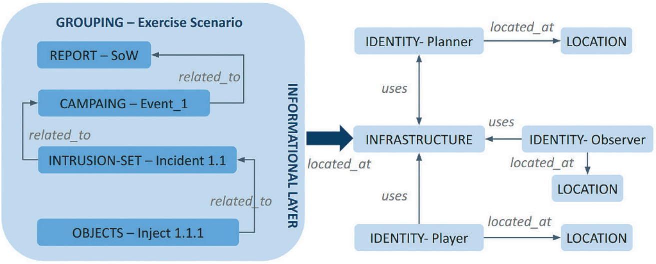
Based on the bottom-up approach, we propose a scenario augmented model (SAM) in two layers. First is the informational layer (see Figure 2) that covers the scenario's context and main attributes.

FIGURE 2: INFORMATIONAL LAYER OF CESO [5]



Second is the operational layer (see Figure 3) that describes an exercise scenario's execution flow, mainly dealing with injects delivery to the intended recipients. The operational layer has two major interrelated parts: 1) the events/injects, which describe the detailed activities of the scenario and the expected actions from the participants, and 2) the participants.

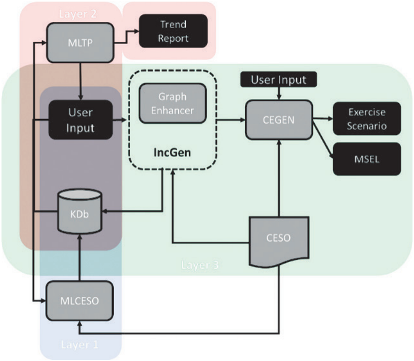
FIGURE 3: OPERATIONAL LAYER OF CESO [5]



Automated Generation of Cybersecurity Exercise Scenarios

The proof-of-concept ML-powered exercise generation framework developed (AiCEF) is illustrated in Figure 4.

FIGURE 4: THE AICEF MODULES [5]



More concretely, to generate a CSE scenario using AiCEF, the EP must perform the following steps:

i. The MLCESO module converts relevant articles or free text into incident breadcrumbs, which are mapped to the CESO ontology and stored in the KDb database.

ii. IncGen generates several incidents based on provided meta tags by merging relevant incident breadcrumbs.

iii. The APT Enhancer module can be used to enhance the generated incidents by simulating known APT activity and filling in the missing information.

iv. CEGen allows the EP to create a CSE scenario by defining various attributes, including the CSE name, number of events and incidents.

v. The scenario is generated in STIX 2.1 format along with a state-of-the-world storyline.

In the following paragraphs, we detail these steps and modules of interest for this work.

Machine Learning to CESO (MLCESO)

The most important step in our methodology is the creation of the ML pipeline that parses free text and extracts objects [46], [47] using named-entity recognition (NER), mapping them to CESO, as defined above. Separate models have been trained to cover the categories of tags as listed in Table I, which are later interlinked to form the STIX 2.1 enhanced incident graph.

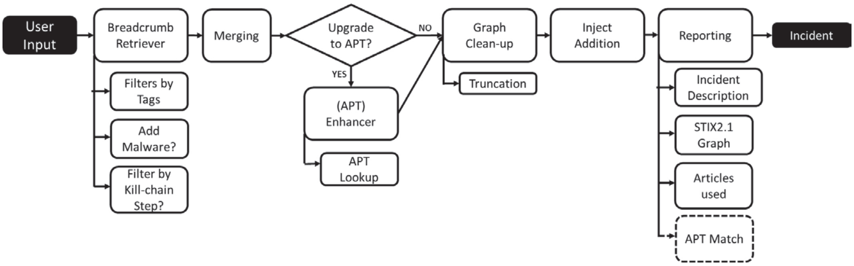
TABLE I: ANNOTATION TAGS PER CATEGORY

Category	(NER) Tag	Link to CESO and STIX 2.1
Attacker	ATTACKER_TYPE	Threat actor (attribute)
	ATTACKER_NAME	Threat actor (attribute), identity
	ATTACKER_ORIGIN	Location
Attack	MALWARE_TYPE	Malware (attribute)
	MALWARE_NAME	Malware (attribute)
	ATTACK_TYPE	Attack pattern
	VULNERABILITY	Vulnerability
Victim	SECTOR	Identity (attribute), scenario
	ASSETS	Threat actor (attribute)
	TECHNOLOGY	Tool

Incident Generation and Enhancement (IncGen) and Knowledge Database (KDb)

Incident creation is the most important step of the scenario generation procedure and consists of several steps to achieve maximum customization (see Figure 5). All the steps can be automated, generating a variety of incidents from which an EP can choose the fittest.

FIGURE 5: INCIDENT GENERATION FLOW [5]



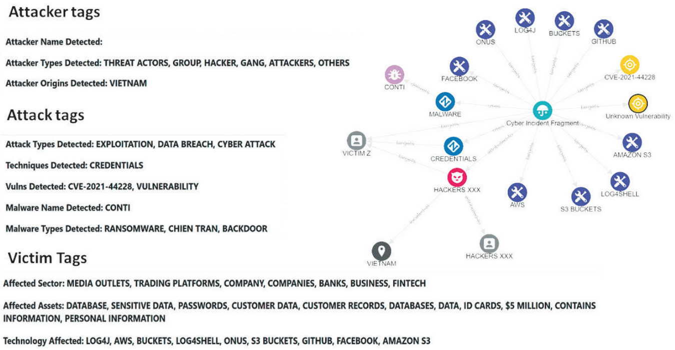
An output report and visualization of IncGen utilizing improved MLCESO tag detection can be seen in Figure 6.

The EP can provide specific text or articles for conversion to incidents or rely on a dynamic generation based on filtering parameters and a search of the existing database. The current knowledge database (KDb) consists of pre-parsed and modeled articles as described in Table II, spanning from 1 January 2020 to 1 March 2022.

TABLE II: KNOWLEDGE DATABASE CONTENT PER SOURCE

Source	Count
bleepingcomputer.com	1,368
securityaffairs.co	169
zdnet.com	495
databreaches.net	938
Total	**2,970**

FIGURE 6: INCIDENT GENERATION OUTPUT AND VISUALIZATION EXAMPLE


Attacker tags

Attacker Name Detected:

Attacker Types Detected: THREAT ACTORS, GROUP, HACKER, GANG, ATTACKERS, OTHERS

Attacker Origins Detected: VIETNAM

Attack tags

Attack Types Detected: EXPLOITATION, DATA BREACH, CYBER ATTACK

Techniques Detected: CREDENTIALS

Vulns Detected: CVE-2021-44228, VULNERABILITY

Malware Name Detected: CONTI

Malware Types Detected: RANSOMWARE, CHIEN TRAN, BACKDOOR

Victim Tags

Affected Sector: MEDIA OUTLETS, TRADING PLATFORMS, COMPANY, COMPANIES, BANKS, BUSINESS, FINTECH

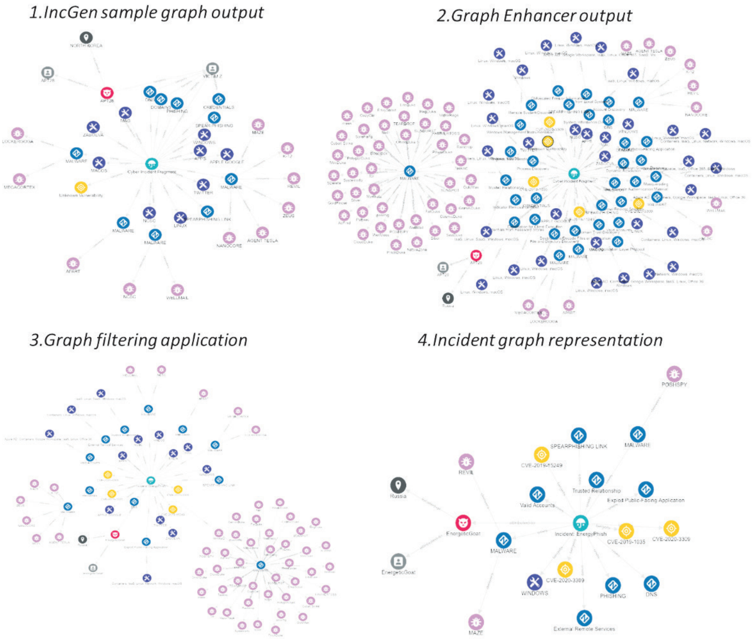
Affected Assets: DATABASE, SENSITIVE DATA, PASSWORDS, CUSTOMER DATA, CUSTOMER RECORDS, DATABASES, DATA, ID CARDS, $5 MILLION, CONTAINS INFORMATION, PERSONAL INFORMATION

Technology Affected: LOG4J, AWS, BUCKETS, LOG4SHELL, ONUS, S3 BUCKETS, GITHUB, FACEBOOK, AMAZON S3

APT Enhancer

Incidents can be enhanced with activity simulating the tactics, techniques and procedures (TTPs) of known APT actors. To simulate the activity of known APT groups, a basic STIX 2.1 structure was created per actor using the groups from MITRE, from which various attributes and TTPs were automatically extracted to populate our database. Thus, we generated a STIX 2.1 graph that can be used to compare and enhance other graphs at will. Currently, 125 APT actors can be simulated. A preview of the intermediate steps is shown in Figure 7.

FIGURE 7: APT ENHANCING AN INCIDENT IN FOUR STEPS (NO INJECTS ADDED)



4. PROPOSED METHODOLOGY

We propose a methodology to measure the effectiveness of AI-generated cyber security incidents for use in CSE scenarios. To do this, sets of incidents that occurred before and during a cyber conflict (pre-conflict and post-conflict sets) are collected and divided into sector-specific pools. The Pre-Conflict Pools are then compared to the actual incidents that took place (covered in the Post-Conflict Pool) to evaluate the methodology and determine the best strategy for generating CSE incidents with AI.

Here is a brief description of the content of the pools generated:

Pre-Conflict Pools (per sector)

> Pool 0 contains a random selection of incidents generated based on reports that have taken place prior to the conflict.
> Pool 1 contains a relevant and targeted selection of incidents generated based on reports published prior to the conflict of interest.
> Pool 2 contains relevant and targeted reports published prior to the conflict of interest, merged to formulate richer incidents.
> Pool 3 contains relevant and targeted reports published prior to the conflict of interest, merged to formulate richer incidents enhanced based on known threat actors' TTPs.

Post-Conflict Pool (per sector)

> Pool 0 contains three incidents: two relevant and targeted incidents generated based on reports published during the conflict of interest, and a merge of a relevant and targeted selection of incidents combined into a single richer incident.

A presentation of the implementation of our methodology using the AiCEF toolset can be found below.

PHASE 1: Incident Modeling (Pre-Conflict)

> Step 1.1: Identify relevant pre-conflict incidents described in various resources and extract text
>
> Step 1.2: Parse text using our MLCESO and store extracted objects in KDb using a PRE-CONFLICT meta tag
>
> Step 1.3: Perform a trend analysis of the parsed incidents in Step 1.2 to discover the most prominent Attack Types & Techniques and Threat Actors and filter by sectors impacted
>
> Step 1.4: Extract from KDb or generate synthetic incidents and form pools per impacted sector:
>
> > Pool 0: Extract related incidents over a threshold based on queries of the original KDb

Pool 1: Extract only rich incidents over a threshold with the PRE-CONFLICT meta tag.
Pool 2: Merge up to three[1] articles describing the same incident, using Incident Generator (IncGen). Incidents must have the PRE-CONFLICT meta tag.
Pool 3: Merge articles using Incident Generator (IncGen) and APT Enhancement based on known Threat Actors extracted in Step 1.3

PHASE 2: Incident Modeling (Post-Conflict)

Step 2.1: Identify relevant post-conflict incidents described in various resources and extract text

Step 2.2: Parse text using our MLCESO and store extracted objects in KDb using a POST-CONFLICT meta tag

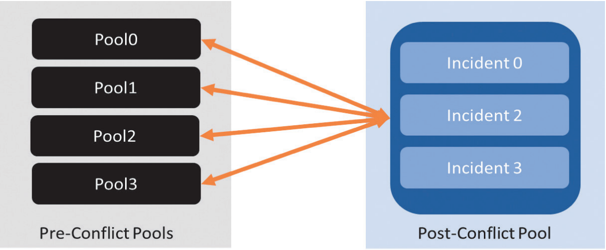
Step 2.3: Generate incidents and form a pool per affected sector:

i. Extract two rich incidents (Incident1-2) over a threshold
ii. Merging up to three relevant articles under a single Incident (Incident0)

PHASE 3: Comparison

Calculate the average coverage percentage of the Post-Conflict Pool versus the Pre-Conflict Pool sets to identify the best strategy in CSE creation using AiCEF.

FIGURE 8: COMPARISON PHASE



The comparison function used (Figure 8) to calculate coverage between modeled incidents can be summarized as follows:

[1] The number of similar articles of the same incident to be merged is a convention based on the average size of the pre-conflict incidents generated so as to be comparable to the one of Incident0 modelled post conflict.

Step 3.1: We are about to compare PreConflictPool(i)= {Inc0, Inc1…} of varying size *i=n* to PostConflictPool = (Inc0,Inc1,Inc2) with *size=m* to calculate its coverage.

Step 3.2: For each IncX in PreConflictPool0 and PostConflictPool, we extract and store in a list the name attributes of the following objects: *Attack_Pattern, Tool, Vulnerability, Malware, Location* objects. We end up with the following lists:

PreConflictPool(i) = {Inc0[Attack_Pattern.name, Tool.name, Vulnerability.name, Malware.name, Location.name], …}

PostConflictPool(j) = {Inc0[Attack_Pattern.name, Tool.name, Vulnerability.name, Malware.name, Location.name], …}

Step 3.3: We perform a one-to-one comparison between the two lists and calculate the total coverage of the PostConflictPool Incident per PostConflictPool incident duplets as follows:

$$TotalCoverage(i) = \frac{AttackPattern_{\ Coverage} + Tool_{\ Coverage} + Vunerability_{\ Coverage} + Malware_{\ Coverage} + Location_{\ Coverage}}{5}$$

where an example coverage of a *Vulnerability* Object between two incidents can be calculated as the (Number of PreConflictPool(i).Inc0[*Vulnerability*] list items that exist in PostConflictPool(i).Inc0[*Vulnerability*]) list / (PostConflictPool(i).Inc0[*Vulnerability*] list Size) * 100.

Then, we calculate the AverageTotalCoverage of all PreConflictPool incidents per PostConflictPool Incident:

$$AvarageTotalCoverage = \frac{\sum_{i=0}^{n} TotalCoverage(i)}{n}$$

5. USE CASE AND RESULTS

Based on the methodology described above, we covered the following use case to help us assess whether the generated scenarios match the attack trends and prove that AiCEF can provide targeted and customized awareness in real life.

We modeled incidents in the recent Russian-Ukrainian cyber conflict along with historical data into incident sets. The sets (pre-conflict and post-conflict) were parsed automatically and consisted of known instances of cyber incidents linked to attacks against Ukraine that have been described in publications and other reliable specialized sources. For this research, we analyzed stories reported by credible publications and reputable security researchers. Before including it in the dataset, each entry was reviewed manually to verify its relevance. We selected a total of 46 articles for the pre-conflict set and 14 for the post-conflict set and classified these into three sectors of interest: public administration, energy and ICT.

Finally, we modeled two exercises, one based on a selected pre-conflict set and the other based on the post-conflict set of incidents or relevant sectors to compare the overall training coverage achieved against future incidents.

PHASE 1: Incident Modeling (Pre-Conflict) in Practice

After implementing all the steps of our methodology, we derived the following outputs per step:

> Step 1.1: We identified 46 articles as the basis for our research describing incidents that took place between January 2020 and January 2022. Although 24 February 2022 is considered the start of the conflict in our specific example, no articles or incidents covering February 2022 were used for our analysis.
>
> Step 1.2 and Step 1.3: Performing a trend analysis of the parsed incidents in Step 1.2, we discovered a set of interesting Attack Types & Techniques and Threat Actors per sectors impacted, as illustrated in Table III.

TABLE III: TOP OBJECTS EXTRACTED

Sector	Attack Types & Techniques	Threat Actor
Public Administration	DOS, DDOS, MALWARE, DATA EXFILTRATION, PHISHING, DEFACEMENT	APT28, SANDWORM, TURLA, APT29
Energy	WIPER, RANSOMWARE, ESPIONAGE	STRONTIUM, KRYPTON, GAMAREDON, BLACKENERGY, TURLA, FANCY BEAR, SOFACY, NOBELIUM, UNC2452, PAWN STORM, SEDNIT
ICT	DOS, DDOS, WIPER, MALWARE	ACTINIUM, NOBELIUM, STRONTIUM

For actors such as APT28, Sandworm, Turla, and APT29, the findings were verified against the CCDCOE study of 2021 [4], and for ACTINIUM, against Microsoft's January 2022 report [48].

Step 1.4: Pools with the following size were then formatted per sector ([Pool0(40), Pool1(6), Pool2(40), Pool3(40)]), with 40 articles generated per Pool2-3 using AiCEF.

PHASE 2: Incident Modeling (Post-Conflict) in Practice

After implementing all the steps of our methodology, we derived the following outputs per step:

> Step 2.1: We identified 14 articles covering the three sectors of interest for the period of March 2022 – June 2022.
>
> Step 2.2 and Step 2.3:
>
> i. Two incidents (Incident1-2) rich in objects were generated per sector.
> ii. Both were merged into a single incident (Incident0).

PHASE 3: Comparison and Results

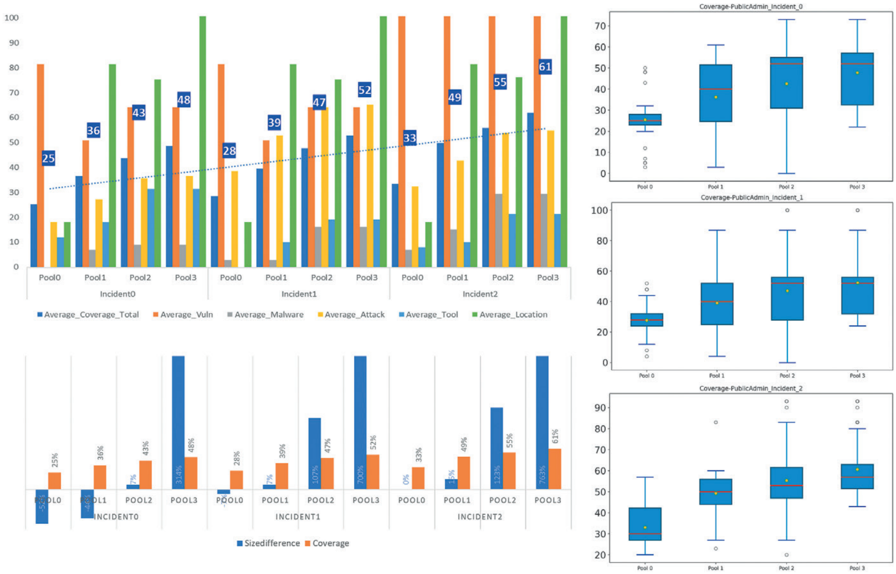
After performing all possible comparisons, both at the incident and exercise level, the following coverage results were retrieved.

Public Administration

For the public administration sector, which was the most impacted in the conflict examined in terms of attack volume, we were able to determine the following, based on Figure 9:

i. Pre-Conflict Pool 2 Incidents covered the three post-conflict incidents selected with an improved rate of an average of approximately 20%. Overall, the coverage achieved was close to 50%, meaning that attacks used during the conflict were not similar to the ones performed before the conflict.
ii. APT enhancement considerably improved the coverage, but the complexity of the incidents and the overall exercise generated rose drastically, providing little training benefit.

FIGURE 9: PUBLIC ADMINISTRATION INCIDENTS COVERAGE RESULTS



Energy Sector

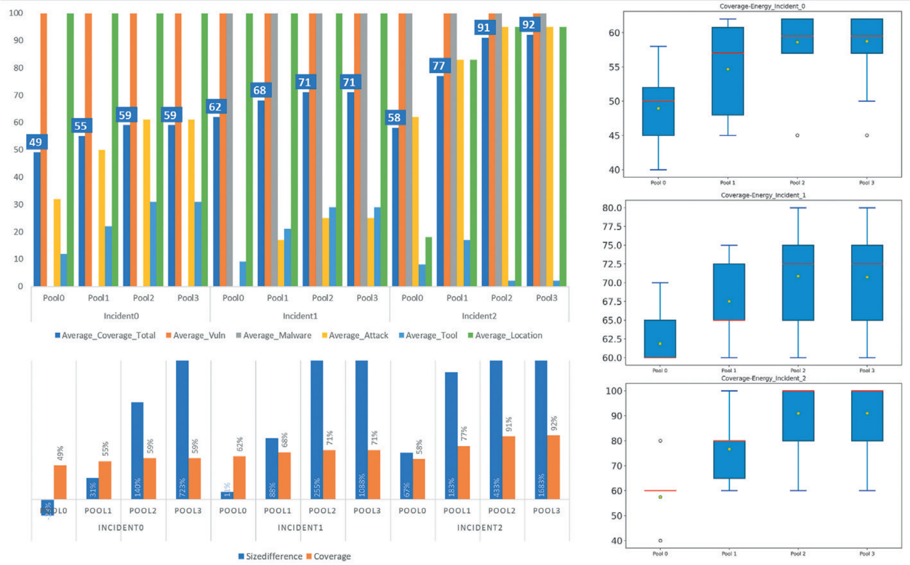
The energy sector is considered the most critical to target during conflict. Taking its importance into account, we concluded that sufficient training coverage was achieved. More specifically as seen in Figure 10:

i. Pre-Conflict Pool 2 Incidents covered the three post-conflict incidents with an improved rate of an average of approximately 17%. Overall, the coverage achieved for Pool 2 was close to 74%, meaning that attacks used during the conflict were similar to the ones performed before the conflict, providing a good space for training and exercises to improve awareness.

ii. Although APT enhancement did not improve the coverage generated, incidents in both Pools 2 and 3 managed to cover 100% for the Post-Conflict Incident 2 with a constant rate of over 60% for all 40 incidents generated.

iii. Malware prediction, although low in the automatic statistics extracted seen below, can be considered satisfactory upon manual analysis. As an example, AiCEF proposed INDUSTROYER as the selected malware to exercise with, when in reality post-conflict incidents of the energy sector were populated by the INDUSTROYER v.2.0, a malware of the same family.

FIGURE 10: ENERGY SECTOR INCIDENTS COVERAGE RESULTS



ICT Sector

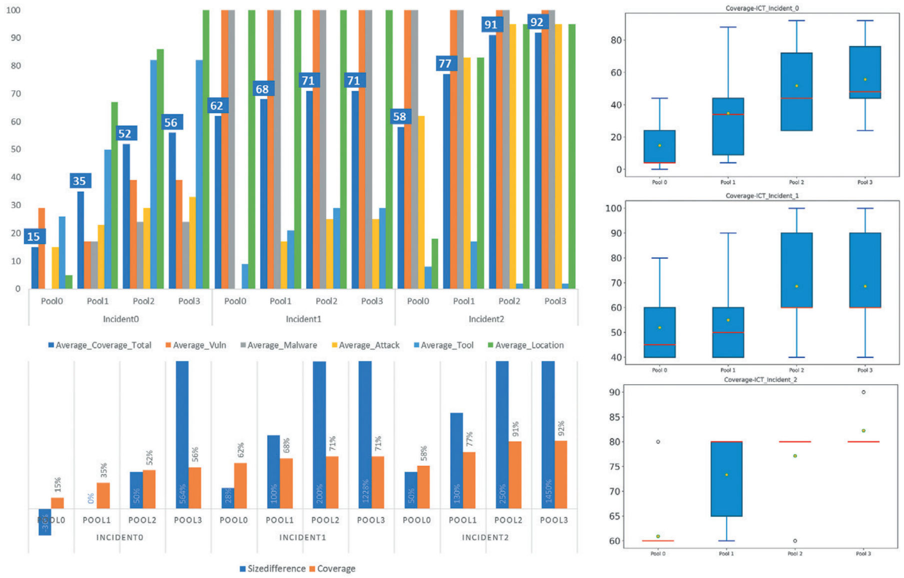
For the ICT sector, based on our analysis represented in Figure 11:

i. Pre-Conflict Pool 2 incidents covered the three post-conflict incidents with an improved rate of an average of approximately 26%.

ii. The coverage achieved for Pool 2 was close to 72%, meaning that attacks used during the conflict were similar to the ones performed before the conflict.

iii. Although APT enhancement significantly improved the coverage, the generated incidents in both Pools 2 and 3 managed to cover 100% of the Post-Conflict Incident 1 with an average of 70% for all 40 generated incidents.

FIGURE 11: ICT SECTOR INCIDENTS COVERAGE RESULTS



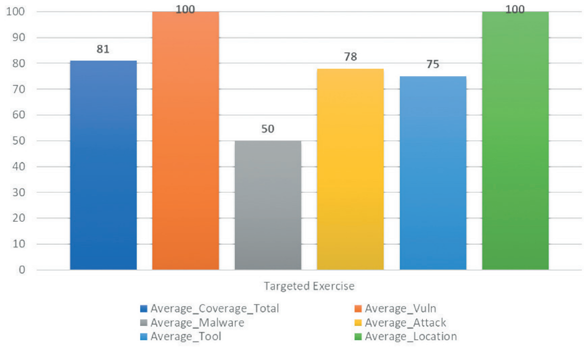
Targeted Exercise Coverage

When comparing pre-conflict incidents with post-conflict incidents in the three sectors, observations made it clear that Pre-Conflict Pool 2 (relevant merged incidents) can achieve the highest coverage with a justifiable graph size difference against Post Conflict incidents. We decided to perform a final comparison, but now at the exercise level, by creating a CSE (named Targeted Exercise) consisting of one event with three incidents, one for each sector. The incident in each sector was a merge of all Incidents of the Pre-Conflict Pool 1 (removing duplicate objects).

We then created a post-conflict exercise by simply modeling a CSE consisting of one event with three incidents, which were the Incident 0's of the Post-Conflict Pool of each sector.

The exercise bundles were compared, and the overall average total coverage achieved was 81% (Figure 12) with a 304% size graph difference, meaning that our proposed exercise graph was three times greater than that of the post-conflict exercise based on real incidents.

FIGURE 12: EXERCISE COVERAGE RESULTS



The specific vulnerabilities used by the attackers, along with the fact that the attackers used unknown vulnerabilities, were predicted by AiCEF at 100%. The framework was also good at detecting the attack techniques and tools used by the attackers in the incidents, paving the way for well-scoped exercise generation.

6. CONCLUSIONS

Based on analysis of the data collected in the above-described use case, we have reached some valuable conclusions on how to create insightful cyber security exercises that can prepare a country or a sector-specific organization ahead of a cyber conflict. Furthermore, with AiCEF we can now generate more realistic incidents and, as a consequence, better-scoped CSEs, following a statistically proven strategy that will offer the best coverage of training objectives in the expectance of future conflicts. The results of this study show that the Pre-Conflict Pool 1 incident generation strategy performs better than a random generation of sector-specific incidents. Using past incidents as a model for exercises can help achieve better coverage of exercise objectives up to an average of 13%.

The Pre-Conflict Pool 2 incident generation strategy, which involves merging multiple resources to describe the same incident, is the most effective in creating incidents using AiCEF, achieving an average of +21% overall coverage of the exercise objective set. The APT enhancement strategy (Pre-Conflict Pool 3) can achieve better overall coverage results but with a high complexity impact. In the future, more innovative methods should be used to enhance a graph based on known actors in a more effective manner.

Additionally, exercises that cover multiple sectors impacted by the same actors can achieve higher coverage (81%) of training objectives than those covering a single sector (public administration: 50%; energy: 74%; ICT: 72%).

As a final remark, we can confidently state that AiCEF can automate the generation of well-scoped CSE scenarios to prepare a specific sector or a country ahead of a cyber conflict. Scenarios can be adapted in real time as the conflict progresses, with little cyber exercise planning expertise, through the power of AI, providing accurate training material.

ACKNOWLEDGMENTS

This work was supported by the European Commission under the Horizon Europe Programme, as part of the project LAZARUS (Grant Agreement no. 101070303).

The information and views set out in this article are those of the author (s) and do not necessarily reflect the official opinion of the European Union Agency for Cybersecurity (ENISA). Neither the European Union institutions nor any person acting on their behalf may be held responsible for any use that may be made of the information contained therein.

REFERENCES

[1] M. Karjalainen, T. Kokkonen, and S. Puuska, "Pedagogical aspects of cyber security exercises," in *2019 IEEE European Symposium on Security and Privacy Workshops (EuroS&PW)*, IEEE, 2019, pp. 103–108.

[2] *Societal security—guidelines for exercises. Standard ISO22398:2013*, ISO Central Secretary, International Organization for Standardization, Geneva, CH, 2013. [Online]. Available: https://www.iso.org/standard/50294.html

[3] A. Conklin, "Cyber defense competitions and information security education: An active learning solution for a capstone course," in Proceedings of the 39th Annual Hawaii International Conference on System Sciences (HICSS'06), IEEE, vol. 9, 2006, pp. 220b–220b.

[4] J. Hakala and J. Melnychuk, "Russia's Strategy in Cyberspace," NATO STRATCOM COE, 2021. [Online]. Available: https://stratcomcoe.org/cuploads/pfiles/Nato-Cyber-Report_15-06-2021.pdf

[5] A. Zacharis and C. Patsakis, "AiCEF: An AI-assisted cyber exercise content generation framework using named entity recognition," *International Journal of Information Security*, 2023.

[6] R. Gurnani, K. Pandey, and S. K. Rai "A scalable model for implementing cyber security exercises," in *2014 International Conference on Computing for Sustainable Global Development (INDIACom)*, IEEE, 2014, pp. 680–684.

[7] S. F. Wen, M. M. Yamin and B. Katt, "Ontology based scenario modeling for cyber security exercise," in *2021 IEEE European Symposium on Security and Privacy Workshops (EuroS&PW)*, IEEE, 2021, pp. 249–258.

[8] M. Mink and F. C. Freiling, "Is attack better than defense? Teaching information security the right way," in *Proceedings of the 3rd Annual Conference on Information Security Curriculum Development*, 2006, pp. 44–48.

[9] W. Schepens, D. Ragsdale, J. R. Surdu, J. Schafer, and R. N. Port, "The cyber defense exercise: An evaluation of the effectiveness of information assurance education," *Journal of Information Security*, vol. 1(2), pp. 1–14, 2002.

[10] G. Vigna, “Teaching network security through live exercises,” in *IFIP World Conference on Information Security Education*, Monterey, CA, USA, 2003, pp. 3–18.

[11] R. Dodge and D. J. Ragsdale, “Organized cyber defense competitions,” in *Proceedings. IEEE International Conference on Advanced Learning Technologies*, 2004, pp. 768–770.

[12] W. J. Adams, E. Gavas, T. H. Lacey, and S. P. Leblanc, “Collective views of the NSA/CSS cyber defense exercise on curricula and learning objectives,” presented at the 2nd Workshop on Cyber Security Experimentation and Test (CSET 09), Montreal, Canada, 2009.

[13] A. Conklin, “The use of a collegiate cyber defense competition in information security education,” in *Proceedings of the 2nd Annual Conference on Information Security Curriculum Development*, 2005, pp. 16–18.

[14] Y. Li, M. Liljenstam, and J. Liu, “Real-time security exercises on a realistic interdomain routing experiment platform,” in *2009 ACM/IEEE/SCS 23rd Workshop on Principles of Advanced and Distributed Simulation*, IEEE, 2009, pp. 54–63.

[15] M. Liljenstam, J. Liu, D. M. Nicol, Y. Yuan, G. Yan, and C. Grier, “Rinse: The real-time immersive network simulation environment for network security exercises (extended version),” *Simulation*, vol. 82(1), pp. 43–59.

[16] B. E. Mullins, T. H. Lacey, R. F. Mills, J. E. Trechter, and S. D. Bass, “How the cyber defense exercise shaped an information-assurance curriculum,” *IEEE Security & Privacy*, vol. 5(5), pp. 40–49, 2007.

[17] B. E. Mullins, T. H. Lacey, R. F. Mills, J. M. Trechter, and S. D. Bass, “The impact of the NSA cyber defense exercise on the curriculum at the air force institute of technology,” in *40th Annual Hawaii International Conference on System Sciences (HICSS’07)*, IEEE, 2007, pp. 271b–271b.

[18] A. Furtună, V. V. Patriciu, and I. Bica, “A structured approach for implementing cyber security exercises,” in *2010 8th International Conference on Communications*, IEEE, 2010, pp. 415–418.

[19] V. V: Patriciu and A. C. Furtuna, “Guide for designing cyber security exercises,” in *Proceedings of the 8th WSEAS International Conference on E-Activities and Information Security and Privacy*, World Scientific and Engineering Academy and Society (WSEAS), 2009, pp. 172–177.

[20] J. Kick, “Cyber exercise playbook,” MITRE Corporation, Bedford, MA, USA, Tech. Rep., 2014.

[21] A. Green and H. Zafar, “Addressing emerging information security personnel needs. a look at competitions in academia: Do cyber defense competitions work?” *AMCIS 2013 Proceedings*, 1:257, 2013.

[22] J. A. Rursch, A. Luse, and D. Jacobson, “IT-Adventures: A program to spark it interest in high school students using inquiry-based learning with cyber defense, game design, and robotics,” *IEEE Transactions on Education*, vol. 53(1), pp. 71–79, 2009.

[23] J. W. Schepens and J. R. James, “Architecture of a cyber defense competition,” in *SMC’03 Conference Proceedings. 2003 IEEE International Conference on Systems, Man and Cybernetics. Conference Theme-System Security and Assurance* (Cat. No. 03CH37483), IEEE, 2003, vol. 5, pp. 4300–4305.

[24] G. B. White, D. Williams, and K. Harrison, “The cyberpatriot national high school cyber defense competition,” *IEEE Security & Privacy*, vol. 8(5), pp. 59–61, 2010.

[25] T. Augustine, R. C. Dodge *et al.*, “Cyber defense exercise: meeting learning objectives thru competition,” in *Proceedings of the 10th Colloquium for Information Systems Security Education*, 2006.

[26] M. Granåsen and D. Andersson, “Measuring team effectiveness in cyber-defense exercises: a crossdisciplinary case study,” *Cognition, Technology & Work*, vol. 18(1), pp. 121–143, 2016.

[27] D. Schweitzer, D. Gibson, and M. Collins, “Active learning in the security classroom,” in *2009 42nd Hawaii International Conference on System Sciences*, IEEE, 2009, pp. 1–8.

[28] G. B. White, G. Dietrich, and T. Goles, “Cyber security exercises: testing an organization’s ability to prevent, detect, and respond to cyber security events,” in *Proceedings of the 37th Annual Hawaii International Conference on System Sciences*, Big Island, HI, USA, 2004.

[29] J. A. Mattson, “Cyber defense exercise: A service provider model,” in *IFIP World Conference on Information Security Education*, Springer, 2007, pp. 81–86.

[30] M. E. Planning, *Directors’s guideline for civil defence emergency management groups, wyd*. (2008). Ministry of Civil Defence & Emergency Management, Wellington.

[31] F. Laamarti, M. Eid, and A. El Saddik, “An overview of serious games,” *International Journal of Computer Games Technology*, 2014, doi: 10.1155/2014/358152.

[32] Y. Chou, *Actionable Gamification: Beyond points, badges, and leaderboards*. Packt Publishing Ltd, 2019.

[33] S. Trepte *et al*., “Do people know about privacy and data protection strategies? Towards the ‘Online Privacy Literacy Scale’ (OPLIS),” in *Reforming European data protection law*, Dordrecht: Springer, 2015, pp. 333–365.

[34] B. D. Cone *et al*., “A video game for cyber security training and awareness,” *Computers & Security*, vol. 26.1, pp. 63–72.52, 2007.

[35] H. Berger and A. Jones, "Cyber security & ethical hacking for SMEs," in *Proceedings of the 11th International Knowledge Management in Organizations Conference on The Changing Face of Knowledge Management Impacting Society*, 2016, pp. 1–6.

[36] M. Hendrix, A. Al-Sherbaz, and V. Bloom, "Game based cyber security training: are serious games suitable for cyber security training?" *International Journal of Serious Games*, vol. 3(1), pp. 53–61, 2016.

[37] D. H. Tobey, "A vignette-based method for improving cybersecurity talent management through cyber defense competition design," in *Proceedings of the 2015 ACM SIGMIS Conference on Computers and People Research*, 2015, pp. 31–39.

[38] N. Wilhelmson and T. Svensson, *Handbook for Planning, Running and Evaluating Information Technology and Cyber Security Exercises*. Försvarshögskolan (FHS), 2011.

[39] U.S. Department of Homeland Security, 2013, "Cyber Tabletop Exercise (TTX)," Homeland Security Digital Library. [Online]. Available: https://www.hsdl.org/?abstract&did=789781

[40] B. Sangster, T. O'Connor, T. Cook, R. Fanelli, E. Dean, C. Morrell, and G. J. Conti, "Toward instrumenting network warfare competitions to generate labeled datasets," in CSET, 2009.

[41] T. Sommestad and J. Hallberg, "Cyber security exercises and competitions as a platform for cyber security experiments," in *Nordic Conference on Secure IT Systems*, Springer, 2012, pp. 47–60.

[42] M. Samejima and H. Yajima, "IT risk management framework for business continuity by change analysis of information system," in *2012 IEEE International Conference on Systems, Man, and Cybernetics (SMC)*, IEEE, 2012, pp. 1670–1674.

[43] K. A. Scarfone, T. Grance, and K. Masone, *Sp 800-61 Rev. 1. Computer Security Incident Handling Guide*, National Institute of Standards and Technology, U.S. Department of Commerce, Gaithersburg, MD, USA, 2008.

[44] MITRE, 2022. [Online]. Available: https://attack.mitre.org/

[45] J. Pastuszuk, P. Burek, and B. Ksieopolski, "Cybersecurity ontology for dynamic analysis of it systems," *Procedia Computer Science*, vol. 192, pp. 1011–1020, 2021.

[46] R. MacIntyre, University of Pennsylvania, PA. 1995. *Penn Treebank Tokenizer*. [Online]. Available: https://www.cis.upenn.edu/~treebank/tokenizer.sed

[47] N. Tsinganos and I. Mavridis, "Building and evaluating an annotated corpus for automated recognition of chat-based social engineering attacks," *Applied Sciences*, vol. 11(22), 10871, 2021, doi: 10.3390/app112210871.

[48] Microsoft Threat Intelligence Center, 2022, "Destructive malware targeting Ukrainian organizations," Microsoft. [Online]. Available: https://www.microsoft.com/en-us/security/blog/2022/01/15/destructive-malware-targeting-ukrainian-organizations/

2023 15th International Conference on Cyber Conflict
Meeting Reality
T. Jančárková, D. Giovannelli, K. Podiņš, I. Winther
2023 © NATO CCDCOE Publications, Tallinn

Request for a Surveillance Tower: Evasive Tactics in Cyber Defense Exercises

Youngjae Maeng
National Security Research Institute
Daejeon, South Korea
brendig@nsr.re.kr

Mauno Pihelgas
Tallinn University of Technology
Tallinn, Estonia
mauno.pihelgas@taltech.ee

Abstract: The cyber defense exercise (CDX) is an emerging live-fire exercise that enables diverse teams with different roles to train in one game. To evaluate the cyber defense capabilities of the training audience, organizers prepare various scores using different scoring methods ranging from technical to non-technical. The technical scores in Locked Shields, for example, consist of an availability check, a usability check, the success of the red team (RT) attack, and forensics.

Immersed in scores due to excessive competition, a blue team (BT) may unnecessarily focus on the scoring process, aiming to perform evasive tactics (ET), which boosts scores unfairly by abusing the weaknesses of the scoring system. ET has occurred in various forms in existing CDXs, and similar cases have been found in the recent iteration of CDXs, meaning that ET is becoming BT's selectable strategy.

Such a phenomenon is undesirable since it will reduce the reliability of the evaluation and the effectiveness of the training. In this paper, we provide an overview of an availability check and examine ET that appeared in both the availability check and RT's evidence-obtaining process, followed by several mitigations to them. We also discuss evidence and usability issues of ET in CDX and conclude by emphasizing the importance of supporting the green team (GT) in researching and implementing a robust scoring system.

Keywords: *cyber defense exercise, evasive tactics, availability check, cheating, scoring*

1. INTRODUCTION

The cyber defense exercise (CDX) is an emerging cyber exercise for training, providing large-scale cyber environments with different types of cyber threats configured to train audiences to respond in technical and non-technical fields. The CDX scores range from technical fields, such as availability, usability, red team (RT) attacks (e.g., client-side, network, and web), and forensics, to non-technical fields, such as law, media, and various types of reports, in order to evaluate the ability to respond in the various fields that cyber threats can affect. The score represents a necessary measure for the evaluation of the exercise; however, both organizers and participants should recognize that the intrinsic value of the exercise extends beyond the score.

Despite such an ideal perception, scores and rankings occasionally lead to overheated competition between the participants, who prioritize score gains over exercise content. This means that participants may try to gain scores unfairly by identifying weaknesses in the rules or systems of the exercise, which we refer to as evasive tactics (ET). For example, suppose that the success of an RT attack is set higher than the total score of the availability and usability scores. Considering score gains and losses, a blue team (BT) may turn off their systems in the middle of the exercise, believing it is advantageous to discard availability-related scores and avoid score loss from possible RT attacks. It is evident that turning off one's systems is not realistic in practice and is unlikely to be the intended content of the exercise.

In addition, BT's overheated competition encourages the application of excessive security policies that are impractical in the real world. In other words, BT's technical knowledge and know-how can be misused as a means of finding and abusing vulnerabilities in scoring systems. Successful ET makes operating a fair evaluation of an exercise challenging, deters training audiences from focusing on the exercise content, and thus lowers the effectiveness of the exercise. Therefore, finding methods to block ET becomes a research topic for the organizers, especially for the green team (GT), which is responsible for developing and operating the technical field of the exercise (cyber range).

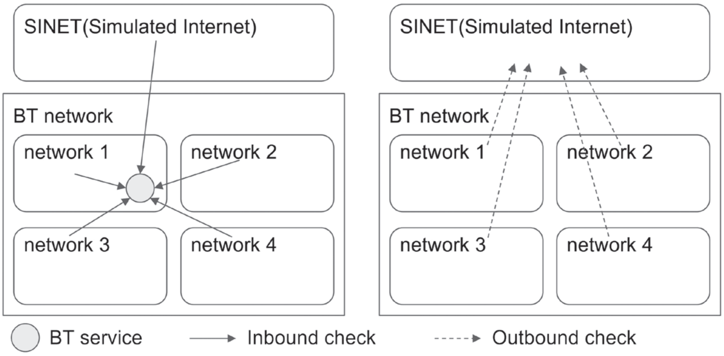
Yamin et al. [1] conducted a comprehensive review of cyber range taxonomies, which can be supplemented by additional related studies [2]–[4]. The paper summarizes unclassified cyber ranges according to architecture, scenarios, capabilities, roles, and tools. There are a couple of ET cases mentioned in the existing literature. Werther et al. [5] mention one example of ET that some teams deleted – the "/bin/rm" command, which caused the failure of a scoring bot's flag rotating task. Considering BT's potential whitelisting of the scoring bot's traffic to a scoring server, the organizers implemented preemptive measures, such as randomizing grading run intervals, adding random content, and random user-agent header strings in the web requests. Pihelgas [6] introduced an availability scoring system and an experience of ET whereby BT tried to trick the system by setting up "fake" services or websites to gain an availability score but avoid a score decrease via RT attack.

2. AVAILABILITY CHECK OVERVIEW

Availability checking can be found uniquely in CDX, in which GT prepares BT infrastructure and periodically monitors the status of the BT services (OK, Warning, Critical, or Unknown) to calculate uptime for scoring.

The availability check can be divided into an inbound check and an outbound check, as shown in Figure 1. The inbound check identifies the availability of a BT service network port, functionalities, user accounts, and network traffic flow (between the source IP address from which the availability check traffic is initiated and the target host's IP). The inbound check could be executed multiple times using different source networks, such as different sub-networks in BT and a simulated internet (SINET), to monitor whether different network routes to the target service are reachable. The outbound check generates traffic from BT hosts by changing their destination domain names or IPs to check whether the traffic can reach SINET.

FIGURE 1: INBOUND CHECK AND OUTBOUND CHECK



The inbound and outbound checks can be divided into usability and availability checks. In the usability check, the user simulation team (UST) intervenes to check whether the target functions operate correctly. Since it is a task that mainly requires an interactive response from the user, the number of usability check targets is proportional to the number of UST members available. In addition, the usability check period should be long enough, considering that exceptions may occur during the manual check process and accounting for the acceptable level of work intensity for UST.

For a more significant number of BT services and more frequent checks for precise measurement, availability checks need to be automated. In the availability check, a script or binary prepared by the GT is executed repeatedly at short intervals to record the check target's state of operation. The availability check in this paper does not assume the use of a separate network (e.g., a management network) for that purpose but checks through BT's network, entailing that BT's network monitoring contains availability check traffic.

While BT concentrates on network monitoring, depending on how availability checks are implemented, the availability check's source IP, execution interval, and scope of functions to be checked may appear distinct from other traffic. If the availability check traffic is easily distinguishable from other traffic, BT's understanding of the availability check process will increase, allowing them to analyze and find the scoring system's weaknesses.

3. EVASIVE TACTICS

This section introduces several evasive tactics depicted in existing CDXs, as well as potential examples. Evasive tactics can take various forms depending on how a CDX is scored, but they can generally be categorized as shown in Table I.

TABLE I: CATEGORIES OF EVASIVE TACTICS

Target score	Availability	Usability	RT's attack
Evasive tactics	Limiting access/function	Degrading service quality	Interfering attack evidence obtainment
Method	A. Whitelisting inbound availability check B. Whitelisting outbound availability check C. Modifying service functionality	D. Adjusting excessive security policy E. Degrading responsiveness	F. Interfering flag obtainment

A. Whitelisting Inbound Availability Check

The inbound checks identify whether the service port is open or further inspect the service's functionality by looking at such actions as login, file download/upload, and custom queries for special systems. Multiple availability checks can be performed on a single service target to determine whether the availability check traffic is reachable from other source networks defined in the exercise rules that BT must follow. The availability check is considered a failure if there is no response within a pre-defined timeout or if the state of the response is not "OK." Any issue with the source IP address, the FQDN (fully qualified domain name) of the destination service, network route, login credentials, or service functionality could result in a check failure.

Finding the cause of the check failure is not always easy in a situation where BT's misconfiguration, GT's configuration error, and RT's attack affect availability are mixed. Therefore, it is normal for BT to try to understand the availability check process to determine the cause of the check failure.

Having BTs with a good understanding of the availability checks has a positive effect by reducing BT's troubleshooting for the checks and an inseparable negative effect that increases the possibility of finding weaknesses in scoring systems. For example, suppose BT investigates availability check traffic and can identify a distinct piece of information. In that case, one can use a whitelisting approach by allowing the identified information to pass the availability check but block RT's access or attack.

Information obtained from network monitoring, such as the continuous use of the same IP, the same check intervals, and any form of packet seen only in availability checks, can help BT's educated guess in locating the availability check traffic. Additionally, a specific user account or function that is frequently used and checked can also be a distinct component of the availability check traffic. BT can apply whitelisting using different security systems like firewalls, IPS, and host-level applications.

B. Whitelisting Outbound Availability Check

Outbound checks examine whether the web traffic from BT's various networks to the SINET (simulated internet within the training network) is reachable. One of the main differences between the outbound and inbound checks is that an agent is required for outbound checks, since the corresponding traffic must originate within BT's host. BT must maintain the agent's process until the exercise ends and ensure the agent's operation to generate the outbound availability checks, for example, by setting an exception so that the antivirus does not block the agent's traffic.

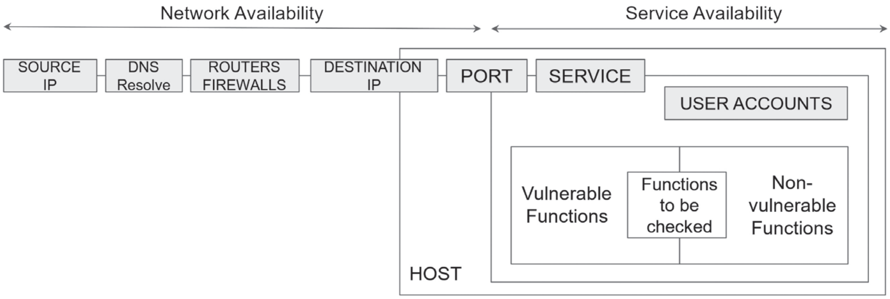
The outbound checks mainly inspect whether the check traffic can reach SINET using FQDN. If the destination FQDN of this traffic is also identifiable or predictable, it can lead to whitelisting, which can be considered safer, as there are many FQDNs prepared and frequently changed for each availability check.

If the agent has all the destinations' addresses to generate the traffic, this list can be easily exposed and used for whitelisting. If authentication is skipped or the server has vulnerabilities, BT will be able to get the list of destination addresses regardless of how a server functions with all the destination addresses.

C. Modifying Service Functionality

The coverage area of inbound availability checks can be divided into network availability and service availability, as shown in Figure 2.

FIGURE 2: SCOPE OF NETWORK AVAILABILITY AND SERVICE AVAILABILITY



Network availability checks examine whether:

1) *DNS queries return the correct destination IP;*
2) *routing has a reachable path to the destination;*
3) *security systems allow the traffic to pass;*
4) *the check target port is open.*

Since network availability checks confirm how alive a target port is, its suitable usage is for checking whether the target host is up.

Service availability checks manage more complex functionalities, such as user account login, sending and receiving email, file uploading and downloading, and browsing web services by using a web browser engine to mimic user behavior. Therefore, service availability checks are recommended for most check targets that have any functions inside.

Suppose a network availability check is incorrectly applied to a service that has further functions. In that case, BT can open a fake service that successfully passes the network availability check to secure the availability score and avoid score deduction from RT attacks. As an example of such an ET, serving static HTML pages or images instead of the original dynamic web pages has been found in several iterations of Locked Shields (LS).

In such an example, ET is possible since parameters in an HTTP request can potentially trigger various vulnerabilities that can be avoided if the web server is configured in a way that does not read the parameters, eliminating the need to search for web vulnerabilities. In other words, a fake daemon with a port opened without any functions can pass the network availability check; however, the RT attack will fail, eliminating the need for BT's defense activities.

Another way to avoid RT's attack is to change the parameter names. In order to attack multiple BTs at the same time for the same attack conditions, RT may have prepared attack scripts with original parameter names to automate the attack processes. These automated attacks assume that the parameter names of the service prepared by GT or RT remain unchanged during the exercise. Yet, BT may attempt to change the parameter names to avoid the automated attacks, assuming that a pre-defined attack requires corresponding parameter names.

Suppose BT collectively changed all HTTP request-related variable names in a web application's source code to other names. If the web application still operates normally after this patch, RT's automated attacks that rely on parameter names will not work. Regardless of the variable names, a usability check involves having the user browse by clicking links on the web user interface. Unless RT identifies the cause of the automated attack failure and resumes the attack manually using updated variable names, the vulnerability may remain intact, preserving the relevant scores.

It is debatable that changing parameter names is one type of ET, as it can temporarily and effectively prevent automated vulnerability scans and attacks in practice. However, since this is a temporary measure and not a fundamental solution, it seems necessary to determine how effective these obscure solutions will be.

D. Adjusting Excessive Security Policy

When a security solution blocks network traffic or processes, it is important for the solution to ensure that the relevant traffic or processes are malicious. Since having a low false-positive rate is critical for the security solution in usability-sensitive environments, such a condition should be applied equally to CDX as much as possible. In terms of usability, measuring the same in CDX is difficult because the number of participants available to simulate user behavior is limited. This means that BT's application of security policies, in consideration of reduced usability but forcing enhanced security, can be an efficient strategy. In other words, it is possible for BT to create a strategy that has no significant impact on usability scores but effectively blocks RT's attack to take advantage of attack-related scores.

For example, BT can set a simple web application firewall script for a web service that loops all parameters in the HTTP request to find a match with any strings containing SQL query keywords and executable names like shell commands, and refuse to process the HTTP request if there is any match. This script can be pre-executed in all web applications (e.g., "auto_prepend_file" in php.ini in the case of PHP). To increase its performance, a pre-process, such as decoding known encodes (e.g., base64), can be added before finding the strings.

Obviously, the above naive approach not only lowers the service's performance, as it uses the web server's resources but also degrades usability by refusing service to a rightful HTTP request that contains non-malicious strings such as "select," "input," "union," "echo," and "cat." In practice, it is absurd to use a usability-degrading firewall. However, in CDX, where usability checks are only performed to a limited extent, it can be an effective filter that passes usability checks but blocks known web application attacks, which becomes a successful ET.

E. Degrading Responsiveness

In reality, keeping a high level of service quality for consumers is essential. The availability and usability scores in CDX account for a large proportion of scores to reflect the importance of quality service. The two scores reflect reliability, but there is another critical factor for measuring the quality of service – responsiveness.

Responsiveness tends to be ignored or considered less important in the evaluations of cyber defense exercises; to the best of our knowledge, no cyber defense exercise reflects the service response time in the score. In most cases, service responsiveness does not significantly impact the exercise content. However, services with deliberately degraded responsiveness can cause inconvenience to other training audiences and, furthermore, become successful ETs. Suppose the availability check timeout is set to 30 seconds, and BT identifies it. Even if BT configures itself to respond to all services after 20 seconds, the availability checks will still pass with ten seconds to spare. However, RT may fail attacks if its timeout is less than 20 seconds, judging an abnormality in the service availability without verifying its actual availability status.

In order to avoid contrived response delays becoming a score-gaining activity in the exercise, service response time should also be taken into account in the score. The average service response time is measured, and the scoring section for the service response time is defined by sufficiently considering a network delay that may not always be constant.

F. Interfering Attack Evidence Obtainment

There are several ways to prove RT's successful attack: capturing a screen, executing a binary file, or obtaining a text-based flag stored in the BT host or service. Having administrative privileges in its system, BT can choose ET to hinder the process of obtaining evidence for the attack instead of conducting defense actions against RT's attack. As a related example, in the 2018 Cyber Conflict Exercise (CCE) [7], where the flags were prepared in the BT host or service, an output command such as "cat" was modified by BT to prevent RT from outputting the flag string. Being suspicious that there was still a vulnerability in the BT host and that the attack was successful (such that one could execute any command in the host except the "cat" command), RT informed the organizers to investigate the situation. The same exercise had another case in which BT registered all flags to IPS to prevent the flags from leaking into outbound traffic. In the above two cases, the organizers would not have known until the end of the exercise if RT had not delved deeply into the issue. This shows that RT serves a necessary role in finding ETs to obstruct the attack evidence acquirement procedure.

4. MITIGATING EVASIVE TACTICS

We introduce several ideas that make it challenging to identify availability check traffic in an environment where BT can monitor its network and system resources. Their primary focus is to increase randomness or reduce the proportion of available traffic to the total traffic, making identification of the availability check traffic a matter of chance.

A. Randomizing IP Addresses and Check Intervals

Randomizing IP addresses and check intervals were introduced by Pihelgas [6] to prevent BTs from identifying availability check traffic. Due to the limitation that off-the-shelf monitoring solutions do not support checking the status of a target service in an evasive manner, such a method will need to be developed separately.

Any form of packet in the check traffic that is identifiable or distinguishable from other traffic can help BT's educated guess to use a whitelisting approach, which allows identified traffic but blocks everyone else in the firewall. In addition, BT may try to compare the frequency of network traffic that includes specific packets, assuming that the frequency of availability check traffic is more frequent than other traffic.

This means that availability check packets need to be configured to look similar to other packets, and the check frequency should also be less than that of other traffic. Finding solutions to avoid the above-mentioned issues requires GT to develop additional functionality or modify existing solutions.

B. Large Number of IPs

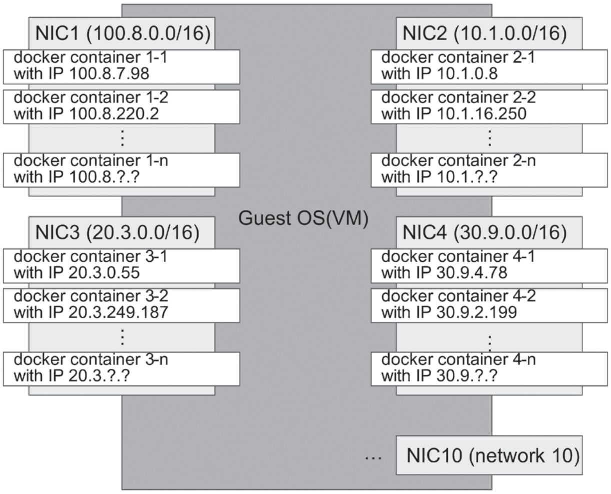
It is necessary to add substantial other traffic in order to make it difficult to identify the availability check traffic, which necessitates having many IPs at the beginning. In the case of VMware, a guest OS allows up to 10 NICs. IPs can be multiplied by adding VMs; however, the number of VMs can be limited by available H/W resources. As many IPs can be required depending on the training content, it is necessary to consider how to secure them while minimizing concerns about H/W resources.

One possible approach is to use Docker's MACVLAN [8] (see Figure 3). With Docker and its MACVLAN, ideally, it is possible to assign as many IPs as the number of Docker containers created in the guest OS. In CCE 2019, for example, GT could run 500 Docker containers with different IPs in a stable manner on a single VM with eight CPU cores and 16GB of memory allocated [9].

The advantage of this method is that it is easy to generate network traffic and maintain relevant network sessions because each container is recognized as an individual host. It is tricky to manage multiple IPs on a single host without having an individual host for each IP, as it requires adjusting the routing table metric to select an NIC when the IPs are in the same network.

The disadvantage is that the above method can only be applied in a network environment that is permitted to be less secure since the vSwitch must allow "promiscuous mode" and "forced transmissions" in its security policy to use MACVLAN.

FIGURE 3: UTILIZING IPS USING A DOCKER CONTAINER AND MACVLAN


NIC1 (100.8.0.0/16)
docker container 1-1 with IP 100.8.7.98
docker container 1-2 with IP 100.8.220.2
docker container 1-n with IP 100.8.?.?
NIC2 (10.1.0.0/16)
docker container 2-1 with IP 10.1.0.8
docker container 2-2 with IP 10.1.16.250
docker container 2-n with IP 10.1.?.?
Guest OS(VM)
NIC3 (20.3.0.0/16)
docker container 3-1 with IP 20.3.0.55
docker container 3-2 with IP 20.3.249.187
docker container 3-n with IP 20.3.?.?
NIC4 (30.9.0.0/16)
docker container 4-1 with IP 30.9.4.78
docker container 4-2 with IP 30.9.2.199
docker container 4-n with IP 30.9.?.?
NIC10 (network 10)

C. Check Coverage of Service Functions

A service consists of many functions. In the case of a web service, there can be login, read and write articles, file upload and download, and many others, which include RT's target functions (RTTFs in short). Checking the availability of RTTF is needed, as those are the basic configurations for the exercise content. However, the naive approach to checking RTTF availability may affect the quality of the exercise.

Suppose the availability check traffic includes only the RTTFs. If the proportion of availability check traffic is noticeably higher than other traffic, the traffic will be sufficient to attract BT's attention. When BT analyses this traffic, they can discover a certain service function that is assumed to be both an availability check and an RT target. BT's successful assumption will significantly lower the scope of its vulnerability analysis.

Including not only RTTF but also non-RTTF in the availability check traffic can be considered for making it a probabilistic problem. While large amounts of traffic with non-RTTF might help conceal availability check traffic, it must be considered that network and host resources are limited. If the service fails to respond to multiple requests, it unintentionally leads to a denial-of-service. Also, GT and BT network monitoring will need faster and larger storage to capture network traffic.

Another temporary countermeasure to this case is to include all functions in the availability check traffic and allow RT or UST to verify the functionality where RT's attack is unsuccessful. Of course, involving RT or UST for availability checks will again be limited to the number of members.

5. DISCUSSION

For a successful CDX, it is necessary to set clear and achievable goals, including feasible procedures to manage operational issues [10]. If ET is suspected during the exercise, evidence proving BT's intentionality is required to link it to a score. However, there is room to consider whether obtaining such evidence is always possible within a limited timeframe.

In addition, usability has different values in reality and in CDX. The difference can make BT's strategy or security policy deviate from reality, undermining the training efficacy of CDX.

A. Finding ET Evidence

Finding clear ET evidence can be a challenging task for GT. Even if ET is suspected, BT may not be directly involved. The BT system is exposed to RT attacks during the exercise, BT or GT mistakes and system errors may be comingled. Therefore, determining the apparent cause of anomalies may be difficult. ET with clear evidence should lead to a score penalty; otherwise, ET must be dispelled if no apparent evidence is found.

Finding anomalies is not a simple task. It is necessary to store and revert snapshots of all systems related to a suspicious moment, and all teams involved must reproduce the same behaviors at the moment of the anomalies to reproduce ET and find evidence of it. Accordingly, the larger the exercise, the more difficult it is to handle the investigation process in an orderly fashion.

If such a verification process is performed in real-time training, then BT may not be able to use related systems, resulting in BT's claim. Therefore, the verification process should be conducted after the end of the exercise. While the evaluation results must be determined in a limited time, restricted manpower and time make the verification process less realistic. Since finding evidence of ETs is an uncertain task requiring significant effort and resources, further research that finds efficient ways to block both known and possible ETs is of great importance.

B. Security and Usability in CDX

Maintaining a balance between security and usability is one of the essential issues in reality, and it is necessary to study whether such a relation is properly reflected in the exercise. This is because a security policy is not easily accepted in reality unless the appropriate level of usability affecting productivity is guaranteed.

Knowing the usability of CDX differs from reality such that BT can construct a secure IT environment by applying unrealistically robust security policies. The robustness of an organization's security policies sometimes directly affects its productivity. For example, if whitelisting is applied to the operating system's app execution, more time and effort will be required to obtain permission to install or run a new app.

More examples are listed below but are not limited to:

1) *network update (Adding/Modifying IP address, VPN, Proxy);*
2) *removable USB device;*
3) *email (attachment);*
4) *install/modify/remove applications;*
5) *network ports accessible from/to SINET;*
6) *shell commands (PowerShell);*
7) *Bluetooth.*

It is necessary to study a list of security policies that affect user productivity and reflect them in the score after evaluation in order to determine whether BT's security policies fall within the range that does not significantly affect productivity.

6. CONCLUSION

Scoring is one of the essential measures that enable exercise evaluation, but it also motivates the use of ET. Successful ET can reduce the reliability of exercise evaluation, hindering the exercise from fulfilling ideal training intentions. Therefore, it is necessary to prepare technical methods to prevent ET.

Availability checking is an essential technique, as interaction occurs actively in the exercise environment (BT system), exercise content (RT attack), and system DevOps (GT), and accounts for a high proportion of training evaluation. The stability and reliability of the availability check are tested in a challenging exercise environment where RT's attacks and BT's defenses occur actively, including potential ET. The availability check has the unique requirement that its checking process needs to be stealthy, but the result should be transparent. Ideally, a successful form of the availability check could be the "panopticon surveillance tower" [11], which creates the illusion of monitoring everything without revealing itself.

Since ET exploits weaknesses in the scoring process, CDXs with different scoring systems may have different ET, meaning that GT's role in scoring becomes more important. The primary training audience of CDX is BT, but BT is also a trainer who tests GT's technical training know-how and skills. As CDX is gaining popularity worldwide and BT's CDX experience increases, evasive tactics are expected to develop, while GT's techniques should also be researched and developed accordingly. The dedication of GT is essential to the technical development of large-scale live-fire CDX. To make the CDX a sustainable festival, it is necessary to maintain an appropriate number of GT members and provide constant attention and support.

In this paper, we introduced BT's ET observed in two CDXs (LS and CCE) with possible examples. Then we proposed a method to use a large number of IPs with low-cost resources as a countermeasure to IP whitelisting. Other countermeasures to ET related to the service function check process and service quality degradation remains for future research.

REFERENCES

[1] M. M. Yamin, B. Katt, and V. Gkioulos, "Cyber ranges and security testbeds: Scenarios, functions, tools and architecture," *Computers & Security*, vol. 88:101636, pp. 1-26, 2019.

[2] J. Almroth and T. Gustafsson, "Cyber range automation overview with a case study of CRATE," in *25th Nordic Conference on Secure IT Systems (Nordsec 2020)*, M. Asplund and S. Nadjm-Tehrani, Eds., Cham: Springer, 2021, pp. 192–209.

[3] M. Sjöstedt, "Monitoring of Cyber Security Exercise Environments in Cyber Ranges," Department of Computer and Information Science, M.S. thesis, Linköping University, 2021.

[4] J. Vykopal, R. Oslejsek, P. Celeda, M. Vizvary, and D. Tovarnak, "KYPO cyber range: Design and use cases," in *Proceedings of the 12th International Conference on Software Technologies*, Madrid, Spain, 2017, pp. 310–321.

[5] J. Werther, M. Zhivich, T. Leek, and N. Zeldovich, "Experiences in cyber security education: The MIT Lincoln Laboratory Capture-the-Flag Exercise," in *4th Workshop on Cyber Security Experimentation and Test (CSET)*, USENIX, 2011

[6] M. Pihelgas, "Design and implementation of an availability scoring system for cyber defence exercises," *14th International Conference on Cyber Warfare and Security (ICCWS)*, pp. 329–337, 2019.

[7] J. Kim, Y. Maeng, and M. Jang, "Becoming invisible hands of national live-fire attack-defense cyber exercise," *2019 IEEE European Symposium on Security and Privacy Workshops (EuroS&PW)*, 2019, pp. 77–84, doi: 10.1109/EuroSPW.2019.00015.

[8] Pipework. 2021. "Software-Defined Networking for Linux Containers." Github.com. [Online]. Available: https://github.com/jpetazzo/pipework

[9] T. Petric. "Running 1,000 Containers in Docker Swarm." Cloudbees.com. 2017. [Online]. Available: https://www.cloudbees.com/blog/running-1000-containers-in-docker-swarm

[10] R. S. Dewar. 2018. "Cybersecurity and Cyber defense Exercises." Center for Security Studies (CSS) Cyberdefense Reports.

[11] "Panopticon." Wikipedia. [Online]. Available: https://en.wikipedia.org/wiki/Panopticon

2023 15th International Conference on Cyber Conflict
Meeting Reality
T. Jančárková, D. Giovannelli, K. Podiņš, I. Winther
2023 © NATO CCDCOE Publications, Tallinn

Towards Generalizing Machine Learning Models to Detect Command and Control Attack Traffic

Lina Gehri
ETH Zurich
Department of Electrical Engineering
and Information Technology
Zurich, Switzerland
lina.gehri@gmail.com

Roland Meier
Cyber-Defence Campus
armasuisse Science and Technology
Thun, Switzerland
roland.meier@ar.admin.ch

Daniel Hulliger
Cyber-Defence Campus
armasuisse Science and Technology
Thun, Switzerland
daniel.hulliger@ar.admin.ch

Vincent Lenders
Cyber-Defence Campus
armasuisse Science and Technology
Thun, Switzerland
vincent.lenders@ar.admin.ch

Abstract: Identifying compromised hosts from network traffic traces has become challenging because benign and malicious traffic is encrypted, and both use the same protocols and ports. Machine learning-based anomaly detection models have been proposed to address this challenge by classifying malicious traffic based on network flow features learned from historical patterns. Previous work has shown that such models successfully identify compromised hosts in the same network environment in which they were trained. However, cyber incident response teams often have to look for intrusions in foreign networks, and we have found that learned models often fail to generalize to different network conditions. In this paper, we analyse the root cause of this problem using five network traces collected from different years and teams of Locked Shields, the world's largest live-fire cyber defence exercise. We then explore techniques to make machine learning models generalize better to unknown network environments and evaluate their accuracy.

Keywords: *machine learning, traffic classification, network security, command and control, Locked Shields*

1. INTRODUCTION

Despite many years of active research, detecting malicious communications from infected hosts in a network remains a challenge. Over the years, attackers have adapted their communication patterns to mimic the protocols and ports of benign traffic, making it difficult to differentiate them in deployed network intrusion detection systems [1]. At the same time, with the wide adoption of HTTPS, network traffic is almost entirely encrypted by default [2], and it is no longer possible for intrusion detection systems such as Suricata, Snort, Bro, or Zeek to analyse the contents in order to look for malicious signatures in the packets' payloads.

As a response, researchers have proposed anomaly detection techniques that use machine learning to identify malicious traffic based on network flow features (cf. surveys in [3,4]). These techniques do not require inspecting the packets' payloads and are thus well-suited for encrypted traffic. However, a major challenge is that available labelled datasets for training are scarce, especially those originating from real environments, because they contain information that the affected organizations do not want to share. Moreover, labelling traffic from real attacks is often impossible due to the lack of ground truth.

One solution to this problem is to use unsupervised learning techniques such as clustering. However, these solutions do not perform well on nonconvex data and are sensitive to initialization and clustering parameters [3]. Another approach is to share machine learning models across networks and use models trained in one environment in order to detect malicious activity in other environments. In this paper, we analyse the feasibility of this approach using five real-world datasets collected from Locked Shields, the largest cyber defence live-fire exercise in the world [5].

First, we analyse the detection performance of command and control (C2) attack flows using machine learning models trained and tested in different environments. We find that models that may work well for a particular environment typically fail to generalize to multiple environments. Second, we investigate the root cause for this effect by analysing which model features work best under which conditions. Then, we explore flow-based and host-based models that generalize to different environments. Our results show that it is possible to train generalized models by carefully selecting time-independent features that are not significantly affected by the environment. However, they also show that training such models is not trivial, and the models generally fail to achieve the same performance as those trained and tested in the same environment.

Our main contribution is the comparison and analysis of supervised machine learning models in five realistic network datasets that include millions of real attack flows generated by security professionals over the course of multiple days at different instances of the Locked Shields cyber defence exercises. Our work presents novel experimentation with new insights made possible using a systematic analysis of these datasets.

2. RELATED WORK

There have been many attempts to exploit machine learning methods for network intrusion detection systems (IDSs), and we refer here to surveys that summarize these techniques. Ahmad et al. [6] and Liu and Lang [3] compare machine learning methods for network-based IDSs and review recent papers on this topic. Khraisat et al. [7] review papers about various kinds of IDSs, and Da Costa et al. [8] concentrate on Internet of Things-related detection. Lastly, Lashkari et al. [9] outline botnet detection methods, including some machine learning-based methods, using various data sources.

Khraisat et al. [10] use the NSL-KDD dataset [11] to compare the classifiers C5, C4.5, SVM, and Naive Bayes. They find that the C5 classifier performs best, with an accuracy of 99.82% and few false positives. Alqahtani et al. [12] compare seven ML-based classification techniques for IDS development using the KDD'99 cup dataset [13]. They find that the random forest model performs best, with an accuracy of 94% and the highest precision and recall score. Jabbar et al. [14] combine a random forest classifier with an average one-dependence estimator to classify traffic. They use the Kyoto dataset [15], and the combined model achieves an accuracy of 90.51% with a false alarm rate of 0.14%.

In this paper, we focus on random forest models trained on Locked Shields datasets to detect malicious flows and hosts. The concept of cyber defence exercises such as Locked Shields is described in [16], and Max Smeets reviews the development and evaluates the achievements of Locked Shields until 2022 in [17]. Our work builds on the work by Känzig et al. [18], which also uses data from Locked Shields to train and test machine learning methods. While their models were developed and tested for only two years of the same team, we generalize the trained classifiers to different years and teams of Locked Shields.

Similar to our work, the authors of [19] and [20] investigate whether it is possible to circumvent detectors of C2 traffic. However, their focus is on the modification techniques that allow circumventing detectors, not on the impact of different environments.

3. LOCKED SHIELDS DATASETS

This section introduces the datasets used in this paper. Locked Shields is a live-fire cyber defence exercise based on realistic scenarios. It is organized once a year by the NATO Cooperative Cyber Defence Centre of Excellence (CCDCOE) [21]. The scenarios involve a cyber incident affecting a fictional country.

Each member nation of the CCDCOE can participate as a Blue Team that assumes the role of the defenders. Blue Teams are typically between 20 and 100 persons, with an average of 40 persons in 2021. The Blue Teams are challenged by a Red Team consisting of professional penetration testers and hackers. The Red Team's goal is to compromise the Blue Team's systems. Attacks include defacing websites, stealing data, denial of service, and compromising hosts by executing malicious payloads [22]. The Red Team uses standard exploitation tools such as Kali Linux [23], Metasploit [24], and Cobalt Strike [25]. The latter is used as the default C2 tool. Custom attacks can be launched if necessary. The whole exercise takes place in Gamenet, which consists of more than 5,000 virtual systems. Every Blue Team is responsible for protecting more than 150 systems over a period of three days. These systems include Linux and Windows machines as well as firewalls, routers, 5G services, drones, industrial control systems, and other systems. To create realistic traffic in the network of the Blue Teams, other teams act as users and use the Blue Teams' services during the whole exercise [22].

We have collected the Locked Shields network traffic of two countries (Country A and Country B) for different years in the form of PCAP files. The network traffic of Country A's Blue Team is from 2017, 2018, 2019, and 2021,[1] and the traffic of Country B's Blue Team is from 2021, resulting in five datasets, as shown in Table I. All datasets are highly imbalanced and heavily skewed towards normal traffic, especially the dataset for 2019, which includes only 0.006% (about 4,000) malicious flows.

In addition, we have auxiliary Red Team activity reports that allow us to label the malicious C2 flows from these PCAP files.

[1] Locked Shields 2020 was cancelled due to COVID-19.

TABLE I: OVERVIEW OF LOCKED SHIELDS DATASETS

Dataset	Size PCAP files	Size CSV files	Number of flows	% malicious
LS17	109 GB	7.1 GB	14,094,546	10.7%
LS18	207 GB	10.7 GB	20,925,882	8.7%
LS19	1.4 TB	34.4 GB	62,955,546	0.006%
LS21A	1.7 TB	24.9 GB	51,699,619	0.5%
LS21B	1.1 TB	19.0 GB	39,903,036	1.1%

4. CHALLENGES OF TRANSFERRING MODELS TO DIFFERENT DATASETS

In this section, we analyse how well a machine learning model trained on Locked Shields data from one year/team performs in detecting malicious flows from another year/team. As a baseline, we consider the machine learning pipeline developed by Känzig et al. in [18].

Model Training

For reasons of space, we analyse only the performance of the best-performing machine learning model developed in [18]. It is a random forest model where the maximum tree depth is 10, and the number of trees is 128. The model was trained with the best 20 features, including custom features and per-flow features extracted using a modified version of CICFlowMeter [26]. (We describe the modifications of CICFlowMeter in Appendix A.) The extracted feature set includes time-based features, such as interarrival times between packets, including average, maximum, minimum, and standard deviation values, as well as time-independent features, such as the number of packets. The best 20 features were selected using a recursive feature elimination algorithm applied to the LS17 dataset. A complete list of the features is provided in [27], and we include a list of the 20 most important features in Appendix B. A flow is defined by its quintuple; it is bidirectional, and the first packet defines the direction.

We trained four separate models on a subsample of 7,000,000 flow instances from the datasets of Country A. To subsample, we randomly sample malicious and normal flows with a ratio of malicious flows as close to 10% as possible. To label the malicious C2 flows, we extract a list of malicious IPs with the help of the Cobalt Strike attack reports from the Red Teams, as suggested by Känzig et al. [18]. Then, we use this list to label the flows extracted by CICFlowMeter. If the source or destination IP is in the list, the flow is labelled as malicious.

Evaluation

We evaluate the models using a fivefold cross-validation of each model to predict the flow labels for all datasets and compare the F1 score of predictions to the labels. The F1 score is calculated as follows:

$$F1 = 2\times \frac{precision \times recall}{precision + recall}$$

With $precision = \frac{TP}{TP + FP}$ and $recall = \frac{TP}{TP + FN}$

where TP, FN, and FP correspond respectively to the true positive, false negative, and false positive rates.

The results are shown in Table II. Training and testing interchangeably on the 2017 and 2018 datasets gave good results, in line with what Känzig et al. obtained. However, the LS17 and LS18 models are bad at classifying flows for Country A's 2019 and 2021 data. Generally, the 2019 dataset performed worst when used for training or testing. Finally, none of the models trained on Country A's data performed well on Country B's data.

TABLE II: F1 SCORES FOR DETECTING C2 MALICIOUS FLOWS

Test data / Training data	LS17	LS18	LS19	LS21A	LS21B
LS17	0.993	0.966	0.007	0.856	0.215
LS18	0.945	0.993	0.060	0.806	0.167
LS19	0.743	0.928	0.791	0.351	0.000
LS21A	0.952	0.918	0.038	0.986	0.158

Possible explanations for the bad performances include the fact that the features were selected using 2017 data only and might not be as relevant for the other years. In addition, the Locked Shields network infrastructure looks different each year, meaning time-dependent features can vary, leading to wrong classifications. Finally, the bad performance of the 2019 model might be because there are only about 4,000 malicious flows, amounting to only a few malicious training instances.

5. CROSS-DATASET FEATURE ANALYSIS AND RANKING

One way to enhance the badly transferable models from Section 4 is by selecting a more suitable set of features. In this section, we use feature elimination and ranking methods on Country A's datasets to select features that generalize better to the different datasets.

Feature Elimination

First, we eliminate irrelevant features across all datasets using the methods described below. A complete list of the eliminated features can be found in Appendix C.

Constant features: We remove the features that are constant over all datasets as they provide no information about whether a flow is malicious or normal.

Feature correlation: Any two highly correlated features contain approximately the same information about the label, and dropping one does not erase information. To remove only features that are inherently correlated and not just because the different gamenets are similar, we include the CIC-IDS2017 dataset [28] in the analysis.

We proceed as follows: First, we calculate the sample Pearson correlation coefficient r for each feature pair of each dataset of Country A and the CIC-IDS2017 dataset and take its absolute value. Then, we find the feature pairs with $|r| > 0.9$ for all datasets. Finally, for each pair, we discard the feature with the lowest relative mutual information (RMI) value with the label.

Relative mutual information: Next, we eliminate all features with less than 15% RMI in all datasets of Country A. The mutual information (MI) between two random variables X and Y measures how much information X contains about Y [29]. We use the sklearn.feature_selection.mutual_info_classif function [30] to calculate the MI between each feature and the discrete label for each dataset. If the feature is also discrete, the function uses the frequencies of the values x and y and the value pairs (x, y) to estimate the probability mass functions. If the feature is continuous, it estimates the MI from k-nearest neighbour statistics, according to [31]. The RMI corresponds to the percentage of uncertainty removed from X when Y is known. It is calculated by dividing the MI by X's entropy:

$$\mathrm{RMI(X;Y)} = \mathrm{MI(X;Y)}/\mathrm{H(X)}$$

We calculate the RMI by dividing each MI score by the entropy of the corresponding year's labels. Finally, we remove the features with less than 15% RMI for all datasets.

Feature Ranking and Selection

Next, we rank the features according to their importance using recursive feature elimination (RFE) and single-feature cross-validation (SF-CV) for each of Country A's datasets.

Feature Ranking with Recursive Feature Elimination

We use RFE to rank the features for each year. We employ the default scikit-learn RFE function [32] on the subsampled LS datasets with 7,000,000 instances. RFE eliminates features one by one. The average ranks of each year's top 10 features can be found in Table III, which shows that features ranked very highly for one year need not rank highly in all the other years. Still, some features are highly ranked for all years; for example, numbers 1 through 10 are, with three exceptions, all ranked in the top 20 for all years. We use these average RFE ranks to choose the features for our random forest models.

TABLE III: RANKS OF THE TOP 10 FEATURES FOR EACH DATASET ACCORDING TO RFE, SORTED BY AVERAGE RANK NUMBER

	No	LS17	LS18	LS19	LS21A
Fwd Pkt Len Max	1	2	4	2	4
Bwd Pkt Len Std	2	5	7	5	6
TotLen Fwd Pkts	3	1	2	16	9
Bwd Pkt Len Max	4	20	6	3	7
Pkt Len Max	5	12	16	10	5
TotLen Bwd Pkts	6	24	3	1	15
Fwd Pkt Len Std	7	3	5	22	16
Pkt Len Mean	8	13	17	11	10
Bwd Pkt Len Mean	9	4	30	4	19
Fwd IAT Max	10	10	12	18	17

Feature Ranking with Single-Feature Cross-Validation

To get a clearer picture of how decisive each feature is, we use SF-CV F1 scores. To compute the scores, we use fivefold cross-validation on an RF model that uses only a single feature and a 7,000,000-instance subsample of a dataset. The average F1 scores over all folds can be found in Table IV for the top 10 features, sorted by average score. For 2017, 2018, and 2021, there are many features with a score higher than 0.9, while for 2019, there is not a single one. Also, only two of the 20 features have a score over 0.1 for 2019.

This could explain the poor performance when using the 2019 data in Section 4 above.

TABLE IV: F1 SCORES OF THE TOP 10 FEATURES FOR EACH DATASET ACCORDING TO SF-CV, SORTED BY AVERAGE SCORE

	No	LS17	LS18	LS19	LS21A	Avg.
Bwd Pkt Len Std	1	0.98	0.99	0.73	0.95	0.91
Pkt Len Var	2	0.98	0.99	0.69	0.96	0.91
Bwd Seg Size Avg	3	0.97	0.99	0.69	0.94	0.90
Bwd Pkt Len Mean	4	0.97	0.99	0.69	0.94	0.90
TotLen Fwd Pkts	5	0.97	0.99	0.66	0.93	0.89
Pkt Len Max	6	0.97	0.98	0.57	0.96	0.87
Fwd Pkt Len Max	7	0.97	0.98	0.57	0.96	0.87
Bwd Pkt Len Max	8	0.98	0.99	0.56	0.94	0.87
Fwd Pkt Len Mean	9	0.97	0.98	0.60	0.89	0.86
Fwd Seg Size Avg	10	0.97	0.98	0.60	0.89	0.86

Time-Independent Features

The last feature selection method is to consider time-independent features only. This includes any features directly influenced by bandwidth changes or packet loss, such as packet interarrival times or byte rates. We expect that these features are most affected by network environment changes. In Appendix D, we provide a ranking of these features using RFE. These time-independent features also have dependencies on network conditions and time, but these are less direct than for time-dependent features.

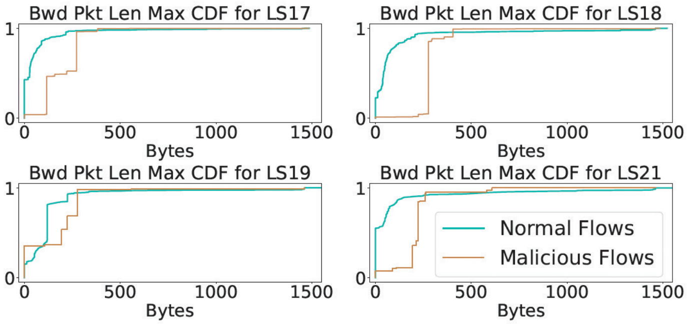
Packet Length-Related Features

In both rankings, all of the top 9 features are related to the packet lengths in a flow. Hence, we analyse one of these features in more depth.

First, we plot the value distributions of Bwd Pkt Len Max representatively for all packet length features as CDFs in Figure 1. Bwd Pkt Len Max is the transport layer payload size of the largest packet in the backward direction. We can see why this feature leads to good performance for most Locked Shields years. The values for malicious flows are clearly higher than the values for normal flows, and many malicious flows have the same feature value. This also explains why the feature does not work as well for 2019, as both CDFs have vertical jumps at 0 and at about 250.

Next, we study some packets to find out why the malicious packets are larger and often have the same size. We stress that the subsequent inferences are primarily based on spot checks and, thus, are not necessarily representative of the entire dataset.

FIGURE 1: DISTRIBUTIONS OF BWD PKT LEN MAX VALUES



We start with the 2018 dataset, as it has the largest vertical jump in the malicious CDF. Almost all communication with a malicious IP is over TLS, implying that the Red Team uses HTTPS connections. While we cannot read the content, we can see exchanges repeating every few seconds. This could be an infected host checking in with a team server. Usually, the team server's answer packet size is 277 bytes, which matches the jump in the CDF. This might be the team server's default answer if there are no new commands. The normal flows with Bwd Pkt Len Max 0 seem to be caused by flows with no backward packets and TCP flows consisting only of zero-length flag packets such as SYNs and ACKs. Therefore, in 2018, there appears to be a lot of beaconing over HTTPS without any new commands.

We also look at the 2019 dataset as its value distributions differ most from the other years. Again, there are many presumed beaconings over HTTPS. The typical answer packets are 223 or 277 bytes in length, which corresponds to two of the jumps in the CDF. It could be that the Red Team is using two Malleable C2 profiles. We can also see more malicious HTTP conversations than in 2018, where the largest packets are

194 bytes. Most malicious flows with zero packet length consist of SYN packets in the forward direction without any answer from the team server. We do not know why the servers were unreachable, but this seems to have been a problem especially in 2019.

6. GENERIC MODELS AND THEIR EVALUATION

In this section, we propose and evaluate two generic model types – a flow-based type and a host-based type – which use a combination of Country A's datasets to select generic features, as explained in Section 5 above.

Flow-Based Models

Description: The goal of the flow-based models is to detect individual malicious flows. The models are:

- Generic, 10 Feat.: a generic random forest model using RFE to select the best 10 features across all features.
- Generic, 10 t.-i. Feat.: a generic random forest model using RFE to select the best 10 time-independent features across all features.
- Generic, 20 Feat.: a generic random forest model using RFE to select the best 20 features across all features.
- Generic, 20 t.-i. Feat.: a generic random forest model using RFE to select the best 20 time-independent features across all features.

Evaluation: We evaluate the flow-based models on Country A's datasets and Country B's dataset to assess their transferability. The F1 scores can be found in Table V.

TABLE V: F1 SCORES OF THE GENERIC FLOW-BASED MODELS WITH 10 OR 20 FEATURES (TIME-INDEPENDENT (T.-I.) OR NOT)

Test data	LS17	LS18	LS19	LS21A	LS21B
Generic, 10 Feat.	0.980	0.991	0.426	0.975	0.116
Generic, 10 t.-i. Feat.	0.985	0.992	0.554	0.971	0.162
Generic, 20 Feat.	0.991	0.992	0.621	0.967	0.135
Generic, 20 t.-i. Feat.	0.992	0.993	0.638	0.989	0.185

First, we look at the results of Country A's datasets. The diversity of the training data leads to better and more consistent results than in Section 4. While the F1 scores for testing on 2017, 2018, and 2021 data fluctuate by a maximum of 0.06, there are more significant differences for the 2019 data. The model that achieves the highest overall F1 scores is the generic model using the top 20 time-independent features. The time-independent features also improve the results for the models using only 10 features, suggesting that models can be generalized by ignoring time-dependent features. Using 20 features generally works better than using 10, which is to be expected, as the model has more data points to make a decision. When we inspect the superior results of the models with 20 features in more detail by looking at precision and recall separately (see Tables VI and VII), we can see that recall is above 0.97 for all models and test datasets, even for 2019, indicating that the models detect a considerable percentage of malicious traffic. However, while precision is always 0.93 or higher for all other years, 0.47 is the highest score for 2019. We must remember that the datasets are very imbalanced; precision would be high even if a model classified all flows as normal. This implies that the models classify many normal traffic flows as malicious in the 2019 dataset.

Unfortunately, the scores for Country B's dataset are all below 0.2. When inspecting precision and recall individually (see Tables VI and VII), we can see that neither is high, though the precision scores are similar to those for the 2019 dataset. Again, this means the models classify many normal traffic flows as malicious. At the same time, the deficient recall scores (below 0.2) indicate that the model also fails to classify genuinely malicious traffic. We can partially explain the problem when we consider the value distribution of Bwd Pkt Len Max for Country B's dataset, which shows that there are many malicious flows with a length of 0. In the CDFs of Country A's datasets (Figure 1), barely any of the malicious flows have a length of 0, which probably means they are classified as normal in Country B's data. On top of that, the non-zero malicious flows consist of far greater packets than any flows in Country A's dataset, making it difficult for the model to classify them correctly.

TABLE VI: THE RECALL OF THE GENERIC MODELS USING 20 FEATURES CHOSEN ACCORDING TO THE RFE RANKING, SELECTING ONLY TIME-INDEPENDENT FEATURES (T.-I.) OR SELECTING FROM ALL FEATURES

Test data	LS17	LS18	LS19	LS21A	LS21B
Generic, 20 Feat.	0.985	0.989	0.975	0.988	0.080
Generic, 20 t.-i. Feat.	0.985	0.989	0.978	0.994	0.114

TABLE VII: THE PRECISION OF THE GENERIC MODELS USING 20 FEATURES CHOSEN ACCORDING TO THE RFE RANKING, SELECTING ONLY TIME-INDEPENDENT FEATURES (T.-I.) OR SELECTING FROM ALL FEATURES

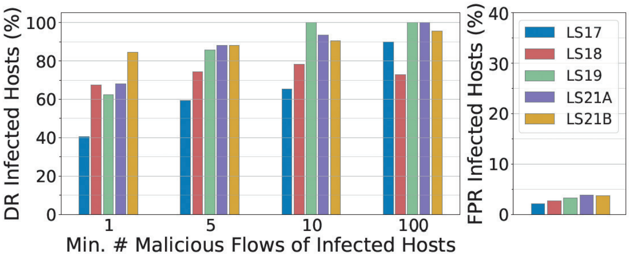
Test data	LS17	LS18	LS19	LS21A	LS21B
Generic, 20 Feat.	0.998	0.994	0.456	0.947	0.420
Generic, 20 t.-i. Feat.	1.000	0.998	0.474	0.985	0.491

Host-Based Models

Description: The purpose of the generic host-based model is to identify infected hosts using the classification of malicious flows. We define infected hosts as IP addresses that are the source IP of at least one labelled malicious flow, and we define detected infected hosts as IP addresses involved in at least $n = \{1, 5, 10, 100\}$ flows predicted as malicious. As the underlying model, we consider the generic model using the top 20 time-independent features, which was the flow-based model with the highest F1 scores in Table V above.

Evaluation: We compare the detected infected hosts to the actual infected hosts to calculate the detection rate (DR) and the false positive rate (FPR). We also determine if infected hosts involved in more malicious flows are more accurately detected. The DR and the FPR of the infected hosts for all datasets can be found in Figure 2.

FIGURE 2: DETECTION RATE (DR) AND FALSE POSITIVE RATE (FPR) FOR ALL DATASETS FOR THE GENERIC MODEL WITH 20 TIME-INDEPENDENT FEATURES



Again, we first consider Country A’s results. For n = 1, the model does not have a very high DR of infected hosts; however, its FPR is below 4% for all years. In absolute numbers, it detected 19 out of 47 infected hosts that were involved in at least one malicious flow in 2017, 33 out of 49 in 2018, 10 out of 16 in 2019, and 15 out of 22 in 2021. The DR improves significantly when we only consider infected hosts that communicate more often ($n > 1$). It is reasonable to consider the DRs for infected hosts with more than five or ten malicious flows, as such hosts tend to be more precarious for a network. They can siphon out more information, act on more commands, or serve as a pivot for other C2 sessions. There were 26 false alarms for a total of 1,193 non-infected hosts in 2017, 33 for 1,233 normal hosts in 2018, 94 for 2,852 normal hosts in 2019, and 135 for 3,553 normal hosts in 2021.

Next, we look at Country B's results. The performances when testing on Country B's dataset are surprisingly good compared with the F1 scores from Table V. The model detects about 33 of the 39 infected hosts with 119 false alarms out of a total of 3,185 normal hosts. These results are as good as and better than those for Country A's datasets. Interestingly, the detection of infected hosts obtained such good results considering that the F1 scores were quite low for the flow-based models. We suspect that while the Red Team used different methods and commands in Country B's case, the initial connection to the team server had comparable network indicators, leading the model to classify these flows as malicious and hence detecting the infected host despite the network conditions being different.

As a result, we can train supervised models on Country A's datasets that can successfully detect a large portion of the infected hosts in Country A's and Country B's datasets with a relatively low FPR, especially when the hosts communicated with a malicious IP multiple times.

7. CONCLUSION

Developing generic machine-learning models that detect malicious traffic in various network environments is challenging. We analysed the flow classification performance of various random forest models depending on the feature selection, model parameters, and training data. We determined that a mix of training data from different environments leads to models vastly outperforming models trained on only one dataset. These mixed models achieve F1 scores over 0.99 when tested on Locked Shields data from Country A's 2017, 2018, and 2021 datasets and over 0.63 for the 2019 dataset. We identified the time-independent features selected by an RFE ranking over all of Country A's datasets as particularly effective in achieving good classification performances. However, we also saw that achieving high scores in completely unfamiliar environments is an open problem for future research.

Further, we demonstrated that models that sum up the number of malicious flows significantly increase the detection rate in Country A's and Country B's networks. Hosts that communicate with a malicious server more than 100 times have an increased detection rate of over 90% and FPR below 4%, even for network environments not used for the training.

REFERENCES

[1] S. Wendzel, S. Zander, B. Fechner, and Ch. Herdin, “Pattern-based survey and categorization of network covert channel techniques,” *ACM Computer Surveys*, vol. 47(3), pp. 1–26, Article 50, Apr. 2015, doi: 10.1145/2684195.

[2] “HTTPS encryption on the web – Google Transparency Report.” Google. 2022. Accessed: Jun. 23, 2022. [Online]. Available: https://transparencyreport.google.com/https/overview

[3] H. Liu and B. Lang, “Machine learning and deep learning methods for intrusion detection systems: A survey,” *Applied Sciences*, vol. 9(20), p. 4396, Oct. 2019, doi: 10.3390/APP9204396.

[4] K. Shaukat, S. Luo, V. Varadharajan, I. A. Hameed and M. Xu, “A Survey on Machine Learning Techniques for Cyber Security in the Last Decade,” in *IEEE Access*, vol. 8, pp. 222310-222354, 2020, doi: 10.1109/ACCESS.2020.3041951.

[5] “Locked Shields.” CCDCOE. 2022. [Online]. Available: https://ccdcoe.org/exercises/locked-shields/

[6] Z. Ahmad, A. S. Khan, Ch. W. Shiang, J. Abdullah, and F. Ahmad, “Network intrusion detection system: A systematic study of machine learning and deep learning approaches,” *Transactions on Emerging Telecommunications Technologies*, vol. 32(1), e4150, Jan. 2021, doi: 10.1002/ETT.4150.

[7] A. Khraisat, I. Gondal, P. Vamplew, and J. Kamruzzaman, “Survey of intrusion detection systems: techniques, datasets and challenges,” *Cybersecurity*, vol. 2(1), pp. 1–22, Dec. 2019, doi: 10.1186/S42400-019-0038-7.

[8] K. A. P. Da Costa, J. P. Papa, C. O. Lisboa, R. Munoz, V. Hugo, and C. De Albuquerque, “Internet of Things: A survey on machine learning-based intrusion detection approaches,” *Computer Networks*, vol. 151, pp. 147–157, 2019, doi: 10.1016/j.comnet.2019.01.023.

[9] A. H. Lashkari, G. D. Gil, J. E. Keenan, K. F. Mbah, and A. A. Ghorbani, “A survey leading to a new evaluation framework for network-based botnet detection,” in *Proceedings of the 7th International Conference on Communication and Network Security*, 2017, pp. 59–66, doi: 10.1145/3163058.3163059.

[10] A. Khraisat, I. Gondal, and P. Vamplew, “An anomaly intrusion detection system using C5 decision tree classifier,” *Lecture Notes in Computer Science*, LNAI vol. 11154, pp. 149–155, Nov. 2018, doi: 10.1007/978-3-030-04503-6_14.

[11] M. Tavallaee, E. Bagheri, W. Lu, and A. A. Ghorbani, “A detailed analysis of the KDD CUP 99 data set,” *IEEE Symposium on Computational Intelligence for Security and Defense Applications, CISDA 2009*, Dec. 2009, doi: 10.1109/CISDA.2009.5356528.

[12] H. Alqahtani, I. H. Sarker, A. Kalim, S. M. M. Hossain, S. Ikhlaq, and S. Hossain, “Cyber intrusion detection using machine learning classification techniques,” *Communications in Computer and Information Science*, vol. 1235 CCIS, pp. 121–131, Mar. 2020, doi: 10.1007/978-981-15-6648-6_10.

[13] “KDD Cup 1999 Data.” UCI. 1999. Accessed: Mar. 10, 2021. [Online]. Available: http://kdd.ics.uci.edu/databases/kddcup99/kddcup99.html

[14] M. A. Jabbar, R. Aluvalu, and S. S. Reddy, “RFAODE: A novel ensemble intrusion detection system,” *Procedia Computer Science*, vol.115, pp. 226–234, Jan. 2017, doi: 10.1016/j.procs.2017.09.129.

[15] Kyoto University. “Traffic Data from Kyoto University’s Honeypots.” Takakura. 2015. Accessed: Mar. 10, 2021. [Online]. Available: http://www.takakura.com/Kyoto_data/

[16] E. Seker and H. H. Ozbenli, “Concept of cyber defence exercises (CDX): Planning, execution, evaluation,” in *2018 International Conference on Cyber Security and Protection of Digital Services (Cyber Security)*, 2018, pp. 1–9, doi: 10.1109/CyberSecPODS.2018.8560673.

[17] M. Smeets, “The role of military cyber exercises: A case study of Locked Shields,” *2022 14th International Conference on Cyber Conflict: Keep Moving! (CyCon)*, Tallinn, Estonia, 2022, pp. 9–25, doi: 10.23919/CyCon55549.2022.9811018.

[18] N. Känzig, R. Meier, L. Gambazzi, V. Lenders, and L. Vanbever, “Machine learning-based detection of C&C channels with a focus on the Locked Shields cyber defense exercise,” *2019 11th International Conference on Cyber Conflict (CyCon)*, Tallinn, Estonia, 2019, pp. 1–19, doi: 10.23919/CYCON.2019.8756814.

[19] C. Novo and R. Morla, “Flow-based detection and proxy-based evasion of encrypted malware C2 traffic,” in *Proceedings of the 13th ACM Workshop on Artificial Intelligence and Security*, Nov. 2020, pp. 83–91, doi: 10.1145/3411508.3421379.

[20] G. Xavier, C. Novo, and R. Morla, Tweaking Metasploit to Evade Encrypted C2 Traffic Detection, 2022, *arXiv:2209.00943*.

[21] "NATO Cooperative Cyber Defence Centre of Excellence." CCDCOE. 2022. [Online]. Available: https://ccdcoe.org/

[22] "Cyber Defence Exercise Locked Shields 2013 After Action Report," CCDCOE, Tallinn, Estonia, 2013. [Online]. Available: https://ccdcoe.org/uploads/2018/10/LockedShields13_ AAR.pdf

[23] OffSec Services. "Kali Linux." Kali.org. 2022. [Online]. Available: https://www.kali.org/

[24] rapid7. "Metasploit Framework." Github.com. 2022. [Online]. Available: https://github.com/rapid7/metasploit-framework

[25] HelpSystems. "Cobalt Strike." Cobaltstrike.com. 2022. [Online]. Available: https://www.cobaltstrike.com/

[26] A. H. Lashkari. "CICFlowMeter." Github.com. 2022. [Online]. Available: https://github.com/ahlashkari/CICFlowMeter

[27] "CICFowMeter Features." Github.com. [Online]. Available: https://github.com/ahlashkari/CICFlowMeter/blob/master/ReadMe.txt

[28] Canadian Institute for Cybersecurity. "Intrusion Detection Evaluation Dataset (CIC-IDS2017)." UNB. 2017. Accessed: Jun. 5, 2021. [Online]. Available: https://www.unb.ca/cic/datasets/ids2017.html

[29] T. E. Duncan, "On the calculation of mutual information," *SIAM Journal on Applied Mathematics*, vol. 19(1), pp. 215–220, 1970, doi: 10.1137/0119020.

[30] "sklearn.feature_selection.mutual_info_classif." Scikit-learn.org. [Online]. Available: https://scikit-learn.org/stable/modules/generated/sklearn.feature_ selection.mutual_info_classif.html

[31] Brian C. Ross. 2014. "Mutual Information between Discrete and Continuous Data Sets." *PLOS ONE*, vol. 9(2), e87357, Feb. 2014, doi: 10.1371/JOURNAL.PONE.0087357.

[32] "sklearn.feature_selection.RFE – scikit-learn 1.1.1 documentation." Scikit-learn.org. [Online]. Available: https://scikit-learn.org/stable/modules/generated/sklearn.feature_selection.RFE.html

APPENDIX

A. CICFlowMeter Tool Modifications

The CICFlowmeter tool[2] extracts flows from PCAP files and exports each flow as a CSV file entry with the 76 CICFlowMeter features and additional metadata, namely source and destination IPs and MAC addresses, the protocol number, and a flow timestamp. The modifications to this tool for this work are:

- preventing memory overflows when processing big PCAPs by regularly flushing to the CSV file;
- adding a new feature Dst IntExt, which has the value 0 if the destination IP address is inside the internal network and 1 if it is outside;
- adding a time filter for PCAPs, making it possible to only process PCAPs from a directory inside a certain time window;
- filtering out flows with TCP SYN count 0, which are flows created by a suboptimal TCP flow tracking logic.

[2] https://github.com/ahlashkari/CICFlowMeter.

B. Top 20 Most Important Features

TABLE VIII: TOP 20 CICFLOWMETER FEATURES FROM [18]

No	Feature
1	Protocol
2	Dst IntExt
3	Flow IAT Max
4	Fwd IAT Tot
5	Subflow Bwd Pkts
6	Subflow Fwd Byts
7	Bwd Header Len
8	Tot Bwd Pkts
9	Fwd Pkt Len Std
10	Fwd Seg Size Min
11	Bwd Pkt Len Std
12	Bwd IAT Mean
13	Active Mean
14	Init Fwd Win Byts
15	FIN Flag Cnt
16	Bwd Pkt Len Min
17	Flow Pkts/s
18	Fwd IAT Max
19	Flow IAT Mean
20	Subflow Fwd Pkts

C. Eliminated Features

TABLE IX: ELIMINATED FEATURES

Constant	High correlation	Low RMI
Bwd PSH Flags	Active Mean	Active Min
Fwd URG Flags	Active Max	Active Std
Bwd URG Flags	Pkt Size Avg	Bwd Blk Rate Avg
URG Flag Cnt	Bwd Byts/b Avg	Idle Mean
	Idle Max	Idle Std
	Fwd Pkts/s	RST Flag Cnt
	Pkt Len Std	Subflow Bwd Byts
	Tot Bwd Pkts	Subflow Fwd Byts

D. Ranking of Time-Independent Features

TABLE X: TIME-INDEPENDENT FEATURES RANKED WITH RFE, SORTED BY AVERAGE RANK

Feature	No	LS17	LS18	LS19	LS21A
Pkt Len Max	1	8	8	2	5
Init Fwd Win Byts	2	1	18	4	1
Fwd Pkt Len Max	3	7	10	9	4
Bwd Pkt Len Std	4	4	17	8	6
Pkt Len Var	5	2	11	17	7
Bwd Pkt Len Max	6	18	14	1	8
Fwd Pkt Len Std	7	3	13	20	10
Pkt Len Mean	8	13	5	15	13
Bwd Header Len	9	9	4	12	23
Init Bwd Win Byts	10	10	19	7	12
TotLen Fwd Pkts	11	12	7	21	9
Bwd Seg Size Avg	12	6	20	11	15

PSH Flag Cnt	13	14	3	25	11
ACK Flag Cnt	14	17	2	19	16
Fwd Header Len	15	22	9	3	21
TotLen Bwd Pkts	16	21	16	6	14
Bwd Pkt Len Mean	17	5	23	13	18
Fwd PSH Flags	18	19	1	22	19
Fwd Seg Size Min	19	20	25	5	17
Fwd Seg Size Avg	20	11	24	14	22
Tot Fwd Pkts	21	24	6	16	26
Fwd Pkt Len Mean	22	16	21	18	24
SYN Flag Cnt	23	25	26	10	20
Down/Up Ratio	24	15	15	26	30
CWR Flag Count	25	29	30	30	2
ECE Flag Cnt	26	27	31	31	3
Fwd Act Data Pkts	27	26	12	28	27
FIN Flag Cnt	28	23	22	27	28
Subflow Fwd Pkts	29	28	28	23	25
Subflow Bwd Pkts	30	32	29	29	29
Pkt Len Min	31	31	33	32	32
Fwd Pkt Len Min	32	34	32	33	31
Bwd Pkt Len Min	33	33	34	34	33

2023 15th International Conference on Cyber Conflict
Meeting Reality
T. Jančárková, D. Giovannelli, K. Podiņš, I. Winther
2023 © NATO CCDCOE Publications, Tallinn

Human-centered Assessment of Automated Tools for Improved Cyber Situational Awareness

Benjamin Strickson
Elemendar
London, United Kingdom

Cameron Worsley
Elemendar
London, United Kingdom

Stewart Bertram
Elemendar
London, United Kingdom

Abstract: Attempts to deploy autonomous capabilities, including artificial intelligence (AI), within cybersecurity workflows have been met with an implementation challenge. Often the impediment is the ability of software engineers to assess and quantify the benefits of machine learning (ML) models for cyber analysts. We present a case study demonstrating the successful testing and improvement of an ML tool through human-centered assessments. For the benefit of researchers in this field, we detail our own wargaming environment, which was tested using members of a government intelligence community. The participants were presented with two cybersecurity tasks: report annotation and a situational awareness assessment. Both of these tasks were statistically assessed for the difference between task completion with and without access to automation tools. Our first experiment – report annotation – showed a task improvement of +14.0 ppts in recall and +9.19 ppts in precision; there was an overall significant positive difference in f1 values for the ML subjects ($p < 0.01$). Our second experiment – cyber situational awareness (CSA) – showed a 66.7% improvement in user scores and a significant positive difference for the ML subjects ($p < 0.01$). The conclusions of our work focus on the need to rebalance the attention of software engineers away from quantitative metrics and toward qualitative analyst feedback derived from realistic wargame testing frameworks. We believe that sharing our wargame scenario here will allow other organizations to either adopt the same testing methodology or, alternatively, share their own CSA testing framework. Ultimately, we are hoping for a more open dialogue between researchers working across the cyber industry and government intelligence agencies.

Keywords: *human-centered AI, cyber situational awareness, autonomous capabilities*

1. INTRODUCTION

Cyber threat intelligence (CTI) analysts are expected to consume multiple reports daily; their aim is to report up the command chain any information that is likely to be of value. In a typical security operations center (SOC), this often involves collating multiple intelligence feeds to compile a list of novel vulnerabilities being used to target victims in the same industry as the organization. In a government intelligence department, this often involves reading reports that track advanced persistent threats (APTs) and their tactics, techniques, and procedures (TTPs). A number of frameworks and taxonomies have been devised to help CTI analysts when they ingest relevant information. The following frameworks allow organizations to communicate and automatically exchange information with each other using the same data structure: Structured Threat Information eXpression (STIX), Malware Information Sharing Platform (MISP), MITRE ATT&CK, and Cyber Kill Chain. Automating the generation of this structured information is an open challenge. For larger organizations with deep pockets, teams of analysts can be employed to read dozens of targeted reports each day. For smaller organizations, or for those who want to consume and sort as much CTI as possible, this requires automation.

For the first CTI challenge we plan to address in this study, automating the translation of unstructured text data into structured data, software engineers commonly examine a blend of rules and ML approaches. The difficulty with relying on ML for cyber information extraction is that the risks associated with errors are significantly higher than in other ML domains, such as the advertising or media industry. One missed cyber campaign targeting your industry has bigger implications than one bad movie recommendation. For the second CTI challenge in this study – generating useful visualization techniques from intelligence – software engineers typically use node-edge diagrams. With thousands of potentially relevant and multi-modal data points, this is often difficult, and while node-edge diagrams have been commonly used, they have rarely been empirically evaluated in published studies. We believe that our work addressing these two software development tasks is especially timely. Automation technologies, and specifically ML, have reached a reasonable state of maturity for cybersecurity; the blocker to further adoption is if the "analyst or downstream decision-maker cannot trust the outputs" [1].

For both of the CTI challenges set out above, we reviewed previously published approaches to user testing. For testing software in the defense domain, a wargame-style user trial is considered one of the better methodologies. There is a lot of variation within the field of wargame design, and they are often uniquely tailored to the end user group and the people who run them. Within the team, we drew on UK government defense experience and attempted to adapt and improve the approach for a cyber context with the ambition of gaining as much feedback as possible from analysts.

Below we summarize the two central outcomes that we are hoping to achieve with our experiments:

1) to establish a human-centered wargaming methodology that maximizes analyst engagement and improves cyber software development outcomes;
2) to empirically test if automated CTI tools improve analyst task completion for report annotation and CSA.

2. RELATED WORK

Human-centered user study assessments for cybersecurity are rarely published. We have also reviewed more general work on situational awareness for reference. Salmon et al. [2] defined seven types of measurement for this; we used the following four in our study: CSA requirement analysis, self-rating techniques, observer rating techniques, and performance measures. Patrick et al. [3] built on this work by developing and validating one of the seven types of measurement, a freeze probe technique, for cyber log analysts. This task is too different from our intended end use case – the longer-term tracking of APTs and TTPs – so we could not use their methodology. We can, however, learn from their conclusions: to develop a task vocabulary familiar to the analysts and adapt the questionnaires to different cyber roles.

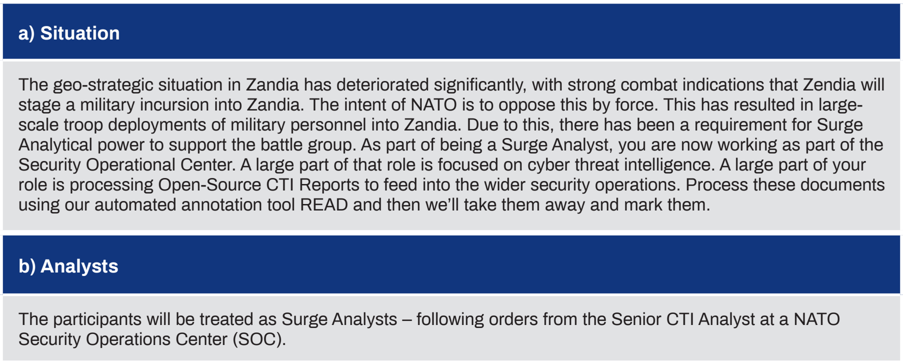
Evaluation studies of automated ML tools for report annotation are numerous, and there are a number relevant to our cyber domain [4], [5]. The problem we have identified, as engineers working to prototype this research, is that your end users must be able to trust the automated outputs. No research, we believe, has to date seriously tackled the issue of trust in ML for cybersecurity and the follow-on consequences for situational awareness. The comprehensive review of the CTI industry by Bouwman et al. interviewed 14 professionals who pay for CTI. They found that actionability, relevance, and confidence were valued over coverage [6]. The review of CTI feeds by Oosthoek et al. concluded that most sources of raw CTI data are undependable [7]. The quantitative review by Griffioen et al. agreed – analyzing 24 CTI feeds for over a year, they found that most indicators are active for at least 20 days before they are listed [8]. The ultimate problem with large quantities of low-quality CTI is that it leads to analyst distrust and can result in decision paralysis [9].

Our second area of focus, automated visualization tools for CSA, had even fewer relevant published studies. Nodal graphs are one solution to the challenge of CSA for CTI reports; they are able to highlight important relationships within a text document and even link multiple CTI documents together. These are sometimes known as cyber knowledge graphs, and there are even a number of commercial providers for this technology. However, we have been unable to find empirical user trial research done on the impact of this technology on situational awareness, and technology companies have little incentive to test and publish their internal assessments and trials.

3. WARGAMING METHODOLOGY

In this section, we will describe the user testing environment or wargame we created to generate task scores and user feedback for our software. A cyber or defense wargame is typically related to real-world infrastructure and actors to provide analysts with an environment that feels familiar. To improve the relatability of our wargame, the Elemendar research team consulted the Laboratory for Analytical Sciences based at North Carolina State University. Working together, we constructed a scenario that was close enough to reality to be familiar to analysts while generalizing some aspects and anonymizing the actors so that analysts would not lean on prior knowledge. We believe that to maximize analyst participation and actionable feedback, a wargame must maximize three design features: it should be topical; it should be immediately familiar through the use of common terminologies; and it should be led in an engaging and gamified manner. In Figure 1, we have reproduced our own wargame scenario that was shown to the analysts.

FIGURE 1: WARGAME SUMMARY SLIDE FOR PARTICIPANTS



a) Situation

The geo-strategic situation in Zandia has deteriorated significantly, with strong combat indications that Zendia will stage a military incursion into Zandia. The intent of NATO is to oppose this by force. This has resulted in large-scale troop deployments of military personnel into Zandia. Due to this, there has been a requirement for Surge Analytical power to support the battle group. As part of being a Surge Analyst, you are now working as part of the Security Operational Center. A large part of that role is focused on cyber threat intelligence. A large part of your role is processing Open-Source CTI Reports to feed into the wider security operations. Process these documents using our automated annotation tool READ and then we'll take them away and mark them.

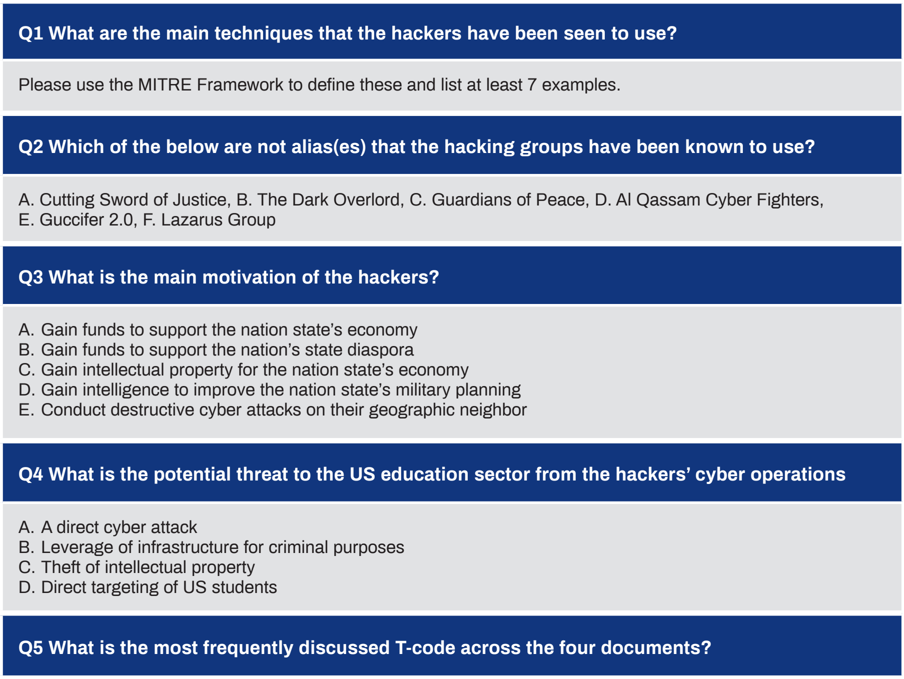
b) Analysts

The participants will be treated as Surge Analysts – following orders from the Senior CTI Analyst at a NATO Security Operations Center (SOC).

For the first task – report annotation – our aim was to find out if an automated annotation tool improves task performance. To construct the CTI reports, we had to consider the following variables: number of annotations, difficulty of annotations, and type of annotations. These were important variables to get right; they would ensure that the conclusions drawn from the data were relevant. The ideal report would give the analyst enough time to find all the entities but not too much time so as to mimic the pressure on real analysts. We decided on thirty minutes to annotate each document and mark it as done, with a five-minute break between reports. Once they had annotated all documents, they were given a questionnaire to complete, providing us with more detail on how they found the tool.

For the second task – situational awareness – our aim was to measure whether or not automated graph visualizations increased the participants' situational awareness when compared to a report-only approach. To measure this, we asked the participants to provide answers to a set of intelligence requirements under the two conditions (Figure 2). For the task, they were provided with a document bundle made up of four open-source CTI reports discussing a nation-state-backed adversary. For condition one, the participants were given 30 minutes to manually read the four documents. Once their time was up, they were asked to answer the intelligence requirements to the best of their ability. For condition two, the participants were given 30 minutes to review automatically produced node-edge diagrams alongside each document (see Figure 3). Again, once their time was up, they were asked to answer the same intelligence requirements to the best of their ability. As well as asking the participants to respond to the questions, we also asked them to provide a document reference to validate their answers. This was to mitigate any external knowledge the participants already possessed of the scenario.

FIGURE 2: INTELLIGENCE REQUIREMENTS SLIDE FOR PARTICIPANTS



Q1 What are the main techniques that the hackers have been seen to use?

Please use the MITRE Framework to define these and list at least 7 examples.

Q2 Which of the below are not alias(es) that the hacking groups have been known to use?

A. Cutting Sword of Justice, B. The Dark Overlord, C. Guardians of Peace, D. Al Qassam Cyber Fighters,
E. Guccifer 2.0, F. Lazarus Group

Q3 What is the main motivation of the hackers?

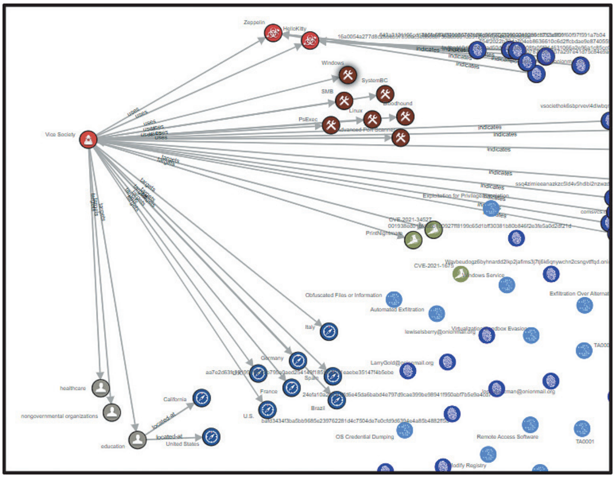
A. Gain funds to support the nation state's economy
B. Gain funds to support the nation's state diaspora
C. Gain intellectual property for the nation state's economy
D. Gain intelligence to improve the nation state's military planning
E. Conduct destructive cyber attacks on their geographic neighbor

Q4 What is the potential threat to the US education sector from the hackers' cyber operations

A. A direct cyber attack
B. Leverage of infrastructure for criminal purposes
C. Theft of intellectual property
D. Direct targeting of US students

Q5 What is the most frequently discussed T-code across the four documents?

FIGURE 3: EXAMPLE STIX NODE-EDGE DIAGRAMS



Recruiting a pool of intelligence analysts is always challenging, given their high workloads. Despite this, it was critical that we got the right mixture of participants to validate any conclusions we would draw. We asked interested parties to complete a short demographic survey so we could gather some information about their background and experience. This demographic survey confirmed that there was a range in age, technical ability, and job specialization. Given the high turnover of analysts within intelligence departments and the relatively junior position of CTI analysts, we were satisfied that the participants were representative. The only assumptions we had were that the participants had basic analytical proficiency and some knowledge of how to use nodal graphs. There were a couple of areas to address ahead of the analyst day: none of these users had used an automated report annotation tool before, and some users had no knowledge of the STIX vocabulary. To give users the same baseline education, we created a video demonstrating how to annotate using the tool and how to identify STIX objects.

4. EVALUATION METHODOLOGY

Empirical user trials for new ML technologies in defense research are rarely published in academic journals, one of the reasons for this is that academics struggle to recruit representative groups for testing. Our domain – government cybersecurity – is no different, and we do not have the financial resources of a large pharmaceutical company or the large user base of a social media company to acquire the sample size that we would like under ideal trial conditions. While acknowledging that our trial could have had more users, we were able to recruit 13 analysts, who were highly representative of our end users as they were drawn from the government intelligence community. We also ensured that the data generated from our trials met any and all criteria required for statistical tests. We used two randomized groups with six and seven individuals under within-subject design conditions. This meant that each group would have two different conditions for both tasks: with automation and without. By having two groups who undergo two different conditions in opposite order, we counteract the possible order effect and minimize transfer learning across conditions.

A. Quantitative Task Metrics

The evaluation metrics we used for the annotation task are considered standard for named-entity recognition (NER) studies: recall, precision, and f1. We have explained these metrics below in non-technical language for reference:

1) Recall – what proportion of all the relevant answers were annotated;
2) Precision – what proportion of all the annotations were relevant answers;
3) F1 – harmonic mean of recall and precision to give a single result.

The evaluation metrics we used for the situational awareness task were question responses, with each question answer being binary. The questions were scored equally, with the exception of question one, which was weighted to account for each correct technique.

We planned to formally quantify the difference between the two treatments for both tasks with a statistical test. The paired sample t-test was chosen, and further analysis showed that our data met the assumptions required for this test: continuous dependent variable, independent observations, and dependent variable normally distributed for both groups. Given the small sample size, a test for normal distribution was important – we used the Shapiro-Wilk test. The results did not show evidence of non-normality for the ML treatment group (W = 0.914, p value = 0.157) or for the non-ML treatment group (W = 0.909, p value = 0.153). Our statistical null hypothesis assumes that there is no difference between users with and without the tool. Our informal expectation gained from our previous experience working with analysts is that we may see an improvement in recall scores, with lower improvements in precision scores. Our experience tells us that human cyber analysts, while slower, are mostly able to match automated labeling models in terms of accuracy of annotation.

B. Qualitative User Interviews

In addition to the quantitative data obtained, we also gathered qualitative data through two approaches. The first questionnaire asked analysts a series of questions focusing on the tool’s automated capabilities, and a second feedback survey gave analysts the chance to discuss their experiences. This allowed us to conduct a thematic analysis of the responses, grouping problem areas into actionable themes for our software development team [10]. This type of feedback has the potential to be significantly more valuable than the task completion metrics. However, it requires that the user interviews are unstructured so as not to prejudice any responses. Importantly, it also requires the analysts to be sufficiently engaged in the exercise so that they have the desire to give detailed responses.

5. RESULTS

The report annotation task results in Table I indicate a noticeable difference in the scores obtained by the two treatments. A paired sample t-test of the f1 scores proved that with very high certainty ($p < 0.01$), we can say there is a significant difference between the groups' f1 scores, and we reject the null hypothesis that there is no difference between the means of the two groups. An unexpected result from this study was the improvement in precision. One suggested explanation of this is the less experienced analysts within the experiment, and our own team of more experienced analysts whose precision would normally equal that of the ML tool.

TABLE I: ANNOTATION TASK RESULTS

Type	Recall (mean)	Precision (mean)	F1 (mean)
ML treatment	0.945	0.891	0.917
No ML treatment	0.829	0.816	0.822

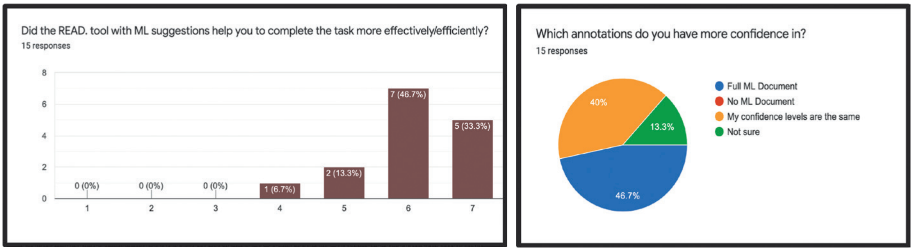
The CSA task results in Table II also indicate a noticeable difference in the scores obtained by the two groups. A paired sample t-test again rejected the null hypothesis that there is no difference between the two group scores. This means that with very high certainty ($p < 0.01$), there is a positive difference for analysts using the graph visualization tools. Examining the results, we find an increase in scores on three out of the five questions (two, three, and four), with the other two remaining the same across both treatments.

TABLE II: CSA TASK RESULTS

Type	Score %
Graph visualization treatment	70.5
No Graph visualization treatment	42.3

The first round of qualitative feedback was conducted immediately following the report annotation task, and we found that the results from the statistical analysis were reflected in the feedback. We first asked the participants to quantify the improvement felt using ML (Figure 4). With above 4 being a positive effect, we can see that most participants noted an improvement in annotation efficiency. Second, we asked the users about their confidence in the ML predictions (Figure 5). The surprising result was that so many analysts (46.7%) were more confident in the READ predictions. One possible factor is the relatively high number of junior analysts on the test. As expected, we found some analysts (40%) who found their confidence levels were the same, meaning they are as confident in their own annotations as they are in READ. Especially positive for our annotation tool was that no analysts were more confident in their own annotations than in READ.

FIGURES 4 AND 5: QUALITATIVE FEEDBACK RESULTS



While these survey questions were useful in providing some quantitative insights, the open-text qualitative feedback provides much-needed context on the participants' experiences. After reviewing the responses to our feedback forms from the participants, we identified the following four themes: modifying incorrect predictions, the ability to trust ML, cross-referencing difficulties, and cascading entity changes. We then selected the first two of these themes to focus on for subsequent software development (Tables III and IV).

TABLE III: ML ERRORS

Themes	Insights
Modifying incorrect predictions	Having to correct ML mistakes was a negative, a “frustrating” experience for users.
1. *“As an annotator I found fixing ML mistakes frustrating. I would rather start with a blank slate [...] All that being said, I understand the value of making the data machine readable and operable so I would use the tool to enable that.”*	
2. *“I really didn’t like having to fix ML recommendations.”*	
3. *“The ML is handy in some aspects, but when it’s off on something multiple times it’s a bit annoying to reject and then re-highlight each of the items.”*	

TABLE IV: CONFIDENCE IN ML

Trust in ML	Users lose trust in ML when they have to correct ML mistakes, even if those mistakes are small. They also feel the need to vet ALL ML suggestions, which can take more time than manually annotating.
1. *“Doc. A (which was Full ML for me) took me more time to annotate than the non ML Doc B. I think because I was less confident in verifying the objects. Subsequently, I was a little more confident in identifying objects on the second document and therefore also a little faster even though it was non ML.”*	
2. *“I felt I had more control in the unannotated document, but I achieved the same results much faster using the ML tagged document. Got hung up a bit tagging ATT&CK techniques in my first doc.”*	
3. *“[H]ow does an analyst have enough confidence that there were no other names that were missed by ML (especially without manually also reviewing the source data)?”*	

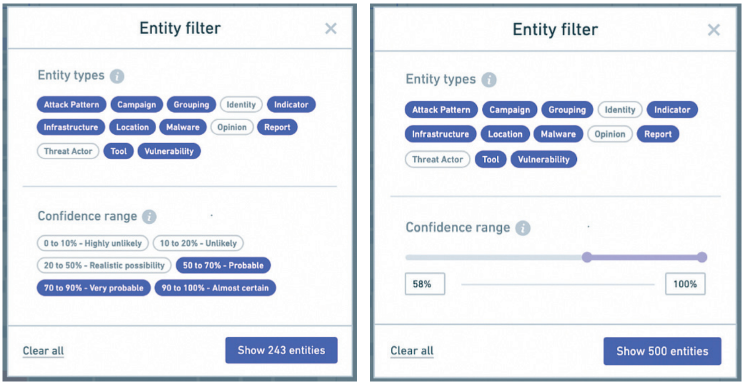
Using this feedback, the next step was to agree on future product development ideas. In the next section, we draw on these two development themes, using the feedback to inform future plans. We also needed to consider re-running these experiments in the future; we designed the next steps with this in mind, considering how the proposed functionality could be tested with users in a repeat trial.

6. IMPACT

One of our stated aims for this project was to improve software development outcomes for automated tools that undergo user prototyping. This section sets out our own company’s attempt to capitalize on the trial results. The objective here is not to present an ideal software design process; instead, we encourage others working in industry and academia to better link published technology research with software outcomes. The first insight we gained from the user feedback was that the lack of model explainability led to analyst dissatisfaction. This was not explicitly stated by the analysts, but it became clear from the feedback; we noticed that when a model predicted an entity incorrectly, this led to users expressing their frustration. If an ML model is wrong enough times, then there is a very real possibility that analysts will not return to using the tool. This led to the first new product feature, visualizations of our model confidence scores. We attempted to improve model explainability by communicating to analysts the uncertainty in our ML models. The intended outcome is the management of analysts' expectations; if analysts can be taught to expect errors in the ML process, then we can reduce user frustration and keep them using the tool for longer.

The design choices for confidence scores needed consideration; when displaying ML confidence, we needed to decide whether they should be a continuous or discrete measurement (see Figure 6). The standard approach for the ML team was a continuous measurement of confidence between zero and one; this could be displayed as a percentage, which the analysts would easily interpret. While this metric felt intuitive, the feedback from our analyst team members and government contacts was that it introduced confusion regarding the appropriate score that should trigger an accept or reject decision. Taking this feedback onboard, the team produced a discrete measure of confidence using semantic buckets. The perceived benefits of such buckets are that they allow analysts to collaboratively agree on thresholds more easily and that the difference between semantic buckets was more interpretable than the difference between integers (e.g. 67% and 77%).

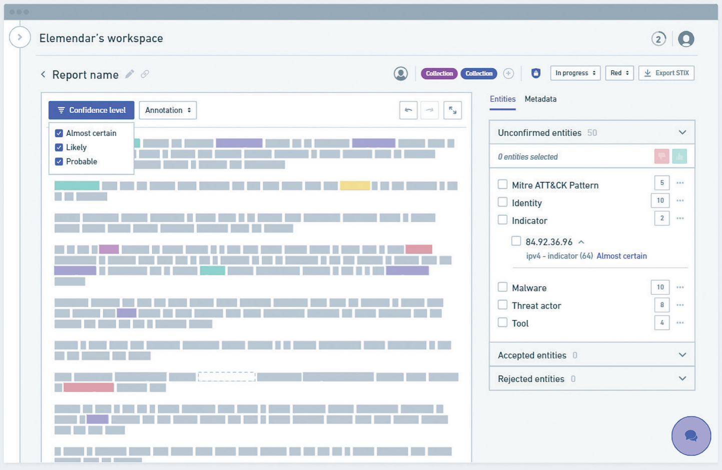
FIGURE 6: POTENTIAL DESIGN VERSIONS OF THE CONFIDENCE SCORE FEATURE



The second insight we gained from the feedback was the need for a method to ignore highly certain predictions, leaving only the uncertain predictions for the analysts to focus on. This led us to design our second new product feature to improve explainability

– ML confidence filters. We pushed this idea to our design team, who were tasked with designing an entity filter button that would allow the users to only see the predictions which had a user's chosen level of confidence. The design decisions for this feature focused on the functionality of the filter. Initial discussions had outlined an alternative approach where analysts could change the models to only extract entities with certain levels of confidence. This approach was highly restrictive and would only be appropriate where there had been a senior-level decision that analysts should not be allowed to see entities with low confidence. The approach we decided on shows all model predictions, allowing the analysts to filter as they work (Figure 7).

FIGURE 7: CONFIDENCE FILTER FEATURE FINAL IMPLEMENTATION VERSION



7. CONCLUSION

In order to retain and improve the confidence that users have in automated tools, either ML models need to achieve near-perfect results or the general approach to software development needs to be adapted. It is clear from our study that there is likely to be significant user frustration when models underperform. Our view on this frustration is that no ML model is going to be right all of the time and that, while improving model predictions should be an ongoing effort, there are other more important measures to consider. We believe there are two realistic ways to add explainability to automated text tools: feature importance and model confidence. While the second approach can only provide limited insight into the ML predictive process, the significant benefit comes from better interpretability, less subjectivity, and increased functionality, such as filtering predictions by confidence scores.

A number of participants found our own ML good for predicting the more obvious entities that appear en masse in CTI documents (e.g. more numerous classes such as malware and indicators), while they expressed lower confidence in the harder classes (e.g. ATT&CK patterns). Our conclusion for cyber ML tools is that, while we may find certain subclasses more difficult to predict, this does not discount the value proposition of an automated annotation tool. If the tool can do the heavy lifting and, importantly, can do it well, that will give analysts the confidence to automatically accept the easier entity predictions and focus on the harder, more subjective entity extraction. Some of the improvements that we felt could have been made to our study are as follows: a larger analyst sample would have been desirable from a frequentist statistical perspective. Second, we focused solely on task scores; to robustly quantify the difference made to analyst efficiency, we would have required additional resources to conduct a more detailed screen recording analysis.

In conclusion, we have established the benefits of an automated tool for cyber analysts on both a cyber annotation task and a situational awareness task. We have shown statistically that pairing analysts with ML tools quantifiably improves task performance. The other significant finding from this work has been the analyst-machine trust dynamic. We received detailed feedback from a number of analysts stating that the lack of trust in predictions would be a blocker to further usage. To gain this level of feedback, it was necessary to design a wargame that was topical, gamified, and immediately understandable for analysts. Our hope is that our trial design process and subsequent development steps can encourage others in the cyber ML field to conduct their own engaging wargame scenarios, gain better insights, report statistically robust user testing, and share software development impacts with the research community.

ACKNOWLEDGMENTS

We would like to thank the Laboratory of Analytical Sciences at North Carolina State University for their support throughout 2021 and 2022. This includes a number of individual analysts who were important contributors to the successful delivery of our wargame.

REFERENCES

[1] S. L. Dorton and S. Harper, “Trustable AI: A critical challenge for naval intelligence,” Center for International and Maritime Security (CIMSEC), White Paper, 2021. [Online]. Available: https://cimsec.org/trustable-ai-a-critical-challenge-for-naval-intelligence/

[2] P. M. Salmon et al., “Measuring Situation Awareness in complex systems: Comparison of measures study,” *International Journal of Industrial Ergonomics*, vol. 39(3), pp. 490–500, 2009.

[3] P. Lif, M. Granåsen, and T. Sommestad, "Development and validation of techniques to measure cyber situation awareness," presented at the International Conference On Cyber Situational Awareness, Data Analytics And Assessment, IEEE, 2017.

[4] S. Dasgupta *et al*., "A comparative study of deep learning based named entity recognition algorithms for cybersecurity," presented at the IEEE International Conference on Big Data, IEEE, 2020.

[5] H. Wu, X. Li, and Y. Gao. "An effective approach of named entity recognition for cyber threat intelligence." in *Proceedings of the IEEE 4th Information Technology, Networking, Electronic and Automation Control Conference (ITNEC)*, pp. 1370–1374, 2020.

[6] X. Bouwman *et al*., "A different cup of {TI}? The added value of commercial threat intelligence," presented at the 29th USENIX Security Symposium, 2020.

[7] K. Oosthoek and C. Doerr, "Cyber threat intelligence: A product without a process?" *International Journal of Intelligence and Counterintelligence*, vol. 34(2), pp. 300–315, 2021.

[8] H. Griffioen, T. Booij, and C. Doerr, "Quality evaluation of cyber threat intelligence feeds," presented at the *International Conference on Applied Cryptography and Network Security*, Springer, 2020.

[9] D. Schlette *et al*., "Measuring and visualizing cyber threat intelligence quality," *International Journal of Information Security*, vol. 20.1, pp. 21–38, 2021.

[10] V. Clarke, V. Braun, and N. Hayfield, "Thematic analysis," in *Qualitative psychology: A practical guide to research methods*, pp. 222–249, 2015.

2023 15th International Conference on Cyber Conflict
Meeting Reality
T. Jančárková, D. Giovannelli, K. Podiņš, I. Winther
2023 © NATO CCDCOE Publications, Tallinn

Leveling the Playing Field: Equipping Ukrainian Freedom Fighters with Low-Cost Drone Detection Capabilities

2LT Conner Bender
Postdoctoral Researcher
Tandy School of Computer Science
University of Tulsa
Tulsa, OK, United States
conner@utulsa.edu

Jason Staggs
Adjunct Assistant Professor
Tandy School of Computer Science
University of Tulsa
Tulsa, OK, United States
jason-staggs@utulsa.edu

Abstract: The unprecedented conflict in Ukraine has seen heavy use of asymmetric warfare tactics and techniques, including the use of drones. In particular, Da-Jiang Innovations (DJI) drones have played a major role in the conflict, supporting tactical military operations for both opponents by providing reconnaissance and explosive ordnance across the battlefield. The same drones have also been leveraged to provide humanitarian aid across Ukraine. However, Ukraine has publicly accused DJI of helping Russia target Ukrainian civilians by allowing Russian military forces to acquire and use a proprietary DJI drone-tracking system called AeroScope. This system has allowed Russian forces to geolocate and target Ukrainian civilians piloting DJI drones, which has often led to kinetic strikes against drone operators. Modern DJI drones beacon telemetry and remote identification information that allows the AeroScope system to identify and track the drone and operator at ranges of up to 30 miles away. Cost and ease of access are the primary factors that have hindered Ukraine's ability to counter this threat with AeroScope systems of their own to identify and locate DJI drones and operators used by Russia. This has provided an asymmetric advantage to Russia on the battlefield. Although cybersecurity researchers have demonstrated that DJI drone identification wireless datalinks are unencrypted, it remains a mystery how to collect and decode these signals over the air in real time using low-cost and widely available software-defined radios. This paper addresses the problem by reverse engineering DJI drone identification signals and message structures to detect drone IDs over OcuSync and Enhanced Wi-Fi datalinks. A functioning open-source prototype is detailed that can detect DJI OcuSync drones using two HackRF One software-defined radios. The methodology can easily be adopted by others to rapidly assemble and deploy low-cost DJI drone and operator detection and geolocation systems that are functionally similar to the AeroScope system.

Keywords: *drone cybersecurity, geolocation, software-defined radio, DJI drones, RF reverse engineering*

1. INTRODUCTION

Throughout the Ukraine conflict, both opponents have widely used guerrilla warfare methods, including small off-the-shelf unmanned aerial vehicles or drones to support military objectives. Small drones have been leveraged for numerous purposes on the battlefield to conduct reconnaissance operations, deliver explosive ordnance, aid in search and rescue missions, and provide humanitarian assistance [3], [4].

Unfortunately, most of these small drones have been plagued by an unexpected feature known as "remote identification," which automatically beacons the location of drones and their corresponding operators. According to numerous public reports, this feature has been abused by Russian forces using a system called AeroScope by Da-Jiang Innovations (DJI) to target and kill innocent Ukrainian civilians. Cost and ease of access hinder Ukraine's ability to counter the threat of AeroScope monitoring systems targeting Ukrainian drone operators with AeroScope systems of their own, thus providing an asymmetric advantage to Russia on the battlefield [3], [4].

To counter the cost-prohibitive nature of DJI's proprietary AeroScope system and the difficulty of acquiring one, we propose an open-source alternative using inexpensive software-defined radios (SDR) and components. Our solution builds on the research of others by reverse engineering the OcuSync wireless datalink and integrating a solution using inexpensive and widely available SDRs to reliably decode DJI identification beacons [1], [2].

DJI originally claimed that their identification beacons were encrypted. However, this was disproven by cybersecurity researcher Kevin Finisterre, who demonstrated that a Mavic Mini SE drone ID beacon could be detected in plaintext via an Enhanced Wi-Fi datalink [5]. We build upon the work of Finisterre by focusing on reverse engineering signals and message structures used by the DJI OcuSync protocol. Furthermore, we validate that neither Enhanced Wi-Fi nor DJI OcuSync wireless datalinks are encrypted. In addition, this paper discusses the three types of drone ID packet structures: license, flight information version 1, and flight information version 2. We demonstrate that these packet types can be recognized and decoded by the DJI OcuSync detection system we have developed. The detection system is equipped with two HackRF Ones and runs a web application responsible for displaying real-time detection data. Lastly, we show how to achieve DJI Enhanced Wi-Fi detection and extend such detection capabilities to non-DJI Wi-Fi drones like Parrot.

The methodology described in this paper can easily be adopted by others to rapidly assemble and deploy low-cost DJI drone detection and location systems that are functionally similar to the AeroScope system. The results of this work have the potential to negate the asymmetric advantages afforded to opponents leveraging AeroScope systems.

2. BACKGROUND

Tensions between Russia and Ukraine have steadily escalated since 2014, when Russia invaded and subsequently annexed Crimea. Russia continued to carry out escalatory actions with its unprovoked cyberattack on a key portion of the power grid in Ukraine, leaving more than 230,000 civilians without power. On February 24, 2022, Russian military forces conducted an unprecedented invasion of Ukraine on a scale that had not been seen in Europe since World War 2. Thus far, the invasion has resulted in countless casualties, including the deaths of innocent civilians. As of this writing, the Russian invasion has displaced over 7.6 million Ukrainians from their homes, leading to a massive refugee crisis. The Russian military campaign has been multi-dimensional, with attacks being waged via land, sea, air and cyberspace. Asymmetric warfare tactics have also been used extensively on both sides, including the use of drones [3], [6].

One of the key enablers of asymmetric warfare in the Russia-Ukraine war has been the Chinese drone technology company Da-Jiang Innovations (DJI), headquartered in Shenzhen, China. DJI essentially monopolizes the off-the-shelf drone market with approximately a 70–80% market share [7]. Since the invasion of Ukraine in February 2022, multiple accusations have been made by Ukrainian government officials that DJI has been showing levels of favoritism to Russia [4], [5]. Furthermore, Ukraine has publicly accused DJI of allowing Russia to target innocent civilians with missiles using its AeroScope drone monitoring technology [4], [5].

AeroScope is the equivalent of a radar system for detecting DJI drones. The identification information shares similar functional characteristics with ADS-B, which is used in the aviation industry. The system can be deployed to monitor for the presence of DJI drones up to 30 miles away. DJI AeroScope was originally intended to be used for public safety purposes if a rogue drone was illegally flying in a restricted area, such as near an airport or other drone-protected area. The system could be used to find the drone and the individual operating it. Additionally, the AeroScope system is only supposed to be sold to law enforcement and security agencies that are actively engaged in such efforts to protect the public [4], [5].

By default, modern DJI drones automatically beacon telemetry and remote identification information revealing information about the drone, including the GPS coordinates of the operator, once every second. One of the reasons DJI began integrating this technology into their drone platforms was to comply with the U.S. Federal Aviation Agency's (FAA) remote identification requirements for small drones [8]. The FAA has outlined a list of remote identification requirements for small drones and has stated that after September 16, 2023, drones that do not comply with the remote identification reporting requirement must not operate in federal airspace [1]. Table I shows the minimum required information elements that must be included in broadcast messages.

TABLE I: REMOTE IDENTIFICATION REQUIREMENTS OF DRONES

Elements	Performance
Drone serial number and/or session identifier	Rate of one message per second
Controller latitude and longitude	±100 feet with 95% probability
Controller geometric altitude	±15 feet with 95% probability
Drone latitude and longitude	±100 feet with 95% probability
Drone geometric altitude	±150 feet with 95% probability
Drone velocity	Rate of one message per second
Timestamp	Synchronized with all other elements
Drone emergency status	On/Off

DJI drones use one of two proprietary communications protocols for command-and-control and broadcasting ID information: (i) Enhanced Wi-Fi or (ii) OcuSync [8].

I. *Enhanced Wi-Fi*: The Enhanced Wi-Fi protocol is used by older DJI Spark and Mavic Air models. The protocol transmission range is limited to a visual line of sight.

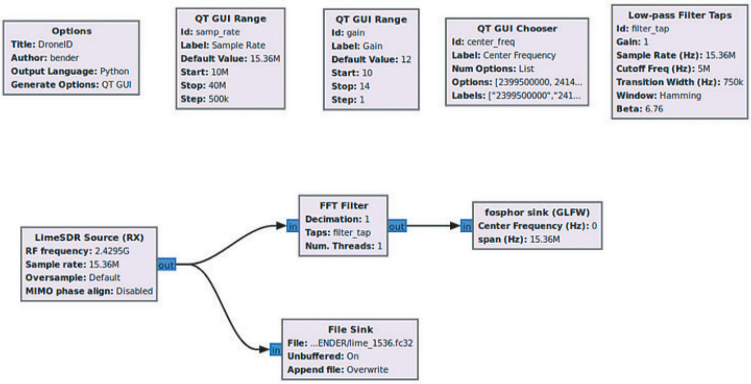
II. *OcuSync*: The OcuSync protocol is used by the DJI Mavic series, Air series and Mini series of drones. This new DJI protocol, which leverages SDR radio technology, has a protocol transmission range of approximately 2.5 miles.

Efforts have been made to reverse engineer DJI communication protocols. Department 13, a company specializing in drone countermeasures, discovered operational details about drone remote identification packets by examining an accidental release of DJI's Mavic Pro drone firmware. The discovery is documented in a white paper that provides the only publicly available information about DJI remote identification frames sent with Enhanced Wi-Fi drones [1].

3. OCUSYNC DRONE IDENTIFICATION AND DEMODULATION METHDOLOGY

In 2022, an open-source effort was initiated to demodulate DJI OcuSync drone ID signals [2]. The repository revealed that OcuSync drone IDs are loosely based on long-term evolution (LTE) cellular standards. This means any SDR capable of sampling up to 15,360,000 samples per second (the LTE sample rate) and up to a bandwidth of 10 MHz is capable of capturing OcuSync drone IDs. Figure 1 shows a GNU Radio flow graph leveraging a LimeSDR to see DJI drone ID signals in real time (via a phosphor sink). Any SDR capable of sampling 32-bit floating point IQ data (.fc32 file) up to 15.35 MSPS can use this GNU Radio flow graph to capture drone ID signals. Figure 2 shows the waterfall output from the GNU Radio flow graph with a drone ID message highlighted.

FIGURE 1: GNU RADIO FLOW GRAPH FOR LIMESDR TO CAPTURE DRONE IDS



Options
Title: DroneID
Author: bender
Output Language: Python
Generate Options: QT GUI

QT GUI Range
Id: samp_rate
Label: Sample Rate
Default Value: 15.36M
Start: 10M
Stop: 40M
Step: 500k

QT GUI Range
Id: gain
Label: Gain
Default Value: 12
Start: 10
Stop: 14
Step: 1

QT GUI Chooser
Id: center_freq
Label: Center Frequency
Num Options: List
Options: [2399500000, 2414...
Labels: ["2399500000","241...

Low-pass Filter Taps
Id: filter_tap
Gain: 1
Sample Rate (Hz): 15.36M
Cutoff Freq (Hz): 5M
Transition Width (Hz): 750k
Window: Hamming
Beta: 6.76

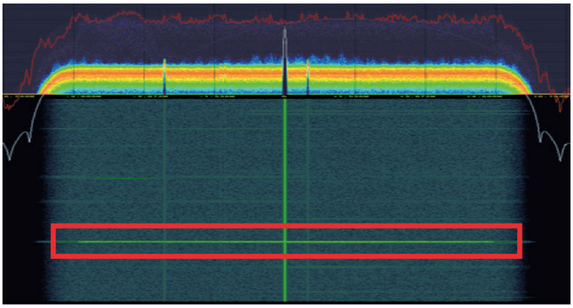
LimeSDR Source (RX)
RF frequency: 2.4295G
Sample rate: 15.36M
Oversample: Default
MIMO phase align: Disabled

FFT Filter
Decimation: 1
Taps: filter_tap
Num. Threads: 1

fosphor sink (GLFW)
Center Frequency (Hz): 0
span (Hz): 15.36M

File Sink
File: ...ENDER/lime_1536.fc32
Unbuffered: On
Append file: Overwrite

FIGURE 2: WATERFALL DISPLAY HIGHLIGHTING CAPTURED DJI DRONE ID



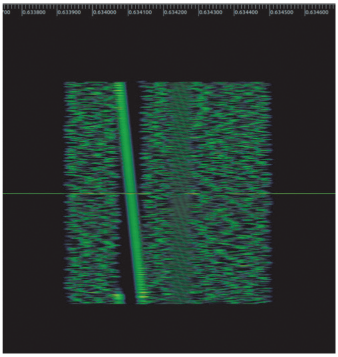
Given a capable SDR, a successful capture must be tuned to center frequencies broadcasted by drone ID signals. Table II shows center frequencies in the 2.4 GHz and 5.8 GHz frequency bands, on which drone IDs are found. This data was collected by observing the 2.4 GHz and 5.8 GHz bands while several DJI Mavic models were powered on. The 2.4 GHz frequency band ranges from 2,399.5 MHz to 2,474.5 MHz, and the 5.8 GHz frequency band ranges from 5,741.5 MHz to 5,831.5 MHz. During testing, we also discovered that even if a user forces the DJI OcuSync communications downlink to be 2.4 GHz or 5 GHz with the DJI GO smartphone application, DJI drone ID signals do not adhere and continue to broadcast out-of-band from the communications link.

TABLE II: 2.4 GHz CENTER FREQUENCIES AND 5.8 GHz CENTER FREQUENCIES FOR DJI DRONE IDS

2.4 GHz Frequency Band	5.8 GHz Frequency Band
2,399.5 MHz	5,741.5 MHz
2,414.5 MHz	5,756.5 MHz
2,429.5 MHz	5,771.5 MHz
2,444.5 MHz	5,786.5 MHz
2,459.5 MHz	5,801.5 MHz
2,474.5 MHz	5,816.5 MHz
	5,831.5 MHz

During testing, it was observed that a signal broadcasts on a frequency until 12–20 drone IDs are transmitted before hopping to a different frequency. Figure 3 shows a sample DJI Mavic Pro drone ID signal. A drone ID signal is approximately 600 milliseconds in duration.

FIGURE 3: SAMPLE DRONE ID SIGNAL FROM A DJI MAVIC PRO CAPTURED ON 2,429.5 MHz



A single drone ID signal contains nine orthogonal frequency-division multiplexing (OFDM) symbols. Although in some instances drone ID signals contain eight OFDM symbols. However, demodulation only needs eight OFDM symbols, as the first symbol in a nine OFDM system can be discarded. The first and last OFDM symbols are 80 milliseconds in size, and the middle symbols are 72 milliseconds (e.g., [80, 72, 72, 72, 72, 72, 72, 72, 80]). Once captured, DJI drone ID signals can be demodulated using digital signal processing. The dji_droneid GitHub repository outlines the following demodulation steps [2]:

1. Identifying the start of a drone ID
2. Creating a low-pass bandwidth filter
3. Applying a coarse frequency offset correction
4. Extracting OFDM symbols (sans cyclic prefixes)
5. Measuring channel impulse rate
6. Quantizing quadrature phase shift key (QPSK) into bits
7. Descrambling bits
8. Turbo decoding and rate matching
9. Deframing bytes

The source code in the GitHub repository is written in MATLAB/Octave, and the captures used an Ettus B205-mini SDR. The SDR retails for around $1,345, and the existing codebase is incompatible with inexpensive and easier-to-come-by SDRs such as the HackRF One. In this work, the demodulation steps were ported over to Python. The proceeding figures contain algorithms outlining each step of the demodulation process in Pythonic pseudocode.

Start of Drone ID

Figure 4 shows the algorithm for generating Zadoff-Chu (ZC) sequences with a root index and sequence length [9]. In a drone ID, there are two OFDM symbols with ZC sequences, symbols 4 and 6, in it. The rootIndex for OFDM symbol 4 is 600, and the rootIndex for OFDM symbol 6 is 147. The seqLen is 601 because the formula only works for an odd number of samples. The middle sample (300) is removed after computation. The sequence is then applied to the data carriers (in buffer). The buffer shifts to have the zero-value placed in the center, and an inverse Fourier transformation then occurs. The result of the algorithm yields a 600-long ZC sequence shifted with a root index (zadoffChuSeq).

FIGURE 4: GENERATION OF ZADOFF-CHU SEQUENCE

```
Input: rootIndex: Root index of ZC symbol
Input: seqLen: Length of ZC sequence
Output: zadoffChuSeq: ZC sequence
dataCarrierIndices ← [$i$ = 212…813, $i \neq 512$]
n ← [$i$ = 0…600]
seq ← $e^{\frac{-1j \times \pi \times \mathrm{rootIndex} \times n \times (n+1+2.\times 0)}{\mathrm{SeqLen}}}$
seq ← delete(seq, 300)
buffer ← zeros(1024)
buffer [dataCarrierIndices] ← seq
zadoffChuSeq ← invFFT(shiftToCenter(buffer))
```

Figure 5 depicts an algorithm performing a normalized cross-correlation that finds the ZC sequence in OFDM symbol 4 (zc4) among 32-bit floating point IQ data (iqData). This is accomplished with the NumPy correlate function. The next step is selecting the greatest peak found in cross-correlation (minimum and maximum parameters vary based on signal strength) using the findPeaks function. After the greatest peak is identified, the start of the drone ID burst (startBurst) can be found by backtracking four OFDM symbols in length. A clean drone ID burst is trimmed from startBurst to burstDuration.

FIGURE 5: NORMALIZED CROSS-CORRELATION

```
Input: iqData: 32-bit floating point IQ data
Output: burst: Drone ID burst
zc4 ← zadoffChuSeq(600, 601)
crossCorrelation ← $|$ correlate(iqData, zc4) $|^2$
peak ← findPeaks(crossCorrelation, height=(1e5,1e6))
fftSize ← 1024
zcOffset ← 80 + (72 × 3) + (fftSize × 3) # burst start
startBurst ← peak – zcOffset
```

```
burstDuration ← (80 × 2) + (72 × 7) + (fftSize × 9)
burst ← burst[startBurst:startBurst+burstDuration]
```

Low-Pass Bandwidth Filter

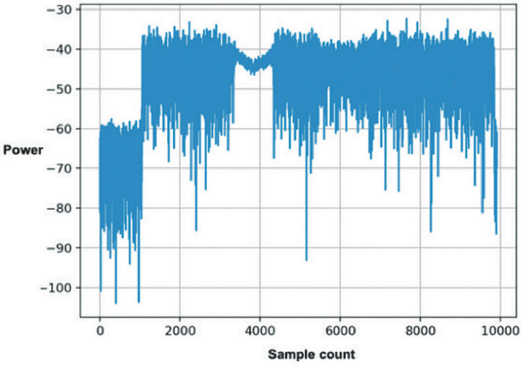
The algorithm in Figure 6 performs a low-pass bandwidth filter on a drone ID burst. The filter function is from the signal package. The filter window firWin is fitted to the drone ID burst bandwidth (bw) and sample rate (sampRate) at a length of n (51). The filterTaps object is a low-pass bandwidth filter applied to the drone ID burst, resulting in filteredBurst.

FIGURE 6: APPLYING A LOW-PASS BAND FILTER

```
Input: burst: Drone ID burst
Output: filteredBurst: Filtered drone ID burst
n ← 51
bw ← 10e6
sampRate ← 15.36e6
filterTaps ← firWin(n, bw/sampRate)
filteredBurst ← digitalFilter(filterTaps, burst, axis=0)
```

Figure 7 shows the graphical result of the low-pass bandwidth filter. The graph plots the magnitude (squared) of the drone ID in a log scale.

FIGURE 7: DJI DRONE ID BURST WITH LOW-PASS BANDWIDTH FILTER



Coarse Frequency Offset Correction

The algorithm in Figure 8 shows the coarse frequency offset calculation. The coarse frequency offset is a minor error—compared to an actual frequency offset—generated by the SDR. The coarse frequency offset is calculated by inspecting the symbol cyclic prefix in the second OFDM symbol. The second OFDM symbol begins with a cyclic prefix and ends with an inverted cyclic prefix. The cp variable denotes the first cyclic prefix in the second OFDM symbol, and the copy variable denotes the second cyclic prefix. In NumPy, a dot product operation is obtained by conjugating cp before multiplying it with the copy variable. The result is the sum of all the elements along axis 0. The offsetRadians is generated on the complex plane in radians, which is inversely applied to the drone ID burst to correct the coarse frequency offset.

FIGURE 8: CORRECTION OF THE COARSE FREQUENCY OFFSET

```
Input: burst: Filtered drone ID burst
Output: newBurst: Frequency offset corrected burst
fftSize ← 1024
cp ← burst [1104:1176] # 1st cyclic prefix
copy ← burst[2128:2200] # 2nd cyclic prefix
offsetRadians ← $\frac{\text{angle(sum(}\bar{\text{start}} \times \text{end, axis=0))}}{\text{fftSize}}$
newBurst ← burst × $e^{\text{-1j} \times \text{-offsetRadians} \times [1\ldots\text{length(burst)+1}]}$
```

OFDM Symbol Extraction

Figure 9 shows the time and frequency domains from the drone ID burst. The process for extracting OFDM symbols in the drone ID burst involves stripping cyclic prefixes and converting the remnants into the time and frequency domains. There are nine iterations that start at the end of each cyclic prefix in an OFDM symbol. Before the algorithm begins, the burst is already in the form of a time domain, so each symbol is stored in a corresponding timeDomain array row. The timeDomain is then converted to the frequency domain (freqDomain) by computing a Fourier transformation, followed by shifting the zero-frequency component to the center. The final array is stored in the corresponding freqDomain array row.

FIGURE 9: GENERATION OF TIME AND FREQUENCY DOMAINS

```
Input: burst: Drone ID burst
Output: timeDomain: Time domain of symbols
Output: freqDomain: Frequency domain of symbols
prefixes ← [72, 80, 80, 80, 80, 80, 80, 80, 72]
fftSize ← 1024
freqDomain ← zeros(length(prefixes), fftSize)
timeDomain ← zeros(length(prefixes), fftSize)
offset ← 0
for i in range(length(prefixes)):
        offset ← prefixes[i] + offset
        timeDomain[i,:] ← burst[offset:offset+fftSize]
        freqDomain[i,:] ← fftShift(fft(timeDomain[i,:]))
        offset ← fftSize + offset
end
```

Channel Impulse Response

The algorithm in Figure 10 shows the channel impulse response, which calculates the average walking phase offset. The channel impulse response represents the distortion of a signal [9]. The algorithm in Figure 4 generates both ZC sequences found in OFDM symbols 4 and 6 (zc4 and zc6). Each ZC sequence is converted to the frequency domain by a Fourier transformation. The Golden references for each ZC sequence are stored in channel1 and channel2 for OFDM symbols 4 and 6, respectively. The channel estimation (est) derives from channel1. The elements that are not data carriers become discarded. The average phase offset (phaseOffset) for each symbol is calculated by computing the angle of channel1 and channel2 and then summing all the elements together. That sum is then divided by the number of data carriers (600). The final phaseOffset value is the average of both channel1 and channel2.

FIGURE 10: MEASUREMENT OF CHANNEL IMPULSE RESPONSE

```
Output: phaseOffset: Average walking phase offset
Output: est: Channel estimation
dataCarrierIndices ← [$i$ = 212…813, $i \neq$ 512]
zc4 ← fftShift(fft(zadoffChuSeq(600, 601)))
zc6 ← fftShift(fft(zadoffChuSeq(147, 601)))
channel1 ← $\frac{\text{zc4}}{\text{freqDomain[3,:]}}$ # OFDM symbol 4
channel2 ← $\frac{\text{zc6}}{\text{freqDomain[5,:]}}$ # OFDM symbol 6
channel1 ← channel1[dataCarrierIndices]
channel2 ← channel2[dataCarrierIndices]
est ← channel1
channel1Phase ← $\frac{\text{sum(angle(channel1), axis=0)}}{600}$
channel2Phase ← $\frac{\text{sum(angle(channel2), axis=0)}}{600}$
phaseOffset ← $\frac{\text{channel1Phase - channel2Phase}}{2}$
```

Quantize QPSK into Bits

The algorithm in Figure 11 shows the process of demodulating QPSK into constellation mappings (bits). The algorithm equalizes the frequency domain to only include data carriers, adjusting the sample to the previously calculated phaseOffset. The absolute phase offset is calculated from multiplying phaseOffset with the distance each OFDM symbol is from the symbol that was used for equalization. Because the phase offset was calculated between both ZC sequences (which are in OFDM symbols 4 and 6), the phaseOffset is directly applied to OFDM symbol 5. The algorithm then loops through dataCarriers and converts the complex samples (representing QPSK constellation points) into bits.

FIGURE 11: QUANTIZATION OF QPSK TO BITS

```
Input: est: Channel estimation
Input: phaseOffset: Average walking phase offset
Output: demodBits: Quantized bits from drone ID
carrierIndices ← [$i$ = 212…813, $i \neq 512$]
demodBits ← zeros(9, 1200)
for a in length(bits):
    dataCarriers ← freqDomain[a, carrierIndices] × est
    dataCarriers ← $e^{1j \times -phaseOffset \times (a-4)}$
    offset ← 0
    quantizedBits ← zeros(1200)
    for b in length(dataCarriers):
        sample ← dataCarriers[b]
        if real(sample) > 0 and imag(sample) > 0 then
            bits ← [0,0]
        end
        elif real(sample) > 0 and imag(sample) < 0 then
            bits ← [0,1]
        end
        elif real(sample) < 0 and imag(sample) > 0 then
            bits ← [1,0]
        end
        elif real(sample) < 0 and imag(sample) < 0 then
            bits ← [1,1]
        end
        else
            bits ← [0,0]
        end
        quantizedBits[offset:offset+2] ← bits
        offset ← offset+2
    end
    demodBits[a,] ← quantizedBits
end
```

Table III shows the QPSK constellation mapping. The algorithm saves each quantized data carrier into demodBits.

TABLE III: QPSK MODULATION MAPPING

b(*i*), b(*i*+1)	I	Q
00	$\frac{1}{\sqrt{2}}$	$\frac{1}{\sqrt{2}}$
01	$\frac{1}{\sqrt{2}}$	$-\frac{1}{\sqrt{2}}$
10	$-\frac{1}{\sqrt{2}}$	$\frac{1}{\sqrt{2}}$
11	$-\frac{1}{\sqrt{2}}$	$-\frac{1}{\sqrt{2}}$

Descramble Bits

Figure 12 depicts an algorithm for descrambling demodulated drone ID bits into complex values. The two initial values for descrambling are hardcoded polynomial values: lsfrX1 (whose value is outlined in 3GPP 36.211 7.2) and lsfrX2 (0x12345678 for drone IDs) [9]. The variable n_c is defined in the LTE standards as 1,600. The first loop generates the m-sequence for lsfrX1. The second loop generates the m-sequence for lsfrX2. The third loop generates the resulting Gold sequence (goldSeq). The final operation performs a bitwise XOR operation between OFDM symbols (2, 3, 5, 7, 8 and 9) and the goldSeq.

FIGURE 12: DESCRAMBLING BITS INTO COMPLEX VALUES

```
Input: demodBits: Quantized bits from drone ID burst
Output: descBits: Complex values from descrambling

lfsrX1 ← $\begin{bmatrix} 1\,0\,0\,0\,0\,0\,0\,0 \\ 0\,0\,0\,0\,0\,0\,0\,0 \\ 0\,0\,0\,0\,0\,0\,0\,0 \\ 0\,0\,0\,0\,0\,0\,0 \end{bmatrix}$

lfsrX2 ← $\begin{bmatrix} 0\,0\,0\,1\,1\,1\,1\,0 \\ 0\,1\,1\,0\,1\,0\,1\,0 \\ 0\,0\,1\,0\,1\,1\,0\,0 \\ 0\,1\,0\,0\,1\,0\,0 \end{bmatrix}$

finalSeqLen ← 7200
$n_c$ ← 1600
x1 ← zeros($n_c$ + finalSeqLen + 31)
x2 ← zeros($n_c$ + finalSeqLen + 31)
goldSeq ← zeros(finalSeqLen, type='int8')
x1[0:31] ← lfsrX1
x2[0:31] ← lfsrX2
for i in length(finalSeqLen + $n_c$)
        x1[i+31] ← (x1[i+3] + x1[i]) % 2
end
```

```
for i in length(finalSeqLen + $n_c$)
        x2[i+31] ← (x2[i+3] + x2[i+2] + x2[i+i] + x2[i]) % 2
end
for i in length(finalSeqLen)
        goldSeq[i] ← (x1[i+$n_c$] + x2[i+$n_c$]) % 2
end
demodBits ← demodBits[[1, 2, 4, 6, 7, 8],:]
descBits ← demodBits  goldSeq
```

Turbo Decoder and Rate Matcher

The turbo decoder is implemented using the turbofec library [10]. The 7,200 descrambled bits from the algorithm in Figure 12 are passed into the turbo code remover C++ program, which was provided by the dji_droneid GitHub repository [2]. The program sets up the necessary structures and buffers to interface with turbofec for turbo decoding and rate-matching logic. The program outputs decoded data when the CRC-24 check returns 0x00. Else, it outputs the calculated CRC-24 error. Poor SDR recordings, interference and frequency offsets often attribute to failed decoding calculations.

Deframe Bytes

This research has identified three types of drone ID packets: license, flight information version 1 and flight information version 2. Software reverse engineering methods were used to understand the contents of the drone ID packets.

Figure 13 shows the structure of a license plate packet. License packets have a packet type of 0x11. After the packet type comes the serial number of the detected drone. Following the drone serial number are the custom license and flight plan values provided by the DJI GO smartphone application user.

FIGURE 13: DRONE ID LICENSE PACKET

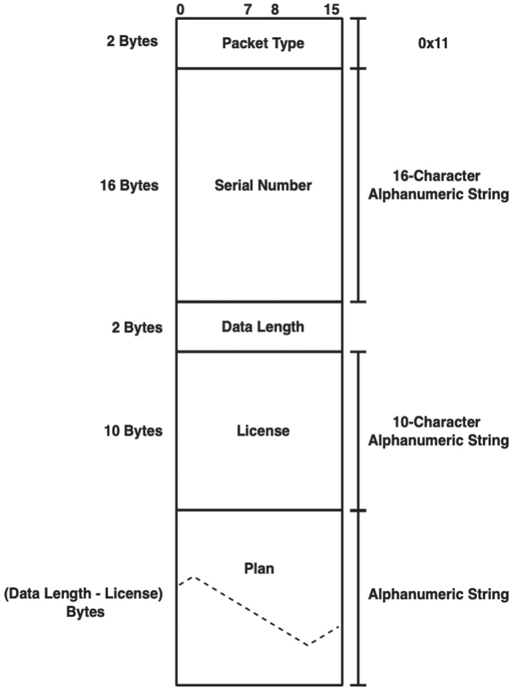


Figure 14 shows the structure of a version 1 flight information packet. Version 1 packets have a packet type of 0x1001. The state information varies depending on whether the drone is operating properly (e.g., motors on, in air, home point set) [1]. State information is followed by the serial number of the detected drone. The drone's GPS coordinates, altitude, height, x speed, y speed, z speed, pitch angle, roll angle, yaw angle and return-to-home GPS coordinates are arranged sequentially. Each GPS coordinate (longitude and latitude) is packed into two bytes using the following computation:

$$\frac{GPS\ Coordinate}{180} \times \pi \times 10^7$$

Next in the packet is the model field that specifies the product type of the drone and the universally unique identifier (UUID), an 18-character string identifier that ties the unmanned aerial vehicle to a DJI user account.

FIGURE 14: DRONE ID FLIGHT INFORMATION PACKET (VERSION 1)

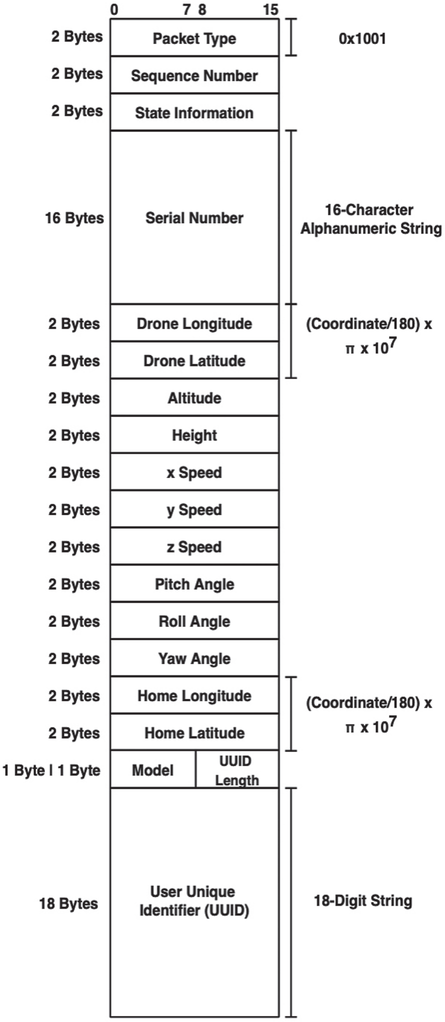
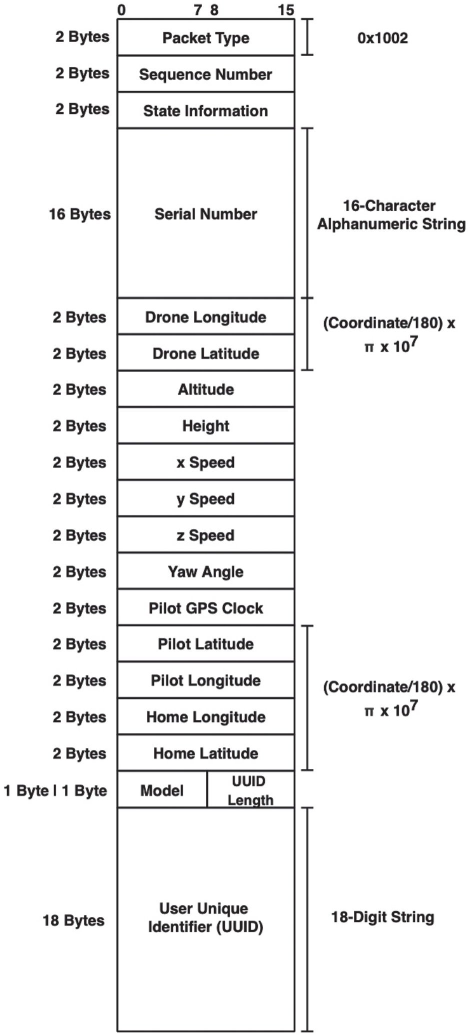


Figure 15 shows the structure of a version 2 flight information packet. Version 2 packets have a packet type of 0x1002. The rest of the packet follows the same arrangement as a version 1 packet, except there are no pitch angles or roll angles, only yaw angles. Next is the pilot GPS clock, which measures the number of milliseconds since epoch (January 1, 1970) to assess phone GPS accuracy. The pilot GPS coordinates, return-to-home GPS coordinates, model, UUID length and UUID are arranged sequentially.

FIGURE 15: DRONE ID FLIGHT INFORMATION PACKET (VERSION 2)



2 Bytes	Packet Type	0x1002
2 Bytes	Sequence Number	
2 Bytes	State Information	
16 Bytes	Serial Number	16-Character Alphanumeric String
2 Bytes	Drone Longitude	(Coordinate/180) x π x 10^7
2 Bytes	Drone Latitude	
2 Bytes	Altitude	
2 Bytes	Height	
2 Bytes	x Speed	
2 Bytes	y Speed	
2 Bytes	z Speed	
2 Bytes	Yaw Angle	
2 Bytes	Pilot GPS Clock	
2 Bytes	Pilot Latitude	(Coordinate/180) x π x 10^7
2 Bytes	Pilot Longitude	
2 Bytes	Home Longitude	
2 Bytes	Home Latitude	
1 Byte I 1 Byte	Model | UUID Length	
18 Bytes	User Unique Identifier (UUID)	18-Digit String

The algorithm in Figure 16 shows the conversion between aerodynamic angles to quantity values. Pitch, roll and yaw angles can be converted to quantities using a set of conditionals. If the angle is 0, the quantity yields 0. If the angle is less than 0, adding 180 to the angle yields the quantity. If the angle is greater than 0 but less than 180, modular dividing the angle by 180 yields the quantity. Otherwise, adding 180 to the angle yields the quantity.

FIGURE 16: CONVERT AERODYNAMIC ANGLES TO QUANTITIES

```
Input: angle: Pitch, roll or yaw angle
Output: quantity: Pitch, roll or yaw
if angle is 0 then
    quantity ← 0
end
elif angle < 0 then
    quantity ← angle + 180
end
elif angle > 0 and angle < 180 then
    quantity ← angle % 180
end
else
    quantity ← angle + 180
end
```

In the event of only detecting a license packet, which contains a serial number and no model information, there is a way to predict the model of the drone by looking at the serial number prefix (universal three-string constants). Table V (see Appendix), which has data derived from the Federal Aviation Administration Aircraft Inquiry Database and DJI Service Request and Inquiry website, shows DJI models corresponding to serial number prefixes [11], [12]. The table reveals all the drones equipped with remote identification capabilities. Also, the table shows all possible AeroScope IDs (value in the model field) for both versions of flight information packets. The AeroScope IDs were found on a DJI storage domain [13]. All the AeroScope IDs are associated with DJI models, except for AeroScope ID 240, which is associated with a non-DJI Yuneec H480 drone. Yuneec is a known DJI competitor. The AeroScope ID associated with the Yuneec H480 model may be present because DJI was attempting to detect Yuneec drones and/or DJI has plans to potentially acquire Yuneec.

4. DJI DRONE DETECTION

This section discusses how we leveraged our knowledge gained from studying and reverse engineering the DJI OcuSync protocol to develop a real-time DJI drone detection system, without purchasing an expensive and proprietary DJI AeroScope unit. We used low-cost radio hardware (HackRF Ones) to demonstrate how this system could be assembled at a minimal cost.

OcuSync Detection

A DJI OcuSync detection system runs effectively using "cheap" SDRs (under $500), such as HackRF Ones, coupled with Intel-based commodity hardware to process the Python software and turbo decoder program. Real-time detection of DJI drones requires reliable RF throughput and rapid frequency hopping across known center frequencies (Table II) to maximize performance.

The algorithm in Figure 5 can be outfitted to include a dynamic cross-correlation that is constantly scanning for drone IDs. The @njit decorator in Numba, a high-performance Python compiler, can be used to aid in faster calculations and yield better performance benchmarks when applied to the cross-correlation function.

The tradeoff with low-cost SDRs (e.g., HackRF Ones) is unreliable crystal oscillators, which occasionally invoke frequency offsets. These offsets can result in failed demodulations. However, there are ways to calculate a frequency offset at any juncture and apply a correction to a capture. One technique for solving this problem for the HackRF One is by using the CellSearch program in the LTE-Cell-Scanner GitHub repository [14].

The algorithm shown in Figure 17 leverages the CellSearch program to ping nearby LTE bands (stored in lteBand) from a range of frequencies: a frequency minimum (freqMin) to a frequency maximum (freqMax). The program can measure the crystal frequency error of the LTE frequency (crystalCorrection). The crystalCorrection is then transformed into a parts per million (ppm) number with a simple equation.

FIGURE 17: CORRECTION OF HACKRF FREQUENCY OFFSET

```
Input: freqBand: Nearby frequency band
Output: ppm: HackRF crystal oscillator error
lteBand ← {1: [2140, 2140.1]…103: [757.5, 757.6]}
freqMin ← lteBand[freqBand][0]
freqMax ← lteBand[freqBand][1]
crystalCorrection ← CellSearch(freqMin, freqMax)
ppm ← 1e6 × (1- crystalCorrection)
```

The ppm number can be inputted into the -C argument of the hackrf_transfer command (which interfaces with the HackRF to receive radio frequency data) to correct the frequency offset:

```
hackrf_transfer file.cs8 -f [CENTER_FREQUENCY] -s 1536000000 -C [PPM]
```

The last step is converting file.cs8, which is a collection of complex 8-bit signed integer samples, into IQ data. The conversion can be performed with the following code snippet:

```
buffer = open(file.cs8, type=int8)
buffer = buffer [::2] + 1j × buffer[1::2]
```

Figures 18 and 19 show the working prototype built for OcuSync drone ID detection. A refurbished Mac Mini, two HackRF Ones and two Altelix 2.4 GHz omni-directional antennas are used as the OcuSync drone ID detection system. The total cost is less than $1,000, making it significantly cheaper than any detection system (including DJI AeroScope) on the market. During operation, one HackRF hops to a different center frequency on the 2.4 GHz frequency band, and another HackRF hops to a different center frequency on the 5.8 GHz frequency band. Each capture consists of 1,500,000 complex 8-bit signed integers samples, which is enough samples to potentially contain a drone ID.

FIGURE 18: MAC MINI AND TWO HACKRF ONES IN A PELICAN CASE

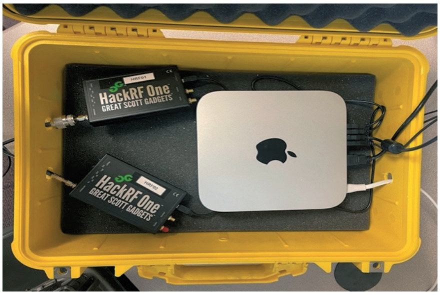
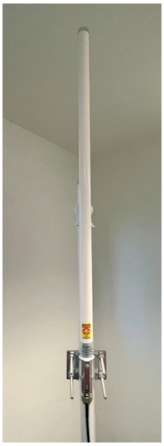


FIGURE 19: ALTELIX OMNI-DIRECTIONAL 2.4 GHZ ANTENNA



Enhanced Wi-Fi Detection

A DJI Enhanced Wi-Fi detection system can be achieved by simply using network adapters with monitor mode and custom clock rate capabilities.

Department 13 revealed that DJI drones use Atheros chips to broadcast beacon frames containing drone IDs [1]. The beacon frames are emitted with a bandwidth of 5 MHz. This is typically not allowed for normal network interface cards; however, Atheros network interface cards are able to be clocked at half rate (10 MHz) or quarter rate (5 MHz) [15].

Figure 20 shows a Qualcomm Atheros QCNFA435 M.2 WLAN/Bluetooth laptop Wi-Fi card that can detect Enhanced Wi-Fi drones. These cards can be purchased from any online retailer for less than $30. An Atheros network interface card can also be used by Kismet, a wireless network device detector and sniffer that operates on Wi-Fi interfaces.

FIGURE 20: QUALCOMM ATHEROS WI-FI M.2 CHIP

The following command allows Kismet to scan Wi-Fi channels 1–177 (consisting of both 2.4 GHz and 5.8 GHz frequencies) at a 5 MHz bandwidth:

```
kismet -c adapter:channels = "1W5, 2W5, 3W5, 4W5, 5W5, 6W5, 7W5, 8W5, 9W5, 10W5, 11W5, 12W5, 13W5, 14W5, 140W5, 149W5, 153W5, 157W5, 161W5, 165W5, 169W5, 173W5, 177W5"
```

The Kismet control panel allows users to filter Wi-Fi devices based on MAC address, advertised SSID or beacon vendor tag. The best way to detect DJI drones is by filtering individualized 802.11 vendor tags. IEEE keeps an open-source record of all registered company 802.11 vendor tags. DJI vendor tags are the following: 60-60-1F, 34-D2-62 and 48-1C-B9. By extension, Kismet supports non-DJI drone detection such as Parrot. Parrot vendor tags are as follows: 00-12-1C, 90-03-B7, A0-14-3D and 00-26-7E.

Such techniques could easily be employed to provide Enhanced Wi-Fi inspecting capabilities in a drone ID detection system. We leave the integration and testing of the Enhanced Wi-Fi detection capability for future work.

5. EXPERIMENTATION AND RESULTS

We performed experimental testing to validate the effectiveness of the DJI OcuSync drone ID detection system. An experiment was conducted in an urban environment to assess the range and capabilities of the DJI OcuSync drone ID detection system over a period of two days, using three different types of DJI drones flown at varying distances away from the DJI OcuSync drone ID detection system.

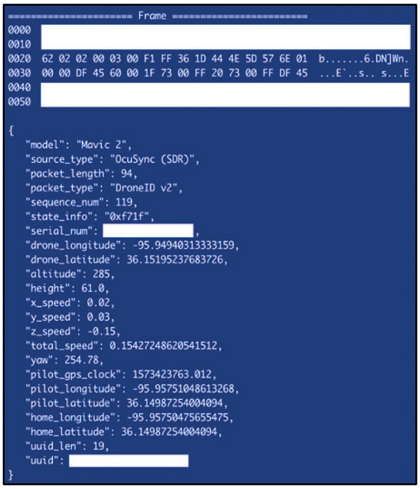
Table IV shows the detection results from the experiment. The detection system was able to capture a Mavic Pro flight information packet and a Mavic Mini license packet, both approximately 0.3 miles away.

TABLE IV: DETECTION RESULTS FROM EXPERIMENT

Model	Range	Packet Type
Mavic Pro	~0.30 miles	Flight information (v1)
Mavic Mini	~0.30 miles	License
Mavic 2	~0.75 miles	Flight information (v2)

The furthest distance the detection system was able to find was a DJI Mavic 2, which was 0.75 miles away, as shown in Figures 21 and 22. Although DJI AeroScope stationary units have the potential to detect drones up to 30 miles (50 km) away, our experimental testing showed that we were only able to reliably detect drones less than a mile away. It should be noted that the DJI AeroScope claims of 30 miles are likely estimates under the best of circumstances with relatively low noise environments compared to our testing in the middle of an urban environment. Performance degradations are expected in such environments, but there is plenty of room to improve our solution. Future work will incorporate testing of different types of budget SDRs and improved digital signal processing techniques to increase the reliability of our solution. Additionally, we will explore improvements to antenna design and placement to achieve optimal performance for our testing environments.

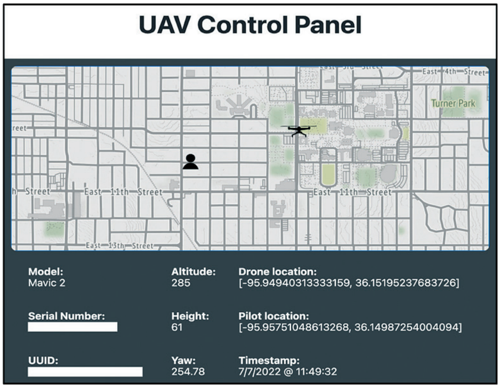
FIGURE 21: SCRIPT OUTPUT OF A DETECTED MAVIC 2 DRONE



```
====================== Frame ========================
0000  [redacted]
0010
0020  62 02 02 00 03 00 F1 FF 36 1D 44 4E 5D 57 6E 01  b.......6.DN]Wn.
0030  00 00 DF 45 60 00 1F 73 00 FF 20 73 00 FF DF 45  ...E`..s.. s...E
0040  [redacted]
0050

{
   "model": "Mavic 2",
   "source_type": "OcuSync (SDR)",
   "packet_length": 94,
   "packet_type": "DroneID v2",
   "sequence_num": 119,
   "state_info": "0xf71f",
   "serial_num": [redacted],
   "drone_longitude": -95.94940313333159,
   "drone_latitude": 36.15195237683726,
   "altitude": 285,
   "height": 61.0,
   "x_speed": 0.02,
   "y_speed": 0.03,
   "z_speed": -0.15,
   "total_speed": 0.1542724862054l512,
   "yaw": 254.78,
   "pilot_gps_clock": 1573423763.012,
   "pilot_longitude": -95.95751048613268,
   "pilot_latitude": 36.14987254004094,
   "home_longitude": -95.95750475655475,
   "home_latitude": 36.14987254004094,
   "uuid_len": 19,
   "uuid": [redacted]
}
```

FIGURE 22: DRONE CONTROL PANEL SHOWING DETECTED MAVIC 2 DRONE



Model: Mavic 2

Altitude: 285

Drone location: [-95.94940313333159, 36.15195237683726]

Serial Number: [redacted]

Height: 61

Pilot location: [-95.95751048613268, 36.14987254004094]

UUID: [redacted]

Yaw: 254.78

Timestamp: 7/7/2022 @ 11:49:32

6. CONCLUSIONS

Asymmetric warfare has been widely employed throughout the conflict in Ukraine, especially the use of small drones for conducting kinetic strikes. In particular, DJI drones have played a major role in the conflict on both sides, supporting military operations in the form of reconnaissance and weapon delivery systems. The delivery of humanitarian aid has been further empowered by using drones as well. Ukrainian officials have publicly accused DJI of helping Russia target innocent civilians by allowing Russian military forces to acquire and use a proprietary DJI drone-tracking system called AeroScope. Cost and ease of access are primary factors that have hindered Ukraine's ability to counter this threat with AeroScope systems of their own to identify and locate DJI drones and operators used by Russia. This has created an asymmetric advantage of sorts on the battlefield for Russia. Our work demonstrates an alternative to AeroScope that is similar in functionality. Compared with AeroScope, the proposed OcuSync drone ID detection system is superior in terms of affordability, cost and flexibility. It is the first demonstrated detection system that offers real-time detection of DJI OcuSync drone IDs using low-cost SDRs with robust packet dissection (license and flight information). The detection system is significantly cheaper than DJI AeroScope, which is marketed anywhere from $20,000 to $40,000 and can predict DJI drone models based on serial numbers. However, our experimental testing has demonstrated that the detection system is currently limited to a 0.75-mile (1.2-km) detection range, whereas the commercial DJI AeroScope stationary units can detect drones up to 30 miles (50 km) away. Future work will focus on increasing the range of the OcuSync drone ID detection system, conducting additional experiments with the detection system and merging both Enhanced Wi-Fi and OcuSync detection streams into a centralized web application for a comprehensive drone detection solution. Given the success achieved with our DJI OcuSync ID detection system, we believe this will offer a path for negating asymmetric advantages for opponents using DJI AeroScope monitoring systems.

ACKNOWLEDGMENTS

Special thanks go to David Protzman for releasing the DJI OcuSync drone ID open-source project. Without that resource, the demodulation steps would not have been possible to recreate.

REFERENCES

[1] "Anatomy of DJI's Drone Identification Implementation," Department 13, Canberra, Australia, White Paper, 2017.

[2] D. Protzman. "DJI DroneID RF Analysis." GitHub. 2022. [Online]. Available: www.github.com/proto17/dji_droneid

[3] A. Chapple. "The Drones of the Ukraine War." Radio Free Europe Radio Liberty. Nov. 17, 2022. [Online]. Available: www.rferl.org/a/ukraine-russia-invasion-drones-war-types-list/32132833.html

[4] S. Hollister, "DJI drones, Ukraine, and Russia – what we know about AeroScope," *Verge*, Mar. 23, 2022. [Online]. Available: https://www.theverge.com/22985101/dji-aeroscope-ukraine-russia-drone-tracking

[5] S. Hollister, "DJI insisted drone-tracking AeroScope signals were encrypted — now it admits they aren't," *Verge*, Apr. 28, 2022. [Online]. Available: https://www.theverge.com/2022/4/28/23046916/dji-aeroscope-signals-not-encrypted-drone-tracking

[6] O. Karasapan. "Ukrainian refugees: Challenges in a welcoming Europe." Brookings Institution. Oct. 14, 2022. [Online]. Available: www.brookings.edu/blog/future-development/2022/10/14/ukrainian-refugees-challenges-in-a-welcoming-europe/

[7] M. McNabb. "Has the U.S.-China trade war changed DJI's drone market share? The latest form Drone Industry Insights." DroneLife. Mar. 5, 2021. [Online]. Available: https://dronelife.com/2021/03/05/has-the-u-s-china-trade-war-changed-djis-drone-market-share-the-latest-from-drone-industry-insights/

[8] Remote Identification of Unmanned Aircraft, Code of Federal Regulations, Title 14, Chapter 1, Part 89, Federal Aviation Administration, Washington, DC, USA, 2022.

[9] "TS 136.211, Group Radio Access Network, Evolved Universal Terrestrial Radio Access (E-UTRA), Physical Channels and Modulation (Release 10), V10.3.0." 2011. 3GPP. [Online]. Available: https://etsi.org/deliver/etsi_ts/136200_136299/136211/10.03.00_60/ts_136211v100300p.pdf

[10] T. Tsou, TurboFEC, GitHub, 2018. [Online]. Available: https://github.com/ttsou/turbofec

[11] "How to check your product's serial number." Da-Jiang Innovations. 2022. [Online]. Available: https://repair.dji.com/product/serial/index

[12] "Aircraft Inquiry, Washington, DC." Federal Aviation Administration. 2022. [Online]. Available: https://registry.faa.gov/aircraftinquiry

[13] "areoscope_type." Da-Jiang Innovations. 2022. [Online]. Available: mydjiflight.dji.com/links/links/areoscope_type

[14] J. Xianjun. "LTE-Cell-Scanner." GitHub. 2022. [Online]. Available: www.github.com/JiaoXianjun/LTE-Cell-Scanner

[15] A. Chadd. "Half and Quarter rate support." freeBSD. 2012. [Online]. Available: wiki.freebsd.org/dev/ath_hal%284%29/HalfQuarterRate

APPENDIX

TABLE V: DJI MODELS, SERIAL NUMBER PREFIXES AND KEYS

Model	Prefix	AeroScope ID
Inspire 1	041, W21	1
Phantom 3 Series	0JX	2
Phantom 3 Series Pro	P76	3
Phantom 3 Std	03Z, P5A	4
M100	M02	5
ACEONE	-	6
WKM	-	7
NAZA	061	8
A2	061	9
A3	067	10
Phantom 4	07D, 07J, 0AX, 0HA, 189	11
MG1	05Y	12
M600	M64	14
Phantom 3 4k	P7A	15
Mavic Pro	08Q, 08R	16
Inspire 2	095, 09Y, 0A0	17
Phantom 4 Pro	0AX	18
N2	-	20
Spark	0AS, 0BM	21
M600 Pro	M80	23
Mavic Air	0K1, 0K4	24
M200	0FZ	25
Phantom 4 Series	CE1	26
Phantom 4 Adv	0HA	27
M210	0N4	28

M210RTK	17U, 1DA	30
A3_AG	-	31
MG2	-	32
MG1A	-	34
Phantom 4 RTK	0UY, 0V2	35
Phantom 4 Pro V2.0	11U, 11V	36
MG1P	0YS	38
MG1P-RTK	0YL	40
Mavic 2	0M6, 163	41
M200 V2 Series	17S	44
Mavic 2 Enterprise	276, 29Z	51
Mavic Mini	1SC, 1SD, 1SZ, 1WG	53
Mavic Air 2	1WN, 3N3	58
P4M	1UD	59
M300 RTK	1ZN	60
DJI FPV	37Q	61
Mini 2	3NZ, 3Q4, 5DX, 5FS	63
AGRAS T10	IEZ	64
AGRAS T30	35P	65
Air 2S	3YT	66
M30	-	67
Mavic 3	F4Q, F45	68
Mavic 2 Enterprise Adv	298	69
Mini SE	4AE, 4DT, 4GM	70
Mini 3 Pro	-	73
YUNEEC H480	YU1	240

2023 15th International Conference on Cyber Conflict
Meeting Reality
T. Jančárková, D. Giovannelli, K. Podiņš, I. Winther
2023 © NATO CCDCOE Publications, Tallinn

Russian Invasion of Ukraine 2022: Time to Reconsider Small Drones?

Aleksi Kajander
Early-Stage Researcher
Department of Law
Tallinn University of Technology
Tallinn, Estonia
aleksi.kajander@taltech.ee

Abstract: In May 2022, an Estonian-Russian man was arrested at the Estonian border with Russia for attempting to supply the Russian armed forces with crowdfunded drones. The case had two intertwined striking aspects: the law under which the individual was prosecuted and the drones themselves. While it is no revelation that drones are dual-use goods, the drones in question were three DJI Mini 2, which, owing to their small size and features, are exempt from the current European Union (EU) restrictions on the export of dual-use goods. Such small commercial drones have proven to be excellent for aerial surveillance and indirect fire correction on the battlefield. Consequently, the individual was prosecuted for 'knowingly supporting a foreign act of aggression' based on a newly added provision to the Estonian Penal Code.

This paper discusses the growing importance of commercial small drones on the battlefield, which are not included in Annex I of the EU Dual-Use Regulation, as well as the implications of this on the EU's dual-use goods export restrictions, and the legal framework that is available to EU member states for preventing the delivery of such drones to support a war of aggression. The paper is divided into three sections, the first dedicated to the role of small drones on the battlefield in Ukraine, the second to the EU's dual-use export restrictions and the third to the role of domestic legal frameworks that may prohibit exports of such drones through laws criminalizing aggression.

Keywords: *drones, dual-use, export control, legal framework, Ukraine conflict, European Union*

1. INTRODUCTION

The further invasion of Ukraine by Russia in 2022 brought military drones such as the Bayraktar TB2 into the spotlight.[1] However, the conflict has additionally highlighted the growing importance of small commercially available civilian drones on the modern battlefield.[2] Small commercial drones have proven to be an excellent platform for aerial surveillance and artillery fire correction.[3] The extensive use of small commercial drones has led to manufacturers both protesting their military use and suspending sales in Ukraine and Russia.[4]

Despite the wide array of civilian uses for drones,[5] their dual-use nature is not a new revelation, as unmanned aerial vehicles (UAVs) were already listed as dual-use goods in the European Union (EU) context prior to the conflict.[6] Nevertheless, the association of drones with the military, especially with legally controversial drone strikes,[7] is so widespread it has caused psychological barriers to the adoption and usage of drones in civilian contexts.[8] The attempts by the EU to address this connection, as well as to reduce knowledge of the fact that civilian developments of drones have military benefits, have invited criticism in the past.[9]

Nevertheless, under the EU definition, for a drone to qualify as 'dual-use', it must either have a maximum endurance of one hour or greater, or an endurance between 30 minutes to less than an hour and, at the same time, be designed to take off and have stable flight in wind gusts equal to or exceeding 25 knots.[10] Consequently, for this paper, small commercial drones refer to UAVs that are not originally intended for military use, weighing less than 250 grams, with a specification below the

[1] Mohammed Eslami, 'Iran's Drone Supply to Russia and Changing Dynamics of the Ukraine War' (2022) 5(2) Journal for Peace and Nuclear Disarmament 507, 509.

[2] Matt Burgess, 'Small Drones Are Giving Ukraine an Unprecedented Edge' (*Ars Technica*, 5 October 2022). <arstechnica.com/information-technology/2022/05/small-drones-are-giving-ukraine-an-unprecedented-edge/> accessed 15 November 2022.

[3] ibid.

[4] Eg 'DJI Reassess Sales Compliance Efforts in Light of Current Hostilities' (*DJI*, 26 April 2022) <www.dji.com/uk/newsroom/news/dji-statement-on-sales-compliance-efforts?utm_medium=network-affiliate&awc=7327_1668509201_a623249e5ecbe883ef7237bc62b1d597&pbc=awin2017> accessed 15 November 2022.

[5] Marcus Schulzke, 'Drone Proliferation and the Challenges of Regulating Dual-Use Technologies' (2019) 21 International Studies Review 497, 506.

[6] Regulation (EU) 2021/821 of the European Parliament and the Council Setting Up a Union Regime for the Control of Exports, Brokering, Technical Assistance, Transit and Transfer of Dual-Use Items (recast) OJ L 206 Annex I, Category 9, 9A012.

[7] Adam Smith, 'Drones as Techno-legal Assemblages' (2022) 4(2) Law, Technology and Humans 152, 153–54.

[8] Mario Mendoza, Mauricio Alfonso, and Stephane Lhuillery, 'A Battle of Drones: Utilizing Legitimacy Strategies for the Transfer and Diffusion of Dual-Use Technologies' (2021) 166 Technological Forecasting & Social Change 120539, 120540.

[9] Philip Boucher, 'Domesticating the Drone: The Demilitarization of Unmanned Aircraft for Civil Markets' (2015) 21(6) Science and Engineering Ethics 1392.

[10] ibid.

requirements for maximum endurance and wind resistance, as defined in Annex I of the EU's dual-use regulation.

As the conflict has demonstrated through donations,[11] drones that do not meet these criteria, such as the DJI Mini 2,[12] are frequently used or desired to be used on the battlefield. Such drones, as exemplified by the DJI Mini 2, while subject to registration, do not require a remote piloting certificate.[13] The rationale for the focus on DJI's models in this paper when examples for technical specifications are needed is that DJI holds the majority of the market share, despite its recent drop from 70 per cent to 54 per cent,[14] for all commercial drones.[15] Moreover, DJI's drones are so popular on both sides that identifying whether a drone is friendly has become challenging and has led to them being described as a 'true symbol of modern warfare' by a former chief of Russia's armed forces.[16] Therefore, while DJI itself does not sanction military usage, its drones are arguably the most prolific due to their widespread use. Thus, when technical specifications must be referred to in order to convey typical attributes of commercial drones, DJI's drones will be used. Nevertheless, regardless of the manufacturer, when considering their battlefield utility, small commercial drones have arguably been misclassified and underestimated from a legal point of view in the EU.

However, individual EU member states have taken steps to prevent the export of such drones through national laws. The best example of this is the case of the individual who was convicted for attempting to donate three DJI Mini 2 drones to the Russian military.[17] Curiously, the law he was convicted under was the then newly added Article 91^1 of the Penal Code of Estonia, which entered into force only in May 2022 and criminalized supporting foreign acts of aggression.[18] This may be held as an example of how an individual member state can prevent the export of small commercial drones that would not otherwise be considered dual-use.

Therefore, besides discussing the growing importance of small commercial drones on the battlefield, this paper explores the current legal framework surrounding them, to

[11] See eg Ishveena Singh, 'Finnish Volunteers Deliver 140 DJI Mavic Mini Drones to Ukraine military' (*Dronedj*, 3 March 2022) <dronedj.com/2022/03/03/finland-140-dji-mini-drone-ukraine-military/> accessed 15 November 2022.

[12] Max flight time: 31 minutes, Max Wind Speed Resistance: 8.5 – 10.5 m/s (20.4 knots) as per DJI, 'Specs' <www.dji.com/ee/mini-2/specs> accessed 15 November 2022.

[13] European Union Aviation Safety Agency, 'FAQ n. 136863' <www.easa.europa.eu/en/faq/136863> accessed 15 November 2022.

[14] Wieber de Jager, 'DJI Commercial Drone Market Share Falls Dramatically in 2021' (*Dronexl*, 20 September 2022) <https://dronexl.co/2021/09/20/dji-commercial-drone-market/> accessed 6 January 2023.

[15] Larisa Kapustina et al, 'The Global Drone Market: Main Development Trends' (2021) 129 SHS Web of Conferences 11004.

[16] Isabelle Khursudyan, Mary Ilyushina, and Kostiantyn Khudov, 'Russia and Ukraine Are Fighting the First Full-Scale Drone War' *Washington Post* (2 December 2022) <www.washingtonpost.com/world/2022/12/02/drones-russia-ukraine-air-war/> accessed 6 January 2023.

[17] 'Estonia Hands Confiscated Crowdfunded Russian Drones to Ukrainian Army' (*ERR News*, 5 October 2022) <news.err.ee/1608738604/estonia-hands-confiscated-crowdfunded-russian-drones-to-ukrainian-army> accessed 15 November 2022.

[18] Penal Code (*Karistusseadustik*) (EE) art 91^1.

discover if the EU legal framework provides options for preventing their export or if it is up to the individual EU member states. As a result, the primary research question for the paper is this: Does the current EU legal framework enable the prevention of the export of small drones that are not classified as dual-use goods? Moreover, the research has additional value outside the context of drones, as the research aims to discover whether there is a general possibility of utilizing the EU legal framework to prevent the export of items that are not classified as dual-use but likely should be. For it is not inconceivable that in the future another item will find its use on the battlefield before it can be classified as a dual-use item.

The paper will be of interest from the perspectives of both the EU and its member states, as it will highlight whether local policy and lawmakers should take similar steps as Estonia did, whether the current EU legal framework could be applied differently, or whether it needs adjustment. The research, which is primarily legal in nature, will be carried out mainly through desk research, with a special focus on the critical legal analysis of the relevant existing EU legal framework.

2. SMALL DRONES – MAJOR IMPACT

The incorporation of small commercial drones by combatants in the 2022 Ukraine-Russia conflict has been swift and significant. Small commercial drones enable soldiers to conduct reconnaissance and correct indirect fire effectively and with less risk to life and limb.[19, 20] Their utility is reflected in their popularity, with Ukraine alone operating over 6,000 drones, most of which are manufactured in China.[21]

While commercial drones are available in many shapes and sizes, a considerable number of small drones weighing less than 250 grams have been utilized for military purposes in Ukraine.[22] These small drones are generally used unarmed to perform

[19] Marek Kohv and Archil Chochia, ‘Unmanned Aerial Vehicles and the International Humanitarian Law. Case Study: Russia’ in Holger Mölder and others (eds), *The Russian Federation in Global Knowledge Warfare: Influence Operations in Europe and Its Neighbourhood (Contributions to International Relations)* (Springer 2021).

[20] David Hambling, ‘Small Quadcopters Rule the Battlefield in Ukraine – Which Makes Their Chinese Manufacturers Very Unhappy’ (*Forbes*, 29 April 2022) <www.forbes.com/sites/davidhambling/2022/04/29/small-quadcopters-rule-the-battlefield-in-ukraine---which-makes-their-chinese-manufacturers-very-unhappy/?sh=7346b06f7685> accessed 6 January 2023.

[21] Patrick Galey, ‘Big Guns and Small Drones: The Devastating Combo Ukraine Is Using to Fight Off Russia’ (*NBC News*, 13 May 2022) <www.nbcnews.com/news/world/ukraine-army-uses-guns-weapons-drone-combo-rcna27881> accessed 6 January 2023.

[22] Hambling (n 20).

fire correction and reconnaissance,[23] unlike their larger counterparts, which can be modified to carry ordnance[24] and are frequently used to drop munitions.[25]

Instead, the main benefit of small commercial drones is their combination of low-cost and reconnaissance capabilities. A small commercial drone can be two to three times cheaper than its over 250-gram counterparts.[26] This is significant on a battlefield where drones are unlikely to survive many flights. Furthermore, this has enabled a wider distribution of drones down the chain of command, all the way to squad leaders.[27] The wider distribution has significantly improved ordinary troops' awareness, reconnaissance, and correction of indirect fire. This represents a major shift from the earlier practice of having fewer purpose-built military drones, operated by a small number of specialists.

These conclusions were confirmed by the Royal United Services Institute (RUSI) report on the 'Preliminary Lessons in Conventional Warfighting from Russia' Invasion of Ukraine: February–July 2022'. The report states not only that drones are critical and 'essential across all branches and at all echelons' for situational awareness but additionally that 90 per cent of drones are lost.[28] Therefore, the drones utilized must be 'cheap and attritable'.[29] Consequently, small commercial drones weighing less than 250 grams are arguably ideal, as they provide reconnaissance capability at a low cost.

Smaller drones are also easier to carry on the field as they are pocket-sized when folded, making them far more convenient to distribute on a squad level than their larger counterparts. Furthermore, close to the action, their range capabilities are less important as the enemy may be presumed to be in proximity. Thus, small drones arguably have a crucial tactical role on the modern battlefield.

However, as commercial drones were not intended for military usage, vulnerabilities soon became evident. Chief among these were systems like AeroScope, under which each DJI drone automatically broadcasts the position of the drone and operator.[30]

[23] ibid.

[24] Pierre Ayad and Pariesa Brody, 'Ukrainian Soldiers Are Turning Consumer Drones into Formidable Weapons of War' (*France 24 The Observers*, 8 August 2022) <observers.france24.com/en/europe/20220808-ukraine-russia-modified-commercial-drones-battlefield-donations-weapons> accessed 6 January 2023.

[25] David Hambling, 'Ukraine Is Fielding A "Heinz 57" Fleet of Heavy Drone Bombers Against Russian Forces' (*Forbes*, 20 December 2022) <www.forbes.com/sites/davidhambling/2022/12/20/ukraine-is-using-a-heinz-57-fleet-of-heavy-drone-bombers-against-russian-forces/> accessed 6 January 2022.

[26] 'Consumer Drones Comparison' (DJI) <www.dji.com/ee/products/comparison-consumer-drones> accessed 6 January 2022.

[27] Hambling (n 20).

[28] Mykhaylo Zabrodskyi and others, 'Preliminary Lessons in Conventional Warfighting from Russia's Invasion of Ukraine: February–July 2022' (*RUSI*, 30 November 2022) <static.rusi.org/359-SR-Ukraine-Preliminary-Lessons-Feb-July-2022-web-final.pdf> accessed 6 January 2023.

[29] ibid.

[30] Sean Hollister, 'DJI Drones, Ukraine and Russia – What We Know About AeroScope' (*Verge*, 23 March 2022) <www.theverge.com/22985101/dji-aeroscope-ukraine-russia-drone-tracking> accessed 6 January 2023.

Intended to let law enforcement detect drones that may, for example, threaten airfields,[31] it is clear how the system is a vulnerability in a military context if the adversary has access to such signals. Furthermore, unlike initially claimed by the manufacturer and other sources, the signals were not encrypted,[32] meaning an adversary capable of electronic warfare (EW) could exploit the system to gain a military advantage.

On the topic of EW, with the average life expectancy of a drone being three flights and 90 per cent of Ukrainian drones being destroyed during the first three phases of the conflict, the impact of EW should be further considered.[33] This was hardly surprising considering that the vulnerability of small drones, in particular to jamming, was known before the conflict.[34] However, based on the reports, the effectiveness of EW increases as the conflict becomes more static, with Russian EW effectiveness considerably increasing after rapid movements ceased, as the EW coverage had trouble keeping up.[35] Moreover, RUSI's report makes it clear that the denial of precision, such as could be achieved through fire corrections using drones, was crucial for the survival of units[36] or, to flip the sentiment, for their destruction.

Consequently, drawing all these factors together, a few suggestions can be made as to the importance of small commercial drones in peer or near-peer conflicts. Firstly, while mini or even nano military drones exist, their high cost, as exemplified by the Black Hornet's $195,000 price tag,[37] makes them currently impractical for widespread distribution on a squad level, even for the wealthiest militaries.[38] For the price of a single Black Hornet, one could purchase over 400 small commercial drones,[39] enabling far wider distribution among troops. Therefore, until the EW resistance of small military drones means they are less affected by the horrific rate of attrition seen in Ukraine, small commercial drones are arguably more cost-effective. Moreover, the effectiveness of EW is reduced when the conflict is not static, so the effectiveness of commercial drones may be even greater during such phases.

[31] ibid.

[32] ibid.

[33] RUSI report (n 28) 38.

[34] Michael Horowitz, Sarah Kreps, and Matthew Fuhrmann, 'Separating Fact from Fiction in the Debate over Drone Proliferation' (2016) 41(2) International Security 7, 17–18.

[35] Bryan Clark, 'The Fall and Rise of Russian Electronic Warfare: The Ukraine Invasion Has Become an Old-Fashioned Slog, Enabling Russia to Unleash Its Electronic Weapons' (*IEEE Spectrum*, 30 July 2022) <spectrum.ieee.org/the-fall-and-rise-of-russian-electronic-warfare#toggle-gdpr> accessed 6 January 2023.

[36] RUSI report (n 28) 38.

[37] Philip Dunne, 'Miniature Surveillance Helicopter to Help Protect Front Line Troops' (*UK Ministry of Defence*, 4 February 2013) <www.gov.uk/government/news/miniature-surveillance-helicopters-help-protect-front-line-troops> accessed 6 January 2023.

[38] Kyle Jahner, 'Army wants mini-drones for its squads by 2018' (*ArmyTimes*, 3 April 2016) <https://www.armytimes.com/news/your-army/2016/04/03/army-wants-mini-drones-for-its-squads-by-2018/> accessed 6 January 2023.

[39] See eg DJI Mini 2, manufacturer listed price of 449 dollars: 'DJI Mini 2' (*DJI*) <https://store.dji.com/product/mini-2?vid=99411&set_region=US&from=store-nav> accessed 6 January 2023.

Additionally, despite the awareness of their limitations, both sides appear undeterred in utilizing commercial drones, as neither side has stopped. On the contrary, Ukraine has called for more donations of commercial drones.[40] Consequently, while the usage of commercial drones, small or otherwise, is not without risk, based on their widespread use in Ukraine in 2022, the risk is evidently worth taking, at least until military drones of similar capacities, cost and quantities are available.

Consequently, whether small commercial drones will continue to play a significant role in future peer or near-peer conflicts will depend largely on when affordable equivalent military drones become available. Nevertheless, commercial drones will likely continue to be utilized by armies or groups that are unable to acquire equivalent military drones, whether due to a lack of funds or other factors, because of their considerable utility on a modern battlefield and ease of purchase, even if it comes with the associated risks and vulnerabilities.

3. EU DUAL-USE REGULATION AND SMALL DRONES

A. Introduction to the EU's Dual-Use Regulation and Drones

The EU regime for the control of exports of dual-use items is regulated by Regulation 821/2021 ('the Recast Regulation'), which was adopted in May 2021. Under the Regulation's Article 3(1), for dual-use goods listed in Annex I, authorization is required for their export. Consequently, to establish whether items that may be used for both military and civilian purposes, as required by the definition for dual-use items in Article 2(1) of the Recast Regulation, Annex I is the logical first step in determining if an item is considered dual-use within the EU.

Drones are found in Annex I in Category 9, with 9A012 being UAVs designed to have controlled flight outside the direct natural vision of their operator. However, 9A012 does not include all drones; instead, there are two paragraphs that establish two distinct criteria for drones to qualify as dual-use. Under paragraph (2), if a drone has a maximum endurance of more than one hour, it is considered dual-use. Alternatively, under (1), two cumulative criteria are required, firstly a maximum endurance of at least 30 minutes and less than 60 minutes (a) and secondly a maximum wind resistance of 25 knots (b).

This is distinct from the EU's classification of drones, under which commercial drones are divided into categories C0–C5, based primarily, albeit not exclusively, on their maximum take-off mass.[41] This difference likely stems from much of the dual-use

[40] Chris Vallance, 'Ukraine sent dozens of "dronations" to build army of drones' (*BBC*, 8 July 2022) <www.bbc.com/news/technology-62048403> accessed 6 January 2023.

[41] Commission Delegated Regulation (EU) 2019/945 on Unmanned Aircraft Systems and on Third-Country Operators of Unmanned Aircraft Systems, OJ L 152/1, Annex.

item list being derived from the Wassenaar Agreement, which uses the same technical requirements as Annex I of the Recast Regulation.[42] Therefore, the conclusions on the limitations of the Recast Regulation's definition of dual-use drones are directly transferable to other instances where the Wassenaar Agreement's definition is utilized. Moreover, this has considerable relevance from a NATO perspective as many of their members are a part of the Wassenaar Agreement, not to mention the EU, so any deficiencies in their export control regimes are likely to be a significant collective defence consideration.

Arguably, in practice, based on battlefield usage, the two most important determinants of a commercial drone's military capabilities are its optics for reconnaissance purposes and the maximum take-off mass, which may enable ordnance drops. A commercial drone with a sufficient maximum take-off mass is far more likely to partake in hostilities by dropping ordnance than lighter ones. Moreover, the endurance limitation is less important for small drones that are expected to be used near the enemy, especially if the trip will likely be one-way due to enemy activity.

Therefore, if a drone lacks both optics and a sufficient take-off mass for ordnance drops, it thereby lacks any military utility. Thus, the dual-use definition for drones in the Recast Regulation is not in line with the reality of potential military use, as demonstrated by the conflict in Ukraine. Considering the prevailing reality, the dual-use drone classification in the EU legislation should arguably shift towards those two characteristics to prevent them from unintentionally reaching combatants due to improper definitions.

B. Regulation 821/2021 in Depth

Despite many small commercial drones with military utility being outside the definition in Annex I of the Recast Regulation, there are additional possibilities for the restriction of items not listed in Annex I. Article 3(2) of the Recast Regulation explicitly states that an authorization 'may' be required for the export of certain items not listed in Annex I in accordance with Articles 4, 5, 9 or 10. Consequently, the procedures encompassed in each ought to be examined to determine if they may be utilized in the context of small commercial drones.

1) Article 4

Proceeding in order, Article 4 provides for the possibility that items not listed in Annex I are subject to an export authorization, provided a competent authority informs the exporter beforehand that the items are or may be intended for uses described in subsections (a), (b), or (c). The Recast Regulation's definition of 'exporter' is subject to some legal controversy, as it may contain a loophole that may enable natural

[42] Wassenaar Arrangement Secretariat, 'List of Dual-Use Goods and Technologies and Munitions List' (*Wassenaar*, December 2022) <www.wassenaar.org/app/uploads/2022/12/List-of-Dual-Use-Goods-and-Technologies-Munitions-List-Dec-2022.pdf> accessed 6 January 2023.

persons with such items in their luggage to escape the definition of 'exporter'.[43] That controversy aside, the first use described in Article 4(a), is generally not relevant in the context of small commercial drones as it refers to use in connection with nuclear, chemical or biological weapons, and there is no indication of such uses in connection with small commercial drones.

However, the second, (b), is relevant for small commercial drones, as it pertains to items intended for military end-use in a country subject to an arms embargo. Currently, with the Ukraine-Russia conflict, Russia would qualify, as it is subject to restrictive measures on both the import and export of arms by the EU.[44] Military end-use is considered to consist of any of the three uses encompassed in points (i) through (iii) of Article 4(1)(b). In the context of drones, points (ii) and (iii) are of limited relevance as they concern the use of goods for the development, production, or maintenance of military items and the use of any unfinished products in a plan to produce military items, respectively. Consequently, as small commercial drones primarily have been pressed into service 'as is', at least in the context of the 2022 Ukraine conflict, these two points are unlikely to be relevant.

However, point (i) appears more readily applicable, as military end-use is defined as the incorporation of dual-use items 'into' military items listed in the military list. UAVs are certainly included in the military list of the member states, under category ML10, regardless of their weight or other characteristics,[45] thereby fulfilling that condition for the definition of military end-use. However, the commercial drone itself, paradoxically, both is and is not a military item. In the colloquial, non-legal sense, a commercial drone that is used 'as is' by the military is arguably a military item simply by virtue of the fact that the military uses it. However, under category ML10, there is a stipulation that such UAVs must be 'specially designed or modified for the military', which does not apply to small commercial drones, as they were intended specifically for civilian use.[46] Moreover, as mentioned before, small commercial drones are generally not modified by the military, unlike larger commercial drones that might be needed, for instance, to be able to carry ordnance. Additionally, even if an end-modification by the military as an end user is planned, it may be difficult to ascertain this in advance for both the exporter and the competent authorities, thereby hindering

[43] Article 2(3)(c) of the Recast Regulation defines a natural person carrying dual-use items in their personal luggage as an exporter with the meaning given in Article 1(19)(a) of Regulation 2015/2446. However, Point (b) of 1(19) refers to a natural person while point (a) is essentially identical to point 2(3)(a) of the Recast Regulation, which pertains to a situation where the exporter holds a contract. Therefore, presumably the correct reference should be to point (b) not point (a) of 1(19). Consequently, a natural person with no contract, transporting dual-use items in their luggage is not an exporter under Article 2(3)(c). This loophole creates a problem as for an authorisation to be required under Article 4 and 5, the exporter must be informed beforehand.

[44] Council Decision 2014/512/CFSP [2014] L 229/13.

[45] Council Common Military List of the European Union [2022] C 100/03.

[46] See eg 'DJI Has Always Opposed Combat Use of Civilian Drones and Is Not a "Chinese Military Company"' (*DJI*, 3 November 2022) <www.dji.com/newsroom/news/dji-has-always-opposed-combat-use-of-civilian-drones-and-is-not-a-chinese-military-company> accessed 6 January 2023.

the possible application of Article 4. Nevertheless, the conclusion remains that in the legal sense of the military list, a small commercial drone is not a military item.

Furthermore, the problem with applying this provision crystallizes with the requirement of 'incorporation **into** military items', under Article 4(1)(b)(i). This requires that the dual-use item be 'incorporated into' a military list item. When a commercial small drone is used 'as is', it is not being incorporated into anything. Therefore, despite its actual use by a military in combat, it does not fit into the Recast Regulation Article 4's definition of military end-use, which is paradoxical. Nevertheless, simple usage of a dual-use item 'as is' by a military is not considered military end-use, which arguably leads to the conclusion that Article 4(b) of the Recast Regulation cannot be used as a legal justification to require authorization for the export of small commercial drones. Article 4(c) is equally unsuitable, at least insofar as the drones are used 'as is', rather than for parts or components of military items, in which case it could apply. Therefore, while Article 4(b) and (c) could be used in specific circumstances to require authorization, they are arguably unsuitable for doing so when unmodified small commercial drones are pressed into service in their unaltered state.

2) Article 5

The second possibility for requiring authorizations for items not listed in Annex I is through Article 5 of the Recast Regulation, if the item in question can be considered a 'cyber surveillance' item. Article 5 contains the same requirement as Article 4 on informing the 'exporter' beforehand.[47] Nevertheless, leaving the information requirement aside, drones have been associated with covert surveillance, with concern for their impact on civilians and their human rights.[48] Consequently, it is not entirely inconceivable that drones could be included in Article 5.

However, the definition of a 'cyber-surveillance item' under Article 2(20) of the Recast Regulation arguably precludes drones. The definition requires that the item is 'specially designed to enable the covert surveillance of natural persons by monitoring, extracting, collecting or analysing data **from** information and telecommunications systems'. Thus, as commercial drones do not collect information 'from' information and communications technology (ICT) systems but rather collect it directly from the real world surrounding them using their optics, they fall outside the definition, and hence Article 5 cannot be applied.

3) Articles 9 and 10

The final possible ways to prohibit or require authorization are Articles 9 and 10. Under Article 9(1), a member state may, on its own initiative, prohibit or impose an authorization requirement on the export of dual-use items not listed in Annex I for

[47] See n 43.

[48] Eliza Watt, 'The Principle of Constant Care, Prolonged Drone Surveillance and the Right to Privacy of Non-Combatants in Armed Conflicts' in Russell Buchan and Asaf Lubin (eds), *The Rights to Privacy and Data Protection in Times of Armed Conflict* (NATO CCDCOE Publications 2022) 161–62.

reasons of public security, which include the prevention of terrorism, or for human rights considerations. Therefore, Article 9(1) represents the broadest ground for preventing the export of goods not listed in Annex I, because it does not exhaustively define reasons of public security, as evidenced by the use of the word 'including' before the examples of terrorism or human rights 'considerations'. Hence, conceivably, the export of drones not listed in Annex I to strengthen the military of a country that is conducting an illegal war of aggression in Europe and threatens various EU member states[49] could be considered a matter of public security. Moreover, compared to Article 5, actual human rights violations are not required, but rather 'considerations', whereby, for example, concerns about privacy, a human right, related to the use of drones could arguably be sufficient.

However, as required by Article 9(2) and (4), if a member state adopts such a measure, the Commission and other EU member states must be notified, along with providing the reasons for the move, which must be published in the *Official Journal of the European Union*. If this entails an update to the national list of controlled items, these too must be communicated under paragraph 3, and then published by the Commission in all official languages under paragraph 4. Furthermore, under Article 10(1), if one member state imposes an authorization requirement pursuant to Article 9, other member states must also require authorization or must inform the Commission and other member states if they refuse under paragraph 2.

Hence, arguably Articles 9 and 10 could be used to effectively prevent or impose an authorization requirement for the export of small commercial drones not listed in Annex I at the initiative of a member state. It is perplexing that this has not taken place, at least according to the *Official Journal*, which contained no such notification at the time of writing as required by Article 9(4) for 2022 regarding drones not listed in Annex I.

4. EU SANCTIONS – THE COMMERCIAL DRONE BLINDSPOT

As a response to the continued illegal invasion of Ukraine by Russia, the EU adopted a series of restrictive measures against Russia which, besides the dual-use items, extend the prohibition of selling and exporting to a variety of other items. It has been demonstrated that small commercial drones are not considered dual-use items in the meaning of the Recast Regulation and dual-use item restrictions thus do not apply to them. Therefore, it is not inconceivable that they may fall under a separate restrictive measure.

[49] Ott Tammik, 'Estonia's Kallas Warns of Existential Russian Threat to Baltics' (*Bloomberg*, 22 June 2022) <www.bloomberg.com/news/articles/2022-06-22/estonia-s-kallas-warns-of-existential-russian-threat-to-baltics?leadSource=uverify%20wall> accessed 6 March 2023.

The list of items included in the sanctions is located in the various annexes of Regulation (EU) 833/2014. The list of goods contained therein is truly extensive, including, for example, household sewing machines,[50] toasters,[51] and smartphones[52] above a certain price. Thus, it is perplexing that small commercial drones are excluded. However, in the EU's ninth sanctions package adopted in December 2022, toy drones were added to the restriction,[53] which still leaves commercial drones that are neither classified as toys nor fall under the criteria of Annex I of the Recast Regulation exempt from restrictions. For example, the DJI Mini drone is not considered a toy,[54] nor is it covered by Annex I of the Recast Regulation, so it perfectly demonstrates this blind spot.

Thus, there appears to be a blind spot for small commercial drones, both in the Recast Regulation and the specific restrictive measures against Russia. Conceivably, that overlap could be caused by the assumption that any drone with battlefield utility would be classified as a dual-use item, whereby its inclusion in other restrictions would be unnecessary. Unfortunately, this is not the case, so there is no specific restriction on exporting small commercial drones either under the Recast Regulation or the specific sanctions regime against Russia.

5. MEMBER STATE DOMESTIC LAW

Despite the apparent oversights in the EU approach to the export of small commercial drones to support Russia, individual EU member states have taken steps to prevent such support through domestic criminal legislation. The most concrete example of this is the criminal conviction of an individual in Estonia for organizing a crowdfunding campaign through which three DJI Mini 2 drones were purchased with the intent to supply them to the Russian armed forces in Ukraine.[55] This conviction was made possible by an amendment to the Estonian Penal Code that had been passed during the summer of 2022, building on an earlier provision that criminalized participation in the 'management, execution or preparation' of acts of aggression. The earlier provision was passed after Russia's initial invasion of Ukraine in 2014.[56]

The Estonian Penal Code was amended to include a new Article 91^1, which took effect on 8 May 2022, criminalizing joining, participation in, and supporting foreign acts of aggression. More specifically, the Article criminalizes 'knowingly supporting a foreign act of aggression', which includes financing. The words 'knowingly supporting' refer to direct support – either physical or material contributions – while mental support,

[50] Council Regulation (EU) 833/2014 [2014] OJ L 229/1, Annex XVIII 8452 10 00.

[51] ibid Annex XVIII 8516 72 00.

[52] ibid Annex XVIII 8517 13 00.

[53] Council Regulation (EU) 2022/2474 [2022] OJ L 322 I/1.

[54] European Union Safety Agency, 'FAQ n. 136863' (EASA, 27 July 2022) <www.easa.europa.eu/en/faq/136863> accessed 6 January 2023.

[55] 1-22-4285 (22913000012) Harju Maakohus.

[56] Penal Code (*Karistusseadustik*) (EE) art 91^1.

such as in the form of propaganda, is outside the scope of Article 91, but may be punishable under provisions such as Article 92, which prohibits war propaganda.[57] Violations of Article 91[1] are punishable by up to five years imprisonment for natural persons and carry a pecuniary punishment for legal persons.

During that same summer, an Estonian-Russian dual citizen decided to obtain drones requested by Russia's 76th Air Assault Division.[58] Notably, the drones in question, besides being excluded from Annex I of the Recast Regulation, also weighed less than 250 grams. The fact that such drones were specifically requested by a unit of the Russian military directly engaged in combat is about the strongest evidence one can get for their utility on the battlefield and the absolute necessity of categorizing such drones as dual-use.

Furthermore, the prosecution did not accuse the defendant of any violation other than that under Article 91[1], such as an export violation for attempting to breach the restrictive measures on Russia. Consequently, it appears that in practice, it is up to the individual EU member states to prevent the export of small commercial drones based on their domestic laws. This is problematic, as it means that the approach is fragmented, which besides helping Russia, risks a repeat of the embarrassment of the continued sales of weapons to the country after the 2014 arms embargo.[59] It must be mentioned that the Commission did propose to harmonize criminal offences for violations of the EU sanctions in December 2022.[60] However, the success of such a measure is entirely dependent on the restrictive measures in place, whether through the prohibition of the export of dual-use items or through specific sanctions on goods, which, as demonstrated in the previous section, are far from infallible.

Moreover, the case of the Estonian-Russian dual citizen was particularly clear cut, as his intentions were evident – he was literally in contact with the 76th Air Assault Division and had made an explicit public appeal on his VKontakte account for funds to support the Russian invasion. This in turn raises the question, would even the Estonian Article 91[1] be able to prevent an individual natural person from ferrying small commercial drones to Russia if they were intelligent enough to claim that the drones were a gift or that they were otherwise not 'knowingly' intending them to support Russia's aggression. In such a hypothetical situation, arguably there may be little for the authorities to rely on, as small commercial drones fall into a blind spot in the EU dual-use item and goods sanctions.

[57] Riikikogu, 'Seletuskiri' (*Riikikogu*, 4 April 2022) < www.riigikogu.ee/download/b24122af-c201-477b-9ded-221eff65918c> accessed 6 March 2023.

[58] 1-22-4285 (22913000012) Harju Maakohus.

[59] Francesco Guarascio, 'EU Closes Loophole Allowing Multimillion-Euro Arms Sales to Russia' (*Reuters*, 14 April 2022) <www.reuters.com/world/europe/eu-closes-loophole-allowing-multimillion-euro-arms-sales-russia-2022-04-14/> accessed 6 April 2023.

[60] European Commission, 'Ukraine: Commission Proposes to Criminalise the Violation of EU Sanctions' (*European Commission*, 2 December 2022) <ec.europa.eu/commission/presscorner/detail/en/ip_22_7371> accessed 6 January 2023.

Estonia is not the only EU member state that has criminalized participation in or support of acts of aggression. What makes the Estonian Article 91^1 stand out is its lowered threshold, whereby an individual is not required to wield state authority to be convicted. Under the old Article 91, for a person to be found guilty, they must have been able to control or direct the activities of the state, thus excluding private citizens. By contrast, Article 91^1 has no such requirement, thus lowering the threshold to anyone 'knowingly supporting a foreign act of aggression', even through financing.

While other EU member states have criminalized acts of aggression, often those provisions require the person to wield state authority. For example, Czechia's Section 405[61] and Finland's Chapter 11 Section 4(a)[62] both require the person to be able to exercise control over a state. Consequently, such provisions are ill-suited to preventing grassroots-level support of the sort attempted by the Russian-Estonian individual. Therefore, while the criminalization of support for an act of aggression may be used in lieu of a specific prohibition for attempting to supply dual-use items that are categorized incorrectly, it is not an ideal EU-level solution, as domestic laws vary from state to state. Nevertheless, even under the current framework, EU member states have the ability to prohibit the export of small commercial drones by relying on Article 9(1) of the Recast Regulation alongside their national control list under 9(3), whereby the loophole could arguably be closed relatively effectively. The reasoning behind not utilizing this possibility ought to be examined in the future to prevent repetition of the current situation, where, during a conflict, items that are clearly dual-use are excluded due to a loophole in the definitions.

6. CONCLUSION

The further illegal invasion of Ukraine by Russia in 2022 has demonstrated the battlefield necessity for small drones that drastically improve the reconnaissance and fire correction of combat units on the tactical level. Both sides have relied on small commercial drones in this capacity, as evidenced by the donation requests for such drones during the conflict. This trend is likely to continue in future conflicts because of the need for widespread distribution of drones and the current low survivability of the drones in such roles, which make an inexpensive small commercial drone ideal. Therefore, the acquisition of commercial drones, even as a temporary measure, is likely to remain a tempting or even necessary action for states during conflicts.

However, the EU legislative framework for the prevention of the export of small commercial drones leaves much to be desired. This is evidenced by the exclusion of small commercial drones from Annex I of the Recast Regulation, which is not an issue exclusive to the EU, as the Wassenaar Agreement utilizes the same definition in

[61] Criminal Code (*Zákon č. 40/2009 Sb., trestní zákoník*) (CZ).

[62] Criminal Code (*Rikoslaki 38/1889*) (FIN).

its 2022 control list. Nevertheless, the situation was exacerbated by small commercial drones also being absent from the specific restrictions on goods that were imposed on Russia. Consequently, these crucial battlefield items were excluded from the EU's legal framework, which may have enabled them to be supplied to Russia during the conflict.

Perplexingly, it cannot be concluded that the EU's legal framework is unprepared for such loopholes. On the contrary, it is well-prepared in principle, with several provisions providing possible ways to prevent the export of items not listed in Annex I of the Recast Regulation. However, for small commercial drones, arguably only one is applicable. Still, no member state has taken the initiative to utilize the possibilities provided by Articles 9 and 10, the reasoning for which ought to be examined in further research. Thus, export control of small commercial drones was left to the legal frameworks of individual EU member states, as exemplified by Estonia's amended Penal Code that was successfully used to prevent the export of such drones to Russia. Regardless, as the member state legal frameworks are not harmonized in this respect, the approach is fragmented across the EU, so no measure of confidence can be attached to the notion that small commercial drones cannot find their way to the Russian military via the EU.

2023 15th International Conference on Cyber Conflict
Meeting Reality
T. Jančárková, D. Giovannelli, K. Podiņš, I. Winther
2023 © NATO CCDCOE Publications, Tallinn

Weaponizing Cross-Border Data Flows: An Opportunity for NATO?

Matt Malone[1]
Thompson Rivers University Faculty of Law
Kamloops, British Columbia, Canada
mmalone@tru.ca

Abstract: On July 12, 2022, following the Russian invasion of Ukraine, the European Data Protection Board (EDPB) issued a warning to data exporters, reminding them Russia did not have an adequacy agreement governing cross-border data flows of Europeans' personal data to Russia. As such, blanket transfers of personal data were not permissible under European data protection law; instead, compliance needed to be assessed by data exporters on a case-by-case basis, and, where it could not be ensured, transfers should be suspended. This article views the EDPB declaration as a shot across the bow and extrapolates it to a future where cross-border data flow restrictions are deployed as an instrument of cooperative security as well as deterrence and defense. Given the potential sensitivity of personal information being transferred across borders, along with the economic value inherent in data flows in the digital economy, restrictions on cross-border data flows have the potential to inflict serious harm. This article explores the broader implications of this potential practice, assessing its security opportunities and drawbacks. The article advocates for reforming North Atlantic Treaty Organization (NATO) members' divergent approaches to the regulation of processing of cross-border data transfers; it suggests these member states can and should overcome their splintered approaches by establishing a "safe data zone" to facilitate cross-border data flows among members, where NATO retains the power to issue embargoes on cross-border data flows to specific jurisdictions while otherwise leaving decisional authority for transfers to supranational entities like the European Union (EU) or sovereign states. This approach would increase cross-border data flows between allies while permitting restrictions with adversaries where doing so achieves security objectives.

Keywords: *cross-border data flows, NATO, personal data transfers, European Data Protection Board, Russia*

[1] The author has no disclosures of financial or non-financial interests that are directly or indirectly related to the work.

1. INTRODUCTION

The democratic world faces many novel security threats arising out of the shift from analog to digital economies. In this new economic reality, cross-border data flows of personal information are frequently touted as an essential commodity—even as the "central nervous system" of the digital economy.[2] While such declarations are frequent, democratic countries have generally failed to acknowledge two important implications of this new reality. First, if cross-border data flows—the movement of digital information between servers around the world—are an essential commodity, prosperity accruing from them is a matter of security. Yet they are rarely examined through a security lens.[3] Second, security risks to the supply chain of cross-border data flows, like those to other essential commodities, may have a devastating impact on citizens' quality of life. Legal and physical obstructions that restrict cross-border data flows, such as data localization efforts and embargoes on data transfers in the absence of adequacy agreements, are such a security threat—one the democratic world should address with the urgency it deserves.

Unfortunately, this is something the democratic world shows little interest in doing, even though the threat to national security—in the form of a threat to prosperity—is already here. For example, the democratic world's two most powerful economic blocs, the United States and the European Union (EU), have been locked in a protracted but redundant dispute about data-sharing frameworks over the past few years. During this time, in a series of judgments handed down in 2015 and 2020,[4] the EU's highest court declared invalid two key frameworks governing the flow of EU members' citizens' personal customer data to the US—the Safe Harbour Privacy Principles[5] and the Privacy Shield.[6] These frameworks had deemed adequate American privacy and data protection laws, thereby facilitating transatlantic exchanges of personal data

[2] Center for Strategic and International Studies, *Cross-Border Data Flows and Digital Trade Post-TPP*, YouTube (Apr. 6, 2017), https://www.youtube.com/watch?v=xU7qnrLca0A&t=3s.

[3] *See, e.g.*, Susan Ariel Aaron, Data Is Different: Why the World Needs a New Approach to Governing Cross-border Data Flows, CIGI Paper No. 197 (Nov. 14, 2018); Alex He, *Trade Deals Might Induce Beijing to Bend on Data Restrictions*, CIGI (Jun. 20, 2022), https://www.cigionline.org/articles/trade-deals-might-induce-beijing-to-bend-on-data-restrictions/; *Cross-border Data Flows: Taking Stock of Key Policies and Initiatives*, OECD, (Oct. 12, 2022), https://www.oecd.org/publications/cross-border-data-flows-5031dd97-en.htm.

[4] Case C-362/14, Maximillian Schrems v. Data Protection, ECLI:EU:C:2015:650 (Oct. 6, 2015), https://eur-lex.europa.eu/legal-content/en/TXT/?uri=CELEX:62014CJ0362; Court of Justice of the European Union, *The Court of Justice Invalidates Decision 2016/1250 on the Adequacy of the Protection Provided by the EU-US Data Protection Shield* (Jul. 16, 2020), https://curia.europa.eu/jcms/upload/docs/application/pdf/2020-07/cp200091en.pdf [hereinafter *The Court of Justice Invalidates Decision 2016/1250*]; European Parliament, *The CJEU Judgment in the Schrems II Case* (Sept. 15, 2020), https://www.europarl.europa.eu/RegData/etudes/ATAG/2020/652073/EPRS_ATA(2020)652073_EN.pdf [hereinafter *CJEU Judgment in the Schrems II Case*].

[5] Court of Justice of the European Union Press Release, Judgement in Case C-362/14 Maximillian Schrems v. Data Protection Commissioner: The Court of Justice declares that the Commission's US Safe Harbour Decision is invalid (Oct. 6, 2015), https://curia.europa.eu/jcms/upload/docs/application/pdf/2015-10/cp150117en.pdf.

[6] European Parliament, *supra* note 4.

for commercial purposes between the two democratic blocs. When the frameworks were struck down by the Court of Justice of the European Union, cross-border transfers of data were subject to enormous legal uncertainty.[7] In these decisions and elsewhere, Europeans have demanded the US bolster its privacy and data protection laws. Americans have largely ignored these calls.[8] If left unresolved, this protracted dispute could obstruct data flows between the two economic blocs, with an impact no different, though less visible, than the container ship Ever Given blocking the Suez Canal in March 2021 and disrupting the global supply chain.[9]

This is unfortunate and unnecessary, since the spat, which is putatively about security, is between allies in an intergovernmental military alliance. Deployed as a means of offense or defense between truly adversarial countries, however, restrictions on cross-border data flows may become a significant weapon. How and when they will be used remains a largely uncharted subject, but a shot was fired across the bow on July 12, 2022, when the European Data Protection Board (EDPB), the body created to monitor the application of the General Data Protection Regulation (GDPR),[10] issued a warning to exporters of personal data.[11] Noting Russia's exclusion from the Council of Europe in light of the ongoing war in Ukraine[12]—in addition to unnamed European countries "already looking into the lawfulness of data transfers to Russia"[13]—the EDPB highlighted that Russia lacked an adequacy agreement for cross-border data flows and asked exporters to evaluate Russian data security laws and practices, including appropriate data handling safeguards.[14] In other words, the EDPB advised the transfer of Europeans' personal data to Russia could not occur unless it "ensure[d] an adequate

[7] After the decision in 2020, the EDPB wasted no time in issuing guidances that cross-border data flow "[t]ransfers on the basis of this legal framework are illegal." *See* Natasha Lomas, *No Grace Period after Schrems II Privacy Shield Ruling, Warn EU Data Watchdogs*, Techcrunch (Jul. 24, 2020), https://techcrunch.com/2020/07/24/no-grace-period-after-schrems-ii-privacy-shield-ruling-warn-eu-data-watchdogs/?guccounter=1&guce_referrer=aHR0cHM6Ly93d3cuZ29vZ2xlLmNvbS8&guce_referrer_sig=AQAAAA1pXR9teT1_ByWmh8Bpdeo70N6kAUyVjZAZbxrWl0oryhTF9VH9XYC7b3OIpmXMmDHj2ce9-5n1_VSH6p72WIMJzo1W5mkvDrfovAtzhstyI3I7CSSrUINMQ1H4TZ_tu55tEpc0HXaEjMDp9yirPyvACa2roO46Nh-Sz49zUnhO.

[8] Derek Hawkins, *The Cybersecurity 202*, Washington Post (May 25, 2018), https://www.washingtonpost.com/news/powerpost/paloma/the-cybersecurity-202/2018/05/25/the-cybersecurity-202-why-a-privacy-law-like-gdpr-would-be-a-tough-sell-in-the-u-s/5b07038b1b326b492dd07e83/.

[9] Ryan Browne, *Facebook's EU-U.S. Data Flows Are under Threat—That May Spell Trouble for Other Tech Giants*, CNBC (May 20, 2021), https://www.cnbc.com/2021/05/20/facebook-eu-us-data-flows-are-under-threat-heres-what-that-means.html.

[10] Regulation (EU) 2016/679 of the European Parliament and of the Council of 27 April 2016 on the Protection of Natural Persons with Regard to the Processing of Personal Data and on the Free Movement of Such Data, 2016 O.J. (L 119), art. 68 [hereinafter GDPR].

[11] European Data Protection Board, *Statement 02/2022 on Personal Data Transfers to the Russian Federation* (Jul. 12, 2022), https://edpb.europa.eu/system/files/2022-07/edpb_statement_20220712_transferstorussia_en.pdf [hereinafter EDPB July 12, 2022, Statement].

[12] Council of Europe, *The Russian Federation is Excluded from the Council of Europe* (Mar. 16, 2022), https://www.coe.int/en/web/portal/-/the-russian-federation-is-excluded-from-the-council-of-europe [hereinafter *The Russian Federation is Excluded*]. See also GDPR, *supra* note 10.

[13] EDPB July 12, 2022, Statement, *supra* note 11.

[14] *The Russian Federation is Excluded, supra* note 12.

level of protection"[15] for the data. Blanket transfers were not permissible.[16] Where assessments concluded European data protection standards could not be ensured, the EDPB ordered parties to "suspend data transfers."[17]

These were strong words. To be sure, restrictions on cross-border data flows are not new. In recent years, Russia, China, and Iran, among others, have all taken significant steps "to preemptively restrict and shape the flow of data at national borders."[18] As noted above, within the democratic bloc of countries in the North Atlantic, there are also long-standing disputes over cross-border data flows. On the very same day the EDPB issued its statement, the Danish Data Protection Authority (DDPA) ordered the Danish municipality of Helsingor to suspend cross-border transfers of personal data to the United States, in light of Google's failure to meet security requirements in Denmark's data protection legislation.[19] The EDPB's statement concerning Russia and the DDPA's statement concerning Google left some American spectators with the impression that European data protection authorities were treating apples like oranges—that is, "treating Russia as any other country," just as the DDPA was scrutinizing the "the risks of American government surveillance to Europeans' privacy."[20]

Ironically, both the US and the EU have clear security motives in their long-standing dispute over cross-border data flows. However, they are focusing on different *aspects* of security. The Americans remain focused on economic espionage and trade secret theft and give relatively little attention to privacy and data protection issues (e.g., there has been a lack of any federal legislative reform in years). The Europeans, for their part, are focused on privacy and data protection laws but have been slow to combat economic espionage and trade secret theft (e.g., whereas the GDPR is a binding regulation transposed on all states, the EU Trade Secrets Directive is not directly transposed and only sets minimum standards for members). Rather than leave the task of hashing out a solution to data protection officers, privacy commissioners, and other trade and commerce bureaucrats, this dispute should be resolved in a place meant to address transatlantic security concerns. The EU is not such a place. It is not

[15] GDPR, *supra* note 10, art. 45.

[16] EDPB July 12, 2022, Statement, *supra* note 11.

[17] *Id*.

[18] Robert K. Knake, *Weaponizing Digital Trade: Creating a Digital Trade Zone to Promote Online Freedom and Cybersecurity*, Council on Foreign Relations (Sept. 2020), https://cdn.cfr.org/sites/default/files/report_pdf/weaponizing-digital-trade_csr_combined_final.pdf.

[19] *See, e.g.*, EDPB, *The Danish DPA Imposes a Ban on the Use of Google Workspace in Elsinore Municipality* (Jul. 19, 2022), https://edpb.europa.eu/news/national-news/2022/danish-dpa-imposes-ban-use-google-workspace-elsinore-municipality_en; Rie Aleksandra Walle, *Danish DPA Bans Google Workspace Use and US Transfers*, NoTies Consulting, https://www.noties.consulting/danish-dpa-bans-google-workspace-and-us-transfers/ (last updated Aug. 18, 2022).

[20] Catherine Stupp, *EU Privacy Regulators Are Scrutinising Data Flows to Russia*, WSJ (Jul. 14, 2022), https://web.archive.org/web/20220807222419/https://www.wsj.com/articles/eu-privacy-regulators-are-scrutinizing-data-flows-to-russia-11657826124?ref=quuu. [hereinafter *EU Privacy Regulators Are Scrutinising Data Flows to Russia*].

a security organization. The US and most EU members already have such a forum at their disposal: the North Atlantic Treaty Organization (NATO).

At first blush, NATO seems like an unworkable forum to address the dispute over transatlantic data flows. There are plenty of reasons to dismiss the idea out of hand. After an unsuccessful 20-year war in Afghanistan, NATO is weary. It has also been challenged by a crisis of legitimacy and a turn against multilateralism driven by movements like Brexit and politicians like former US President Donald Trump, who famously evoked a possible American withdrawal from NATO.[21] The President of the EU Commission has even asserted a desire to create the EU's own armed forces.[22] And organizations like the D-10 and the Five Eyes have been assuming greater importance in recent years, to NATO's chagrin. Efforts like the Data Free Flow with Trust also activate important buzzwords, moving the paradigm toward trust networks vital to creating an exchange of cross-border data flows, though such efforts appear more symbolic than substantive.[23] Finally, the EU Commission, through power delegated to bodies like the EDPB, and under the auspices of the GDPR, has taken the lead on greenlighting data-sharing frameworks based on "adequacy" determinations.[24] But the Commission's failure to do so (twice) presents not only an opportunity but an imperative for NATO to take the lead.

The EDPB's pronouncement on July 12, 2022, used cross-border data flow restrictions in a manner tantamount to sanctions in a time of war—as an instrument not only of cooperative security (securitizing trade within a particular zone) but also of defense and deterrence (creating an adverse economic impact on external zones or specific actors). Using this episode as a springboard for a broader discussion about security consequences and implications, this article examines how restricting cross-border data flows presents novel opportunities for NATO, in particular when it comes to achieving the alliance's objectives in cooperative security as well as deterrence and defense; at the same time, it hypothesizes using NATO to establish a "safe data zone" that enables cross-border data flows, which would specifically obviate the scope and framework of the GDPR with security framing, to achieve such objectives.

[21] Philip Rucker & Robert Costa, *Trump Questions Need for NATO, Outlines Noninterventionist Foreign Policy*, Washington Post (Mar. 21, 2016), https://www.washingtonpost.com/news/post-politics/wp/2016/03/21/donald-trump-reveals-foreign-policy-team-in-meeting-with-the-washington-post/.

[22] Cain Burdeau, *EU President Makes the Case for a Pan-European Army*, Courthouse News Service (Sept. 15, 2021), https://www.courthousenews.com/eu-president-makes-the-case-for-a-pan-european-army/.

[23] World Economic Forum, *Every Country Has Its Own Digital Laws. How Can We Get Data Flowing Freely Between Them?* (May 20, 2022), https://www.weforum.org/agenda/2022/05/cross-border-data-regulation-dfft/.

[24] GDPR, *supra* note 10, art. 45.

2. THE PROMISES OF WEAPONIZING DATA FLOW RESTRICTIONS

This section discusses the advantages of, and security opportunities in, using cross-border data flow restrictions as a tool of cooperative security as well as deterrence and defense.

A. Addressing the Limitations of Economic Sanctions

Economic sanctions, such as asset freezes, export and import restrictions, financial prohibitions, and other economic limitations, are key ways that nation-states seek to force specific behavior change or enforce foreign policy.[25] Economic sanctions are not synonymous with trade wars (the threatened or actual infliction of economic harm to coerce a state), which can take the form of leveling import tariffs or engaging in industrial espionage to inflict broad damage.[26] By contrast, economic sanctions often focus on regime change in most cases.[27] But, ironically, economic sanctions often do little to impact the leaders they target. Sanctions also are critiqued for their inefficacy and, in some cases, for aggravating human rights violations.[29] In many weak nation-states, writes Professor Robert Paper, "external pressure is more likely to enhance the national legitimacy of rulers than undermine it."[30] Implementation is often not smooth either. Emblematic of this problem, in response to Russia's invasion of Ukraine, Canada has imposed aggressive sanctions on Russia, including on over 1,500 Russians[31]—even though over 1,200 individuals subject to those sanctions were still free to enter Canada for a significant period of time.[32] The long-time President of the Council on Foreign Relations Richard Haass has noted economic sanctions often "turn out to be little more than expressions of US preferences that hurt American economic interests without changing the target's behavior for the better."[33]

Some observers warn the frequent recourse to sanctions has led state actors to develop "sanction resistance."[34] For example, the use of bilateral currency swaps connecting

[25] *See, e.g.*, *Canadian Sanctions Related to Russia*, Government of Canada, https://www.international.gc.ca/world-monde/international_relations-relations_internationales/sanctions/russia-russie.aspx?lang=eng (last modified Dec. 16, 2022).

[26] Robert A. Pape, *Why Economic Sanctions Do Not Work*, 22:2 International Security 94 (1996).

[27] Gary Clyde Hufbauer et al., Economic Sanctions Reconsidered 67 (3rd ed. 2008).

[28] Pape, *supra* note 26, at 107.

[29] Ania Bessonov, *What Are Sanctions – And Do They Even Work?* CBC (Dec. 22, 2022), https://www.cbc.ca/news/ask-faq-sanctions-1.6693984.

[30] Pape, *supra* note 26, at 107.

[31] Global Affairs Canada, *Canada Imposes New Sanctions On Russian, Iranian and Myanmar Regimes*, Government of Canada (Dec 9, 2022) https://www.canada.ca/en/global-affairs/news/2022/12/canada-imposes-new-sanctions-on-russian-iranian-and-myanmar-regimes.html.

[32] Anja Karadeglija, *Hundreds of Russians Sanctioned over Ukraine War Can Still Enter Canada*, National Post (Oct. 26, 2022), https://nationalpost.com/news/politics/russians-sanctioned-over-ukraine-war-are-not-barred-from-entering-canada.

[33] Richard N. Haass, *Economic Sanctions: Too Much of a Bad Thing*, Brookings Institute (Jun. 1, 1998), https://www.brookings.edu/research/economic-sanctions-too-much-of-a-bad-thing/.

[34] Agathe Demarais, *The End of the Age of Sanctions?* Foreign Affairs (Dec. 27, 2022), https://www.foreignaffairs.com/united-states/end-age-sanctions.

national banks directly has eliminated the need for third currencies to trade—weakening the US dollar as targets of sanctions shift to non-Western payment systems.[35] Because different sanction resistance methods have arisen, unilateral sanctions have had, and risk having, less impact when implemented.[36]

Turning to cross-border data flow restrictions in lieu of economic sanctions appears enticing in light of these shortcomings and flaws. Indeed, they already produce the same outcome—economic harm. Moreover, given that the impact of cross-border data flow restrictions cannot be concealed by shifting economic burdens within a society, they result in broader harm, which is difficult for leaders to evade. While they are similar in nature to measures used in trade wars, they are also far more passive; whereas industrial espionage requires significant breaches of international law and custom, simply adjusting the valve on cross-border data flows does not. For these reasons, they appear to be a useful corollary to, or replacement for, economic sanctions.

B. An Existing Trend

Cross-border data flow restrictions are already a feature of the global economy today. For example, most data protection laws focus on the "adequacy" of recipient states' data protection laws as a prerequisite for transferring data.[37] Determinations about adequacy—or a lack of it—already lead nation-state data protection authorities to issue directives, in addition to taking other protective moves. For example, Russia has been weaponizing its own adequacy determinations and pushing for data localization. Russia's Federal Law No. 152-FZ was amended in July 2014 to require the personal data of Russians to remain localized in Russia, although this requirement did not "prescribe limitations on [...] subsequent cross-border transfer."[38] However, in July 2022, Russia amended the law to impose requirements on the transfer of Russians' personal data. This new law permitted transfers only to jurisdictions with "adequate" data protection for personal subjects' data and only after notifying the Federal Service for Supervision of Communications, Information Technology and Mass Media, or Roskomnadzor, which could prohibit the transfer within ten business days.[39] The transfer of Russians' personal data to jurisdictions with "inadequate" protections is currently only legal with the express permission of Roskomnadzor and "is restricted, until the permission is obtained."[40] These amendments took effect in March 2023.[41]

[35] *Id.*

[36] *Id.*

[37] GDPR, *supra* note 10, arts. 44–49; Federal'nyĭ Zakon RF o Personl'nykh Dannykh [Federal Law of the Russian Federation on Personal Data] 2006, No. 152-FZ, arts. 12 and 18(5).

[38] KPMG, *The "Localization of Russian Citizens"* Personal Data (2018), https://assets.kpmg/content/dam/kpmg/be/pdf/2018/09/ADV-factsheet-localisation-of-russian-personnal-data-uk-LR.pdf.

[39] Federal Law of the Russian Federation on Personal Data 2006, No. 152-FZ, art. 21(3).

[40] Alrud, *Newsletter: The Major Reform of Russian Data Protection and Information Laws In July*, 2022 (Jul. 18, 2022), https://www.alrud.com/upload/Файлы/2022_Информационные_письма/Newsletter_Reform_of_Data_Protection_laws_July_2022_ENG_(1)_(002).pdf.

[41] Stanislav Rumyantsev, *Russian Federation: Navigating Amendments to Russian Personal Data Law*, Mondaq (Sept. 2, 2022), https://www.mondaq.com/russianfederation/data-protection/1227070/navigating-amendments-to-russian-personal-data-law.

Russia is not alone in implementing such measures. Other countries, such as China, are also tightening data flows.[42]

While Russia's actions risk drastically tightening the valves, it is the EU Commission that originally took the lead in developing adequacy as the standard for cross-border data flow transfers. Article 45 of the GDPR sets forth that the Commission shall determine the "adequacy" of protections for the transfer of the personal data of member countries' citizens to third-party countries like the US. As noted above, the Commission tried to do so twice, without lasting success. In 2000, the Commission deemed "adequate" the Safe Harbour Principles framework,[43] which the highest court in Europe subsequently invalidated.[44] In 2015, the Commission deemed "adequate" the Privacy Shield,[45] which Europe's highest court also invalidated.[46] There is currently no framework in place at all, although the Commission and the United States just negotiated the skeleton of a new framework, the EU–US Data Privacy Framework,[47] which, interestingly, was announced shortly after the 2022 Russian invasion of Ukraine during a trip by President Joseph Biden to Brussels.[48] History is certain to repeat itself, however, and the framework is likely to be invalidated in the future. Meanwhile, data protection offices across Europe are already cracking down on data flows in the absence of a framework.[49]

These squabbles between allies are lamentable. They are increasingly resulting in a more splintered internet in the democratic world. Existing security alliances present a better and more suitable solution than the EU Commission's continuously failing efforts to protect the data supply chain. Such an approach is not illogical. Prosperity in the modern economy—and by extension, security—hinges on protecting the data supply chain. On this front, NATO already has the legitimacy and authority to act. NATO's governing body, the North Atlantic Council, is empowered by Article 9 of the North Atlantic Treaty to settle any disputes involving its members by any peaceful

[42] *See, e.g.*, Elizabeth C. Economy, *The Great Firewall of China: Xi Jinping's Internet Shutdown*, Guardian (Jun. 29, 2018), https://www.theguardian.com/news/2018/jun/29/the-great-firewall-of-china-xi-jinpings-internet-shutdown; Barbara Li, *What to Know About China's New Cross-Border Data Transfer Security Assessment Guidelines*, IAPP (Sept. 27, 2022), https://iapp.org/news/a/what-to-know-about-chinas-new-cross-border-data-transfer-security-assessment-guidelines/.

[43] Commission Decision 2000/520/EC of 26 July 2000 pursuant to Directive 95/46/EC of the European Parliament and of the Council on the Adequacy of the Protection Provided by the Safe Harbour Privacy Principles and Related Frequently Asked Questions Issued by the US Department of Commerce (notified under document number C(2000) 2441) 2000 O.J. (L 215).

[44] *The Court of Justice Invalidates Decision 2016/1250*, *supra* note 4.

[45] European Commission, *EU Commission and United States Agree on New Framework for Transatlantic Data Flows: EU-US Privacy Shield* (Feb. 2, 2016), https://ec.europa.eu/commission/presscorner/detail/en/IP_16_216.

[46] *CJEU Judgment in the Schrems II Case, supra* note 4.

[47] European Commission, *Questions & Answers: EU-U.S. Data Privacy Framework, Draft Adequacy Decision* (Dec. 13, 2022), https://ec.europa.eu/commission/presscorner/detail/en/qanda_22_7632.

[48] Daniel Michaels & Sam Schechner, *U.S., EU Reach Preliminary Deal on Data Privacy*, Wall Street Journal (Mar. 25, 2022), https://www.wsj.com/articles/u-s-eu-reach-preliminary-deal-on-data-privacy-11648200085.

[49] Caitlin Fennessy, *Schrems II' DPA Investigations and Enforcement: Lessons Learned*, IAPP (Jun. 17, 2021), https://iapp.org/news/a/schrems-ii-dpa-investigations-and-enforcement-lessons-learned/.

means.[50] In effect, Article 9 gives the Council the power to implement the treaty's provisions, including Article 2's commitment to "promoting conditions of stability and well-being" through the elimination of "conflict in [members'] international economic policies" and "encourag[ing] economic collaboration between any or all [members]."[51]

3. THE DRAWBACKS OF WEAPONIZING CROSS-BORDER DATA FLOW RESTRICTIONS

This section discusses the disadvantages and security challenges in using cross-border data flow restrictions as a tool of cooperative security as well as deterrence and defense.

A. Implementation Challenges

One significant challenge pertaining to the use of cross-border data flows arises at the level of implementation. Without movement from the World Trade Organization, it falls to free trade agreements to regulate cross-border data flows—something they do unevenly. For example, the Comprehensive Economic and Trade Agreement (2016) is silent on cross-border data flows,[52] while the EU–Japan Economic Partnership Agreement (2018) and the EU–Mexico Global Agreement (2016) commit merely to "reassess" cross-border data flow policy.[53] More recent treaties from the EU focus on the nature of privacy as a fundamental right, aspiring to a normative environment whereby "high standards in this regard contribute to trust in the digital economy and to the development of trade."[54]

This inattention from recent trade treaties reveals the degree to which uncertainty and confusion around *how* and *if* cross-border data flows can be regulated is widespread. Indeed, it is hard to know when data is an import or export, since data flows facilitating the operation of a business are often transmitted so quickly. Some question whether cross-border data flows are even "trade" in the classic sense at all—for example, cloud services do not, per se, trade data, but the flow of data with those services is vital to their function; sometimes data is neither an import nor an export, since transmission does not implicate either, and, indeed, is better seen as a byproduct of

[50] North Atlantic Treaty art. 9, Apr. 4, 1949, 63 Stat. 2241, 34 U.N.T.S. 243.

[51] *Id*. art. 2.

[52] Comprehensive Economic and Trade Agreement (CETA) between Canada, of the one part, and the European Union and its Member States, of the other part 2017 O.J. (L 11).

[53] European Commission, *EU-Japan Economic Partnership Agreement* (Jul. 2018), https://trade.ec.europa.eu/doclib/docs/2017/july/tradoc_155725.pdf; European Union, *Modernisation of the Trade part of the EU-Mexico Global Agreement* (Apr. 21, 2018), https://trade.ec.europa.eu/doclib/docs/2018/april/tradoc_156811.pdf.

[54] European Commission, *EU-New Zealand agreement: Documents—Digital Trade*, art. 6(1), https://policy.trade.ec.europa.eu/eu-trade-relationships-country-and-region/countries-and-regions/new-zealand/eu-new-zealand-agreement/documents_en (last visited Jan. 1, 2022).

the trade. The plethora of "data"[55] covered by such provisions is not just *personal* data but also confidential business data, public data, metadata, and machine-to-machine data. The regulation of cross-border data flows in an array of vectors involving myriad actors—private actors, cloud servers, government regulators, third parties, and intermediaries—also provides no shortage of leakage problems when it comes to regulation. In this reality, regulation might be no more than establishing a "non-peeing section of the pool."[56]

This paradigm suggests that legally embedded protections and regulations may be insufficient to regulate "digital borders." Private companies' continuing transfer of personal data between the two largest blocs of the democratic world for the five years following the striking down of the Privacy Shield *in the absence* of an adequacy agreement already suggests these agreements are not generally forceful in nature. Although agreement on substantive rules seems unlikely, even as data localization appears to be gaining traction in certain jurisdictions,[57] any path forward appears to premise transmission based on paradigms of security and trust. Moreover, democratic countries appear increasingly focused on measures like data localization laws; privacy laws; and investment control regimes (e.g., the Committee on Foreign Investment in the United States or the Canada Investment Act and that country's Director of Investment). While these are all significant instruments, they do not directly address cross-border data flows, even if they are related.[58]

B. In-Kind Responses

Adversaries are clearly aware of the advantages of exerting control over data flows. Shortly before the 2022 Russian invasion of Ukraine, Russia conducted targeted cyber attacks on Ukraine's internet infrastructure[59] and unleashed concerted disruption efforts.[60] After connectivity in Kherson was completely knocked out, it was only restored via the Russian Rostelecom infrastructure—not Ukrainian infrastructure.[61] This mirrored the Russian construction of a submarine link to Crimea following its

[55] Karine Bannelier & Anais Trotry, *What is "Data"? Definitions in International Legal Instruments on Data Protection, Cross-Border Access to Data & Electronic Evidence*, Cross Border Data Forum (Jan. 6, 2023), https://www.crossborderdataforum.org/what-is-data-definitions-in-international-legal-instruments-on-data-protection-cross-border-access-to-data-electronic-evidence/?utm_source=mailpoet&utm_medium=email&utm_campaign=a-new-article-has-been-added-by-the-cross-border-data-forum_1.

[56] Canada School of Public Service, *The New Economy Series: Governing Cross-Border Data Flows*, YouTube (Feb. 8, 2022), https://www.youtube.com/watch?v=d1RoboA8vO0.

[57] *See, e.g.*, data localization efforts in China. Cybersecurity Law, art. 37 (China).

[58] Nigel Cory & Luke Dascoli, *How Barriers to Cross-Border Data Flows Are Spreading Globally, What They Cost, and How to Address Them*, Information Technology & Innovation Foundation (Jul. 19, 2021), https://itif.org/publications/2021/07/19/how-barriers-cross-border-data-flows-are-spreading-globally-what-they-cost/.

[59] Patrick Howell O'Neill, *Russia Hacked an American Satellite Company One Hour Before the Ukraine Invasion*, MIT Technology Review (May 10, 2022), https://www.technologyreview.com/2022/05/10/1051973/russia-hack-viasat-satellite-ukraine-invasion/.

[60] *Internet Disruptions Registered as Russia Moves In On Ukraine*, Netblocks (Feb. 24, 2022), https://netblocks.org/reports/internet-disruptions-registered-as-russia-moves-in-on-ukraine-W80p4k8K.

[61] *Id*.

invasion of the region.[62] In 2022, following the Russian occupation in the regions of Kherson and Donetsk, Russians sold cell phones to Russian numbers only on presentation of proof of a passport, essentially facilitating the tracking of individuals.[63] All of these moves facilitated the collection of data from individuals in a manner that enhanced tracking, surveillance, and intelligence-gathering. Such maneuvers are likely to continue—especially as the democratic world responds in kind.

As one observer noted, personal data such as location data and the content of communications—"all of this is personal data. If [data processors] make this data accessible to countries or servers in Russia, which is actively targeting civilians at your border, I daresay this is actually a change that should be taken into account."[64] Shortly after the conflict commenced, Russia also unleashed a wave of policies requiring government agencies to undertake localization efforts.[65] As noted above, Russia has been tightening its internet policy and experimenting with disconnection of the Russian internet from the rest of the internet—even as it seeks to consume data flows from foreign jurisdictions.[66] Ultimately, these policies should not be feared. The entrenchment of authoritarianism in Russia indeed contributes to the fragmentation of cross-border data flows, but a greater risk comes from states in the democratic bloc permitting the personal data of their citizens to enter Russia in the first place. As the Russian invasion of Ukraine demonstrates, Russia is deeply aware of the benefits accruing from unidirectional cross-border data flows—even as the democratic bloc commits to a model where nations decide on cross-border data flows on a transfer-by-transfer, not state-by-state, basis. When it comes to transferring personal data, a paradigm shift toward a "safe data zone" would make states safer. Unless otherwise restricted, data will flow. As the world is rapidly becoming one where "geography ultimately determines what data is allowed to flow,"[67] it is incumbent on the democratic world to address this new reality.

C. Lack of Clarity

There is confusion around what is actually being regulated by cross-border data flows, including e-commerce, monopoly power, and privacy. With countries like the United States lacking broad, meaningful privacy or data protection legislation, there are

[62] Sebastian Moss, *How Russia Took Over the Internet in Crimea and Eastern Ukraine*, Data Center Dynamics (Feb. 25, 2022), https://www.datacenterdynamics.com/en/analysis/how-russia-took-over-the-internet-in-crimea-and-eastern-ukraine/.

[63] Adam Satariano & Scott Reinhard, *How Russia Took Over Ukraine's Internet in Occupied Territories*, New York Times (Aug. 9, 2022), https://web.archive.org/web/20220902022609/https://www.nytimes.com/interactive/2022/08/09/technology/ukraine-internet-russia-censorship.html.

[64] *EU Privacy Regulators Are Scrutinising Data Flows to Russia, supra* note 20.

[65] Luca Bertuzzi, *Russia Reportedly Seeks Tighter Control over the Internet*, Euractiv (Mar. 7, 2022), https://www.euractiv.com/section/digital/news/russia-reportedly-seeks-tighter-control-over-the-internet/.

[66] Alla Naglis & Xenia Melkova, *Data Localization in Russia: Now Backed with Big Fines*, King & Spalding (Jun. 19, 2019), https://kslawemail.com/283/5445/uploads/ca.pdf.

[67] Robert K. Knake, *Weaponizing Digital Trade: Creating a Digital Trade Zone to Promote Online Freedom and Cybersecurity*, Council on Foreign Relations (Sept. 2020), https://cdn.cfr.org/sites/default/files/report_pdf/weaponizing-digital-trade_csr_combined_final.pdf.

significant challenges in establishing networks of trust that ensure the maintenance of certain policy objectives. Clarity challenges are one of the reasons why figures like Rob Knake, currently Deputy National Cyber Director for Budget and Policy in the Biden Administration's Office of the National Cyber Director, have called for "digital trade zones."[68] Knake points to recent multilateral trade agreements as exhibiting an attempt to create such zones (e.g., the Canada–United States–Mexico Agreement expressly requires parties not to prohibit or restrict cross-border data flows of personal information if the activity is business-related).[69] However, this plea for a multilateral approach toward data localization, privacy protection, and data flows keeps the issues squarely within the framework of trade, not security. Moreover, trade treaties have always permitted derogations for security.[70]

4. A NEW ROLE FOR NATO?

The above discussion points to a new potential role for NATO in articulating cross-border data flows as a matter of cooperative security as well as deterrence and defense.[71] The North Atlantic Council, the alliance's main decision-making body, makes NATO different from the other international organizations listed above. For example, the Five Eyes alliance is circumscribed to exchanging signals intelligence. NATO also differs from the G-7 and the D-10, which are merely exchange forums. Article 9 of the North Atlantic Treaty allows the Council to create "subsidiary bodies as may be necessary" to achieve the goals of the alliance.[72] NATO has already set up multiple committees under this power (most famously, its Military Committee).[73] There is nothing to prevent NATO from setting up a committee—or endowing an existing one, such as the Cyber Defence Committee—with the purpose of laying out a cross-border data flows framework for its members, with an agreement that obviates the Safe Harbour Principles, Privacy Shield, or Data Privacy Frameworks with the purpose of establishing a "safe data zone" to facilitate data flows among members. Such a body could retain the power to issue embargoes on cross-border data flows to specific jurisdictions (e.g., Russia), while otherwise leaving transfer decisional authority to the EU or sovereign states.

One might think that the main external obstacle might be how such rules or regulations interact with EU law and, specifically, the GDPR. After all, the GDPR is a significant regulatory instrument: It has been called "the most consequential

[68] *Id.*

[69] Canada–United States–Mexico Agreement, art. 19.11(1) ("Definitions").

[70] *See, e.g.*, General Agreement on Tariffs and Trade, art. XXI ("Security Exceptions") and General Agreement on Trade in Services, art. XIV bis ("General Exceptions").

[71] North Atlantic Treaty Organization, NATO 2022 Strategic Concept, https://www.nato.int/strategic-concept/.

[72] *Committees*, North Atlantic Treaty Organization (Oct. 7, 2022), https://www.nato.int/cps/en/natohq/topics_49174.htm.

[73] *Id.*

regulatory development in information policy in a generation."[74] This is because it is loaded with rights provisions regarding the lawfulness of data processing and conditions for consent to such processing. Since the Snowden revelations about egregious overreaches in American information-gathering—effectively, spying—on its own citizens and pretty much everyone else, the Europeans have renewed their focus on privacy and data protection. European courts have used the language in the GDPR to focus on how data-sharing agreements with the US have lacked a limit on "the powers conferred upon US authorities and [the] lack [of] actionable rights for EU subjects against US authorities" when it comes to American surveillance practices.[75]

For several reasons, though, the GDPR does not apply to NATO. As the GDPR notes in Article 2,[76] its provisions do not apply to activities that "fall outside the scope of [EU] law," as NATO does. The founding treaties of the EU establish a particularly strong exemption to NATO from EU law, with the Treaties of the European Union stating that EU law "shall not prejudice the specific character of the security and defence policy of certain Member States."[77] Those treaties also set forth that EU law "shall respect the obligations of certain Member States, which see their common defence realised in the North Atlantic Treaty Organization (NATO), under the North Atlantic Treaty."[78] The GDPR also explicitly states its provisions do not apply to data processing when EU member states are carrying out activities that fall within the scope of their obligations on common foreign and security policy.[79] The EU's founding treaties make clear EU law shall also not affect "the specific character of the security and defence policy of certain Member States."[80] In the case of NATO, this provision is paramount. Although the EU is a supranational organization with certain legislative power assigned to it by its member states, in areas of defense, the EU's founding treaties respect the sovereignty of individual actors and their commitments to NATO. Treating cross-border data flows as a security matter could be justified under this provision. Supranational regulation of cross-border data flows is not entirely hypothetical, given the willingness of European states to let the EU do just that under the auspices of the GDPR.

The EU's formative treaties provide that agreements like the North Atlantic Treaty shall not be affected or undermined by other requirements of the EU's constituting legal documents.[81] The EU's founding treaties speak for themselves regarding the

[74] Chris Jay Hoofnagle et al., *The European Union General Data Protection Regulation: What It Is and What It Means*, Information & Communications Technology Law (Feb. 10, 2019), https://www.tandfonline.com/doi/full/10.1080/13600834.2019.1573501?cookieSet=1.

[75] *CJEU Judgment in the Schrems II Case, supra* note 4.

[76] GDPR, *supra* note 10, art. 2.

[77] Consolidated Versions of the Treaty on European Union and the Treaty on the Functioning of the European Union art. 42, Oct. 26, 2012, 2012 O.J. (C 326) [hereinafter Treaties of the European Union].

[78] *Id*.

[79] GDPR, *supra* note 10, art. 2.

[80] Treaties of the European Union, *supra* note 77, art. 42.

[81] Consolidated Version of the Treaty on the Functioning of the European Union, art. 351, May 9, 2008, 2008 O.J. (115), https://eur-lex.europa.eu/LexUriServ/LexUriServ.do?uri=CELEX:12008E351:EN:HTML.

supremacy of the realization of common defense objectives. They also enshrine a respect for, and "strict observance" of, international law—a fundamental principle of which is *pacta sunt servanda*. This principle requires parties to perform their international treaty obligations—such as carrying out their obligations under NATO—in good faith. As for more recent documents like the EU Charter of Fundamental Rights, which also contains provisions protecting data rights, these have limited reach. Like the GDPR, the Charter only applies to "institutions, bodies, offices and agencies of the Union... when they are implementing [EU] law."[82] NATO, an intergovernmental military alliance with a membership list largely overlapping with (but nonetheless different from) the EU, is not such a body. The Charter does not apply to it. Finally, and perhaps most importantly, there is a realpolitik dimension at play: The EU has recently shrunk. Since the advent of Brexit, the power of European courts has diminished with the very real loss of a member state, and the possibility that others might also leave. Already, the UK has taken a more flexible approach to data flows than the EU courts tolerated,[83] reversing undesired decisions pertaining to cross-border data flows from EU courts.

5. CONCLUSION

By moving discussions around data flows to a forum where members share security interests, like NATO, the US and most European countries may be able to move past their narrow conceptions of security. Such a move would not be out of step for the alliance, since NATO is already attempting to adopt a cyber-conscious posture. For example, in July 2016, its members signed a Cyber Defense Pledge[84] committing to enhancing their cyber defenses (an action likely catalyzed by cyber attacks against many of NATO's own websites during the Russo-Ukrainian War in 2014).[85]

The hardest challenge in turning NATO into a "safe data zone" will not be EU law. It will be finding the morale among NATO members to reform the organization to tackle novel security challenges as such. The end of the Afghanistan project was highly mediatized and bitter. But this fact should not demoralize multilateralism. Instead, it should convey the urgency of the need to act and reform the alliance, since the alternative to revitalization is vulnerability to new threats.[86] NATO has always been designed to protect common security interests. In the past, NATO found its raison

[82] Charter of Fundamental Rights of the European Union, art. 51, Oct. 26, 2012, 2012 O.J. (C 326), https://fra.europa.eu/en/eu-charter/title/title-vii-general-provisions.

[83] Peter Swire, *U.K.'s Post-Brexit Strategy on Cross-Border Data Flows*, Lawfare (Sept. 1, 2021), https://www.lawfareblog.com/uks-post-brexit-strategy-cross-border-data-flows.

[84] *Cyber Defence*, North Atlantic Treaty Organization (Mar. 23, 2022), https://www.nato.int/cps/en/natohq/topics_78170.htm.

[85] Adrian Croft & Peter Apps, *NATO Websites Hit in Cyber Attack Linked to Crimea Tension*, Reuters (Mar. 15, 2014), https://www.reuters.com/article/us-ukraine-nato-idUSBREA2E0T320140316.

[86] Rachel Ellehuus, *NATO Futures: Three Trajectories*, Center for Strategic & International Studies (Jul. 21, 2021), https://www.csis.org/analysis/nato-futures-three-trajectories.

d’être in roles like standing up to the Soviet Union during the Cold War and tackling terrorism after 9/11. Today, the shared security interests of its members are increasingly channeled through dependence on civil and critical infrastructure, which includes the free flow of data. If data truly is an essential commodity, allowing restrictions on cross-border data flows between allies endangers their future collective prosperity—a security threat of potentially existential proportions. The corollary is that they are such a threat to adversaries, too.

2023 15th International Conference on Cyber Conflict
Meeting Reality
T. Jančárková, D. Giovannelli, K. Podiņš, I. Winther
2023 © NATO CCDCOE Publications, Tallinn

Limits on Information Operations Under International Law

Talita Dias
Senior Research Fellow
International Law Programme
Chatham House
The Royal Institute of International Affairs
tdias@chathamhouse.org

Abstract: Information or influence operations have been part and parcel of domestic and international life for centuries, having been used for a range of private and public purposes – from commercial advertisement to political propaganda. Yet, given their unprecedented scale and speed, digital information operations carried out by states and non-state actors have given rise to new international legal challenges. Notably, they have played an increasingly significant role in several offline harms – from health misinformation and disinformation hampering the fight against COVID-19 to online hate paving the way for acts of violence around the world. This calls into question the orthodox view that information operations do not violate international law. The purpose of this paper is to assess the extent to which existing international law – including general rules and principles and those specific to broadcasting and telecommunications – limits the digital deployment of information operations by states and non-state actors. It does so by first addressing the vexing yet overlooked question of factual and legal causation between those operations and some of the harmful consequences attributed to them. The paper then turns to how key international rules and principles, such as the principle of non-intervention, obligations of due diligence, and international human rights law, apply together to four key categories of information operations: propaganda, misinformation and disinformation, malinformation, and online hate speech.

Keywords: *information operations, information and communications technologies, propaganda, misinformation and disinformation, causation, international law*

1. INTRODUCTION

Information or influence operations can be defined as 'any coordinated or individual deployment of digital resources for cognitive purposes to change or reinforce attitudes or behaviours of the targeted audience'.[1] Prime examples are (a) propaganda (the selective presentation of information, facts or views in order to emotionally influence and/or manipulate audiences), (b) misinformation (the dissemination of false information without knowledge of its inaccuracy and/or the intention to deceive) and disinformation (the dissemination of knowingly or deliberately false information), (c) malinformation (the dissemination of verifiable information, personal views or opinions to cause harm, such as doxing), and (d) hate speech (the use of rhetoric to attack, denigrate or dehumanize individuals or groups on the basis of protected characteristics, such as race, ethnicity, nationality, religion, gender, sexual orientation, or disability).[2]

Although each type of information operation has distinctive traits and legal consequences, the four categories often overlap. Most recent, real-world examples of information operations combine traits of more than one of these categories. Moreover, with the rise of information and communications technologies (ICTs), such as the Internet, and the proliferation of personal computers, smartphones, and easily accessible online platforms, such as search engines and social media, the vast majority of information operations have now moved online.[3] Given the ease, speed, and scale with which content can be disseminated online, information operations have also increased in number, sophistication, and pervasiveness.[4]

[1] Oxford Institute For Ethics Law and Armed Conflict (ELAC), 'The Oxford Process: The Oxford Statement on International Law Protections in Cyberspace: The Regulation of Information Operations and Activities' <www.elac.ox.ac.uk/the-oxford-process/the-statements-overview/the-oxford-statement-on-the-regulation-of-information-operations-and-activities/> accessed 7 March 2023; Tsvetelina van Benthem, Talita Dias, and Duncan Hollis, 'Information Operations under International Law' (2022) 55 Vanderbilt Journal of Transnational Law 1217, 1219, n 1.

[2] *See* van Benthem, Dias, and Hollis (n 1) 1228–29; Claire Wardle and Hossein Derakhshan, 'Information Disorder: Toward an Interdisciplinary Framework for Research and Policymaking' (*Council of Europe*, 27 September 2017) 5 <https://rm.coe.int/information-disorder-report-version-august-2018/16808c9c77> accessed 7 March 2023; UNGA, 'Disinformation and Freedom of Opinion and Expression – Report of the Special Rapporteur on the Promotion and Protection of the Right to Freedom of Opinion and Expression, Irene Khan' (2021) UN Doc A/HRC/47/25, [10]–[12], [15].

[3] *See eg* Jacob T Rob and Jacob N Shapiro, 'A Brief History of Online Influence Operations' (*Lawfare*, 28 October 2021) <www.lawfareblog.com/brief-history-online-influence-operations> accessed 7 March 2023; Samantha Bradshaw, 'Influence Operations and Disinformation on Social Media' (Centre for International Governance Innovation, 23 November 2020) <www.cigionline.org/articles/influence-operations-and-disinformation-social-media/> accessed 7 March 2023.

[4] For a comprehensive data-driven study of information operations, see Diego A Martin, Jacob N Shapiro, and Julia G Ilhardt, 'Online Political Influence Efforts Dataset – Version 3.0' (*Empirical Studies of Conflict Project*, 3 February 2022) <https://esoc.princeton.edu/publications/trends-online-influence-efforts>; Diego A Martin, Jacob N Shapiro, and Julia G Ilhardt, 'Trends in online foreign influence efforts – Version 2.0' (*Princeton University*, 5 August 2020) <https://scholar.princeton.edu/sites/default/files/jns/files/trends_in_online_influence_efforts_v2.0_aug_5_2020.pdf> accessed 7 March 2023.

For instance, the COVID-19 pandemic saw a tsunami of false or misleading information about health treatments, preventive measures, and the origins of the virus.[5] Some of these campaigns were carefully orchestrated to cause physical harm, social discord, or political polarization, while others were simply forwarded by users who naively thought they were acting for the public good.[6] These operations were often combined with political propaganda, such as claims originating from Russia or China that Western measures to tackle the pandemic were ineffective or unwarranted.[7] There were also instances of misinformation, disinformation, or malinformation relating to the origin of the virus that were combined with online hate messages. An example was the labelling of coronavirus as the 'Chinese' or 'Asian' virus, and the ensuing stigmatization and targeting of Asians.[8]

Similarly, foreign and domestic information operations during the 2016 and 2020 US presidential elections and, more recently, the 2022 US mid-term elections combined political propaganda, misinformation and disinformation, and online hate speech.[9] Notably, the United States House Select Committee on the January 6 Attack recently concluded that former US president Donald Trump's tweets, containing, at once, false claims of electoral fraud, incitement to hatred and violence, and carefully orchestrated political propaganda, were instrumental in spurring on the Capitol riots.[10]

[5] Nick Robins-Early, 'Desperation, Misinformation: How the Ivermectin Craze Spread Across the World' (*Guardian*, 24 September 2021) <www.theguardian.com/world/2021/sep/24/ivermectin-covid-peru-misinformation> accessed 7 March 2023; Rick Rouan, 'Fact Check: Study Falsely Claiming Face Masks are Harmful, Ineffective is Not Linked to Stanford' (*USA Today*, 24 April 2021) <www.usatoday.com/story/news/factcheck/2021/04/24/fact-check-study-falsely-claiming-masks-harmful-isnt-stanfords/7353629002/> accessed 7 March 2023; Jon Cohen, 'Scientists "strongly condemn" rumors and conspiracy theories about origin of coronavirus outbreak' (*Science*, 19 February 2020) <www.science.org/content/article/scientists-strongly-condemn-rumors-and-conspiracy-theories-about-origin-coronavirus> accessed 7 March 2023.

[6] *See* Julie Posetti and Kalina Bontcheva, 'Disinfodemic: Dissecting Responses to COVID-19 Disinformation' (UNESCO, 2020) 5 < https://unesdoc.unesco.org/ark:/48223/pf0000374416:> accessed 7 March 2023; European Commission, 'Joint Communication to the European Parliament, the European Council, the Council, the European Economic and Social Committee and the Committee of the Regions: Tackling Covid-19 Disinformation – Getting The Facts Right' (*EUR-Lex*, 2020) < https://eur-lex.europa.eu/legal-content/EN/TXT/?uri=CELEX:52020JC0008> accessed 7 March 2023.

[7] Reid Standish, 'Study Shows How Russian, Chinese Disinformation About COVID-19 Evolved During The Pandemic' (*Radio Free Europe*, 2 December 2021) <www.rferl.org/a/russia-china-covid-disinformation-campaigns/31590996.html> accessed 7 March 2023.

[8] Mark Townsend and Nosheen Iqbal, 'Far Right Using Coronavirus as Excuse to Attack Asians, Say Police' *The Guardian* (29 August 2020) <www.theguardian.com/society/2020/aug/29/far-right-using-coronavirus-as-excuse-to-attack-chinese-and-south-east-asians> accessed 7 March 2023.

[9] *See eg* National Intelligence Council, 'Foreign Threats to the 2020 US Federal Elections' (ICA 2020-00078D, 10 March 2021) < www.dni.gov/files/ODNI/documents/assessments/ICA-declass-16MAR21.pdf> accessed 7 March 2023; 'Threat Report: The State of Influence Operations 2017–2020' (Facebook, May 2021) < https://about.fb.com/wp-content/uploads/2021/05/IO-Threat-Report-May-20-2021.pdf>; Abigail Abrams, 'Here's What We Know So Far About Russia's 2016 Meddling' (*Time*, 18 April 2019) <https://time.com/5565991/russia-influence-2016-election/> accessed 7 March 2023; Insikt Group, 'Malign Influence During the 2022 US Midterm Elections' (*Recorded Future*, 13 October 2022) <www.recordedfuture.com/malign-influence-during-the-2022-us-midterm-elections-disinformation-misinformation> accessed 7 March 2023.

[10] Select Committee to Investigate the January 6th Attack on the United States Capitol, 'Final Report' (117th Congress Second Session House Report 117–663, 22 December 2022) 55, ch 6 <www.govinfo.gov/content/pkg/GPO-J6-REPORT/pdf/GPO-J6-REPORT.pdf> accessed 7 March 2023.

Even before the Russian invasion of Ukraine on 24 February 2022 and at least since the annexation of Crimea in 2014,[11] the hybrid war in Ukraine has also been marked by an intricate confluence of different information operations.[12] In particular, Russia continues to dub the invasion a 'special military operation'.[13] And this has been justified by false allegations spread online as well as offline that Ukraine is run and inhabited by 'neo-Nazis' who have subjected ethnic Russians in Eastern Ukraine to genocide and war crimes.[14] Several online posts, including the infamous RIA Novosti op-ed setting out a clear plan to 'de-Nazify' Ukraine, contained explicit messages of hate and direct threats of violence against Ukrainians.[15] Deployed alongside kinetic and cyber operations, such influence operations have been instrumental in fuelling and sustaining the conflict by garnering support from soldiers, policymakers, and the Russian public.[16]

As this brief account shows, public discourse in the media and policy circles has been quick to link online information operations with real-world consequences ranging from more abstract threats against democracy to physical harm to individuals. There is indeed anecdotal and increasing scientific evidence that influence operations have *contributed* to many of the consequences mentioned above.[17] At the same

[11] *See eg* Kim Zetter, 'Inside the Cunning, Unprecedented Hack of Ukraine's Power Grid' (*Wired*, 3 March 2016) <www.wired.com/2016/03/inside-cunning-unprecedented-hack-ukraines-power-grid/> accessed 7 March 2023.

[12] Joe Tidy, 'Ukraine Says It Is Fighting First "Hybrid War"' (*BBC News*, 4 March 2022) <www.bbc.co.uk/news/technology-60622977> accessed 7 March 2023.

[13] 'Ukraine Conflict: Russian Forces Attack from Three Sides' (*BBC News*, 24 February 2022) <www.bbc.co.uk/news/world-europe-60503037> accessed 7 March 2023; E Eduardo Castillo and Jamey Keaten, 'Putin Says Ukraine "Special Military Operation" Is Taking Longer Than Expected' (*PBS News Hour*, 7 December 2022) <www.pbs.org/newshour/world/putin-says-ukraine-special-military-operation-is-taking-longer-than-expected> accessed 7 March 2023.

[14] Jack Goodman and others, 'War in Ukraine: The Making of a New Russian Propaganda Machine' (*BBC News*, 29 May 2022) <www.bbc.co.uk/news/world-europe-61441192> accessed 7 March 2023; BBC Reality Check Team, 'Ukraine Crisis: Vladimir Putin Address Fact-Checked' (*BBC News*, 22 February 2022) <www.bbc.co.uk/news/60477712> accessed 7 March 2023; Alexey Kovalev, 'Russia's Ukraine Propaganda Has Turned Fully Genocidal' (*Foreign Policy*, 9 April 2022) <https://foreignpolicy.com/2022/04/09/russia-putin-propaganda-ukraine-war-crimes-atrocities/> accessed 7 March 2023.

[15] Mariia Kravchenko, 'What Should Russia Do with Ukraine?' [Translation of a propaganda article by a Russian publication] (*Medium*, 4 April 2022) <https://medium.com/@kravchenko_mm/what-should-russia-do-with-ukraine-translation-of-a-propaganda-article-by-a-russian-journalist-a3e92e3cb64> accessed 7 March 2023.

[16] 'Defending Ukraine: Early Lessons from the Cyber War' (*Microsoft*, 22 June 2022) 3–4, 12–22 <https://query.prod.cms.rt.microsoft.com/cms/api/am/binary/RE50KOK> accessed 7 March 2023; TS Allen and AJ Moore, 'Victory without Casualties: Russia's Information Operations' (2018) 48 Parameters 59 <https://press.armywarcollege.edu/cgi/viewcontent.cgi?article=2851&context=parameters> accessed 7 March 2023.

[17] *Eg* 'The Online Information Environment: Understanding How the Internet Shapes People's Engagement with Scientific Information' (*Royal Society*, January 2020) <https://royalsociety.org/-/media/policy/projects/online-information-environment/the-online-information-environment.pdf?la=en-GB&hash=691F34A269075C0001A0E647C503DB8F>; Jon Bateman and others, 'Measuring the Effects of Influence Operations: Key Findings and Gaps From Empirical Research' (*Carnegie Endowment for International Peace*, 18 June 2021) <https://carnegieendowment.org/2021/06/28/measuring-effects-of-influence-operations-key-findings-and-gaps-from-empirical-research-pub-84824> accessed 7 March 2023; Laura Courchesne, Isra M Thange, and Jacob N Shapiro, 'Review of Social Science Research on The Effects of Influence Operations' *Empirical Studies of Conflict Report* (17 July 2021) <https://scholar.princeton.edu/sites/default/files/cts_2021_effects_of_ios_evidence_review.pdf> accessed 7 March 2023; Jon Agley and Yunyu Xiao, 'Misinformation About COVID-19: Evidence for Differential Latent Profiles and a Strong Association with Trust in Science' (2021) 21 BMC Public Health 89.

time, information operations are speech or verbal acts. As such, they cannot *in and of themselves* or *directly* cause any of the harmful outcomes that they are usually associated with. After all, addressees still need to act upon an information operation for it to have any effect on the outside world. This raises both factual and legal questions.[18]

The factual conundrum lies in how to properly measure and label the actual link between influence operations and the different effects with which they are somehow connected. And the key legal question addressed in this paper is to what extent, if any, states can be held responsible under existing international law for 'mere' speech acts or their alleged consequences, whether these originate from their own agents or private entities. Any answer to this question requires a careful assessment of the complex interaction between different international legal rules or regimes at play.[19] To complicate things further, which rules or regimes apply depends on the type of information operation in question and other factual circumstances, such as who the speaker is, where the speech act occurs, and the consequences to which it may lead. In particular, one must not only assess the rules applicable to state behaviour online, such as the principle of non-intervention or the prohibition of certain forms of propaganda but also factor in the right of individuals to receive and impart information under international human rights law.[20]

It is important to reiterate that existing international law applies *by default* and *in its entirety* to ICTs, just as it does to other technologies.[21] This is true insofar as conventional and customary rules have a general scope of application and can be interpreted to accommodate new phenomena.[22] Though helpful in clarifying how international law applies online, 'domain-specific' state practice and *opinio juris* (i.e., the formative elements of customary international law) are not necessary to prove that existing rules and principles apply to digital information operations. Several rules of international law – general and specific – already regulate different types of such operations, as will be explained below.[23] Though drafted decades before digital technologies emerged, the content of these rules is sufficiently broad and flexible to cover both analogue and digital information operations.

[18] van Benthem, Dias, and Hollis (n 1) 1270; Henning Lahmann, 'Infecting the Mind: Establishing Responsibility for Transboundary Disinformation' (2022) 33 European Journal of International Law 411, 421.

[19] van Benthem, Dias, and Hollis (n 1) 1275–76.

[20] ibid 1237, 1240.

[21] Dapo Akande, Antonio Coco, and Talita de Souza Dias, 'Drawing the Cyber Baseline: The Applicability of Existing International Law to the Governance of Information and Communication Technologies' (2022) 99 International Law Studies 4.

[22] ibid 15.

[23] *See* van Benthem, Dias, and Hollis (n 1); Henning Lahmann, 'Information Operations and the Question of Illegitimate Interference under International Law' (2020) 53 Israel Law Review 189; Lahmann, 'Infecting the Mind' (n 18); Björnstjern Baade, 'Fake News and International Law' (2018) 29 European Journal of International Law 1357.

Against this background, this paper offers a doctrinal analysis of the extent to which existing international law limits different types of information operations. It first examines the issue of factual and legal causation between influence operations and the various outcomes that have been attributed to them, focusing on the relevant rules and principles of international law. It then assesses how various international legal rules or regimes apply concurrently to the four main types of information operations identified above: propaganda, misinformation and disinformation, malinformation, and online hate speech.

The paper argues that, while there is growing scientific evidence of a causal link between information operations and different harmful outcomes, there are no general standards of causation in international law, including for speech acts. Each rule of international law contains its own standard of causation, if any. Several rules and principles of international law that are found to apply to information operations do not require a causal link between the prohibited or required conduct, including speech acts, and any results thereof. Instead, only an intention to cause or constructive knowledge of a certain harm is usually required. This means that many information operations are well within the scope of international law, whether or not they actually lead to any real-world consequences. This paper also posits that whether international law prohibits, permits or somehow restricts different information operations depends on the close interaction between various international legal rules and principles, including their causation or knowledge standards. This has led to some confusion about the exact extent to which international law limits those operations. Yet a careful analysis of applicable rules and principles reveals that information operations are significantly limited by existing international law. This is true insofar as their deployment by states or non-state actors is done with an intention or constructive knowledge that online or offline harms to protected subjects or objects, such as other states or individuals, may occur.

2. INFORMATION OPERATIONS AND THE CAUSATION CONUNDRUM IN INTERNATIONAL LAW

Causation or causality is a notoriously vexing question in different legal fields.[24] It may be straightforward to establish a causal link between events or behaviours and their temporally and physically close results, such as the shooting that actually kills someone. But it is difficult to do so for acts that occur further up and earlier on in the chain of events. This is especially the case with speech acts, including information operations, that may have an *influence* yet not a direct impact on the behaviour that eventually causes a certain harmful result 'in the physical world'.[25] Even greater

[24] *See generally* HLA Hart, 'Causation in Legal Theory' in HLA Hart and Tony Honoré, *Causation in the Law* (OUP 1985) 84.

[25] Lahmann, 'Infecting the Mind' (n 18) 421–22; Richard Ashby Wilson, *Causation in International Speech Crimes. In Incitement on Trial: Prosecuting International Speech Crimes* (CUP 2017) 71–72.

uncertainty surrounds the extent to which *online* content affects offline behaviour or events. Key factors include the distance, anonymity, and often automated nature of online speakers, many of whom are simply bots.[26] The unprecedented amount of information available online also makes it difficult to pinpoint which pieces of content affect multiple addressees.[27]

To be sure, there is growing research, especially in the fields of behavioural economics, psychology, and anthropology, that empirically supports the existence of causal links between online information operations and the consequences associated with them, such as their impact on human health and violence.[28] But this work is still in the early stages. One of the difficulties lies in distinguishing between causation and correlation (the latter may be evidence of the former but the two remain separate concepts).[29] Likewise, it is often hard to measure the exact effects of massive information campaigns on both particular individuals and society at large, especially in the case of diffuse or non-physical harms such as threats to democracy or trust in science.[30]

International law does not escape these conundrums. And the fragmentation between different sub-fields of the discipline has compounded the uncertainty. Not only is causation an under-explored issue in international legal scholarship and practice but there is also great confusion about applicable standards of causation.[31] As Plakokefalos points out, there is first and foremost conflation between the concepts and tests for factual (or natural) and legal causation (also called 'scope of responsibility'): while the former relates to the material or empirical cause-and-effect linkage between an act or omission and a result, the latter has to do with moral and policy considerations that narrow down the scope of state responsibility, exempting it for certain factual causes.[32] Examples of such considerations include reasonableness, fairness, proximity, and foreseeability.[33]

Even when the two concepts or stages of the causal analysis are separated, there is a notorious lack of consistency in the tests or standards applied by different international courts and scholars.[34] In particular, some insist on an exacting 'but for' test for factual

[26] Tal Orian Harel, Jessica Katz Jameson, and Ifat Maoz, 'The Normalization of Hatred: Identity, Affective Polarization, and Dehumanization on Facebook in the Context of Intractable Political Conflict' (2020) 6 Social Media + Society.

[27] *See* Lahmann, 'Infecting the Mind' (n 18) 436; Courchesne, Thange and Shapiro, 'Review of Social Science Research on The Effects of Influence Operations' (n 17) 2.

[28] *See* ibid 437–38 and works cited in n 17.

[29] Lahmann, 'Infecting the Mind' (n 18) 426–27; Naomi Altman and Martin Krzywinski, 'Association, correlation and causation' (2015) 12 Nature Methods 899.

[30] Lahmann, 'Infecting the Mind' (n 18) 436.

[31] Vladyslav Lanovoy, 'Causation in the Law of State Responsibility' (2022) British Yearbook of International Law 3–4 <https://doi.org/10.1093/bybil/brab008> accessed 3 January 2022; Ilias Plakokefalos, 'Causation in the Law of State Responsibility and the Problem of Overdetermination: In Search of Clarity' (2015) 26 European Journal of International Law 471, 472–73.

[32] Plakokefalos (n 31) 475. See also Lahmann, 'Infecting the Mind' (n 18) 426.

[33] Plakokefalos (n 31) 478; Lanovoy (n 31) 61–63.

[34] Plakokefalos (n 31) 490–91.

causation, which includes only the necessary or essential causes of a certain result.[35] This is often coupled with legal requirements of sufficient directness[36] or proximity.[37] Conversely, others advance less exacting factual and legal causation standards, such as those based on considerations of reasonableness and foreseeability,[38] or requiring a 'substantial contribution' to the result.[39]

The confusion primarily arises because, as others have pointed out, there is no overarching standard, principle, or rule of factual or legal causation across international law.[40] The general rules of state responsibility reflected in the International Law Commission's Articles on the Responsibility of States for Internationally Wrongful Acts, particularly Articles 2 and 12 and their commentaries, say nothing about causation, referring the question back to particular rules of international law.[41] This means that, when it comes to information operations, relevant standards of factual and legal causation will depend on the applicable rules at hand, which in turn depends on the type of operation, as will be discussed below. Arguably, such diversity of standards is not necessarily a bad thing,[42] since different rules have different aims and scopes of application. Specific standards of causation should be assessed in the light of each rule's text, purpose, and context.[43] Nevertheless, as a general matter, three sets of rules may be identified.

First, certain rules of international law hold, implicitly or explicitly, that a state must refrain from engaging in a certain act that *itself* causes a certain result. Given the close link between conduct and result, it is unlikely that speech acts will engage a state's responsibility for a breach of those rules, except in cases of complicity in another state's (principal) conduct.[44] This is arguably the case with the prohibition on the use of force, which requires states to refrain from using military force against other states

[35] Plakokefalos (n 31) 476; Lanovoy (n 31) 14. See eg *Application of the Convention on the Prevention and Punishment of the Crime of Genocide (Bosnia and Herzegovina v Serbia and Montenegro)* (Merits) [2007] ICJ Rep 43, para 462.

[36] Lanovoy (n 31) 47–54, citing eg *Bosnian Genocide* (n 35) para 462; *The M/V 'Saiga' (no 2) (Saint Vincent and Grenadines v Guinea)* (Judgment) ITLOS Reports 1999, 10, para 172; German-US Mixed Claims Commission, Administrative Decision No II, 29.

[37] Lanovoy (n 31) 54–57, citing eg *Dix Case (United States v Venezuela)* (1903–1905) 9 RIAA 119, 121.

[38] Lanovoy (n 31) 57–60, 78–79; Plakokefalos (n 31) both citing eg *Responsibility of Germany for Damage Caused in the Portuguese Colonies in the South of Africa (Portugal v Germany) ('Naulilaa Arbitration')* (1930) 2 RIAA 1013.

[39] Lahmann, 'Infecting the Mind' (n 18) 427–28, citing eg *Ndindabahizi* (Judgment) ICTR-2001-71-I (15 July 2004) para 463.

[40] Lahmann, 'Infecting the Mind' (n 18) 426; Lanovoy (n 31) 4–5; International Law Commission (ILC) 'Draft Responsibility of States for Internationally Wrongful Acts, with commentaries' (2001) A/56/10 (DARSIWA) 92–93, comm (10) to art 31.

[41] DARSIWA (n 40).

[42] *Contra* Lanovoy (n 31) 6, 83.

[43] Similarly, Plakokefalos (n 31) 471–73.

[44] *See* DARSIWA (n 40) art 16 and commentary, especially commentaries (5) and (10).

without consent.[45] The same is true for certain negative human rights obligations, such as the duty to respect life.[46]

Secondly, other rules admit more than one form of participation in the wrongful conduct and/or result, that is, participation other than 'principal' perpetration or commission. In those cases, speech acts may well amount to a breach of international law. Examples include the principle of non-intervention and the prohibition of specific forms of propaganda, which cover state *suppor*t for the activities of non-state groups, as argued below.

Thirdly, some rules do not require any assessment of causation between state conduct and result, at least in the traditional sense of factual or natural causation. Instead, considerations of knowledge, reasonableness, foreseeability, and/or probability govern the connection between wrongful conduct and any result arising thereof.[47] This is arguably the case of certain rules of international law that can be breached by an omission – that is, rules that require states to take positive action or engage in a certain course of conduct, such as general obligations of due diligence[48] and positive human rights obligations.[49] The lack of a causation requirement may be attributed to these rules' focus on a state omission rather than any particular result.[50] In fact, some rules do not even require an actual result to occur. For instance, the duty to protect the right to life does not require actual deprivation of life to occur; it only needs to be objectively foreseeable.[51]

The remaining sections will assess how existing international law applies to different information operations in light of these general remarks and specific standards of causation applicable under each relevant primary rule.

[45] *See* art 2(4), Charter of the United Nations (adopted 26 June 1945; entered into force 24 October 1945) 1 UNTS XVI. Similarly, Lahmann, 'Infecting the Mind' (n 18) 425.

[46] UN Human Rights Committee (HRC), 'General Comment No 31 [80], The Nature of the General Legal Obligation Imposed on States Parties to the Covenant' (2004) UN Doc CCPR/C/21/Rev.1/Add.1 (GC 31) para 6.

[47] *See* Antonio Coco and Talita de Souza Dias, '"Cyber Due Diligence": A Patchwork of Protective Obligations in International Law' (2021) 32 European Journal of International Law 771, 778. In the context of positive human rights obligations, see Vladislava Stoyanova, 'Causation between State Omission and Harm within the Framework of Positive Obligations under the European Convention on Human Rights' (2018) 18 Human Rights Law Review 309, 315–16; Vladislava Stoyanova, 'Fault, Knowledge and Risk Within the Framework of Positive Obligations Under the European Convention on Human Rights' (2020) 33 Leiden Journal of International Law 601, 618–19.

[48] On the structure of these rules, see generally Coco and de Souza Dias (n 47); Heike Krieger and Anne Peters, 'Due Diligence and Structural Change in the International Legal Order' in Heike Krieger, Anne Peters and Leonard Kreuzer (eds), *Due Diligence and Structural Change in the International Legal Order* (OUP 2020).

[49] HRC, GC 31 (n 46) paras 7–8.

[50] *Bosnian Genocide* (n 35) paras 429–30.

[51] *See* HRC, 'General Comment No 36 on Article 6 of the International Covenant on Civil and Political Rights, on the Right to Life' (2018) UN Doc CCPR/C/GC/36 (GC 36) paras 6–7.

3. PROPAGANDA

For centuries, propaganda and other strategies to influence or convince others have pervaded social and political life, domestically and internationally, in peacetime and war. Thus, they have not been generally prohibited under international law. However, certain types of propaganda may cross a threshold of dangerousness and are likely to incite civil unrest, domestic regime change, or war. These so-called 'hostile' or 'subversive' forms of propaganda have been prohibited under customary international law at least since the eighteenth century, when revolutionary France withdrew its public call to support independence movements abroad.[52] Subversive propaganda carried out by state organs directly, or non-state groups acting with the support of a state, that results in interference or *seeks to* interfere in the internal or external affairs of the targeted state has been traditionally considered a violation of the principle of non-intervention. This is true insofar as it removes the victim state's freedom to make key governmental decisions within its exclusive domain.[53]

The assumption is that, just like the use or threat of force, subversive propaganda directed at foreign audiences may force a state to steer its internal or external affairs against its will.[54] Thus, the Declaration on the Inadmissibility of Intervention in the Domestic Affairs of States and the Protection of their Independence and Sovereignty, adopted by the UN General Assembly in 1965, specifically included within the scope of the principle of non-intervention the duty of states to abstain 'from any defamatory campaign, vilification or hostile propaganda *for the purpose* of intervening or interfering in the internal affairs of other States'.[55] As noted by the International Court of Justice (ICJ) in the *Nicaragua* case, the principle of non-intervention:

> forbids all States or groups of States to intervene directly or indirectly in internal or external affairs of other States. […] Intervention is wrongful when it uses *methods of coercion* in regard to such choices, which must remain free ones. The element of coercion, which defines, and indeed forms the very essence of, prohibited intervention, is particularly obvious in the case of an intervention

[52] Michael G Kearney, *The Prohibition of Propaganda for War in International Law* (OUP 2007) 11–12.

[53] *See* Vernon Van Dyke, 'The Responsibility of States for International Propaganda' (1940) 34 American Journal of International Law 58, 58, 60–65, 73; Lawbence Preuss, 'International Responsibility for Hostile Propaganda against Foreign States' (1934) 28 American Journal of International Law 649, 652; Arthur Larson, 'The Present Status of Propaganda in International Law' (1966) 31 Law and Contemporary Problems 439, 445–47; John B Whitton, 'Hostile International Propaganda and International Law' (1971) 398 The Annals of the American Academy of Political and Social Science 14,15–18; Eric De Brabandere, 'Propaganda' (*Max Planck Encyclopedias of International Law*, August 2019) paras 12–16 <https://opil.ouplaw.com/display/10.1093/law:epil/9780199231690/law-9780199231690-e978> accessed 7 March 2023.

[54] *See Military and Paramilitary Activities in and against Nicaragua (Nicaragua v United States)* (Merits) [1986] ICJ Rep 14, para 205.

[55] UNGA, 'Declaration on the Inadmissibility of Intervention and Interference in the Internal Affairs of States' UN Doc A/RES/36/103 (1981) lit. j (emphasis added) <https://digitallibrary.un.org/record/27066?ln=en> accessed 7 March 2023.

> 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.[56]

Some have argued that a prohibited intervention must produce some coercive effect in the victim state, meaning that a direct causal link would have to exist between the intervening act and the coercive effect.[57] However, this view would go against the ICJ's framing of the principle, which speaks of 'coercive methods' as opposed to effects. Thus, neither propaganda nor other forms of direct or indirect interference need to be successful in causing any particular results for the principle of non-intervention to be breached.[58] In its own formulations of the principle, the UN General Assembly has referred to '*attempted threats* against the personality of the State or against its political, economic and cultural elements',[59] and behaviours that '*seek to* disrupt the unity or to undermine or subvert the political order of other States', '*designed to* intervene or interfere in the internal and external affairs of third States' and with the '*purpose* of intervening or interfering in the internal affairs of other states'.[60]

Furthermore, as noted earlier, conduct that *supports* the coercive acts of third parties is also clearly covered. This means that causation is arguably not a requirement of the principle of non-intervention insofar as particular effects need not be caused by a state for it to breach the principle. Instead, the better view is that interferences that either have a coercive purpose, employ coercive methods, or produce coercive effects are prohibited.[61] Online messages issued by a state that incite its own population, a foreign country, or its population to overthrow another state's government are examples of information operations that would likely violate the principle of non-intervention.

Propaganda that constitutes advocacy for violations of international humanitarian law (IHL), such as indiscriminate attacks against civilians, also runs contrary to states'

[56] *Nicaragua* (n 54) para 205 (emphasis added). See also Case Concerning Armed Activities in the Territory of the Congo (Democratic Republic of Congo v Uganda) (Merits) [2005] ICJ Rep 168, paras 162–64.

[57] Lahmann, 'Infecting the Mind' (n 18) 423–24; Harriet Moynihan, 'The Application of International Law to State Cyberattacks: Sovereignty and Non-Intervention' (*Chatham House*, 2 December 2019) paras 101–4 <www.chathamhouse.org/2019/12/application-international-law-state-cyberattacks> accessed 4 January 2022; Michael Schmitt (ed), *Tallinn Manual 2.0 on the International Law Applicable to Cyber Operations* (CUP 2017) 320, 322.

[58] *See also* Mohamed Helal, 'On Coercion in International Law' (2019) 52 NYU Journal of International Law and Policy 1, 43–45; van Benthem, Dias and Hollis (n 1) 40–41.

[59] UNGA, 'Declaration on Principles of International Law concerning Friendly Relations and Cooperation among States in accordance with the Charter of the United Nations' (1970) UN Doc A/RES/2625(XXV) para 1, principle 25 (emphasis added).

[60] UNGA, 'Declaration on the Inadmissibility of Intervention and Interference in the Internal Affairs of States' UN Doc (1981) A/RES/36/103, letters f, h, and j, respectively (emphasis added).

[61] Similarly, van Benthem, Dias, and Hollis (n 1) 40–41; 'Letter from the Minister of Foreign Affairs to the President of The House of Representatives on the International Legal Order in Cyberspace – Appendix: International Law in Cyberspace' (5 July 2019) 3 <www.government.nl/documents/parliamentary-documents/2019/09/26/letter-to-the-parliament-on-the-international-legal-order-in-cyberspace> accessed 7 March 2023; Suella Braverman, 'International Law in Future Frontiers, Speech at Chatham House' (19 May 2022) <www.ukpol.co.uk/suella-braverman-2022-speech-at-chatham-house/> accessed 7 March 2023.

duty to ensure respect for IHL,[62] enshrined in Article 1 Common to the Geneva Conventions,[63] Article 1 of Additional Protocol I to the Conventions,[64] and customary international law.[65]

More controversial is the positive duty of states *to prevent or prohibit* subversive propaganda by private entities, given potential clashes with the rights of individuals to freedom of expression and information.[66] This is reflected in the reluctance of some states to fully embrace Article 20(1) of the International Covenant on Civil and Political Rights (ICCPR),[67] which requires states to prohibit by law propaganda for aggressive war. Several states, including the UK, the US, France, Australia, and the Netherlands have made reservations to this provision.[68] However, these reservations focus on the need to ensure consistency with the right to freedom of expression and thus reject the need to enact *domestic legislation* specifically prohibiting war propaganda.[69] No state has questioned the *unlawfulness* of propaganda for aggressive war under international law, whether carried out by states or non-state actors.[70] If aggression is itself prohibited and criminalized under international law, so is incitement to engage in it.[71] This is why complicity in aggression is also criminal under international law.[72]

A duty to prevent and punish subversive and war propaganda by private entities is also specifically recognized in Articles 1 and 2 of the 1936 Convention concerning the Use of Broadcasting in the Cause of Peace.[73] Furthermore, in its General Comment 36 on the right to life, the UN Human Rights Committee asserted that failure to punish war propaganda might amount to a failure to protect the right to life under Article 6 ICCPR.[74] These and other rules arguably require states to adopt a basic

[62] Laurence Boisson de Chazournes and Luigi Condorelli, 'Common Article 1 of the Geneva Conventions revisited: Protecting collective interests' (2000) 837 International Review of the Red Cross <www.icrc.org/en/doc/resources/documents/article/other/57jqcp.htm> accessed 7 March 2023.

[63] *See* International Committee of the Red Cross (ICRC), 'Commentary to Article 1, Convention (I) for the Amelioration of the Condition of the Wounded and Sick in Armed Forces in the Field. Geneva, 12 August 1949' (2016) <https://ihl-databases.icrc.org/en/ihl-treaties/gci-1949/article-1/commentary/2016> accessed 7 March 2023.

[64] Commentary to Article 1, 'Protocol Additional to the Geneva Conventions of 12 August 1949, and relating to the Protection of Victims of International Armed Conflicts (Protocol I), 8 June 1977' (1987) <https://ihl-databases.icrc.org/en/ihl-treaties/api-1977/article-1/commentary/1987?activeTab=undefined> accessed 7 March 2023.

[65] ICRC, 'Rule 144' <https://ihl-databases.icrc.org/customary-ihl/eng/docs/v1_rul_rule144> accessed 7 March 2023.

[66] Larson (n 53) 449–50; van Dyke (n 53) 65–68, 72.

[67] International Covenant on Civil and Political Rights (adopted 16 December 1966, entered into force 23 March 1976) 999 UNTS 171.

[68] ibid.

[69] Hersh Lauterpacht, 'Revolutionary Activities by Private Persons Against Foreign States' (1928) 22 American Journal of International Law 105, 123.

[70] Kearney (n 52) 123, 148–49; Manfred Nowak, *UN Covenant on Civil and Political Rights: CCPR Commentary* (NP Engel 2005) 473.

[71] Whitton (n 53) 21; Larson (n 53) 443–45.

[72] Nikola Hajdin, 'Responsibility of Private Individuals for Complicity in a War of Aggression' (2022) 116 American Journal of International Law 788.

[73] Adopted 23 September 1936, entered into force 2 April 1938, 186 UNTS 301.

[74] GC 36 (n 51) para 59.

legal framework for online content moderation by various platforms with a view to preventing or mitigating the prevalence of content that might amount to foreign interference or war propaganda.

Scholarly writings[75] and international jurisprudence[76] also lend support to the view that subversive propaganda – irrespective of any results – is covered by broader ‘due diligence’ obligations under international law, such as the duties to prevent acts contrary to the rights of other states[77] and significant transboundary harm or injury to persons, property, or the environment.[78] As seen earlier, since these obligations focus on state omissions, no causal link between the lack of diligence and any ensuing harms is required for a breach to occur; constructive knowledge or objective foreseeability thereof is sufficient. Nevertheless, any prohibition, criminal or civil, of incitement to or propaganda for war by individuals must be subject to the strict requirements of legality, legitimacy, and necessity and proportionality for limiting the rights to freedom of expression and information under international human rights law.[79] Such requirements are found in Article 19(3) ICCPR and its regional counterparts, such as Article 10 of the European Convention on Human Rights.[80]

4. MISINFORMATION AND DISINFORMATION

Although ‘fake news’ has become a buzzword in recent years, it is not a new phenomenon. The intentional or non-intentional dissemination of false or misleading information has been a key feature of warfare and peacetime political strategy for centuries. Think of the staged border incidents Nazi Germany used to justify its 1939 invasion of Poland.[81] Think also of the US’ unfounded claims that Saddam Hussein was manufacturing weapons of mass destruction to justify the 2003 Iraq invasion.[82]

[75] *See eg* Kearney (n 52) 16; Larson (n 53) 450; Whitton (n 53) 23–25; John C Novogrod, ‘Collective Security under the Rio Treaty: The Problem of Indirect Aggression’ (1969–1970) 3 JAG Journal 99, 104.

[76] Eg *Island of Palmas Case (or Miangas), United States v Netherlands*, Award, 4 April 1928, II RIAA 829 (1928) ICGJ 392 (PCA 1928) 839; *Trail Smelter Case (USA v Canada)* (1941) 3 RIAA 1911, 1963.

[77] *Corfu Channel Case (United Kingdom v Albania)*, Judgment, 9 April 1949, ICJ Reports (1949) 4, 22.

[78] *See* ILC, ‘Draft Articles on Prevention of Transboundary Harm from Hazardous Activities’, with commentaries, Report of the International Law Commission on the work of its fifty-third session (23 April–1 June and 2 July–10 August 2001), UN Doc A/56/10, 2001.

[79] UN High Commissioner for Human Rights, ‘Expert Workshops on the Prohibition of Incitement to National, Racial or Religious Hatred’ UN Doc (2013) A/HRC/22/17/Add.4 (‘Rabat Plan of Action on the prohibition of advocacy of national, racial or religious hatred that constitutes incitement to discrimination, hostility or violence’) para 18.

[80] European Convention for the Protection of Human Rights and Fundamental Freedoms, as amended by Protocols Nos. 11 and 14 (adopted 4 November 1950; entered into force: 3 September 1953) ETS 5.

[81] Cody K Carlson, ‘This Week in History: Nazis Stage Fake Attack at the Start of WWII’ (Deseret News, 4 September 2014) <www.deseret.com/2014/9/3/20547775/this-week-in-history-nazis-stage-fake-attack-at-the-start-of-wwii> accessed 7 March 2023.

[82] Glenn Kessler, ‘The Iraq War and WMDs: An Intelligence Failure or White House Spin?’ Washington Post (22 March 2019) <https://webcache.googleusercontent.com/search?q=cache:s6OPNy1LOmcJ:www.washingtonpost.com/politics/2019/03/22/iraq-war-wmds-an-intelligence-failure-or-white-house-spin/+&cd=1&hl=en&ct=clnk&gl=uk> accessed 7 March 2023.

More recently, COVID-19 disinformation campaigns have led to vaccine hesitancy, serious illness, and death.[83]

The international regulation of misinformation and disinformation is marked by uncertainty and controversy. This is so for several reasons. As noted earlier, the causal link between the dissemination of false or misleading information and any harmful consequences is only indirect and requires further action on the part of their multiple addressees. Moreover, deception is not the most traditional form of interference in a state's internal or external affairs or its inherently governmental functions. There is also a clear tension between securing a peaceful and stable information space among *states*, and *individual* rights to freedom of expression and information.[84] After all, the latter rights are not limited to accurate or innocuous information.[85]

Nevertheless, three key conclusions can be reached. First, states must respect and protect the right of individuals to freedom of information.[86] This means that they may not fabricate, sponsor, encourage, or further disseminate statements or information that they should reasonably know are false, irrespective of any causation between speech acts and actual or potential results.[87] This is true insofar as the false statements undermine the right of individual addressees – whether at home or abroad – to be properly and freely informed.[88]

Secondly, the cross-border dissemination of false or misleading information by a state may breach the victim's right to non-intervention in its internal or external affairs.[89] As seen earlier, the meaning of *coercive* interference is not limited to the threat or use of force but extends to deception that is intended to force, or effectively forces, a state to adopt a course of action that it otherwise would not, irrespective of any particular results. Thus, like subversive propaganda, false statements aimed at regime change or foreign electoral processes, whether online or offline, may be contrary to the principle

[83] World Health Organization, 'Fighting Misinformation in the Time of COVID-19, One Click at a Time' (WHO, 27 April 2021) <www.who.int/news-room/feature-stories/detail/fighting-misinformation-in-the-time-of-covid-19-one-click-at-a-time> accessed 7 March 2023.

[84] De Brabandere (n 53) paras 3, 8–11.

[85] *Handyside v the United Kingdom* App no 5493/72 (ECtHR, 7 December 1976) para 49; HRC, 'General Comment No 34 – Article 19: Freedoms of Opinion and Expression' (2011) UN Doc CCPR/C/GC/34 (GC 34) paras 11–12; UN Special Rapporteur on Freedom of Opinion and Expression, the Organization for Security and Co-operation in Europe (OSCE) Representative on Freedom of the Media, the Organization of American States (OAS) Special Rapporteur on Freedom of Expression and the African Commission on Human and Peoples' Rights (ACHPR) Special Rapporteur on Freedom of Expression and Access to Information, 'Joint Declaration on Freedom of Expression and "Fake News", Disinformation and Propaganda' (3 March 2017) preambular para 7 <www.osce.org/fom/302796> accessed 7 March 2023.

[86] GC 34 (n 85) para 7.

[87] 'Joint Declaration on Freedom of Expression and "Fake News", Disinformation and Propaganda' (n 85) para 2(c); UNGA (n 2) para 88.

[88] Karl Joseph Partsch, 'Freedom of Conscience and Expression, and Political Freedoms' in Louis Henkin (ed), *The International Brill of Rights: The Covenant on Civil and Political Rights* (Columbia University Press 1981) 209, 219; Nowak (n 70) 446 and 459; *The Sunday Times v United Kingdom* (1979) Series A no 30, para 66; *NIT SRL v Moldova* App no 28470/12 (ECtHR, 5 April 2022) para 192; *Manole and others v Moldova* App no 13936/02 (ECtHR, 17 September 2009) para 100.

of non-intervention. This includes, for example, online content seeking to mislead the population of another state about the location of voting centres or about polls on which candidate is predicted to win an election.

Thirdly, if false or misleading statements that may cause harm or injury in another state cannot be attributed to states but emanate from individuals, the freedoms of expression and information take centre stage. Obligations requiring states to exercise due diligence to prevent, stop or redress foreseeable harm to other states must not overstep permissible limitations to those rights. This applies, for instance, to the obligation contained in Article 3 of the 1936 Broadcasting Convention. Thus, as with subversive and war propaganda, any action to tackle the spread of online misinformation or disinformation must carefully balance individuals' rights to freedom of expression and information against states' duties to prevent, stop, and redress the foreseeable harms of misinformation and disinformation, including their obligation to protect human life and health.

This balance can be achieved when restrictions on the online dissemination of false or misleading information are clearly grounded in law, as well as necessary and proportionate to achieve a legitimate aim, in line with Article 19(3) ICCPR and similar provisions. This will usually require the enactment of a basic legal framework governing the publication and moderation of online misinformation and disinformation. In this author's view, while the *intentional* dissemination of disinformation may be subject to limitations, such as content moderation and civil liability, the *unintentional* spread of such content should not be sanctioned.[90] Criminal sanctions should be reserved for only the most serious forms of disinformation, such as defamation and libel.[91]

5. MALINFORMATION

The dissemination of accurate information or opinions with the intent to cause harm is not per se prohibited under international law. However, malinformation may violate the principle of non-intervention insofar as the leak or publication is intended to coerce or effectively coerces a state in matters within its internal or external affairs. This could happen if, for example, leaked confidential information about the identity or location of undercover state operatives compromises law enforcement or military operations. Likewise, states must exercise due diligence in preventing or redressing the release of such information if it is of such a nature as to foreseeably contravene the rights of the victim state or to cause harm or injury to persons, property, or the environment therein.

[89] Larson (n 53) 442, 447–49; De Brabandere (n 53) paras 17–20, Baade (n 23) 1362–65.

[90] 'Joint Declaration on Freedom of Expression and "Fake News", Disinformation and Propaganda' (n 85) preambular paras 4 and 5, operative para 1(e).

[91] UNGA (n 2) paras 41–43.

Malinformation is usually preceded by cyber espionage and/or electronic surveillance operations. While many would argue that the exfiltration of governmental data is not itself unlawful under international law, the *method* by which such information is extracted may violate one or more rules of international law, depending on the consequences of the operation in question.[92] Thus, covert campaigns to extract information or those seeking to mislead the victim state with respect to who is behind the extraction are not in and of themselves prohibited by international law. Such operations are only limited insofar as their method foreseeably causes harm to a protected object or subject, such as by infringing an individual's right to receive information. In the same vein, electronic surveillance against individuals may violate the right to privacy under international human rights law, unless the operation is carried out in accordance with the law and is necessary and proportionate to achieve a legitimate aim.[93]

6. ONLINE HATE SPEECH

'Online hate speech' is an umbrella term encompassing a multitude of digital content[94] – from hateful emojis[95] to direct and public incitement to commit violence of the kind seen during the Rwandan genocide.[96] This means that international legal responses to online hate will vary depending on the content, as well as the speaker, audience, medium, and context. Three legal categories may be proposed, taking into account applicable rules of international human rights law and international criminal law.[97]

The first comprises the most serious types of online hate speech which amount to international crimes and give rise to both individual criminal liability and state responsibility under international law.[98] Within this category falls the inchoate offence of direct and public incitement to commit genocide, prohibited under Article III(c)

[92] Antonio Coco, Talita Dias, and Tsvetelina van Benthem, 'Illegal: The SolarWinds Hack under International Law' (2022) European Journal of International Law <https://doi.org/10.1093/ejil/chac063> accessed 4 January 2022.

[93] *See eg* International Covenant on Civil and Political Rights, art 17; European Convention on Human Rights, art 8; UN Human Rights Council, 'The Right to Privacy in the Digital Age: Report of the United Nations High Commissioner for Human Rights' (15 September 2021) UN Doc A/HRC/48/31, paras 8 and 39.

[94] *See* '"Hate Speech" Explained: A Toolkit' *Article 19* (2015) <www.article19.org/data/files/medialibrary/38231/'Hate-Speech'-Explained---A-Toolkit-%282015-Edition%29.pdf> accessed 7 March 2023.

[95] Hannah Rose Kirk and others, 'HATEMOJI: A Test Suite and Adversarially-Generated Dataset for Benchmarking and Detecting Emoji-based Hate' (*University of Oxford*, 2021) <https://ora.ox.ac.uk/objects/uuid:0570eaf5-e729-4ef5-b27a-b6d511abcdc3/download_file?file_format=&safe_filename=Kirk_et_al_2021_Hatemoji_a_test_suite--.pdf&type_of_work=Working+paper> accessed 7 March 2023.

[96] *Nahimana et al case* (Appeal Judgment) ICTR-99-52-A (28 November 2007) paras 673–715.

[97] Talita Dias, 'Tackling Online Hate Speech through Content Moderation: The Legal Framework Under the International Covenant on Civil and Political Rights' (*SSRN*, 30 June 2022) <https://ssrn.com/abstract=4150909> accessed 7 March 2023.

[98] *Bosnian Genocide* (n 35) paras 160–69.

of the Genocide Convention.[99] Context is key in determining whether seemingly neutral expressions are in fact coded language directly inciting genocide. An example is the labelling of individuals or groups as animals that ought to be killed against a background of inter-communal violence,[100] as with cartoons and radio broadcasts in Rwanda.[101] Serious forms of online hate speech may also amount to instigation to or aiding and abetting international crimes, including genocide, crimes against humanity, war crimes, and the crime of aggression.[102] However, in those instances, participation in crime requires not only an intention to instigate or assist in the commission of the relevant crime but also a causal link between the criminal conduct and the result (i.e., the speech acts must have 'substantially contributed' to the commission of the principal crime).[103] 'Mere hate speech' below this threshold is unlikely to amount to the commission of an international crime, especially given the lack of a sufficiently clear criminalization of the speech acts in question.[104]

The second category consists of hateful expressions that constitute incitement to certain types of unlawful action. A prominent example is found in Article 20(2) ICCPR, which stipulates that '[a]ny advocacy of national, racial or religious hatred that constitutes incitement to discrimination, hostility or violence shall be prohibited by law'. Like the prohibition of war propaganda, the wording and the very inclusion of this provision in the Covenant were heavily debated. Some states, like the US, the UK, and the Netherlands, still reserve their right not to enact domestic legislation giving effect to this provision,[105] citing freedom of expression concerns.[106] This suggests that Article 20(2) ICCPR is not part of customary international law, despite the Human Rights Committee holding otherwise.[107] A similar debate[108] exists over the scope and status of Article 4(a) of the International Convention on the Elimination of All Forms of Racial Discrimination.[109] In any event, the right of individuals to be free from incitement to discrimination, recognized in Article 7 of the Universal Declaration of Human Rights, does reflect international custom.[110] Though this provision does

[99] Convention on the Prevention and Punishment of the Crime of Genocide (adopted 9 December 1948; entered into force 12 January 1951) 78 UNTS 277.

[100] See *Nahimana et al case* (n 96) paras 477–672. See also UN Human Rights Council, 'Report of the Detailed Findings of the Independent International Fact-Finding Mission on Myanmar' (2018) UN Doc A/HRC/39/CRP.2, paras 1316–18.

[101] 'Rwanda: From Hatred to Reconciliation' (*Al Jazeera World*, 29 September 2015) <www.aljazeera.com/program/al-jazeera-world/2015/9/29/rwanda-from-hatred-to-reconciliation> accessed 7 March 2023.

[102] *See eg* UN Human Rights Council, 'Report of the Independent International Fact-Finding Mission in Myanmar' (2018) UN Doc A/HRC/39/64, paras 83–89.

[103] *Nahimana et al case* (n 96) 480, 482.

[104] *Nahimana et al case* (n 96), Partly Dissenting Opinion of Judge Theodor Meron, paras 5–8.

[105] International Covenant on Civil and Political Rights.

[106] Jeroen Temperman, *Religious Hatred and International Law: The Prohibition of Incitement to Violence or Discrimination* (CUP 2015) 72–74.

[107] HRC, 'CCPR General Comment No 24: Issues Relating to Reservations Made upon Ratification or Accession to the Covenant or the Optional Protocols thereto, or in Relation to Declarations under Article 41 of the Covenant' (1994) UN Doc CCPR/C/21/Rev.1/Add.6, para 8.

[108] Temperman (n 106) 69, 122, n 1.

[109] Adopted 21 December 1965; entered into force 4 January 1969, 660 UNTS 195.

[110] Hurst Hannum, 'The Status of the Universal Declaration of Human Rights in National and International Law' (1996) 25 Georgia Journal of International and Comparative Law 287, 342–43.

not impose a legal prohibition of incitement, it arguably requires states to exercise their best efforts to protect individuals from prohibited discrimination and incitement thereto, in line with the rights to freedom of expression and information.

The third category of online hate is 'limited speech', or hateful expressions that are in principle protected but may be limited pursuant to the three-part test described in Article 19(3) ICCPR and equivalent provisions – that is, legality, legitimacy, and necessity and proportionality. Thus, denial of historical facts such as the Holocaust, while not necessarily amounting to incitement within the meaning of Article 20(2) ICCPR, may nonetheless be limited by law in contexts where this would be a necessary and proportionate step to safeguard the rights or reputations of others or public order.[111] All other forms of online hate falling below this threshold (i.e., that cannot be limited via the three-part test) must be protected. There is no definitive category of 'protected speech', since all speech acts, including hate speech, may be subject to limitations. Yet, some types of speech should receive heightened protection given the public interest in their dissemination, not the least to ensure individuals' right to receive information. This includes political speech[112] and statements whose dissemination is in the public interest, such as impartial journalistic reporting on public affairs.[113]

7. CONCLUSION

The rise of online information operations, both domestic and cross-border, has prompted different responses from governments, corporations, civil society, and other stakeholders around the world. But the extent to which international law limits such operations remains uncertain. On the one hand, many still hold on to the misconception that international law is altogether indifferent to the phenomenon. On the other, the scope of applicable rules is often misunderstood, particularly the need for any causal link between speech acts and their possible effects. This paper has tried to debunk those myths and misconceptions by providing a doctrinal assessment of how different rules of international law interact to limit four key types of information operations: propaganda, misinformation and disinformation, malinformation, and online hate speech.

First, it argued that, while international law lacks a *general* standard of factual and legal causation, different standards may be clearly identified for *particular* rules. The

[111] *See eg* HRC, *Faurrison v France*, 'Communication No. 550/1993' (1996) UN Doc CCPR/C/58/D/550/1993 paras 9.4–9.7.

[112] *Case of Mouvement Raëlien Suisse v Switzerland* App no 16354/06 (ECtHR, 13 July 2012 para 61.

[113] Council of Europe, 'Recommendation CM/Rec(2022)4 of the Committee of Ministers to member States on promoting a favourable environment for quality journalism in the digital age' (17 March 2022) <https://search.coe.int/cm/pages/result_details.aspx?objectid=0900001680a5ddd0> accessed 7 March 2023.

paper has also contended that important rules and principles that apply to information operations, such as non-intervention and due diligence, do not seem to require any causal link between prohibited or required conduct, including speech acts, and any of their alleged effects. Instead, they cover intended or foreseeable harms to certain protected subjects and objects that might arise from different types of conduct, including information operations.

Crucially, this paper has argued that, while complex, the interplay between different applicable rules of international law in the context of each type of information operation can be worked out in practice. This requires consideration of the factual features of each operation, an understanding of the scope of applicable rules – including any required causal link or knowledge threshold – as well as their careful balancing. These rules include the principle of non-intervention, various due diligence obligations, rules and principles of IHL, and different human rights, most notably the rights to freedom of expression, information and privacy. Such rules do not exhaust the scope of applicable international law on the matter and more research is needed into other relevant rules or principles, such as sovereignty and self-determination. They are also shrouded in legal controversies and enforcement challenges.[114]

Yet existing international legal rules and principles already provide a workable legal framework that significantly limits the deployment of information operations by states and non-state actors. These rules seek to balance two overarching considerations: the right of individuals and states to a free information space, and the need to mitigate some of the harms that information operations may cause online and offline.

[114] van Benthem, Dias, and Hollis (n 1) 1268–84.

2023 15th International Conference on Cyber Conflict
Meeting Reality
T. Jančárková, D. Giovannelli, K. Podiņš, I. Winther
2023 © NATO CCDCOE Publications, Tallinn

Seeing Through the Fog: The Impact of Information Operations on War Crimes Investigations in Ukraine

Lindsay Freeman
Director of Technology, Law & Policy
Human Rights Center
UC Berkeley School of Law
Berkeley, California, USA
lfreeman@berkeley.edu

Abstract: As Russian forces closed in on Kyiv, a MiG-29 Fulcrum swooped in and took down six Russian jets. The next day, the same MiG shot down ten more. Stories of the hero fighter pilot spread like wildfire throughout Ukraine and across the internet, turning the "Ghost of Kyiv" into a living legend.

But he was not living, or even real. The pilot and his exploits were a total fiction created as part of an influence campaign spread via social media to strike terror into Russian forces, fortify the resolve of Ukrainian citizens, and amaze the world with Ukraine's unexpected strength and courage.

The strategic use of the online information environment is only one facet of intangible warfare between Russia and Ukraine that makes this contemporary conflict particularly unique and complex. Propaganda, disinformation, and psychological operations are as old as warfare itself, but advanced digital technologies now reshape conflicts in often unanticipated, unforeseen, and surprising ways. These changing dynamics inevitably have an impact on those tasked with investigating war crimes and establishing the truth of what occurred on the battlefield.

This paper examines the strategic use of digital information and communications technologies in the Russia–Ukraine conflict to better understand how they are changing the dynamics of war, war narratives, and war crimes investigations. The first section of the paper briefly explains how war crimes investigators and prosecutors are increasingly relying on digital material as evidence in their cases. The second section considers how digital information operations are being deployed and how these operations impact the investigation of war crimes. Finally, the third section highlights some of the tools that can help war crimes investigators fight back against a complex and chaotic information environment.

Keywords: *information warfare, disinformation, influence operations, war crimes investigations, digital evidence, Berkeley Protocol*

1. INTRODUCTION

> *"The great uncertainty of all data in war is because all action, to a certain extent, planned in a mere twilight—like the effect of a fog—gives things exaggerated dimensions and unnatural appearance."*
>
> *Carl von Clausewitz*[1]

Throughout history, military generals and strategists have espoused the importance of information in warfare, characterizing the ways in which information operations can be used to further military objectives and gain a competitive edge over the opposition. As Napoleon Bonaparte once stated, "War is ninety percent information."[2] Information operations—a term that encompasses a range of activities from disseminating propaganda to spreading disinformation to blocking access to communication channels—can have a profound impact on the course of conduct on the battlefield, as well as the narratives that surround it. While information warfare is not a new concept, the adoption of digital information and communications technologies (ICTs) has changed the nature of wartime information operations in interesting and unforeseen ways. These new dynamics are currently playing out in Ukraine, where the largest international armed conflict in Europe since World War II is taking place.

While historical analysis of information operations can provide some insight into what the world is witnessing in Ukraine today, there are many novel elements of modern information warfare that cannot be fully understood within traditional frameworks. The teachings of Sun Tzu and Carl von Clausewitz never explained how to crowdfund weapons or troll the enemy on social media. These traditional military theorists could not have imagined the speed and scale of modern digital communications, nor could they have envisioned a world in which ordinary citizens across the globe could monitor the battlefield in near real time with high-resolution satellites and live drone

[1] Carl von Clausewitz, *On War*, vol. 1 (Altenmünster: Jazzybee Verlag, 1950)

[2] Napoleon I, *Military Maxims of Napoleon* (New York: Wiley and Putnam, 1845).

feeds. But this novel information universe is precisely the reality being experienced in Ukraine's fight against Russian aggression and occupation.

New technologies can shrink time and space, blur borders, and alter the ways in which information travels from the battlefield to the outside world, and from the outside world to the battlefield. Although social media, smartphones, and other digital ICTs have played critical roles in past armed conflicts over the years—from their use in documenting war crimes in Syria to their role in furthering crimes against humanity in Myanmar—their application in the Russia–Ukraine conflict is unprecedented. As policy analysts Christian Perez and Anjana Nair explain, "Throughout the ongoing conflict, social media has served as a battleground for states and non-state actors to spread competing narratives about the war and portray the ongoing conflict in their own terms."[3] In a hybrid war between state militaries with far-reaching implications for the rest of the world, the information environment has become contested, corrupted, and dizzyingly complex. The lessons of the past can only take us so far in understanding this new digital world.

Advanced digital technologies are not only changing the nature of warfare and facilitating the weaponization of information but also transforming how war crimes investigations are conducted. Over the past decade, in response to hostile governments blocking access to crime scenes, the international criminal justice community has built up the capacity to conduct remote investigations using ICTs. In recent years, satellite imagery, call data records, wire transfers, and social media content have all been recognized as admissible evidence in international criminal trials.[4] Investigators, lawyers, and judges are also becoming more accepting of the use of video conferencing platforms to conduct witness interviews or provide testimony. The ability to gain virtual access to witnesses and allow them to share their experiences with investigators thousands of miles away in The Hague is a truly groundbreaking development in international legal practice. However, relying on information coming out of a conflict zone without having been there in person raises fresh concerns about the ability of investigators to assess the credibility of witnesses and the reliability of information viewed through a digital prism. These developments raise a key question: When the information environment is itself a domain of battle, how can investigators separate fact from fiction to establish the truth?

[3] Christian Perez and Anjana Nair, "Information Warfare in Russia's War in Ukraine: The Role of Social Media and Artificial Intelligence in Shaping Global Narratives," *Foreign Policy*, 22 August 2022, https://foreignpolicy.com/2022/08/22/information-warfare-in-russias-war-in-ukraine/.

[4] *Prosecutor v. Jean-Pierre Bemba Gombo, Aimé Kilolo Musamba, Jean-Jacques Mangenda Kabongo, Fidèle Babala Wandu and Narcisse Arido (Bemba et al.)*, Decision on Requests to Exclude Dutch Intercepts and Call Data Records, ICC-01/05-01/13-1855, TC VII, 26 April 2016; *Bemba et al.*, Decision on Requests to Exclude Western Union Documents and other Evidence Pursuant to Article 69(7), ICC-01/05-01/13-1854, TC VII, 29 April 2016; *Prosecutor v. Ahmad Al Faqi Al Mahdi*, Judgement and Sentence, ICC-01-12-01/15-171, TC VIII, 27 September 2016; *Prosecutor v. Salim Jamil Ayyash, Hassan Habib Merhi, Hussein Hassan Oneissi, Assad Hassan Sabra (Ayyash et al.)*, Trial Judgment, STL-11-01/T/TC, 18 August 2020.

This paper begins with an overview of the role of growing importance that digital evidence is playing in war crimes investigations and prosecutions. This section is followed by an analysis of ICT-enhanced information operations in the Russia–Ukraine conflict and the various challenges that are emerging for war crimes investigators as a result. It then introduces the reader to some of the tools that are emerging to help war crimes investigators grapple with and overcome the challenges, including the *Berkeley Protocol on Digital Open Source Investigations*, which is currently being used by war crimes investigators in Ukraine.

2. THE CHALLENGE OF INVESTIGATING WAR CRIMES AND THE PROMISE OF DIGITAL EVIDENCE

Investigating war crimes has always been challenging for a variety of reasons, from the complexity of cases involving thousands of victims and witnesses to the difficulty in obtaining battlefield evidence. The first trial at the International Criminal Court (ICC), *Prosecutor v. Lubanga*, brought these challenges into stark focus.[5] Due to security issues in the locations relevant to their investigation, ICC investigators were unable to visit many of the crime scenes. Instead, they relied on intermediaries—locals from the area—who could help them find witnesses to the events.[6] At trial, it was revealed that several of the witnesses had been paid to give false testimony by the intermediaries. Without realizing it, the prosecution had brought unreliable narrators and a tainted witness pool before the court.

The prosecution was also overly dependent on witness statements taken by United Nations (UN) investigators, who had prior access to relevant individuals.[7] These statements had not been taken for the purposes of a prosecution, and investigators had been unable to find the original witnesses, validate their statements, and get their consent for use in court. This posed a problem at trial, since the prosecution did not have the authority to disclose the statements to the defense, although they were already relying on them in their case. The trial was stayed for several months as a result. Moreover, ICC investigators did not visit and forensically examine the crime scenes until they were nearing the trial phase. When they did finally visit the relevant geographic locations, they discovered that some of what their witnesses had testified to was not accurate. The very first ICC case came close to getting dismissed, which raised larger questions about the ICC Office of the Prosecutor's (OTP) ability to do fulfill its mandate. The inability of ICC investigators to visit the territory where the crimes occurred was not unique to the Democratic Republic of Congo in the *Lubanga*

[5] Aliza Shatzman, "The Prosecutor v. Thomas Lubanga Dyilo: Persistent Evidentiary Challenges Facing the International Criminal Court," *George Mason International Law Journal* 12, no. 2 (2021).

[6] Caroline Buisman, "Delegating Investigations: Lessons to be Learned From the Lubanga Judgment," *Northwestern University Journal of International Human Rights*11, no 3 (2013).

[7] Christodoulos Kaoutzanis, "A Turbulent Adolescence Ahead: the ICC's Insistence on Disclosure in the Lubanga Trial," *Washington University Global Studies Law Review* 12, no. 2 (2013).

case. For years, investigators could not enter the territory of Sudan to investigate its head of state, Omar al Bashir, and other accused persons. Today, investigators face similar circumstances in Myanmar and Burundi.

These immense obstacles led the OTP to look for other solutions in its investigative work. Advanced digital technologies that could facilitate remote investigations, such as high-resolution satellite imagery or call data records, seemed like an answer to the problem. With the introduction of smartphones, internet connectivity, and social media, suddenly individuals within the conflict zone could capture what was happening on the ground and share it with the outside world. This was incredibly promising for the prosecutor's office, which could hire analysts to collect and review the data, thus moving the investigation forward even while their access was blocked.

The appeal of remote, technology-enabled investigation tactics was not limited to ICC investigators. About a decade ago, the international criminal justice community began to recognize the potential for utilizing user-generated content in their work.[8] This interest in digital open-source information, especially social media content, was initially driven by the conflicts in Syria, Iraq, and Libya, where smartphones and social media platforms became primary tools for war documentation. These tools allowed, and continue to allow, witnesses and first responders to record atrocities as they unfold, often in places where it is difficult, if not impossible, for international investigators to gain access.

These contemporary conflicts raised new and important questions about the possibility of investigating entities to acquire, analyze, and authenticate large volumes of digital information. The internal conflict in Myanmar, in which social media was used as a tool to incite violence against the Rohingya minority, further demonstrated how this type of digital information could provide evidentiary value by establishing the criminal intent of the perpetrators. In such cases, the propaganda and hate speech itself served as critical evidence in building a case against the Myanmar military for several crimes against humanity, including persecution.

However, this new source of potential evidence came with its own challenges. The volume of digital information online was immense, and the anonymous nature of the internet made it difficult to verify or even identify the information source. Thus, during these conflicts, conversations in the international criminal law community focused on how new technologies could improve investigative practice, leading to a series of workshops and the drafting of the *Berkeley Protocol on Digital Open Source Investigations*,[9] a UN manual co-published by the Office of the High Commissioner for Human Rights and Berkeley Law's Human Rights Center. The issues also led

[8] Rebecca J. Hamilton, "User Generated Evidence," *Columbia Journal of Transnational Law* 57 (2018): 1–61.

[9] The author of this paper led the drafting of the *Berkeley Protocol*.

to conversations about how new technologies, such as machine learning (ML) and artificial intelligence (AI), could be used to improve and enhance investigative practice—for example, the use of natural language processing to assist in document review and the use of object recognition technology to assist with the imagery analysis. The field was evolving, and then Russia invaded Ukraine, the digital battlefield exploded, and a whole new set of challenges began to emerge.

3. INFORMATION OPERATIONS IN THE RUSSIA–UKRAINE CONFLICT AND EMERGING CHALLENGES FOR INVESTIGATORS

As Russian forces closed in on Kyiv, launching the Kremlin's initial offensive against the capital city in late February 2022, a MiG-29 Fulcrum shot down six Russian planes. The next day, the same Ukrainian pilot shot down ten more Russian jets. Stories of this Ukrainian pilot spread and amplified on the internet, quickly turning him into a living legend. But he was not living, or even real. Rather, he was a fiction created to strike terror into Russian forces, fortify the resolve of Ukrainian citizens, and amaze the world with Ukraine's unexpected strength and courage. Now recognized as propaganda, stories of the "Ghost of Kyiv" were believed by many, until the Ukrainian Air Force ultimately admitted that this character was created as part of an influence campaign.[10] Nevertheless, his legend lives on in murals and other artwork commemorating the hero pilot.[11]

The Ghost of Kyiv is one of several examples of influence operations in the Russia–Ukraine conflict, which are the focus of this section. Since a significant amount of scholarship and commentary has already been written on the topic generally, this section focuses on a few of the emergent trends seen in the conflict that directly affect war crimes investigations. This includes the way in which military forces are engaging in content creation to influence or deceive not only enemy forces but individuals beyond the battlefield; the ability of governments to create chaos through the proliferation of competing narratives, while also controlling information flows to specific audiences; and the parallel trends of exposing closed-source information through hack-and-leak operations on the one hand and censoring information through internet shutdowns on the other.

A. Influence and Deception

With the global popularity of social media, parties to a conflict have a much larger potential sphere of influence, with a multitude of platforms through which they can

[10] Lateshia Beachum, "The 'Ghost of Kyiv' Was Never Alive, Ukrainian Air Force Says," *Washington Post*, 1 May 2022, https://www.washingtonpost.com/world/2022/05/01/ghost-of-kyiv-propaganda/.

[11] "Ghost of Kyiv Mural Unveiled in Ukrainian Capital in Celebration of Aviation Day," Yahoo! News, 27 August 2022, https://news.yahoo.com/ghost-kyiv-mural-unveiled-ukrainian-200800183.html.

reach a broad audience. To exploit social media successfully, however, the parties need to create interesting, engaging, and emotionally driven content to capture the public's attention. The ability to produce this type of online content requires specialized skills that are not traditionally associated with the armed forces.

From the start of the full-scale invasion in February 2022, the Ukrainian government and military have demonstrated an adeptness for conducting creative information operations, successfully using social media to win the hearts and minds of the Western world and, in so doing, gaining the political and financial support necessary to sustain their fight against Russian forces. From the start of the conflict, President Volodymyr Zelensky has used digital media to connect with the people of Ukraine. His nightly addresses have played an important role in boosting the morale of the Ukrainian people and soliciting support from the rest of the world. The Ukrainian president's videos represent only a small fraction of the videos, images, and memes generated by the Ukrainian government and military, distributed across a range of social media platforms, including Twitter, Facebook, Telegram, and TikTok. For example, the official Twitter account of Ukraine's Ministry of Defense posts a persistent stream of content that can be sincere and heart-breaking one minute and sassy and biting the next.[12] The account posts well-produced, professional-looking videos set to music that encourage sympathy for Ukrainians while antagonizing their Russian invaders. In one case, they adroitly used close-up imagery of a bombed-out playground for a campaign that generated millions of dollars from the public to buy kamikaze drones.[13] Similar imagery depicting the destruction of schools has been widely circulated online, leading some investigators to quickly conclude that these attacks were war crimes. However, such photographs can be misleading based not on what the photographer captures in the frame but on what they leave out. In some instances, panning out reveals a military base or indicators that the school was being used for military purposes. Thus, while the content from these accounts may not be manipulated or altered, the framing is far from neutral and objective. Rather, it is intended to influence the consumer. War crimes investigators are not immune from this influence.

The Ukrainian approach to propaganda, using real imagery in clever ways, stands in contrast to the favored Russian tactic, which relies on falsified narratives, fake content, and disinformation to amplify emotions, stoking fear and fueling hatred. Russia has been engaging in these tactics for many years as part of its geopolitical agenda against the West, but the resources put into it and the sophistication have grown with the use of information operations troops within the Russian military apparatus. Generally, rather than focusing on the quality of content with catchy phrases, popular songs, and humor, the Russian government and military propaganda apparatus patently produces false

[12] See "Defense of Ukraine," Twitter, https://twitter.com/DefenceU.

[13] Daniel Boffey, "Ukraine Crowdfunding Raises Almost $10m in 24 Hours to Buy Kamikaze Drones," *The Guardian*, 12 October 2022, https://www.theguardian.com/world/2022/oct/12/ukraine-crowdfunding-kamikaze-drones-russian-attack-cities-military.

claims and conspiracy theories intended to convey an inaccurate account of events to the consumer. Rather than sharing this information through official channels, it often disseminates it through proxies, such as fake websites made to look like traditional and reputable news sources. As with influence campaigns, even well-trained war crimes investigators are susceptible to being fooled by these deception tactics.

Both parties in the Russia–Ukraine conflict are creating social media content that is designed to go viral. The speed at which information spreads across the internet exacerbates the challenges for investigators in a variety of ways. First, false information travels faster than facts, as one Twitter-based study revealed.[14] This phenomenon means that investigators monitoring social media are likely to see the false version of events before seeing the accurate version. In addition, thoroughly fact-checked news stories take longer to be published than unverified ones based on speculation rather than hard facts. This is problematic because even trained investigators are susceptible to anchor bias, which describes "people's tendency to rely too heavily on the first piece of information they receive on a topic."[15] In addition to issues of bias, the speed at which online information is shared and the constant stream of content tend to create a sense of urgency and anxiety that may cause investigators to shorten or altogether skip the verification process. This means that international criminal investigators need a high degree of digital literacy, skepticism, and understanding of digital culture to do their job effectively. It also means that digital investigators need time to do their jobs well.

B. Chaos and Control

Russia's longtime go-to tactic for information warfare has been to create chaos and confusion by overwhelming the information space with a high volume of conflicting stories about a single event. Newer technologies, such as automated botnets paired with artificial intelligence, now generate content to inundate online platforms with conflicting narratives.[16] The ease with which digital information can be quickly created, altered, repurposed, or amplified is unique to our modern world in which Hollywood special effects are affordable and commercially available to everyone on a smartphone, botnets can be used to control thousands of accounts at once, and artificial intelligence has been optimized to generate fake videos that are indistinguishable from real ones. Fake imagery and audio recordings have advanced so much that it is often difficult to tell them apart from the real thing. The quality of fake imagery and the amplification of false narratives online is not intended to deceive but to undermine trust more generally so that people begin to believe that nothing is real and nothing

[14] Larry Greenemeier, "False News Travels 6 Times Faster on Twitter than Truthful News," *Scientific American*, 9 March 2018, https://www.pbs.org/newshour/science/false-news-travels-6-times-faster-on-twitter-than-truthful-news.

[15] Kassiani Nikolopoulou, "What is Anchoring Bias? Definition and Examples," *Scribbr*, 16 December 2022, https://www.scribbr.com/research-bias/anchoring-bias/#:~:text=Anchoring%20bias%20describes%20people's%20tendency,anchor%2C%20to%20make%20subsequent%20judgments.

[16] Paul Szoldra, "Deepfakes Are Russia's New 'Weapon of War'," *Ruck*, 20 November 2022.

can be trusted. This lack of trust can create an even greater problem for investigators and lawyers who must convince judges to trust the evidence. Thus, while skepticism is important, investigators need to find a way to properly convey when content is reliable and when it is not. Deepfakes also raise concerns about "the liar's dividend,"[17] which might allow war criminals to evade accountability by claiming that real content is fake.

In contrast to the everything, everywhere, all the time approach to information operations, the architecture of the internet and diversification of platforms provide for very precise and selective information targeting. New digital technologies increase the ability to design messaging to target specific audiences. With traditional media, the same news or information was generally distributed to all recipients equally, but digital media operates differently. As the world learned from the Cambridge Analytica scandal involving digital consultants to Donald Trump's 2016 presidential campaign, anyone can pay social media platforms to micro-target messages to specific users. Micro-targeting is defined as "a marketing strategy that uses consumer data and demographics to identify the interests of specific individuals or very small groups of like-minded individuals and influence their thoughts or actions."[18] This ability to distribute targeted information to specific communities gives parties unprecedented control not only over the information shared but over how it is shared and who sees it. For example, Russian information operations troops share different messages with Western countries than they do with the BRICS countries (Brazil, India, China, and South Africa alongside Russia) or with the Russian people.[19] Depending on where you live—both geographically and on the internet—a person may have very different perceptions of what is happening in the Russia–Ukraine conflict.

The parties to this conflict are using parallel tracks—one that uses voluminous, fast-paced, and chaotic distribution of content to overwhelm internet users and another that uses information silos and micro-targeting to send precisely crafted messages to very specific audiences. These dual tactics are confounding war crimes investigators, who, on one hand, must sort through an unmanageable firehose of information to find the "signal in the noise" and, on the other hand, must actively go hunting in different online communities and forums to ensure they are getting a full picture of what is happening. Thus, the volume of information requires that investigator to search for evidence in an endless ocean of information, while the siloing of information necessitates that investigators search for evidence in the equivalent of a thousand rivers.

[17] Kaylyn Jackson Schiff, Daniel S. Schiff, and Natalia Bueno, "The Liar's Dividend: Can Politicians Use Deepfakes and Fake News to Evade Accountability?" SocArXiv Papers, 10 May 2022, https://osf.io/preprints/socarxiv/q6mwn/.

[18] Linda Tucci, "Microtargeting," *TechTarget*, February 2013, https://www.techtarget.com/searchcio/definition/microtargeting.

[19] "The GRU's Galaxy of Russian-Speaking Websites," *Open Facto*, 27 January 2022, https://openfacto.fr/2022/01/27/the-grus-galaxy-of-russian-speaking-websites/.

C. Exposure and Concealment

Two other notable trends in information operations in the Russia–Ukraine conflict are the hacking, leaking, and exposure of private information versus the censoring of information by blocking websites or internet connectivity. Hack-and-leak operations are defined as operations in which "malicious actors use cyber tools to gain access to sensitive or secret material and then release it in the public domain."[20] Internet shutdowns are understood as "state-enforced disruptions of internet access aimed at controlling the flow of information."[21]

Since the February 2022 invasion, on an almost daily basis, there have been new online leaks of documents and datasets alleged to be from Russian government agencies and private businesses. One month into the invasion, the Secret Service of Ukraine published the names of 620 alleged agents of Russia's Federal Security Service, presumably obtained through hacking.[22] Similarly, a website called Distributed Denial of Secrets (DDoSecrets) started releasing regular document dumps to their email subscriber list. In less than two months, two million emails from Russian government and private entities were leaked, making them accessible to any member of the public. These online leaks are high volume and unlikely to have been reviewed in full by anyone before their publication. In some cases, the parties to the conflict have openly leaked private documents themselves, while at other times they have used proxies. There has also been a significant amount of leaking coming from anonymous sources and third parties. In addition to the questions around the legality of acquisition, which could lead to the exclusion of evidence in court, online leaks are extremely difficult to authenticate and can be laced with malware or contain strategically placed false information.[23]

If leaked documents are, in fact, authentic, they could serve as a fruitful source of evidence for war crimes investigators. However, like everything else on the internet, leaked documents must be handled with caution and viewed with skepticism. These document dumps could easily contain false information designed to mislead. As DDoSecret explains, datasets released during war have "an increased chance of malware, ulterior motives and altered or implanted data, or false flags / fake personas." There have already been examples of tainted leaks, in which hackers manipulated the

[20] James Shires, "Hack-and-Leak Operations and U.S. Cyber Policy," *War on the Rocks*, 24 August 2022, https://warontherocks.com/2020/08/the-simulation-of-scandal/.

[21] "The Impact of Internet Shutdowns on Human Rights Defenders in India," *American Bar Association*, 14 November 2022, https://www.americanbar.org/groups/human_rights/reports/india-internet-shutdowns/#:~:text=Internet%20shutdowns%20are%20state%2Denforced,controlling%20the%20flow%20of%20information.

[22] "Ukraine Intelligence Publishes Names of 620 Alleged Russian Agents," *Reuters*, 28 March 2022, https://www.reuters.com/world/europe/ukraine-intelligence-publishes-names-620-alleged-russian-agents-2022-03-28/.

[23] Lindsay Freeman "Hacked and Leaked: Legal Issues Arising from the Use of Unlawfully Obtained Digital Evidence in International Criminal Cases," *UCLA Journal of International Law and Foreign Affairs* 25, no. 2 (2021): 45.

documents before sharing them publicly online.[24] In addition to the very real risks of embedded malware and implanted false information, these documents come without any of the contextual information or metadata needed to authenticate them for use in legal proceedings. With all these new types of digital evidence that have not been tested in court, war crimes investigators face a mammoth challenge in determining what information will be required to authenticate the evidence and whether it will be found admissible by a future court.

In contrast to the approach of openly sharing information for strategic advantage, governments can also do the opposite. As an authoritarian regime, the Kremlin has used its monopoly over the media in Russia and Russian-occupied parts of Ukraine as its main propaganda tool. By controlling the content distributors and regulating what they share, the government can manipulate its audience. This carefully curated information is strengthened by the elimination of competing views, which can be achieved by buying out or shutting down independent news sources, blocking access to certain websites, and causing internet blackouts at opportune times. Russia uses its control over the information infrastructure, including radio, television, and internet access, to tactically deprive people of access to competing views and ensure its propaganda is the only information available to its intended audience.

The censorship of information, particularly through government control of the internet, which can be shut down relatively easily, creates an issue for war crimes investigators since these shutdowns can cut off the distribution of real-time information sharing from inside the conflict to the outside world. If a government shuts down the internet at the same time as its military forces are overtaking a village and killing civilians, then witnesses and journalists are unable to share photographs, videos, and accounts of what is unfolding, leaving a dearth of evidence for events that have occurred during digital blackouts. Since investigators are led by the evidence, due to internet shutdowns they may focus too heavily on big events with lots of documentation and ignore atrocities that are not captured digitally.

The use of the internet as a domain of battle illuminates the potential pitfalls and digital tripwires that can ensnare and confound modern war crimes investigators. These traps include the problem of investigators getting caught in information silos and failing to account for cognitive or algorithmic biases when sorting through and analyzing digital content. War crimes investigators are not immune from entrapment in these information silos, a hazard that is especially dangerous if they lack self-awareness. Therefore, while investigators should recognize the value of digital open-source information for intelligence, lead information or even evidence, it is necessary to temper enthusiasm for this supply of data with a healthy skepticism and an active awareness of the potential pitfalls of relying too much on digital information sources.

[24] Adam Hulcoop et al., “Tainted Leaks: Disinformation and Phishing with a Russian Nexus,” Citizen Lab, 25 May 2017, https://citizenlab.ca/2017/05/tainted-leaks-disinformation-phish/.

4. DEVELOPING TOOLS FOR WAR CRIMES INVESTIGATORS TO SEE THROUGH THE DIGITAL FOG OF WAR

While there is no easy solution to the above-described issues, several tools could help war crimes investigators in their fight against these growing challenges. The tools include both the creation of standards and guidelines, along with training for investigators and experimentation and application of technological solutions. This section focuses on the *Berkeley Protocol*, the first international standard and guidance for using open-source digital information in the investigation of war crimes,[25] and the use of machine learning and artificial intelligence in investigations.

A. International Investigative Standards

The English-language version of the *Berkeley Protocol* was published in December 2020, and many international organizations and civil society groups received training on it the following year. While there were several ongoing non-international armed conflicts during this period, it was not until Russia's full-scale invasion of Ukraine on 24 February 2022 that the protocol's dissemination and adoption picked up steam.

As the first major conflict to break out since the advance publication of the *Berkeley Protocol* in December 2020 (it will not be officially launched until it is available in all six UN languages, which will occur in mid-2023), the Russia–Ukraine conflict serves as a primary test case as to whether such standards can help address the investigation and legal challenges of a contested information battlefield.

As soon as Russia invaded, the Office of the Prosecutor General of Ukraine took the initiative to translate the protocol's text into Ukrainian. The translated document was distributed to others engaging in the documentation and investigation of what was unfolding in Ukraine. Two weeks into the conflict, the prosecutor general announced on Twitter that her office was using the *Berkeley Protocol* in their investigative work.[26] Soon after, the National Police of Ukraine and, separately, a consortium of Ukrainian civil society groups called the 5 AM Coalition received training on digital

[25] United Nations Office of the High Commissioner for Human Rights and Human Rights Center, UC Berkeley School of Law. *Berkeley Protocol on Digital Open Source Information*, (December 2020); Alexa Koenig, *The New Forensics: Using Open Source Information to Investigate Grave Crimes* (Berkely, CA: Human Rights Center, UC Berkeley School of Law, 2018); Stefano Trevisan, "Open-Source Information in Criminal Proceedings: Lessons from the International Criminal Court and the Berkeley Protocol," *Giurizprudenza Penale* 4 (2021): 9–10; Sam Dubberley, Alexa Koenig, and Daragh Murray, eds. *Digital witness: using open source information for human rights investigation, documentation, and accountability* (New York, NY, Oxford University Press, 2020); Daragh Murray, Yvonne McDermott, and Alexa Koenig, "Mapping the Use of Open Source Research in UN Human Rights Investigations," *Journal of Human Rights Practice* 14, no. 2 (2022): 554-81; https://www.ohchr.org/en/publications/policy-and-methodological-publications/berkeley-protocol-digital-open-source.

[26] Edward Lempinen, "In Ukraine, Berkeley Experts Are Shaping the Legal Fight Against War Crimes," *Berkeley News*, 21 February 2023, https://news.berkeley.edu/2023/02/21/in-ukraine-berkeley-experts-are-shaping-the-legal-fight-against-war-crimes/.

open-source investigations based on the protocol's methodology. In September, the ICC Prosecutor and Eurojust launched "practical guidelines for documenting and preserving information on international crimes," which endorsed the *Berkeley Protocol*.[27]

The speed of the *Berkeley Protocol*'s adoption and the near-unanimous and immediate consensus around its use is a success story in and of itself, aligning a diverse and complex ecosystem of actors who traditionally have not always worked well together.[28] Common standards and definitions are an important way for different groups with different approaches and goals to communicate successfully. Therefore, rather than engaging in crosstalk or remaining in silos, international investigative entities—from civil society documenters and human rights researchers to police and prosecutors—are now, with the guidance of the protocol, getting on the same page. The availability of the document to civil society organizations, which increasingly want to support prosecutors in their pursuit of justice and accountability for war crimes, was also an important watershed in the professionalizing of their work and getting the work on civil society organizations recognized by prosecutors.

In terms of the protocol's substantive guidance, it is too early to assess definitively whether it has succeeded in improving the quality and accuracy of digital investigations in Ukraine. That test will come when the evidence collected today is introduced into court in future trials.

While it has been helpful in this regard, the protocol provides a broad framework that must be adapted to specific operational contexts. To reach a diverse audience in different jurisdictions, the protocol was written as high-level guidance and, in order to future-proof the document, it was intentionally designed to be technology agnostic. Therefore, to be fully effective, the protocol needs to be supplemented with standard operations procedures that are context-specific and technology systems and tools to support the process. Digital evidence collection, preservation, and analysis processes perform best when calibrated to the unique requirements of specific environments and circumstances.

B. Advanced Digital Technologies

The mass adoption of the *Berkeley Protocol* has launched a new dialogue about the most appropriate and effective digital tools to assist prosecutors with these challenges. In particular, there has been a growing desire for technology solutions like the use of

[27] "ICC Prosecutor and Eurojust Launch Practical Guidelines for Documenting and Preserving Information on International Crimes, *International Criminal Court*, 21 September 2022, https://www.icc-cpi.int/news/icc-prosecutor-and-eurojust-launch-practical-guidelines-documenting-and-preserving-information.

[28] Stephen J. Rapp, "Bridging The Hague - Geneva Divide." *The Hague Institute for Global Justice*, 13 January 2017, https://thehagueinstituteforglobaljustice.org/nding-accountability_0bovycq8ok8pjy3pcggx3v/.

artificial intelligence—mainly natural language processing, object recognition, and facial recognition—to sort through the vast quantities of material.

While the application of natural language processing, object recognition, and facial recognition have been experimented with in criminal investigations for some time now, it has taken a while to develop the technology for the context of war crimes investigations. While natural language processing, a branch of artificial intelligence "concerned with giving computers the ability to understand text and spoken words in much the same way human beings can,"[29] has worked well in the more widely spoken languages for some time, it still struggles with rarer languages, localized dialects, and languages written in other scripts like Cyrillic. Similarly, object recognition works well for everyday objects for which there is a lot of training data, like cars, but it is less reliable when it comes to tanks, drones, and weapons in the field. While facial recognition on CCTV works well, it is far less effective when used on hand-held footage or video with occluded faces, which is generally the type of material handled and used by war crimes investigators. There are also many current efforts to develop deepfake detection and other technology tools that will assist in the verification process.

These technologies show promise for assisting investigators in their tasks, but they cannot and should not be seen as something that can replace the work of human investigators. Prosecutors might rely too heavily on them and trust them too readily. There can be bias in the training itself, in the collection of data, and in the fact that sometimes the technology simply gets it wrong. More importantly, many of these tools are still experimental and have not advanced to the stage in which full confidence can be placed in them when a person's life and the legitimacy of the justice system are on the line.

5. CONCLUSION

In less than a decade, the use of digital evidence in international criminal investigations and trials has evolved significantly, and so too have the challenges of making this type of evidence effective in court.[30] Digital technologies are developing at such a rapid pace that there are already complicated new issues arising in the Ukraine conflict that are not addressed in any currently accepted guidance.

[29] "What is natural language processing?" IBM, https://www.ibm.com/topics/natural-language-processing#:~:text=Natural%20language%20processing%20(NLP)%20refers,same%20way%20human%20beings%20can, accessed 8 April 2023.

[30] Lindsay Freeman and Raquel Vazquez Llorente, "Finding the Signal in the Noise: International Criminal Evidence and Procedure in the Digital Age," *Journal of International Criminal Justice* 19, no. 1 (2021): 163–88.

The International Criminal Court and other international justice mechanisms are often criticized for the length of their proceedings. As a result, the use of automated tools, artificial intelligence, and other technology hacks becomes an appealing option for sorting through the unprecedented volume of potentially relevant digital material. However, the very complexities of the information environment necessitate a slow, deliberate, and thorough approach to reviewing digital evidence. War crimes investigators should view technological assistance as providing a useful support function, not as a shortcut that minimizes their effort.

While the *Berkeley Protocol* and the increasing sophistication of investigators mark positive progress in the field, the ongoing conflict in Ukraine reveals the growing need to also recognize the harms and dangers raised by the use of and reliance on new technologies. There is no one solution that will be able to address the multitude of complex ways in which digital technologies are exploited to advance the military and political agendas of Russia, Ukraine, and all the third parties with a stake in this conflict. As a result, investigators need to understand the online environment in which they work as dynamic, constantly changing, and requiring a level of flexibility from war crimes investigators.

ACKNOWLEDGMENTS

The author sincerely thanks Dr. Alexa Koenig, Taťána Jančárková, and the other CyCon reviewers for their thoughtful, candid, and constructive feedback, and the Ukraine investigation team at Berkeley Law’s Human Rights Center, whose research and engaging conversations inspired and contributed to this article.

2023 15th International Conference on Cyber Conflict
Meeting Reality
T. Jančárková, D. Giovannelli, K. Podiņš, I. Winther
2023 © NATO CCDCOE Publications, Tallinn

From Cyber Security to Cyber Power: Appraising the Emergence of 'Responsible, Democratic Cyber Power' in UK Strategy

Joe Devanny
Lecturer
Department of War Studies
King's College London
London, United Kingdom
joseph.devanny@kcl.ac.uk

Andrew C. Dwyer
Lecturer in Information Security
Information Security Group
Royal Holloway, University of London
London, United Kingdom
andrew.dwyer@rhul.ac.uk

Abstract: Across three successive strategies (2009, 2011 and 2016) 'cyber security' was the umbrella concept for United Kingdom (UK) cyber strategy. Conceptual continuity belied changes in substance, as the state played an increasingly active role, particularly domestically. Cyber security remains a top priority in the UK's most recent (2022) strategy, but it was superseded as the umbrella concept by 'cyber power'. We argue that this was a deliberate decision, global in outlook, and with complex and contestable strategic implications. The UK's concept of 'responsible, democratic cyber power' (RDCP) responds to significant changes between 2016 and 2022 in the geopolitics and threat environment affecting (but not confined to) cyberspace. The UK's new cyber strategy promises to align domestic and international actors under an integrated approach, addressing perceived strategic vulnerabilities and exploiting opportunities to pursue national interests. We investigate RDCP's conceptual coherence and strategic utility, tracking its emergence as UK strategic discourse shifted from one of cyber security to cyber power. RDCP offers one avenue for states to coordinate cyber strategy, integrating the various components branded under 'cyber' as an instrument of national strategy – pursuing security, prosperity, and projection of national values and influence. However, there are different potential interpretations of RDCP and an even greater number of potential ways to implement it. In the UK, as elsewhere, effective cyber power requires prioritization about what a state values, whether in developing a resilient and competitive cyber ecosystem or in meeting the challenges and threats posed by systemic competitors. We conclude by reflecting on what it means to be a 'medium-sized, responsible and democratic cyber power' in an era of increasing inter-state competition in cyberspace.

Keywords: *cyber power, inter-state competition, national cyber strategy, United Kingdom*

1. INTRODUCTION

Cyber security is central to state strategy amidst the return of overt geopolitical competition, recent avowals of cyber forces by several states, and recognition that cyberspace supports both economic prosperity and the projection of national values. Many states have published national strategies addressing (in)security in cyberspace and exploiting its perceived opportunities. Scholarship on cyber strategy reflects the mainstreaming of cyber security within wider national strategy (Fischerkeller, Goldman, and Harknett 2022), alongside debates about the broader concept of 'cyber power' (Kramer, Starr, and Wentz 2009; Betz and Stevens 2011; Smeets 2022). Meanwhile, the focus of research is increasingly broader than the state as researchers are mindful of the private sector and civil society as actors and targets in cyberspace (Maschmeyer, Deibert, and Lindsay 2021). Analysis of published cyber strategies offers an important method to investigate how states perceive and intend to address the strategic implications of threats and opportunities in cyberspace, what audiences they seek to influence, and to what end.

In recent years, as cyber strategies have expanded in scope, they have presented 'whole-of-government' approaches as insufficient, instead advocating for greater inclusion of industry and civil society with formulations such as 'whole-of-system', 'whole-of-society' and 'whole-of-cyber' (Devanny 2021). This has occurred alongside growing public acknowledgement over the past 20 years of the role of cyber operations in national strategy (Healey 2013).[1] Corresponding institutional arrangements, in intelligence agencies and military cyber commands, should therefore be explored as part of an integrated national strategy. This challenges institutions more accustomed to secrecy – the 'Ronan Keating doctrine' of 'saying it best when saying nothing at all' (Dwyer and Martin 2022) – and now increasingly expected to strike a delicate balance between saying too much (and risking compromising equities) and saying too little to satisfy the imperatives of strategic communications (Buchan and Devanny 2022).

The United States looms large in the literature about cyber strategy and its institutional arrangements. This is understandable, given the power and influence of the US and

[1] We use cyber operations to encompass a range of activities that are sometimes referred to as 'offensive' but may not be exclusively characterized as such.

the comparatively greater public availability of information about its activities and doctrine. But cyber strategy is relevant globally. More states are publishing cyber strategies and establishing cyber security centres. This surface similarity masks significant national variation in the experience of devising and implementing cyber strategy. Partly, variation stems from inevitable asymmetries of states' capabilities. However difficult it is to define 'cyber power', some states clearly have greater capabilities than others (Willett 2019; Voo, Hemani, and Cassidy 2022). Variation also stems from the difficulty of achieving strategic objectives in and through cyberspace, including the challenge (in any government) of navigating the inter-institutional and bureaucratic politics of cyber strategy (Harknett and Smeets 2022; Valeriano, Jensen, and Maness 2018; Lindsay 2021).

The United Kingdom (UK) is one 'medium-sized' state that has tried to develop and implement a distinctive cyber strategy for over a decade against the same backdrop of geopolitical competition facing other states. This article examines UK cyber strategy since 2009, contributing to the growing literature aimed at understanding how different states experience and address challenges in cyberspace. This turn in scholarship mirrors the increasing participation of more states (and non-state stakeholders) in global diplomacy about the future of the Internet and norms of responsible state behaviour in cyberspace (Kavanagh 2017). Published UK cyber strategies, alongside public interventions by senior officials, comprise the national perspective of one capable state actor on contemporary developments in cyberspace. The UK government plays an active role in global cyber diplomacy and has recently portrayed itself as a 'cyber power'. In proposing and accepting the challenge of developing a narrative of 'responsible, democratic cyber power' (RDCP), the UK highlighted many of the issues facing states in formulating, communicating (domestically and internationally), and implementing cyber strategy.

This article argues that the UK's adoption of 'cyber power' as a strategic umbrella concept was a deliberate decision, global in outlook, with complex and contestable strategic implications. In the first section, we track the development of the UK's four national cyber (security) strategies (2009, 2011, 2016 and 2022). We note how RDCP responded to significant geopolitical and cyber security changes between 2016 and 2022. The UK's 2022 strategy promised to better integrate domestic and international efforts, address perceived strategic vulnerabilities, and exploit opportunities to pursue national interests. The second section of the article investigates RDCP's conceptual coherence and strategic utility. RDCP – construed broadly as the impact of democratic values and accountability arrangements on the responsible exercise of power in cyberspace – offers a framework for national cyber strategy. It integrates the various components branded under 'cyber' as instruments of national strategy pursuing security and prosperity, and projecting national values and influence. RDCP is, however, open to several different interpretations and methods of implementation. In the UK, as elsewhere, effective application of cyber power requires decisions about what is valued and prioritized, whether in developing resilient and competitive cyber ecosystems or meeting the challenges and threats posed by geopolitical competitors. We conclude by reflecting on what it means to be a 'medium-sized, responsible and democratic cyber power' in this era of increasing inter-state competition in cyberspace.

2. A BRIEF HISTORY OF UK CYBER STRATEGY

In the first three successive UK strategies (2009, 2011 and 2016), 'cyber security' was the framing concept, only replaced by 'cyber power' in 2022. This section addresses the principal similarities and differences between these four iterations, situating them in the context of wider national security strategy and politically across six successive UK premierships. We identify key themes in the UK's emerging approach, including the rising prominence of cyber operations, and the increasing size and ambition of each published strategy.

The 2009 Strategy

The UK government published its first National Cyber Security Strategy (NCSS) in 2009 (HM Government 2009a). This followed a wider trend, towards the end of a long period of Labour government (1997–2010), in which several national security documents and initiatives were created, including the first National Security Strategy (2008). Before the 2010 general election, there was cross-party recognition that national security coordination needed to improve. The Brown government's cyber security strategy is an example of that trend, which drew some inspiration from contemporary US practice (Devanny and Harris 2014). The UK's first strategy was several years behind the US National Strategy to Secure Cyberspace (2003), and unsurprisingly behind the UK's first counter-terrorism strategy (2003 – published in 2006). Cyber security has lagged counter-terrorism – and, indeed, other 'cyber' priorities such as intelligence collection – as a national security priority (Hannigan 2019, 10), and the UK often historically lags the US in national security transparency. In retrospect, 2009 was the start of a period in which cyber security was steadily elevated as a UK government priority, vis a vis both other cyber, and non-cyber, priorities.

The 2009 strategy was relatively short (32 pages). It was much shorter than the government's contemporaneous 'Digital Britain' report (HM Government 2009b) on the economic impact of digital technology that did not mention cyber security. The 2009 strategy's subtitle indicated its priorities: 'Safety, Security and Resilience'. It was co-published by two then relatively new institutions: the Cabinet Office's Office for Cyber Security; and the Cyber Security Operations Centre, led by the cyber, signals intelligence and security agency, Government Communications Headquarters (GCHQ). One former top UK cyber official described the 2009 strategy, not unfairly, as a ‘scoping’ phase (Hannigan 2019, 3). This scoping phase had enduring impact, establishing the analytical foundations for subsequent strategies. Not all its reforms lasted, and its basic approach acknowledged – and left unchallenged – GCHQ’s primacy as the UK’s most cyber-capable institution. This remains true today, but GCHQ’s primacy is offset by the progress made by other institutions. Digital policy and regulatory responsibilities grew in other departments, such as the Department of Digital, Culture, Media, and Sport (and by 2023 the new Department for Science, Innovation and Technology). Successive strategies highlighted the Foreign, Commonwealth and Development Office’s (FCDO) growing role in cyber diplomacy, providing evidence that UK strategy is not monolithic, but is produced through institutional plurality.

The 2009 NCSS frames the problem of cyber security in high-level brushstrokes. It coined three priorities: reducing risk from the UK’s use of cyberspace; exploiting opportunities in cyberspace; and improving the underlying knowledge, capabilities, and decision-making necessary for successful strategy (HM Government 2009a, 3). The strategy obliquely mentioned the need to ‘intervene against adversaries… to exploit cyber space to combat threats from criminals, terrorists and competent state actors’ (HM Government 2009a, 4, 19). It, therefore, started to address the problems of governmental cyber security coordination. It identified the need to improve governance, capability, and doctrine as well as to facilitate the growth of an increasingly digital, secure, and resilient economy and society. Notwithstanding its enduring logic and analysis, the strategy’s impact was inevitably affected by the impending general election. In May 2010, the Brown premiership was replaced by a Conservative-led coalition government.

The 2011 Strategy

The Conservative-Liberal Democrat coalition government entered office determined to handle national security issues differently from its Labour predecessors in the context of controversy regarding the ‘war on terror’ and military operations in Afghanistan and Iraq. Under the coalition, cyber security became more prominent. This reflected the increasing salience of cyber operations in international security, e.g., with the first public reporting about Stuxnet (2010), Shamoon (2012), and Edward Snowden’s allegations about US and wider Five Eyes digital intelligence (2013). The UK government prominently featured cyber security in speeches, strategies, and initiatives. It was presented as an example, domestically, of the new government’s security credentials and investment and, internationally, of UK leadership in multilateral cyber diplomacy (e.g., the 2011 London Conference on Cyberspace). At a time of significantly reduced public expenditure under ‘austerity’ fiscal consolidation, cyber security benefited from a growing budget, innovations in coordination, and institutional reform.

The coalition's 2011 NCSS should be interpreted within a wider, five-yearly framework of National Security Strategies (NSSs) and Strategic Defence and Security Reviews (SDSRs). Both the 2011 and 2016 Cyber Security Strategies followed the top-level priorities of this NSS/SDSR process. The coalition's 2010 NSS identified cyber security as one of four top-tier risks, noting the UK's 'comparative advantage' to achieve 'economic and security opportunities' in and through cyberspace (HM Government 2010a, 30). The SDSR highlighted the negotiation of a UK-US Memorandum of Understanding (MoU) to facilitate information-sharing and joint military cyber operations (HM Government 2010b, 48). The NSS/SDSR process overall committed to investing £650m over four years (eventually uplifted to £860m) in a new National Cyber Security Programme (NCSP). Almost two-thirds of this investment went to the intelligence agencies, primarily GCHQ (HM Government 2010b, 47; HM Government 2011, 25). Reform placed the Office for Cyber Security and Information Assurance, as well as (from 2013) the UK's national Computer Emergency Response Team (CERT UK), under a Deputy National Security Adviser for Intelligence, Security and Resilience. The coalition further reshaped the cyber security institutional landscape, creating: a Joint Forces Cyber Group in the new Joint Forces Command (2012); a Joint Forces Cyber Reserve (2013); a Centre for Cyber Assessment (CCA) (2013); and a national CERT based in the Cabinet Office (2013–14).

The 2011 NCSS, we argue, should be understood as an effort to reshape – and to reshape the public narrative about – the governmental cyber agenda. It was slightly longer than its 2009 precursor and re-framed the UK's top priorities as: tackling cyber-crime and being one of the most secure places in the world to do business online; improving resilience to cyber-attacks; helping shape an open, vibrant, and stable cyberspace to support open societies; and building UK cyber security knowledge, skills, and capability (HM Government 2011, 8). Its subtitle, 'Protecting and promoting the UK in a digital world' highlighted the strategy's broad remit and multiple audiences. It was an exercise in explaining and promoting the UK's agenda domestically and internationally. In continuity with the previous strategy, the 2011 NCSS: emphasized the need to collaborate with business and civil society to improve cyber security awareness and best practice; recognized the continuing need to build domestic cyber security capacity; and discreetly mentioned the requirement to develop sovereign capability 'to detect and defeat high-end threats' (HM Government 2011, 9). It also prominently embraced international engagement and the emerging field of cyber diplomacy, noting the London Conference to develop multilateral and multistakeholder 'rules of the road' for cyberspace.

Overall, the coalition government's cyber strategy did not break fundamentally with the 2009 strategy. It continued the UK's growing recognition of the need to improve national and international coordination to address cyber threats. It also demonstrated that senior figures in government perceived cyber security as a key component of a wider public narrative about security. In former GCHQ Director Robert Hannigan's later assessment, the 2011–16 period highlighted the limits of what could be achieved with the existing approach. Its limitations showed the need for a more active, shaping role for government in cyber security (Hannigan 2019). This was an incremental shift, subsequently embedded in the 2016 NCSS.

The 2016 Strategy

At the 2015 general election, the Conservative party won an outright majority, ending the coalition and establishing the Conservatives as the sole party of government. In this context, the UK produced a new NSS/SDSR in 2015, providing a new framework for the next NCSS in 2016. However, the NCSS was published in November 2016, four months into Theresa May's premiership, five months after the Brexit referendum that had led to Cameron's resignation. Beyond the period's unsettled domestic politics, there were likewise geopolitical changes, including continuing repercussions for European security of the 2014 Russian invasion of Ukraine. This was accompanied by the increasing salience of great power competition as an international security theme, especially in changing attitudes about how the UK and other states should address China's rising power and influence. Each of these factors was bigger than cyber security, but each affected the way that the UK made and implemented cyber strategy.

The 2016 NCSS emerged in this new context. The government adopted a more active role, shaping the national effort to improve cyber security. It devoted more resources to cyber security, increasing the NCSP to £1.9bn. The 2016 creation of the National Cyber Security Centre (NCSC) was the new approach's most prominent institutional manifestation. It provided more clarity, visibility, and leadership to the government's cyber security agenda (Hannigan 2019). The strategy re-phrased the top national cyber priorities as defend, deter, and develop (HM Government 2016, 9). More streamlined than the four 2011 priorities, they closely followed earlier approaches, including the emphasis on cyber diplomacy (described as 'international action').

The NCSC amalgamated several precursor bodies, most prominently GCHQ's information assurance arm (CESG), and the cyber aspects of the Centre for the Protection of National Infrastructure (CPNI), and absorbed newer entities, such as CERT UK and the CCA, created during the previous implementation period. Still formally part of GCHQ, the NCSC adopted a more public-facing profile. Its first Chief Executive, Ciaran Martin, was an articulate and visible cyber-security leader. This was an important part of the new approach, improving cyber security coordination across government, engaging with the private sector, and offering an internationally leading approach to cyber security organization.

This was a period of incremental progress towards what became (in 2020) the National Cyber Force (NCF). Its precursor entity, Defence-GCHQ collaboration under the National Offensive Cyber Programme (NOCP), progressed slowly through inter-institutional deliberations about how best to proceed (Devanny et al. 2021, 11–12; Blessing and Austin 2022, 30). The 2016 NCSS referred to cyber operations in a guarded manner, emphasizing that: 'The principles of deterrence are as applicable in cyberspace as they are in the physical sphere... the full spectrum of our capabilities will be used to deter adversaries and to deny them opportunities to attack us' (HM Government 2016, 47). Shortly before the NCSS's publication, in September 2016 the UK and US finally signed the MoU first mooted in 2010. Shortly after that, both the UK and US commenced cyber operations against the so-called Islamic State group (Devanny et al. 2021, 11).

Despite the domestic political context (Brexit) and the wider geopolitical currents since the 2011 strategy, the 2016–21 implementation period saw a settled effort to build on previous strategies, creating new institutions like the NCSC and increasing funding of the NCSP. There were also increasingly public statements by government officials and ministers about the role of offensive capabilities in UK cyber strategy, and a commitment (in 2018, realized in 2020) to create the NCF (Devanny et al. 2021, 10–12). Indications of the shift towards 'cyber power' as a framing concept for UK strategy could be seen in these increasingly public mentions of cyber operations, as well as in the UK's increasing ambition about the size of the NCF (from a proposed size of c. 500 in 2015 to a target of 2,000 personnel in 2018 and a 3,000 target in 2020). This was, arguably, reflective of growing global unease about cyber security threats, not least in the form of the wave of ransomware crime, and a need to demonstrate a capacity to respond more effectively than before.

The 2022 Strategy

The 2022 National Cyber Strategy (NCS) was published in December 2021, two years after Boris Johnson's emphatic general election victory of December 2019 – and just six months before the end of Johnson's turbulent premiership. Politically and geopolitically, Brexit, COVID-19, and the Russian invasion of Ukraine set the context in which this iteration of UK cyber strategy was developed and implemented. The NCS was also launched alongside growing public awareness of state hacking of IT supply chains following the SolarWinds (2020) and Microsoft Exchange (2021) incidents. The NCS likewise followed the Johnson premiership's flagship national security strategy, the Integrated Review of Security, Defence, Development and Foreign Policy (UKIR) (HM Government 2021a).

The UKIR re-framed cyber strategy under five pillars: strengthening the UK cyber ecosystem; building a resilient and prosperous digital UK; taking the lead in the technologies vital to cyber power; advancing UK global leadership and influence; and detecting, disrupting, and deterring adversaries in and through cyberspace, 'making more integrated, creative and routine use of the UK's full spectrum of levers' (HM Government 2021a, 41; HM Government 2021b, 11–13). Whilst the list is re-ordered and additional aspects are elevated compared to the 2016 NCSS, the top-level prioritization remains remarkably consistent with previous strategies. The NCS is more explicit than its precursors about the role of cyber operations in wider strategy, but with an important caveat that, to date, the UK had not achieved its intended deterrent outcomes with its adversaries (HM Government 2021b, 25). What is, however, less clear from reading the NCS and other UK statements is whether this implies that the UK thinks that more successful deterrence will come from new approaches to cyber operations, or from intensifying existing efforts.

The most notable change in the NCS was its replacement of cyber security – removed from the strategy's title – with the new framing concept of cyber power. This was not a surprise development (Devanny 2021, 64). Cyber power had already appeared in speeches by senior UK officials, notably in a 2019 speech by GCHQ Director Jeremy Fleming (Fleming 2019). It then featured prominently in the UKIR. Even prior to the UKIR's publication, Johnson had pre-announced that the NCF had attained operational capacity. The notional rationale for the NCS's title change was to highlight that a national cyber strategy needed to encompass more than cyber security. As one commentator explained: '[the NCS] elevates the cyber domain from a security concern for technology specialists to a wide-ranging theme of grand strategy—one that will no longer be a "whole-of-government" initiative but will expand into a "whole-of-society" effort' (Beecroft 2021).

The NCS describes cyber power as 'an ever more vital lever of national power and a source of strategic advantage... [It] is the ability to protect and promote national interests in and through cyberspace' (HM Government 2021b, 11). Evident in the UK's understanding of cyber power is its capacity to be more than the deployment of cyber capabilities, extending to cyber diplomacy and capacity-building, as well as the contribution of digital technologies to national prosperity. This expanded view in the UKIR unfolds in a more complex, somewhat under-developed concept: 'Responsible, Democratic Cyber Power' (RDCP) (HM Government 2021a, 40). Regrettably, the UKIR did not precisely define RDCP. It interchangeably referred to RDCP and 'responsible cyber power' (the latter reminiscent of a common phrase in multilateral cyber diplomacy, ‘responsible state behaviour in cyberspace’). The precise definition of the ‘democratic’ element of RDCP is elusive throughout the UKIR (Devanny 2021). The short passage devoted to RDCP (HM Government 2021a, 40–42) implies that it is conceived principally as an operational concept. It states that the UK conducts responsible, targeted, and proportionate operations in cyberspace, in explicit contrast with its adversaries’ less responsible behaviour (HM Government 2021a, 42).

The NCS expands on the UKIR’s development of RDCP, developing its diplomatic aspects, combining the promotion of international stability, upholding the rules-based international order, and championing values such as ‘human rights, diversity, and gender equality’ (HM Government 2021b, 95). It likewise emphasizes collaboration ‘with like-minded nations to promote our shared values of openness and democracy’ (HM Government 2022b, 33). The NCS clarifies that cyber diplomacy, aimed at opposing ‘digital authoritarianism’ and defending citizens’ rights in cyberspace, including advocating ‘democratic values’ in international technology standards, is an important addition to RDCP’s emphasis on responsible operations (HM Government 2022b, 34, 88). RDCP’s expansion to include a wider range of foreign-policy objectives reflects the FCDO’s growing contribution to UK cyber strategy – evident in the significantly-increased size of its Cyber Policy Department (Center for Strategic and International Studies 2022).

Much of the Strategy’s first year of implementation was dominated by the Russian invasion of Ukraine, continuing efforts to address the ransomware crisis, and domestic turbulence across three successive premierships (Johnson, Liz Truss, and Rishi Sunak). This was not the ideal political context for stable stewardship of national strategy, but the mechanics of cyber strategy appeared (publicly, at least) to proceed largely unaffected. The UK’s rhetoric about RDCP was by now well-established, but there was still an open question about its longevity. It is reasonable to speculate about whether RDCP will long survive the May 2023 retirement of Jeremy Fleming (Nicholls 2023), how it will fare under Fleming’s successor as GCHQ Director, Anne Keast-Butler – and how it would translate into a coordinated programme of action across government (e.g., cyber diplomacy led by the FCDO, cyber operations by the NCF). There was clear evidence of ongoing engagement and outreach to academia, industry, and other states, for example, in a Wilton Park conference sponsored by FCDO in November 2022 (Buchan 2022). This engagement fed into wider UK government efforts to develop a vision of RDCP in practice. But questions remained about how best to implement this vision and persuade other states of its merits. This is the subject of the final section.

3. TRANSLATING RESPONSIBLE, DEMOCRATIC CYBER POWER

The longevity of RDCP as a framing concept for UK cyber strategy will depend on whether it continues to be championed by advocates within government and whether it can achieve the (primarily international) objectives of UK strategic communications regarding its cyber strategy. The concept of RDCP is very broad and can serve multiple objectives. It entails the use of hard and soft power in pursuit of the 'national interest' (security, prosperity, values), manifesting both a 'power-based' and 'rules-based' approach to international security (Libicki 2021). Specifically, we suggest that the UK's vision of RDCP can be summarized across four elements:

1) Integration of the UK's cyber ecosystem, including sovereign assets, in the pursuit of the national interest.
2) Diplomatic efforts to shape the future of cyberspace in accord with national interests and democratic values, including through efforts to support cyber capacity-building.
3) 'Operating responsibly' through practising restraint, proportionality, and upholding applicable international law, rules, and norms.
4) Emphasizing liberal democratic processes of accountability and oversight (involving the executive, legislature, judiciary, and wider stakeholders), and including (a degree of) transparency about how these activities are enacted.

There are potential benefits, both domestic and international, from adopting a strategic narrative about RCDP. A recent example of shaping the public narrative domestically was GCHQ's then director, Jeremy Fleming, accepting (not without controversy) an invitation to be a guest editor of the BBC's flagship current affairs programme, *Today* (Targett 2023). This suggests a calculation that the more visible and transparent the UK's cyber actors become – by emphasizing their responsible and democratic attributes – the more public confidence and support there will be for their activities. Another example of this approach was the NCF's publication in April 2023 of a document articulating the UK's approach to offensive cyber operations and identifying three guiding principles of responsible cyber operations – accountability, precision, and calibration (HM Government 2023, 14).

In this respect, the RDCP strategic communication campaign can be seen as a kind of insurance policy against the impact of possible future adverse headlines, e.g., of the Snowden variety. It is an example of pre-emptively, pro-actively shaping the conversation, rather than simply waiting to react in a crisis. For liberal democratic states, effective public communication is not just prudent but essential to accountability and oversight. RDCP also directly intersects with key government actors outside the realm of cyber security, whether in the setting of international technical standards, elaborating legal safeguards over foreign investment, or developing exports that promote interoperable standards embedding strong security and privacy.

However, it is the broader concept of cyber power – rather than its elaboration in RDCP – that underpins the coherence of this diverse range of UK government actors beyond cyber security. The success of RDCP will therefore depend on how effectively the different institutions within the UK government are coordinated internally – both through strategic narrative and policy development – to exert an active leadership role in driving the 'whole-of-system' agenda. Correctly calibrating RDCP's bureaucratic politics will be crucial. It will require a clear sense of leadership and purpose. This is, of course, true of any cyber strategy, whether pursued under the umbrella of RDCP or another organizing principle.

RDCP's prudential logic scales up internationally through cyber diplomacy. Developing a strong narrative about responsible and democratic behaviour could serve UK foreign policy by influencing 'middle ground' states beyond the like-minded group in multilateral cyber negotiations (HM Government 2021b, 94). Here, the 'R' and 'D' in RDCP clearly situate the UK against the behaviour in cyberspace of less responsible, less democratic adversaries. We argue that this was the original intention of departing from the more concise notion of 'responsible state behaviour' that is commonly associated with multilateral cyber diplomacy. In adding 'democratic', the UK implicitly criticizes the behaviour of those states that use cyber power in ways that undermine democracy and democratic values. It also enables the UK to develop a distinctive space between the few most powerful cyber actors and the many states with less capability. However, several commentators have noted that this is not an easy task, given concerns amongst other states over the potential impact and implications of, for example, the US strategy of persistent engagement (Shires and Smeets 2021). Likewise, some have noticed a latent tension between 'cyber power' and 'cyber security' as UK strategic priorities (Dwyer and Martin 2022).

RCDP faces challenges in how it can be effectively promoted as a model for other states to emulate. This is not a problem unique to RDCP. It reflects a context in which progress in cyber diplomacy is difficult, and states like the UK try to play an active, constructive role (Buchan and Devanny 2022). Put simply, the RDCP concept is only likely to succeed if the core concept of 'cyber power' is received favourably by the states it is intended to influence. If there is confusion about the UK's references to 'democratic' uses of cyber capabilities, or allergic reaction to the language of 'power' in the rhetoric of persuasion, then RDCP might need to be reconsidered. Ultimately, the objectives of UK strategy could be pursued under the more traditional rubric of cyber diplomacy – 'responsible state behaviour in cyberspace'. Why would other states, seeking to promote the norms-based order in cyberspace, embrace RDCP?

The answer implicit in RDCP is that states could find in it a foundation to engage internationally and to allay any fears that the avowal (by the UK and other states) of offensive cyber capabilities is tantamount to militarizing cyberspace. Yet, the translation of RDCP for other states to embrace raises at least three issues. First, RDCP, like the broader concept of cyber power, is contested and difficult to quantify (Voo, Hemani, and Cassidy 2022). States possess diverse interpretations of what 'responsibility' and 'democracy' mean in the context of international security. This would make assessing RDCP across states – if it were widely adopted – exceptionally difficult. Second, given that the political independence of states is a fundamental precept of international relations, it is not obvious that the 'democratic' element of RDCP improves upon the more concise, more established, and less domestically prescriptive UN concept of 'responsible state behaviour in cyberspace'. As Johanna Weaver has noted, in cyber diplomacy most states are more preoccupied with international stability than with liberal democratic values. Consequently, it arguably makes more diplomatic sense to carefully calibrate the extent to which UK diplomacy prioritizes concentration on the impact of state behaviour on international stability, rather than on the promotion of democratic values (Weaver 2022). The interchangeable language of the UKIR regarding RDCP and 'responsible cyber power' perhaps suggests that the UK recognizes this nuance. Notably, perhaps, the NCF's recent publication follows this approach, referring to RDCP in the body of the document but to 'Responsible Cyber Power' in its title. This could suggest an on-going refinement of the UK's strategic communications and use of the RDCP concept. Finally, third, there is some real doubt about whether 'power' is a useful trope of cyber diplomacy, in that it could potentially alienate some states which might perceive their own lack of power or powerlessness (Buchan 2022). Might the rhetoric of cyber power be counterproductive for UK policy? It is possible that UK strategy would be better articulated by placing greater emphasis on the UK's cooperative, collaborative role as a partner to many states in global cyber diplomacy, pursuing the incremental gains of quiet leadership in cyber diplomacy and capacity building. This approach might be more attractive and successful diplomatically because it is softer – selling the benefits of what the UK has to offer as a partner, rather than emphasizing the image of its strength as a 'power'. Such a critique echoes in some ways the debates about and reception of the 'responsibility to protect' (R2P) (Crossley 2018). RDCP could even be interpreted – although this interpretation moves beyond the UK's explanation of the concept – to suggest that states have an obligation to use cyber operations to protect populations, or to pursue collective countermeasures. Such implications are likely to appeal to some states more than others (Buchan and Devanny 2022).

Whatever its rhetoric and choice of framing concept for strategic communications, the UK is committed to an active, international cyber strategy: engaging in multilateral (and promoting multistakeholder) efforts to shape global norms of responsible behaviour in cyberspace; providing, funding and sharing best practice regarding cyber capacity building; and working to ensure that regulations and standards for next-generation technologies work for democracies and do not benefit illiberal, authoritarian states. It cannot match the scale of US cyber operations – and it will continue to work as closely as possible to align its operations with those of the US and other allies. Nonetheless, the UK is making a significant investment in offensive cyber operations, offering its growing capability to NATO under Article V and via the Sovereign Cyber Effects Provided Voluntarily by Allies (SCEPVA) mechanism (Devanny et al. 2021, 16). Cumulatively, you might say that this classifies the UK as an 'upper-middle power' in the arena of global cyber cooperation, competition, and conflict. The case study of RDCP demonstrates that the UK is self-aware about the potential latent in its national combination of active diplomacy, convening power, and thought leadership, alongside the hard power of its cyber capabilities. Its advocacy of RDCP could be seen as one interpretation – amongst many – of how the UK should play these cards on behalf of the rules-based international order, the promotion of stability, democratic values, and, of course, its national interest. It offers one – to date under-elaborated – model for other 'middle power' states to follow.

4. CONCLUSION

Published national cyber strategies serve an important function in strategic communication and public diplomacy. They are an opportunity to demonstrate transparency, persuade diverse audiences, and shape opinion. They can also be a 'fudge' – a compromise between different institutional actors, reflecting their respective equities and viewpoints. Four iterations of UK cyber strategy have effectively elevated the priority of cyber security in public debate and have explained for multiple audiences the key points of UK strategy. Since 2009, the UK has maintained a broadly consistent focus on developing its national cyber ecosystem, improving cyber security, and developing its resilience. The rise of (responsible, democratic) cyber power as the framing concept of UK cyber strategy embraced two developments in particular – the rising salience of cyber operations as a publicly-avowed aspect of state strategy, and recognition of the relevance and role of other stakeholders (and increasingly inter-state geopolitical competition).

This article has provided an overview of RDCP's emergence as the UK's new umbrella concept for cyber strategy. It identifies the major challenges facing RDCP, particularly in terms of its international appeal. This is consequential, as RDCP is best interpreted as a framework for translating cyber power for cyber diplomacy, not as a structuring concept for the domestic elements of UK cyber strategy. The language of 'power' poses difficult questions for medium-sized states engaging in cyber operations whilst upholding the increasingly challenged rules-based international order. The evolving interpretation of RDCP indicates that it is still a fluid concept. Its next steps are still to be determined three years after UK officials first publicly invoked 'cyber power'. The reception of the NCF's recent document on Responsible Cyber Power will likely help to shape the next phase of this process. The durability of the 'democratic' element may be limited. It presents challenges and perhaps restricts RDCP's appeal in international engagement. The focus on 'responsibility' may better serve UK objectives and align more closely with wider international discourse about responsible state behaviour in cyberspace. The longevity of the RDCP phrase is therefore much less significant than the effectiveness of the underlying diplomacy, policy, and operations encompassed by it.

ACKNOWLEDGEMENTS

We thank our colleagues – and fellow researchers in other institutions – from whom we have benefited in discussing UK cyber strategy. We are immensely grateful for the constructive feedback provided by the CyCon reviewers, and also to those practitioners who have engaged with us and other academics in discussing UK strategy. Joe Devanny is also grateful to the British Academy for the Innovation Fellowship during which he was able to undertake this research.

REFERENCES

Beecroft, Nick. 2021. 'The UK's Cyber Strategy Is No Longer Just About Security'. Carnegie Endowment for International Peace. 17 December 2021. https://carnegieendowment.org/2021/12/17/uk-s-cyber-strategy-is-no-longer-just-about-security-pub-86037.

Betz, David and Tim Stevens. 2011. *Cyberspace and the State: Toward a Strategy for Cyber-Power.* London: Routledge for the International Institute for Strategic Studies.

Blessing, Jason, and Greg Austin. 2022. *Assessing Military Cyber Maturity: Strategy, Institutions and Capability*. London: International Institute of Strategic Studies.

Buchan, Russell, and Joe Devanny. 2022. 'Clarifying Responsible Cyber Power: Developing Views in the U.K. Regarding Non-intervention and Peacetime Cyber Operations'. *Lawfare Blog*, 13 October 2022. https://www.lawfareblog.com/clarifying-responsible-cyber-power-developing-views-uk-regarding-non-intervention-and-peacetime.

Buchan, Russell. 2022. 'Summary Report: Acting Responsibly in Cyberspace WP3146'. FCDO/Wilton Park. 14–16 November 2022. https://www.wiltonpark.org.uk/app/uploads/2023/01/WP3146-Summary-Report-Acting-Responsibly-in-Cyberspace-Nov-2022-1.pdf.

Center for Strategic and International Studies. 2022. ‘Placing Cyber Diplomacy at the Top of the Agenda’. *Inside Cyber Diplomacy*, 15 August 2022. https://www.csis.org/node/66558.

Crossley, Noele. 2018. ‘Is R2P Still Controversial? Continuity and Change in the Debate on “Humanitarian Intervention”’. *Cambridge Review of International Affairs* 31(5): 415–36. https://doi.org/10.1080/09557571.2018.1516196.

Devanny, Joe and Josh Harris. 2014. *The National Security Council: National Security at the Centre of Government*. London: Institute for Government.

Devanny, Joe, Andrew Dwyer, Amy Ertan, and Tim Stevens. 2021. *The National Cyber Force that Britain Needs?* London: King’s Policy Institute, Cyber Security Research Group and UK Offensive Cyber Working Group.

Devanny, Joe. 2021. ‘The Review and Responsible, Democratic Cyber Power’. In *The Integrated Review in Context: Defence and Security in Focus*, edited by Devanny, Joe and John Gearson, 62–64, London: Centre for Defence Studies.

Dwyer, Andrew, and Ciaran Martin. 2022. ‘A Frontier Without Direction? The U.K.’s Latest Position on Responsible Cyber Power’. *Lawfare Blog*, 1 August 2022. https://www.lawfareblog.com/frontier-without-direction-uks-latest-position-responsible-cyber-power.

Fischerkeller, Michael P., Emily O. Goldman, and Richard J. Harknett. 2022. *Cyber Persistence Theory: Redefining National Security in Cyberspace*. Oxford: Oxford University Press.

Fleming, Jeremy. 2019. ‘Director’s Speech on Cyber Power – As Delivered’. GCHQ. 25 February 2019. https://www.gchq.gov.uk/speech/jeremy-fleming-fullerton-speech-singapore-2019.

Hannigan, Robert. 2019. *Organising a Government for Cyber: The Creation of the UK’s National Cyber Security Centre*. London: Royal United Services Institute (RUSI).

Harknett, Richard J., and Max Smeets. 2022. ‘Cyber Campaigns and Strategic Outcomes’. *Journal of Strategic Studies* 45(4): 534–67.

Healey, Jason, ed. 2013. *A Fierce Domain: Conflict in Cyberspace 1986–2012*. Arlington, VA: Cyber Conflict Studies Association (CCSA).

HM Government. 2009a. *Cyber Security Strategy of the United Kingdom: Safety, Security and Resilience in Cyber Space (Cm 7642)*.

HM Government. 2009b. *Digital Britain: Final Report (Cm 7650)*.

HM Government. 2010a. *A Strong Britain in an Age of Uncertainty: The National Security Strategy (Cm 7953)*.

HM Government. 2010b. *Securing Britain in an Age of Uncertainty: The Strategic Defence and Security Review (Cm 7948)*.

HM Government. 2011. *The UK Cyber Security Strategy: Protecting and Promoting the UK in a Digital World*.

HM Government. 2016. *National Cyber Security Strategy 2016–21*.

HM Government. 2021a. *Global Britain in a Competitive Age: The Review of Security, Defence, Development and Foreign Policy (CP 403)*.

HM Government. 2021b. *National Cyber Strategy*.

HM Government. 2023. *The National Cyber Force: Responsible Cyber Power in Practice*.

Kavanagh, Camino. 2017. *The United Nations, Cyberspace and International Peace and Security: Responding to Complexity in the 21st Century*. Geneva: United Nations Institute for Disarmament Research.

Kramer, Franklin D., Stuart Starr, and Larry K. Wentz, eds. 2009. *Cyberpower and National Security*. Washington, DC: National Defense University Press.

Libicki, Martin. C. 2021. Obnoxious Deterrence. *14th International Conflict on Cyber Conflict: Keep Moving*. Tallinn: NATO CCD COE, 65–77.

Lindsay, Jon R. 2021. 'Cyber Conflict vs. Cyber Command: Hidden Dangers in the American Military Solution to a Large-Scale Intelligence Problem'. *Intelligence and National Security* 36(2): 260–78.

Maschmeyer, Lennart, Ronald J. Deibert, and Jon R. Lindsay. 2021. 'A Tale of Two Cybers – How Threat Reporting by Cybersecurity Firms Systematically Underrepresents Threats to Civil Society'. *Journal of Information Technology & Politics* 18(1): 1–20.

Nicholls, Dominic. 2023. 'Spy Chief Sir Jeremy Fleming to Step Down as Director of GCHQ'. *Telegraph*, 26 January 2023. https://www.telegraph.co.uk/news/2023/01/26/sir-jeremy-fleming-step-director-gchq/.

Nye, Joseph S. 2010. *Cyber Power*. Cambridge, MA: Belfer Center for Science and International Affairs.

Shires, James, and Max Smeets. 2021. 'The U.K. as a Responsible Cyber Power: Brilliant Branding or Empty Bluster?' *Lawfare Blog*, 23 November 2021. https://www.lawfareblog.com/uk-responsible-cyber-power-brilliant-branding-or-empty-bluster.

Smeets, Max. 2022. *No Shortcuts: Why States Struggle to Develop a Military Cyber-Force*. London: Hurst.

Targett, Ed. 2023. 'Opinion: GCHQs' Director Should not Be Playing Guest Editor on the BBC'. *Stack*, 3 January 2023. https://thestack.technology/gchq-director-sir-jeremy-fleming-bbc-today-programme/.

Valeriano, Brandon, Brian Jensen, and Ryan C. Maness. 2018. *Cyber Strategy: The Evolving Character of Power and Coercion*. Oxford: Oxford University Press.

Voo, Julia, Irfan Hemani, and Daniel Cassidy. 2022. *National Cyber Power Index 2022*. Cambridge, MA: Belfer Center for Science and International Affairs.

Weaver, Johanna. 2022. 'The Rules-Based International Order: Some Hard Truths'. Keynote presented at the 14th International Conference on Cyber Conflict: Keep Moving (CyCon 2022), Tallinn, Estonia. https://www.youtube.com/watch?v=O8eFiJaNzRU&ab_channel=natoccdcoe.

Willett, Marcus. 2019. 'Assessing Cyber Power'. *Survival* 61(1): 85–90.

2023 15th International Conference on Cyber Conflict
Meeting Reality
T. Jančárková, D. Giovannelli, K. Podiņš, I. Winther
2023 © NATO CCDCOE Publications, Tallinn

Sharpening the Spear: China's Information Warfare Lessons from Ukraine

Nate Beach-Westmoreland
Head of Strategic Cyber Threat Intelligence
Booz Allen Hamilton[1]
McLean, VA, USA

Abstract: This paper examines the lessons about information warfare (IW) that the People's Republic of China (PRC) is likely to be drawing from the war in Ukraine. To do so, it first analyzes how the People's Liberation Army (PLA) has developed its conception of states contesting the information environment (IE), formed by studying wars and protest movements since the Gulf War. The paper describes the PLA's evolving assessment of the growing importance, scope, and features of this contest. Because PRC strategic analysts typically frame the war in Ukraine as a proxy conflict between the United States (U.S.) and Russia, the paper then briefly compares all three states' doctrinal beliefs about IW. Second, the paper analyzes PRC theorists' assessments of the information conflict dimension of the Russia–Ukraine war. Principally, these insights concern narrative setting around conflicts, the initial war's long-term impact on the IE, and the role of cyberattacks in IW. Finally, the paper offers recommendations to a strategic-level NATO audience concerning IE engagement with the PRC from defensive and offensive perspectives. This paper's main sources are journal and newspaper articles by leading PLA-affiliated IW theorists written for an internal national security audience.

Keywords: *information warfare, People's Republic of China, Ukraine, People's Liberation Army*

[1] The opinions herein are the author's alone and do not represent the official positions of Booz Allen Hamilton or its officers, directors, or shareholders.

1. INTRODUCTION

"The bloody lessons left to us by past wars should be learned emphatically… learning from war—this is our main method."[2] Today, the People's Liberation Army (PLA) still points to this 1936 quote by Mao Zedong as a dictum for its military theory development process.[3]

Lacking any meaningful combat experience since the 1979 Sino–Vietnamese War, the PLA has relied on other countries' wars to develop its theories of modern interstate conflict. A major focus of its theoretical analysis has been the significance of information in conflict. Since the Gulf War, the PLA's concept of "information warfare" (IW) has expanded from a narrow focus on information technology in the physical battlespace (e.g., smart weapons, command and control) to controlling narratives in order to influence the perceptions and decisions of leaders, societies, and the international community.

The war in Ukraine is the latest case study for the PLA to test its IW ideas and refine its strategy. Its analysts have found that expansively contesting the information environment (IE) has become a critical aspect of modern international relations and that their growing concerns about the ability of the People's Republic of China (PRC) to compete in this domain are justified. Furthermore, they have identified useful IW tactics and strategies that the PRC will likely employ in future conflicts.

In 2022, NATO's *Strategic Concept* declared for the first time that the PRC is a "strategic challenge," using "coercive" policies and "malicious" cyberattacks to challenge the Alliance's "interests, security, and values."[4] This declaration is fundamentally a characterization of PRC IW, including its use of cyberattacks to control the IE. Given this context, NATO and its members should understand the lessons that the PRC is drawing from the war in Ukraine as they relate to the theories driving and shaping its coercive policies and cyberattacks. Based on this analysis, this paper suggests ways for NATO and its members to engage with the PRC defensively and offensively in the IE throughout the competition continuum.

[2] Mao Zedong, *Strategic Issues in China's Revolutionary War* (1936), ch. 1, sec. 4, https://www.marxists.org/chinese/maozedong/marxist.org-chinese-mao-193612.htm.

[3] Xiaosong Shou et al. *Science of Strategy* (translated by Project Everest, 2013), 30–32, https://www.airuniversity.af.edu/Portals/10/CASI/documents/Translations/2021-02-08%20Chinese%20Military%20Thoughts-%20In%20their%20own%20words%20Science%20of%20Military%20Strategy%202013.pdf.

[4] NATO, *NATO 2022 Strategic Concept*, Madrid, June 29, 2022, https://www.nato.int/nato_static_fl2014/assets/pdf/2022/6/pdf/290622-strategic-concept.pdf.

2. DEFINITIONS

Recent PLA strategic discussions conceive of warfare as occurring concurrently in three interconnected "spaces" (空间) or "domains" (域): the physical (物理) space where tangible aspects of warfare (e.g., forces, materiel) exist; the information (信息) space where information generation, transmission, and sharing occurs; and the cognitive (认知) space, which is the psychological realm, comprising "knowledge, beliefs, and capabilities."[5] In NATO terminology, "information space" is analogous to IE. Information, in the PLA's model, serves as a "medium between physical and cognitive spaces."[6] "Information warfare" (信息战) therefore is the contest over information to control the physical and cognitive "spaces," serving to "cover ears, blind eyes, and confuse the mind."[7]

Two interrelated aspects of the "information space" are "network space" (网络空间) and "cyberspace" (赛博空间). In strict PRC definitions, "network space" refers to the technology-centric networked systems environment and "cyberspace" refers to the human-centric online environment created by global interconnected networked information and communications technology used to "create, share, store, modify, exchange, and utilize information."[8] Authoritative sources have long bemoaned that PRC and PLA elites often use the two terms interchangeably, which can create confusion when analyzing PRC IW writings.[9] This imprecision is reflected in the expansive term "cyberspace operations" (赛博空间作战), which covers any actions using or targeting "cyberspace" in order to contest the IE, including network attack and defense, electromagnetic attack and defense, public opinion influence, and coercion.[10]

[5] Yi Li, "Cognitive Confrontation: A New Frontier for Future Warfare," January 28, 2020, *PLA Daily*, http://www.81.cn/jfjbmap/content/2020-01/28/content_252969.htm; Xu Yanhou and Hou Qinggang, "What Kind of Combat Concept Should be in the Information War?" *PLA Daily*, September 24, 2020, http://www.81.cn/jfjbmap/content/2019-09/24/content_244013.htm.

[6] Ibid.

[7] Baojun Wang, "Information Warfare: The First Game of Modern Warfare," *People's Daily Online*, December 18, 2012, http://theory.people.com.cn/n/2012/1218/c40531-19932490.html; Yuan Tian and Huang Ming, "Analysis of 'Media War' Reports in the Media Convergence Era," *PLA Daily*, November 26, 2019, http://www.81.cn/jsjz/2019-11/26/content_9683612.htm.

[8] Mingxi Wu, "Virtual Space Technology: Tao is Invisible but Tangible," *PLA Daily*, May 15, 2020, http://www.81.cn/jfjbmap/content/2020-05/15/content_261490.htm.

[9] "Shou Bu Talks about the Technical Basis and Logical Starting Point of Our Cyberspace Security Legislation," Cyber Security Association of China, n.d., http://www.cybersac.cn/News/getNewsDetail/id/1731/type/53.

[10] Teng Wu, Ding Xinxin, Zhang Zixing, and Wang Wenbo, "Demystifying Cyberspace Operations," *PLA Daily*, January 14, 2021, http://www.81.cn/big5/bq/2021-01/14/content_9894076.htm; Wu Mingxi. 2020. "Virtual Space Technology: Tao is Invisible but Tangible." *PLA Daily*. May 15, 2020. http://www.81.cn/jfjbmap/content/2020-05/15/content_261490.htm.

3. THE PRC'S EVOLVING VIEWS ON INFORMATION IN INTERSTATE CONFLICT

In mainstream PRC strategic thought, the core of modern interstate conflict is contesting information control.[11] Over the past 30 years, the PRC has often derived lessons from others' conflicts to develop this conceptualization. This section assesses several key developments in PRC strategic thinking about the purpose and function of information operations in modern interstate conflict. Then the section compares these concepts to similar ones shaping the U.S. and Russia, as PRC strategic discussions conceive of the war in Ukraine as a U.S.–Russia proxy war.

Early Developments

The U.S.-led coalition's swift, overwhelming victory over the Iraqi army in February 1991 sent shockwaves through the PLA, calling into question its capabilities and doctrine. Instead of cutting-edge technology, the PLA had, until then, emphasized the value of mass mobilization and protracted interior defenses—an updated version of the Maoist "People's War" strategy. Several wars involving successful resistance by countries with inferior technology, such as the U.S.–Vietnam War and the Soviet–Afghan War, had seemingly validated this strategy. Before the Gulf War, Iraq's president, Saddam Hussein, had espoused a similar strategy, leading many in China to predict that a quagmire awaited the coalition.[12] Yet after less than 100 hours of direct engagement, the coalition overwhelmed the world's fourth-largest army while suffering almost no casualties. Beijing was stunned.[13]

In June 1991, Central Military Commission (CMC) chairman Jiang Zemin told an early meeting on the lessons of the war that technological superiority—from smart weapons to jammers—had emerged as a key factor in modern conflict.[14] China's outdated military urgently needed to close its technological gap. He declared that China needed to develop asymmetric capabilities for defeating more powerful, technologically advanced enemies (understood to mean the U.S.), evoking what is commonly translated as the "assassin's mace" (杀手锏) of Chinese folklore. In November 1992, Jiang observed at another military meeting that the "rapidly developing international situation" demanded that China "correctly determine [its] military strategy."[15]

[11] Zhifeng Yu, "New Means of Cyber Warfare are Subverting the Rules of the Game in Warfare," *PLA Daily*, December 29, 2017, http://www.81.cn/2017xsdqjzxk/2017-12/19/content_7872999.htm.

[12] Harlan W. Jencks, "Chinese Evaluations of 'Desert Storm': Implications for PRC Security," *The Journal of East Asian Affairs* 6, no. 2 (1992): 453–57, http://www.jstor.org/stable/23253951.

[13] David Shambaugh, *Modernizing China's Military: Progress, Problems, and Prospects* (Berkeley: University of California Press, 2002), 69; Sheryl WuDunn, "After the War: War Astonishes Chinese And Stuns Their Military," *New York Times*, March 20, 1991, https://www.nytimes.com/1991/03/20/world/after-the-war-war-astonishes-chinese-and-stuns-their-military.html.

[14] Jiang Zemin, "On Military Strategic Guidelines and National Defense Science and Technology Issues," *Selected Works of Jiang Zemin Volume 1* (1991), http://reformdata.org/1991/0608/5595.shtml.

[15] David M. Finkelstein, "China's National Military Strategy: An Overview of the 'Military Strategic Guidelines,'" *Asia Policy*, no. 4 (2007): 67–72, http://www.jstor.org/stable/24904602.

Subsequently, leading military strategists assembled at a seminar to discuss how to realign PLA strategy with the dynamics of modern warfare. CMC vice chairman General Zhang Zhen outlined the state of modern warfare, which he called "warfare under high-technology conditions."[16] Control over information—"information warfare"—was fundamental to all aspects of modern warfare: intelligence collection, analysis, and dissemination; smart weapons; and command and control. The People's War strategy, he declared, "urgently needed to be innovated."[17] Jiang reportedly approved of these perspectives, which were adopted the following year in the next iteration of the PLA's official strategy document, *Military Strategic Guidelines for the New Era*.

Early PLA analysis of IW tended to focus on the physical battlespace. According to Wang Baocun and Li Fei of the Academy of Military Science in 1995, IW was the use of information technology in military conflicts: the incorporation of digital information technology in command-and-control or weapons systems (e.g., precision missiles) and the attacks against these information systems (e.g., signal jammers and malware).[18] These analysts also observed that technological superiority offered an opportunity for countries conventionally and technologically outclassed in asymmetric conflict. Information systems would become vital to conflict, but this meant that their disruption could decisively weaken an opponent's warfighting ability.[19]

PLA analysts soon, however, posited that IW would increasingly occur on a broader, society-wide scale. In a representative 1996 *PLA Daily* article, strategist Wei Jincheng observed that to function, society increasingly depended on computerized systems, making entire countries vulnerable to "a paralyzing blow through the internet."[20] Wei argued that the growing potential impact of cyberattacks meant that future conflicts might be bloodless, employing disruptive cyberattacks and "deterring and blackmailing the enemy with dominance in the possession of information."[21]

To the PLA, Yugoslavia's IW with NATO during the 1999 Kosovo War showed how a conventionally outclassed military could compete in the IE. Commentators observed that NATO had IW advantages in data collection and smart weapons but argued that Yugoslavia had mounted effective information offensives. For example, the country established websites to publish its own version of events and disrupted NATO websites. According to one PLA officer writing at the time, these actions "denied NATO complete success, and enabled [Yugoslavia] to preserve its strength and to

[16] Hui Ling, "Admiral Zhang Zhen in the Position of Vice Chairman of the Military Commission," *XiangChao*, no. 1 (2005), https://www.thepaper.cn/newsDetail_forward_1371740.

[17] Ibid.

[18] Baocun Wang and Li Fei, "Information Warfare," *PLA Daily*, June 1995, translated by the Federation of American Scientists Intelligence Resource Program, https://irp.fas.org/world/china/docs/iw_wang.htm.

[19] Ibid.

[20] Jincheng Wei, "Information War: A New Form of People's War," *PLA Daily*, June 25, 1996, https://irp.fas.org/world/china/docs/iw_wei.htm.

[21] Ibid.

maintain some degree of effective command and control."[22] The war had reinforced Beijing's belief that IW would be critical to compete with more powerful countries, like the U.S.[23]

Initial PLA analysis of U.S. military operations in Iraq and Afghanistan prompted the 2003 official endorsement of an emerging three-part conceptualization of IW, called the "three warfares" (三战).[24] The concept contends that states increasingly advance their political interests through control of the IE in three interconnected ways: by influencing decision-makers ("psychological warfare"), shaping popular opinion ("media warfare"), and legitimizing their actions ("legal warfare"), states can achieve their political goals. Ultimately, the strategy aims to erode an opponent's leadership's will or ability to resist and to gather domestic and global support for one's political position. American analyst Dean Cheng observed that while the three-warfares concept was not "established due to the second Gulf War… it would seem that additional impetus was imparted to their development by the recently concluded conflict."[25]

A decade later, PRC analysts held more mixed views on the performance of the U.S.'s IW during the war in Iraq, again seen through the lens of the three warfares. They still admired the U.S.'s initial "shock and awe" strategy as highly effective psychological warfare; according to American scholar Stephan Halper, PRC analysts assessed that such overwhelming displays of force had "precondition[ed] the battlefield and influenc[ed] tactical and operational outcomes."[26] These analysts also credited the dominance of U.S. media warfare. The U.S. had commandeered Iraqi mass communications infrastructure (as opposed to destroying it, like during the Kosovo War), controlling the local IE. The U.S. had also embedded domestic and foreign journalists with its forces, shaping domestic and global perceptions. On the other hand, PRC analysts critiqued U.S. legal warfare. They argued that the U.S.'s perceived flouting of international laws and norms had damaged its soft power and economy.[27] Therefore, the war showed that future planners needed to consider the narrative framing of warfighting in the context of its long-term cognitive impacts.

[22] James Perry, "Operation Allied Force: The View from Beijing," *Air and Space Power Journal* (2000), https://www.airuniversity.af.edu/Portals/10/ASPJ/journals/Chronicles/Perry.pdf.

[23] Zhang Wannian, "Biography of Zhang Wannian," quoted in *People's Liberation Army Modernization: Mid-1990s to 2025*, Michael Chase et al. (RAND Corporation, 2015), 14–15, https://www.jstor.org/stable/10.7249/j.ctt13x1fwr.8.

[24] Stefan Halper, *China: The Three Warfares* (Office of Net Assessment, U.S. Department of Defense, 2013), 31, https://cryptome.org/2014/06/prc-three-wars.pdf.

[25] Dean Cheng, "Chinese Lessons from the Gulf War," in *Chinese Lessons from Other Peoples' Wars*. ed. Andrew Scobell, David Lai, and Roy Kamphausen (Carlisle, PA: U.S. Army War College, 2011), 170–71, https://www.jstor.org/stable/pdf/resrep11966.8.pdf.

[26] Halper, *China*, 348.

[27] Halper, *China*, 347.

Recent Developments

In the past decade-and-a-half, a major new area of analytical focus has been IW's ability to alter countries' political environments. In this period, PLA writers increasingly described IW as occurring outside the bounds of hot war, a shift from earlier analysis. For example, they alleged that Western countries had used IW to instigate or enable many revolutions in the decades of the 2000s and 2010s, such as Ukraine's Revolution for Dignity and Hong Kong's Umbrella Revolution.[28] Analysts also found that Russian IW had influenced the 2016 U.S. presidential election.[29]Within this context, PLA analysts saw several lessons.

Social media platforms now play a central role in contesting the IE. As two IW scholars observed in the *PLA Daily*, internet users now comprise "the largest, most active, and most easily agitated groups in modern society," whose tendencies and behaviors "directly affect social stability and national security."[30] States can exploit this dynamic to create outsized political outcomes. For example, these scholars alleged that Russian hack-and-leak operations and online disinformation had created "national turmoil" that influenced the 2016 U.S. presidential election.[31]

Local control of dominant online media platforms may be critical for contesting the IE. In 2014, PLA professor Dai Xu claimed that the social movements in Hong Kong and Ukraine that year had demonstrated that local information resistance was futile against an adversary—the U.S. in this case—that controlled the leading traditional and social media platforms used in a targeted country. This finding may have further strengthened the PRC's resolve to block domestic access to foreign social networks. Seemingly acknowledging China's historically insular internet culture, Dai fretted that the U.S. had "deceived some countries" into investing in low-tech industries while it "established a 'technical mountain,'" controlling the leading global search engine, web portal, video, messaging, and social networking websites.[32] However, it is unclear whether PLA assessments that Russia effectively manipulated the U.S.'s political environment in 2016 led analysts to the logical conclusion that local control of social networks is not a surefire defense against IW.[33]

Cyberattacks could create information effects that prompt regime change. For example, two analysts writing in the *PLA Daily* claimed that the U.S. and unspecified European

[28] Xu Dai,"How Did the United States Instigate a 'Color Revolution' around the World?" *PLA Daily*, October 29, 2014, http://www.81.cn/mjjt/2014-10/29/content_6626768_2.htm; Chengjun Yang, "Cyber Struggles in Ukraine's Upheaval," *PLA Daily*, March 14, 2014, http://www.81.cn/jwgd/2014-03/14/content_5811404.htm.

[29] Ke Zhang and Yu Zhifeng, "Gain Insights into New Changes in Strategic Cyberwarfare," *PLA Daily*, January 3, 2019, http://www.81.cn/jfjbmap/content/2019-01/03/content_224461.htm.

[30] Ibid.

[31] Ibid.

[32] Xu Dai,"How Did the United States Instigate a 'Color Revolution' around the World?" *PLA Daily*, October 29, 2014, http://www.81.cn/mjjt/2014-10/29/content_6626768_2.htm.

[33] Ke Zhang and Yu Zhifeng, "Gain Insights into New Changes in Strategic Cyberwarfare," *PLA Daily*, January 3, 2019, http://www.81.cn/jfjbmap/content/2019-01/03/content_224461.htm.

countries had eroded popular support for Ukraine’s Yanukovych administration with disinformation and hack-and-leak operations. These countries had also allegedly hindered the administration’s counter-messaging by disrupting its websites, both directly and by supporting opposition cyber groups.[34] As a matter of national security, analysts argued that China needed to be able to control its IE—and implicitly, that of others—with robust defensive and offensive cyber capabilities.[35] This position has evidently held, as can be seen in the 2015 unveiling of China’s unified IW capabilities as part of the PLA Strategic Support Force.

Going further, instigating regime change via cyber operations appeared superior to direct military action. Using Afghanistan and Iraq as examples, analysts argued that invasions—no matter how overwhelmingly victorious they might seem at first—often fail to establish stable or friendly governments. According to PLA academic Lin Dong, Confucian “just war” theory explained this unexpected outcome: a lasting peace requires minimizing lethal and destructive violence.[36] He wrote that, by seeking to dominate its enemies with overwhelming force (e.g., shock-and-awe strategy), the U.S. had exposed itself to criticism during postwar reconstruction, a finding also found in three-warfares assessments of U.S. strategy in the Iraq War.[37] Alternatively, Lin argued, a country should degrade its opponent’s economic, political, and social ability to resist by employing economic, cognitive, and cyberwarfare and minimize hard force.

As evidence of IW’s superiority over conventional warfare at achieving regime change, PLA writers pointed to Russia’s swift, low-casualty annexation of Crimea in 2014. According to two PLA professors, Russia had deftly controlled the IE; they cited examples that would be classified in PLA theory as psychological, media, and legal warfare.[38] Ukraine allegedly lost control of its military because Russia had used sleeper agents and civilian hackers to paralyze Ukrainian command-and-control networks. Russia had directed authoritative experts and state media to push approved narratives, which allegedly garnered support inside Russia and Crimea for the annexation. According to the professors, this propaganda led to high voter turnout and approval for the annexation referendum, thereby legitimizing Russia’s actions worldwide.

As in early IW analysis, PLA authors in the 2010s and onward also continued to consider the utility and limitations of IW in asymmetric conflict. Although the military

[34] Chengjun Yang and Jiang Zheng, “Ukraine Was First Dismantled Online,” *Global Times*, March 21, 2014, http://news.sina.com.cn/pl/2014-03-21/072629758835.shtml.

[35] Ibid.

[36] Dong Lin, “The Violence of War from Destroying the Enemy to Dominating the Enemy,” *Guangming Daily*, November 20, 2022, https://epaper.gmw.cn/gmrb/html/2022-11/20/nw.D110000gmrb_20221120_1-07.htm.

[37] Halper, *China*, 347.

[38] Yuanpu Xia and Yuan Zongyi, “An Analysis of Russia’s Mobilization against Crimea,” *PLA Daily*, September 24, 2021, http://www.81.cn/gfbmap/content/2021-09/24/content_299658.htm.

rapidly matured during this period—modernizing its hardware, reorganizing its force structure, developing a robust local defense industry, and building overseas bases—the PRC leadership publicly assesses that the "PLA still lags far behind the world's leading militaries."[39] Likely for this reason, understanding how to conduct effective asymmetric warfare apparently remained a priority.

Several conflicts seemed to show the huge potential for cyber operations to control the IE. PRC analysts vaunted the intelligence value of U.S. persistent interceptions on Iraqi telecommunications infrastructure in the first decade of the 2000s, the "paralysis" of Georgia by Russian distributed denial-of-service (DDoS) attacks during the 2008 invasion, and the alleged substantial harm to Ukrainian government credibility caused by multiple cyberattacks on electricity distributors.[40] The digital revolution had made societies and economies reliant on networked systems, leaving them more vulnerable to cyberattacks, as earlier assessments had predicted.[41] Yet these examples involved countries conducting cyberattacks against arguably technologically and conventionally outclassed competitors, which may have limited their applicability.

Thus, some analysts doubted whether outclassed countries could truly gain an advantage with IW. As analyst Wei Song wrote in the *PLA Daily*, "the weak can often only gain temporary advantages and small tactical victories through cyberwarfare, while the strong often hold the strategic initiative."[42] Existing cyber powers like the U.S. can maintain the strategic initiative, he argued, because they possess systemic advantages that weak states cannot overcome with cyberattacks. For decades, the dynamics of globally leading technology sectors (e.g., high barriers to entry, success feeding success) created self-perpetuating power accumulation, leaving other states unable to catch up, echoing Dai's concept of the U.S.'s insurmountable "technical mountain." Wei argued that therefore, only the most advanced countries have the foundation to leverage next-generation information technology—AI, big data, and quantum computing—for cyberattack or defense. Even if weak states could launch tactically successful cyberattacks, they would be unable to resist conventional forms of state power like sanctions and kinetic strikes, such as Israel's declared retaliatory airstrike on Palestinian hackers in 2019.

[39] *China's National Defense in the New Era*. The State Council Information Office of the People's Republic of China, July 2019, translated by China Aerospace Studies Institute, https://www.airuniversity.af.edu/Portals/10/CASI/documents/Translations/2019-07%20PRC%20White%20Paper%20on%20National%20Defense%20in%20the%20New%20Era.pdf?ver=akpbGkO5ogbDPPbflQkb5A%3D%3D; Shou et al., *Science of Strategy*, 30–32.

[40] Zhang and Zhifeng, "Gain Insights into New Changes in Strategic Cyberwarfare."

[41] Jincheng Wei, "Information War: A New Form of People's War," *PLA Daily*, June 25, 1996, https://irp.fas.org/world/china/docs/iw_wei.htm.

[42] Song Wei, "A Clear Understanding of the Asymmetry of Cyberwarfare," *PLA Daily*, March 23, 2021, http://www.81.cn/xue-xi/2021-03/23/content_10009053.htm.

Comparison to Concepts from Russia and the U.S.

PRC scholars and analysts often cast the 2022 invasion of Ukraine as ultimately being a proxy war between Russia and the U.S., with NATO serving as an "instrument of American expansionism."[43] PLA analysts typically contend that the U.S. orchestrated NATO's eastward expansion, infringing on Russia's core security interests and compelling its defensive response.[44] Analysts' fundamental explanations for this dynamic vary (e.g., Marxist emphasis on economic enrichment, realist emphasis on power-seeking).[45] Consequently, this proxy war framing fundamentally shapes PRC analysis of the war in Ukraine.

Differences in terminology and concepts confound many comparisons of the PRC, U.S., and Russian militaries' discussions of network, cyber, and information operations. The U.S., the PRC, and Russia use several terms around this topic with similar direct translations whose definitions and concepts can be vastly different or even lack equivalent terms in different languages.[46] Even within countries, terminology may be inconsistent, partly due to an absence of official formal definitions.[47] That said, as can be seen in several areas, there is increasing conceptual alignment among the three militaries about the nature and role of IW in international relations.

The PRC, Russia, and the U.S. treat the IE as an important, contested aspect of modern interstate conflict. The Russian military's official encyclopedia notes that "information confrontation" (*информационное противоборство*) has always been part of international relations, but the development of information technology has dramatically increased the conflict's "scale, content, and forms."[48] The 2014

[43] Grzegorz Stec and Francesa Ghiretti, "How China Views the EU amid the Russia-Ukraine War," Mercator Institute for China Studies, August 4, 2022, https://merics.org/en/merics-briefs/how-china-views-eu-amid-russia-ukraine-war-global-gateway-departing-eu-ambassador; Iliya Kusha, "China's Strategic Calculations in the Russia-Ukraine War," Wilson Center, June 21, 2022, wilsoncenter.org/blog-post/chinas-strategic-calculations-russia-ukraine-war.

[44] Shen Jun, "Exporting Turmoil with Harmful Color Diplomacy," *PLA Daily*, April 23, 2022, http://www.81.cn/jfjbmap/content/2022-04/23/content_314227.htm.

[45] Jiansong Yang and Xu Shiwei, "Why the United States Is Keen to Arm Ukraine," *PLA Daily*, December 2, 2021, http://www.81.cn/bq/2021-12/02/content_10112048.htm; Xiangying Li et al., "US-Russia Wrestling in a Hybrid War through the Lens of the Russia-Ukraine Conflict," PLA Daily, April 23, 2023, http://www.81.cn/jfjbmap/content/2022-04/23/content_314227.htm.

[46] Keir Giles and William Hagestad, "Divided by a Common Language: Cyber Definitions in Chinese, Russian and English," in *2013 5th International Conference on Cyber Conflict*, eds. K. Podins, J. Stinissen, M. Maybaum (Tallinn: NATO CCD COE Publications, 2013), 413–29, https://ccdcoe.org/uploads/2018/10/CyCon_2013_Proceedings.pdf.

[47] Catherine Theohary and John Rollins, "Cyberwarfare and Cyberterrorism: In Brief," Congressional Research Service, March 27, 2015, https://sgp.fas.org/crs/natsec/R43955.pdf; Catherine Theohary, "Information Warfare: Issues for Congress," Congressional Research Service, March 7, 2018, https://sgp.fas.org/crs/natsec/R45142.pdf; "Defense Primer: Information Operations," Congressional Research Service, December 9, 2022, https://crsreports.congress.gov/product/pdf/IF/IF10771/9; Herb Lin, "Doctrinal Confusion and Cultural Dysfunction in the Pentagon Over Information and Cyber Operations," Lawfare, March 27, 2020, https://www.lawfareblog.com/doctrinal-confusion-and-cultural-dysfunction-pentagon-over-information-and-cyber-operations.

[48] "Information Confrontation," Ministry of Defense, n.d., accessed January 6, 2023, https://encyclopedia.mil.ru/encyclopedia/dictionary/details.htm?id=5221@morfDictionary.

Military Doctrine of the Russian Federation contends that national security threats have shifted to the IE, making large kinetic wars less likely.[49] Russia understands that competition in this domain—"information warfare" (*информационная война*)—comprises cyber operations, psychological operations, information operations, and electronic warfare.[50] The U.S. military similarly contends that the IE is a long-contested "operational environment" and the digital revolution has created "new and complex challenges" therein.[51] According to the 2016 *Strategy for Operations in the Information Environment* (SOIE), "throughout the history of warfare, militaries have sought advantage through actions intended to affect the perception and behavior of adversaries."[52]

Unlike Russia and the PRC, the U.S. military has been until recently fixated on conventional kinetic warfare. The 2018 *Joint Concept for Operations in the Information Environment* acknowledged the need for an organizational mindset shift from treating the IE as an "afterthought" to a "foundational concept of all military activities."[53] As scholar Cathy Downes also observed in 2018, "[the U.S.] military['s] understandings of cyberspace, cyber power, and strategy options have been preoccupied with tactical and technical responses to [its own] computer networks and systems."[54]

The PRC, Russia, and the U.S. now all argue that the IE may be contested during or outside armed hostilities. Russia has for many years portrayed itself as existing in a constant state of information conflict, besieged by information threats externally and within.[55] The U.S., however, has until recently tended to speak about information operations as discrete activities related to time-demarcated conflicts.[56] The 2019 Competition Continuum doctrine shows a shift from this thinking. In it, the U.S. acknowledged that a great deal of competition—such as IW—occurs below armed conflict and falls outside the peace-and-war model.[57]

Russian strategists, like many in China, have long valued cyber operations' utility in state-level asymmetric warfare. Russian strategists often argue that such operations may help a technologically weaker state neutralize a technologically and economically

[49] Presidential Administration of Russia, *Military Doctrine of the Russian Federation* (2013), http://static.kremlin.ru/media/events/files/41d527556bec8deb3530.pdf.

[50] Michael Connell and Sarah Vogler, "Russia's Approach to Cyber Warfare," *CNA*, March 2017, https://www.cna.org/archive/CNA_Files/pdf/dop-2016-u-014231-1rev.pdf.

[51] U.S. Department of Defense, *Strategy for Operations in the Information Environment* (2016), https://dod.defense.gov/Portals/1/Documents/pubs/DoD-Strategy-for-Operations-in-the-IE-Signed-20160613.pdf.

[52] Ibid.

[53] U.S. Joint Chiefs of Staff, *Joint Concept for Operating in the Information Environment (JCOIE)*, July 25, 2018, https://www.jcs.mil/Portals/36/Documents/Doctrine/concepts/joint_concepts_jcoie.pdf. Viii.

[54] Cathy Downes, "Strategic Blind–Spots on Cyber Threats, Vectors and Campaigns," *Cyber Defense Review* 3, no. 1 (2018): 84, http://www.jstor.org/stable/26427378.

[55] Connell and Vogler, "Russia's Approach to Cyber Warfare".

[56] U.S. Department of Defense, *Strategy for Operations*.

[57] U.S. Department of Defense, *JDN 1-19 Competition Continuum*, June 3, 2019, https://www.jcs.mil/Portals/36/Documents/Doctrine/jdn_jg/jdn1_19.pdf.

stronger opponent.[58] Therefore, for Russia, like the PRC, strategies useful in asymmetric warfare are extremely attractive—and perhaps prone to hype—given the country's professed strategic technological inferiority to the United States.

Meanwhile, the U.S. understands that this dynamic shapes Russia and China's IW but acknowledges that it has not sufficiently assessed the use of such asymmetric strategies by states. In 2020, the U.S. Department of Defense noted that the U.S. has an "enduring strategic advantage," prompting its adversaries to employ the "indirect and asymmetric" strategies of "irregular warfare" to "erode" its "power, influence, and will."[59] However, the Department of Defense admitted that it needs to develop a "revised understanding of [irregular warfare]" as a part of interstate conflict, having historically focused on irregular warfare by and against substate actors, like terrorists.[60]

4. THE PRC'S ASSESSMENT OF THE 2022 INVASION OF UKRAINE AS IW

PRC strategists have carefully examined the 2022 escalation of the war in Ukraine to find lessons about IW. In May 2022, Fan Yongpeng, deputy director of the China Research Institute at Fudan University, said on a nationally aired political talk show that "for China, the Russia–Ukraine conflict is an important case study, and we must learn from it to be invincible in the information war that is very likely to occur in the future."[61] Public analysis of the war's IE by PRC elites—especially from the PLA—has been limited, but several of their major initial assessments have surfaced.

The U.S. and Ukraine have a superior narrative framing and agenda-setting ability, building or solidifying anti-Russia sentiment within Ukraine and internationally. Analysts concluded that, by using inflammatory rhetoric prior to February 2022 to suggest the war's near-inevitability, the U.S. had stood to benefit from any outcome: appearing prescient (predicting the war) or influential (deterring Russia).[62] During the war, the U.S. and Ukraine created a powerful, sweeping emotional narrative of justice with evocative stories and iconography, such as the Ghost of Kyiv, the Snake Island martyrs, and Volodymyr Zelensky's fatigues.[63] One analyst found that the West's

[58] Bilyana Lilly and Joe Cheravitch, "The Past, Present, and Future of Russia's Cyber Strategy and Forces," in *2020 12th International Conference on Cyber Conflict 20/20 Vision: The Next Decade*, eds. T. Jančárková, L. Lindström, M. Signoretti, I. Tolga, G. Visky (Tallinn: NATO CCD COE Publications, 2020), 129–55, https://ccdcoe.org/uploads/2020/05/CyCon_2020_book.pdf.

[59] U.S. Department of Defense, *Summary of the Irregular Warfare Annex to the National Defense Strategy* (2020), 2–4, https://media.defense.gov/2020/Oct/02/2002510472/-1/-1/0/Irregular-Warfare-Annex-to-the-National-Defense-Strategy-Summary.PDF.

[60] Ibid.

[61] "Cyber-Information Warfare in the Russia-Ukraine Conflict," This is China, 2022, episode 141, http://cifu.fudan.edu.cn/c1/af/c412a442799/page.htm.

[62] "The United States Resorted to Six Public Opinion War Tactics in the Ukrainian Crisis, at Least These Enlightenments for China," *China Daily*, March 29, 2022, http://cn.chinadaily.com.cn/a/202203/29/WS62425c94a3101c3ee7acdd28.html.

[63] *China Daily*, "The United States Resorted to Six Public Opinion War Tactics;" This is China, "Cyber-Information Warfare in the Russia-Ukraine Conflict."

alleged total fabrication of scandalous stories defaming Russia (e.g., Russian soldiers have massacred civilians) is a near-ironclad strategy; wholly falsified stories can rarely be disproven before the public interest and perceptions shift[64] (i.e., the "proving a negative" challenge). Based on these assessments, the PRC might use these IW tactics and strategies in future conflicts and develop specific countermeasures for its persistent confrontation with the United States and its allies.

The U.S. has effectively limited Russia's counter-messaging ability. In recent years, the U.S.—as well as many of its allies—has restricted state-controlled traditional media like the news network Russia Today (RT), thereby severely limiting the direct dissemination of Russian government narratives and rebuttals to foreign audiences. PRC analysts contend that, during the war, the U.S. has allegedly "organized" cyberattacks that disrupted Russian government and media websites, limiting Russia's ability to communicate directly online.[65] Nevertheless, Russia's greatest counter-messaging challenge, according to many analysts, has been its reliance on social media sites that are overwhelmingly U.S.-headquartered.[66] The key takeaway, then, for many PRC analysts is that China must develop alternative, globally popular messaging platforms, especially China-headquartered social media sites. As party-controlled, foreign-facing outlet *China Daily* described this imperative, "giving up the autonomy of public opinion platforms is tantamount to building a fortress on a sandy beach"—suggesting that China's national security and its attempts to conduct IW might be fundamentally doomed without locally controlled, globally relevant media platforms.[67]

Although public PRC analysis has generally found that the U.S. and Ukraine decisively won the initial "information war," some argue that the U.S. may face long-term negative informational effects.[68] For example, an article from PRC propaganda outlet Xinhua republished in the *PLA Daily* pointed to the ripple effects of the sanctions regime targeting Russia's energy sector, which the author described as U.S.-led (again, reflecting the PRC's framing of the Ukraine conflict as a U.S.–Russia proxy war). Originally designed to convey a message of unity, the sanctions led to soaring global energy prices that allegedly "exposed [the U.S.'s] selfish nature" to its European allies

[64] This is China, "Cyber-Information Warfare in the Russia-Ukraine Conflict."

[65] *China Daily*, "The United States Resorted to Six Public Opinion War Tactics."

[66] Jun Liu, "Analysis of the Impact of Social Media on the Conflict between Russia and Ukraine," *People's Forum*, August 3, 2022, http://www.rmlt.com.cn/2022/0803/653242.shtml; *China Daily*, "The United States Resorted to Six Public Opinion War Tactics."

[67] "The United States Resorted to Six Public Opinion War Tactics in the Ukrainian Crisis, at Least These Lessons for China," *China Net*, March 29, 2022, http://news.china.com.cn/2022-03/29/content_78135249.htm.

[68] Minghao Zhao, "Russia-Uzbekistan Conflict Intensifies Global 'Digital Competition,'" Center for International Security and Strategy at Tsinghua University, April 22, 2022, https://ciss.tsinghua.edu.cn/info/zlyaq/4783; Yuan Zhang, "Hot Insights: How Does the Ukrainian Crisis Accelerate the Evolution of the International Landscape?" *PLA Daily*, December 22, 2022, http://www.81.cn/ss/2022-12/22/content_10207234.htm; This is China, "Cyber-Information Warfare in the Russia-Ukraine Conflict."

and emphasized its "selfishness and hegemony" to developing countries.[69] Though this propaganda article is not of PLA origin, its republication in the internally oriented *PLA Daily* suggests a degree of military endorsement for its fundamental perspective. This assessment echoes the PLA's earlier IW critiques of the U.S. execution of the wars in Iraq and Afghanistan. Yet it goes further to argue that the PRC can use such perceived missteps to advance its positive narratives about a more "multipolar" world, neutralizing the influence of the U.S. and its allies and partners.

The role of cyberattacks in the war has rarely been publicly analyzed in the PRC by the military or elite foreign policy community members. Writing in March 2022, Peking University's Sun Yilin observed that there had been many in Ukraine, Russia, and elsewhere, but thus far they had had little observable impact on the war.[70] This outcome ran counter to Zhang and Yu's 2019 prediction that disrupting key industries and telecommunications infrastructure with cyberattacks would be primary elements of modern war. Sun offered several possible explanations for this outcome: victims may have not disclosed their impact; the most destructive attacks required too much time to prepare and so were unsuitable for battlefield needs; and kinetic strikes are much more destructive. Echoing Lin's discussion of "just war," Sun considered whether Russia might have attempted to minimize direct harm to civilians, a Russian Ministry of Defense messaging point at the start of the conflict. Perhaps, he hypothesized, Ukrainian defenses had been well prepared—it is now known to have received preemptive U.S. hunt-forward support[71]—or, as party outlet *China News* suggested, benefiting from foreign technology companies' support[72] (also true[73]). At a minimum, this insight will likely justify increased PRC efforts to boost its national cyber defenses. On a final note, as this paper shows, the PRC's elite strategic thinkers' relative silence on cyberattacks in the Russia–Ukraine war is uncharacteristic; perhaps they have concerns about the quality of evidence, or intentional obfuscation is at play.

[69] Yuan Zhang, "Hot Insights: How Does the Ukrainian Crisis Accelerate the Evolution of the International Landscape?" *PLA Daily*, December 22, 2022, http://www.81.cn/ss/2022-12/22/content_10207234.htm.

[70] Yilin Sun, "Cyberattacks in Russia-Ukraine Conflict: Has the Cyberwar Begun?" Peking University, March 14, 2022, https://www.igcu.pku.edu.cn/info/1242/3734.htm.

[71] "Before the Invasion: Hunt Forward Operations in Ukraine," Cyber National Mission Force Public Affairs (USCYBERCOM), November 28, 2022, https://www.cybercom.mil/Media/News/Article/3229136/before-the-invasion-hunt-forward-operations-in-ukraine.

[72] "From commercial satellites to social media, Western tech companies are deeply involved in the Russia-Ukraine conflict," *China News*, November 2, 2022, https://www.chinanews.com.cn/gj/2022/11-02/9885010.shtml.

[73] Ines Kagubare, "Russia-Ukraine War Has Improved US Cyber Cooperation, Says Key Official," *Hill*, February 2, 2023, https://thehill.com/policy/cybersecurity/3841444-russia-ukraine-war-has-improved-us-cyber-cooperation-says-key-official/.

5. CONCLUSION

These lessons from Ukraine will likely shape the PRC and PLA's IW in several ways. The PRC will likely seek to develop grand narratives around its conflicts by promoting simple, positive messages about itself, its values, and its vision while demonizing and degrading its opponents. It will likely engage in this perception management long before initiating kinetic hostilities. Building social media platforms with global reach will also likely remain a priority for Beijing. The PLA will probably increasingly shape its actions in all domains to manage their impact on perceptions and larger narratives. Furthermore, the PLA may be more inclined to use cyberattacks to coerce leaders and societies persistently, beyond narrow traditional concepts of wartime, rather than to meaningfully shape the physical battlefield.

6. RECOMMENDATIONS

These takeaways, combined with the growing competition between NATO and the PRC, demand that NATO members take steps to manage this challenge. Several recommendations thus follow for NATO members to enhance their interaction with Beijing in the IE.

Develop Common Terminology and Concepts

The Alliance and its members should develop common, consistent terminology and operational concepts to describe IW to include the operating environment, offensive and defensive activities, and their relationship with other domains and types of warfare. Such terminology will facilitate effective intelligence sharing around PRC IW threats, strategic development, and the unified organization of relevant capabilities.

Manage IW Risk Facing Alliance and Member Decision-Making Processes

The PRC's IW ultimately seeks to coerce or mislead key decision-makers to act in ways more conducive to PRC leadership interests. NATO's consensus decision-making requires the buy-in of all members on policy and operational questions, which creates a vulnerability in that targeted information operations need to compel only a single member to act in the PRC's interests.

To prepare for this threat, the Alliance should conduct risk assessments to identify the vulnerabilities facing key elements of NATO and member decision-making processes and recommend ways to mitigate unacceptable risk. Risk assessments should reflect the diversity of members' formal and informal decision-making processes and key internal and external sources of influence (e.g., mass media, social media, think tanks, business sector).

Focus intelligence collection on identifying efforts to manipulate decision-making and share findings between members to raise awareness of methods currently being used to target processes, people, and organizations, and update IW threat models accordingly.

War-Game Information Conflicts Involving the PRC

NATO already possesses war-gaming experience, including that related to information conflicts. Even so, the Alliance should conduct war games specifically reflecting PRC IW tactics, strategies, and intentions to increase NATO, national, and sector resilience to this growing challenge. These war games should include both direct IW confrontation with the PRC, as well as scenarios where the PRC attempts, in terms of the narrative, to capitalize on NATO operations and exercises where it is not directly involved. Likewise, war-gaming should also test whether NATO IW capabilities could deter or respond to PRC IW if warranted. Effective war-gaming of the PRC IW threat demands that NATO applies sufficient intelligence collection and analysis resources to understand current PRC tactics, strategy, and decision-making calculus, as well as projections of future developments.

Preemptively and Persistently Promote Pro-Alliance, Value-Based Narratives

NATO's strategic communications activities and capabilities should preemptively promote messages tailored to key audiences in member states designed to counter expected PRC narratives. PRC strategists often argue that successfully controlling the IE begins well before the outbreak of armed hostilities. The PLA's long-held interest in the "three warfares" concept suggests that PRC IW capabilities will variously target specific Alliance and members' policy elites, the public, and relevant international legal venues. In the Ukraine–Russia conflict, PLA theorists identified the most successful narratives as value-based portrayals of good and evil, just and unjust, and so forth. The PRC will likely therefore preemptively develop narratives that promote itself and its interests and negatively characterize NATO and its members and their policies. Therefore, NATO messages should emphasize the importance of the Alliance with inspiring stories showcasing its values in action. Messaging should be memorable, easily digestible, and tailored to organic promotion across traditional and social media.

Raise Awareness of General and Specific PRC Influence Efforts

NATO should raise awareness in its members' elite, public, and legal venues of general and specific PRC influence efforts. Alliance members should be especially aware of the PRC's intentions to increase its global messaging ability, especially through China-based social media platforms with a global reach. The global traditional and social media environment overwhelmingly favors the Alliance, as media companies with global reach are disproportionally headquartered with Alliance members. PRC strategists' analysis of the Russia–Ukraine war revealed that developing similar platforms is a national security imperative.

Invest in IW Research and Monitoring

NATO should invest in scholarly research and technological solutions to increase the quality and speed of Alliance awareness about IW threats. Understanding the PRC's IW terminology and perspective on threats and security in the IE will improve intelligence analysis, reduce the likelihood of unintended escalation, and enable meaningful dialogue on these matters. Research should examine IW efforts globally to identify lessons learned for increasing military, government, and societal resilience in the face of these efforts. Technological solutions should seek to rapidly identify PRC information operations, especially on social media networks, and aim to characterize key messages, target audiences, and impact.

2023 15th International Conference on Cyber Conflict
Meeting Reality
T. Jančárková, D. Giovannelli, K. Podiņš, I. Winther
2023 © NATO CCDCOE Publications, Tallinn

Cyber Diplomacy: NATO/EU Engaging with the Global South

Eduardo Izycki
PhD Candidate / Researcher
University of Brasília (UnB)
International Relations Institute (IRel)
Brasília, DF, Brazil
eduardo.izycki@aluno.unb.br

Brett van Niekerk
Senior Lecturer
Durban University of Technology
Department of Information Technology
Durban, South Africa
brettv@dut.ac.za

Trishana Ramluckan
Honorary Research Fellow
University of KwaZulu-Natal
School of Law
Durban, South Africa
ramluckant@ukzn.ac.za

Abstract: Since the end of the Cold War, there has been a movement towards a multipolar world as the geopolitical tectonic plates shift. The Russian invasion of Ukraine is likely to be treated by future historians as the turning point ushering in this new multipolar era. In this new context, (cyber) neutrality seems challenging for regions such as Latin America and Africa. These countries, which sit outside the geopolitical fault lines, naturally tend to strive for a balanced, neutral position. Both regions have strong economic ties with China, while maintaining cultural and historical connections with Europe and the US, despite the complex legacy of the colonial and Cold War eras. However, this equilibrium might lean towards the Chinese and Russian positions regarding cyber policy. It is particularly relevant to address this question given that the regions contain numerous swing states. We will present evidence that NATO and the EU are losing ground to China and Russia's views on cyberspace, based on three subjects of study: (i) Global South voting patterns in the UN; (ii) the absence of Global South countries in the roster of like-minded countries in the collaborative attribution of advanced persistent threats and recent Russian cyber campaigns against Ukraine; (iii) the use of offensive cyber capabilities by Global South countries to exert information control and surveillance (mostly enabled by Western companies). This paper argues that NATO and the EU must face reality and engage with the Global South – particularly Africa and Latin America – to maintain a competitive advantage in cyber policy. We suggest a more straightforward values-based approach that involves NATO and the EU engaging in capacity-building and information-sharing with the Global South.

Keywords: *cyber policy, Global South, cyber capabilities, cyber diplomacy*

1. INTRODUCTION

Geological eras are defined in extended periods, and the exact moment of transition is usually unclear. Conversely, for eras in human history, we tend to choose a date or event to mark the changes. Usually, such a choice raises intense debates among historians. Since the end of the Cold War, there has been a shift towards a multipolar world as the geopolitical tectonic plates move and the US's relative power declines, though the extent to which that is happening is the subject of ongoing debate (Nye 2010; Trubowitz and Harris 2019; Layne 2018). The Russian invasion of Ukraine is likely to be treated by future historians as the harbinger of this new multipolar era, which some say will be a post-American century (Acharya 2018; Cohen 2022) or even a Chinese century (Scott 2008). In this new context, (cyber) neutrality seems challenging for regions such as Latin America and Africa. As these countries are located outside the geopolitical fault lines, they tend to strive for a balanced, neutral position.

Both Latin America and Africa receive considerable foreign investment from China, and China is the destination for most of the raw-material exports from the two regions. In addition, China is the number-one commercial partner to both regions (Roy 2022; Regissahui 2019). Conversely, Latin America and Africa have deep cultural and historical ties with Europe and the US. Even so, this legacy is a fraught one, involving colonialism and, more recently, political turmoil and intelligence operations stretching from the Cold War to the recent Snowden revelations (Cohen et al. 2014; Canabarro and Borne 2015).

From the cyber policy debate perspective, we contend that NATO and the EU both assume that they can keep Latin America and Africa within their digital sphere of influence based on cultural and historical connections alone. However, we will argue that there is evidence to suggest otherwise and that both regions are leaning towards China and Russia's cyber policy perspectives and state practice.

In the next section, we will analyse African and Latin American voting patterns in cyber policy issues and how they relate to Chinese and Russian cyber policies. Section 3 discusses the fact that the Global South has not been a part of the *like-minded* countries in collaborative attribution over the last five years and has not endorsed the attributions made in the context of the Russian invasion of Ukraine. The third and last corpus of evidence relates to the use of private offensive cyber capabilities (OCCs) in the Global South. Section 4 presents a few cases where OCCs purchased from Western companies have been systematically used for political persecution. The fact that many of those capabilities are provided by Western companies erodes the argument for a 'clean network' or a safer Internet, especially when those capabilities are deployed for surveillance and information control against domestic targets.

In our concluding remarks, we will suggest how information-sharing can help NATO and the EU to pursue a values-based cyber policy. Furthermore, we suggest that promoting *responsible state behaviour* in the use of OCCs should be part of that policy. This would heighten the contrast with the offensive behaviour displayed by China and Russia over the last decade in cyberspace.

2. GLOBAL SOUTH VOTING PATTERNS

Collett (2021), Dietrich and Pawlak (2022), and Martin (2022) discuss voting at the UN on cybersecurity and Internet governance. Collett (2021) notes two major opposing views that emerged from the 2021 World Conference on International Telecommunications, and Martin (2022) indicates the challenges faced by the UK and US. Collett (2021) and Martin (2022) both indicate the existence of support for the views of Russia and China, although Dietrich and Pawlak (2022) indicate that support is decreasing in the context of some of the cybercrime votes (the proposed amendments, which are not considered below). Subsequently, another vote occurred in 2022 on a Programme of Action on state behaviour in cyberspace; this process was initiated by France and Egypt and co-sponsored by 60 nations (CyberPeace Institute 2022; Weber 2022). Table I lists the six proposals considered here.

TABLE I: VOTING CONSIDERED

Year	Vote number	Proposer	Purpose
2018	A/73/266	US	Group of Governmental Experts (GGE)
2018	A/73/27	Russia	Open Ended Working Group (OEWG)
2018	A/73/187	Russia	Cybercrime
2019	A/74/247	Russia	Open Ended Cybercrime Ad Hoc Committee
2020	A/75/240	Russia	Second OEWG
2022	A/C.1/77/L.73	France and Egypt	Programme of Action

The analysis presented here differs from the previous literature described above in that the 2022 Programme of Action vote is considered and that the focus is on NATO and also includes the BRICS grouping (Brazil, Russia, India, China, and South Africa). Table II presents the overall voting for the six proposals and the general NATO and Russia/China positions. All NATO members except Turkey voted identically across the proposals; Turkey abstained on the first OEWG and the two cybercrime proposals. Russia and China voted opposite to the NATO members on all proposals. The voting bloc aligned with Russia and China includes North Korea, Iran, Nicaragua, and Syria, all of whom voted identically across the six proposals. In addition, Cuba, Venezuela, and Zimbabwe only diverged from Russia regarding the 2022 Programme of Action; Cuba abstained, Venezuela did not vote, and Zimbabwe voted 'yes'. Belarus, which might have been expected to vote alongside Russia given its support for Moscow's invasion of Ukraine, varied by abstaining in both the GGE and Programme of Action proposals.

TABLE II: OVERALL VOTES AND THE NATO AND RUSSIA/CHINA POSITIONS

Year	Vote number	Yes	No	Abstain	No vote	NATO	Russia/China
2018	A/73/266	135	12	16	30	Yes	No
2018	A/73/27	119	46	14	14	No	Yes
2018	A/73/187	94	59	33	7	No	Yes
2019	A/74/247	79	60	33	21	No	Yes
2020	A/75/240	92	50	21	30	No	Yes
2022	A/C.1/77/L.73	157	6	14	16	Yes	No

As this paper is focused on Africa and Latin America, Figure 1 illustrates the voting of these regions in comparison to NATO, with the Russia/China bloc consisting of the 10 nations described above. In addition, the voting of the five BRICS countries is shown.

As is evident, the positions of African and Latin American nations tend to vary, with significant numbers not voting (Africa) or abstaining (Latin America) in some of the votes. The BRICS countries do not vote consistently either, with Brazil, India, and South Africa not always siding with Russia and China. As mentioned above, the voting bloc appearing to support Russia and China also did not always vote consistently, but four nations (North Korea, Iran, Nicaragua, and Syria) voted unwaveringly with Russia and China.

FIGURE 1: VOTING OF THE VARIOUS GROUPS

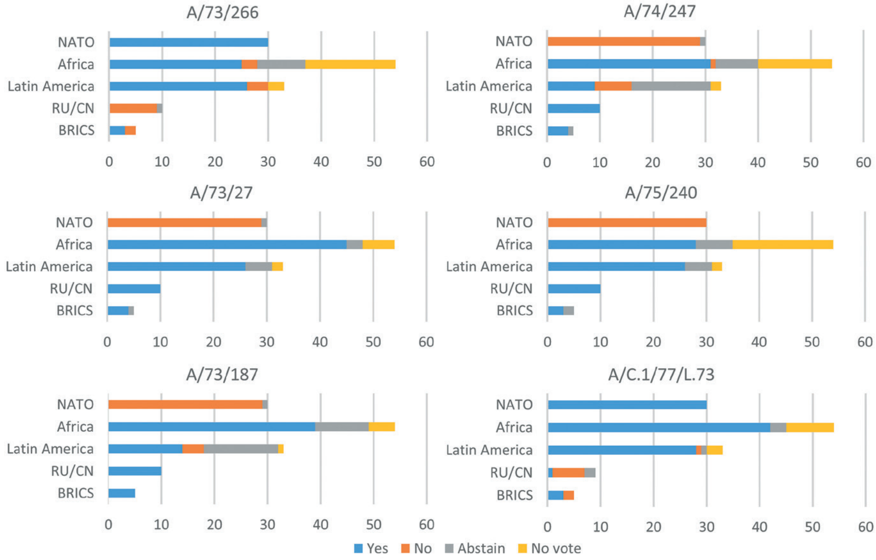
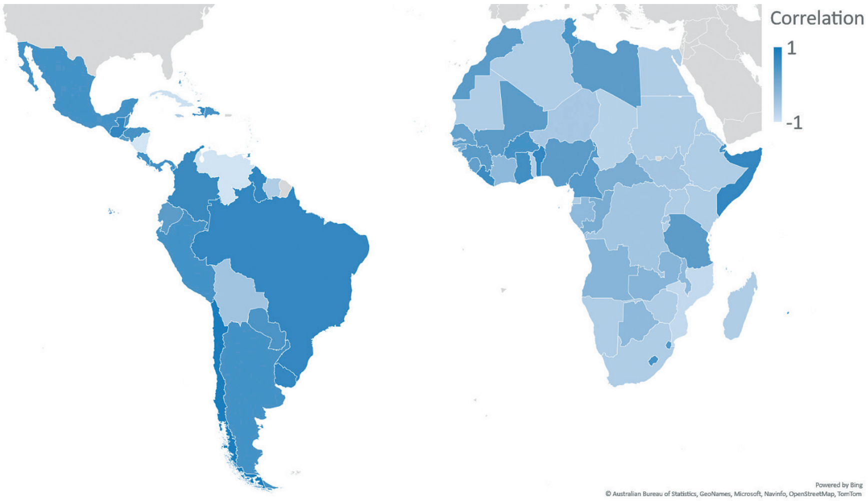


Figure 2 is a map generated using Microsoft Excel, illustrating the correlation of the national votes from Africa and Latin America to the NATO position. Dark blue (1) indicates strong alignment with NATO, and light blue (-1) indicates strongly opposing votes, i.e. those aligned with the Russian and Chinese position. This was done by assigning values to the votes (yes = 3, abstain = 2, no = 1, and no vote = 0).

The given regions displayed an overall inclination towards Chinese and Russian positions: 27 countries had a correlation of below -0.5, while only 14 countries had a correlation of above 0.5. In addition, Latin American countries (10 of which had over 0.5 and seven of which had under -0.5) were evidently more aligned with the NATO position than African ones (20 of which had under -0.5 and four of which had over 0.5). The varying positions in Africa and Latin America, combined with abstentions and lack of voting, indicate that these regions can be considered a swing vote in international cyber diplomacy.

FIGURE 2: MAP OF VOTING CORRELATION TO THE NATO POSITION


Correlation
1
-1
Powered by Bing
© Australian Bureau of Statistics, GeoNames, Microsoft, Navinfo, OpenStreetMap, TomTom

While the 2022 vote on the Programme of Action may have been affected by the Russian invasion of Ukraine, another trend may be emerging: that of inclusivity. The GGE process only had a limited number of participating nations, whereas the Russian proposals allowed participation for all nations, as well as multi-stakeholder modalities. The 2022 Programme of Action was not only inclusive but also aimed at capacity-building (Weber 2022), which will ultimately benefit those nations that are struggling to participate. Therefore, broader initiatives sponsored by NATO members to aid inclusivity and capacity-building within Africa and Latin America could provide greater support in cyber diplomacy.

3. NOT LIKE-MINDED COUNTRIES?

As shown in the last section, African and Latin American countries are included in the UN's cyber policy discussions on cybersecurity and Internet governance. But there is still a divide that broadly distances countries from those regions from the group of *like-minded* countries that include NATO and European countries.

NATO is primarily a political and military alliance, with the primary objective being the collective defence of its members 'against all threats, from all directions. To do this, the Alliance fulfils three core tasks: deterrence and defense; crisis prevention and management; and cooperative security' (NATO 2022). In addition, a 2016 fact sheet on cyber defence provided by the Public Diplomacy Division indicated cooperation with partners and industry (NATO 2016). We argue that engaging in cyber diplomacy is an activity that supports the three core tasks and should not be seen as an additional responsibility outside of the main strategic mandate of the Alliance. Cyber diplomacy includes aspects of cooperation, capacity-building, and confidence-building measures, which align with cooperative security and crisis management related to cyberspace. In addition, open statements can be made in the common forums in support of deterrence. For example, NATO and NATO-affiliated centres of excellence (e.g. CCDCOE, Strategic Communication COE, and Hybrid COE) can participate in the UN OEWG through the multi-stakeholder sessions; this provides a way to engage and foster collaboration with industry and regional organizations (e.g. the African Union and the Organization of American States [OAS]), as well as providing a platform to provide a narrative to support NATO's deterrence task.

EU member states have a range of cyber diplomacy initiatives. The EU is the most advanced region in terms of cyber diplomacy (Lațici 2020) and includes the Cyber Diplomacy Toolbox (European Commission 2020; Borrel 2020; Lațici 2020). The Paris Call for Trust and Security in Cyberspace has been particularly successful, with two workstreams including engagement with the Global South: a working group on engaging emerging countries in 2021 (LetsTalkCyber 2020), and scoping meetings for a South–South cooperation in capacity-building workstream were held in 2022. However, during these discussions, limitations and shortcomings emerged, particularly in approaches to capacity-building: engagement is often narrow, with only a few stakeholders; capacity-building does not align with the needs of the recipient country; and capacity-building efforts could undercut small and medium-sized enterprises whose niche is in cybersecurity training and consulting (Paris Call 2022). Ifeanyi-Ajufo (2022) equally indicates the challenge of 'fragmented and divergent cybersecurity cooperation models and visions'. Other forums for cybersecurity cooperation among NATO, the EU, Africa and Latin America include the Sixth European Union – African Union Summit and the Africa–EU Partnership on Policy and Regulation Initiative for Digital Africa (Ifeanyi-Ajufo 2022). In March 2023, collaboration in the digital sphere between the EU and Latin America was enhanced with the launch of the EU–LAC Digital Alliance (European Commission 2023). In the Americas, there is the Working Group on Cooperation and Confidence-Building Measures in Cyberspace (CBM) within the OAS, where 'member-states have incrementally added new CBMs to the list' (Hurel 2022). However, these initiatives have not been enough for African and Latin American countries to be included in the roster of *like-minded* countries when discussing concrete cyber defence mechanisms. The Global Action on Cybercrime and engagement with the Economic Community of West African States and an MoU with the Dominican Republic are focused on cybercrime and incident response (Borrel 2020; Ifeanyi-Ajufo 2022). This does not necessarily help provide a broader ideological perspective on the use of the Internet for military and geopolitical purposes. EU Cyber Direct has engaged with both Latin America and Africa. However, Ifeanyi-Ajufo (2022) indicates that Africa still has not embraced cybersecurity and may be resistant to engagements over the fear of digital colonialism. This suggests that a different approach is required.

Chinese engagement in Africa includes promoting the digital sovereignty of African nations and funding technology projects, known as the Digital Silk Road (DSR). Of the 90 technology loan projects provided by China worldwide, 74 are being provided to African governments (Hicks 2022; Tugendhat and Voo 2021). Efforts have been made to expand the DSR to Latin America. However, prior to the DSR initiative, China had invested US $1.5 billion into Latin American technology projects from 2009 to 2015 (Malena 2021). In addition, surveillance technology is a major export from China, as evidenced by its backing of digital authoritarianism in Latin America (Moreno 2022) and its investment of over US $200 million in 2023 for Zimbabwe to implement a surveillance system (Africa Defence Forum 2023).

In the last five years, NATO and EU members have undertaken the collaborative attribution of cyber campaigns. Notable examples include the US Democratic National Committee hacks (DHS 2016), NotPetya (CFR 2018), the Russian cyber attacks against Georgia (European Council 2020; Roguski 2020), and the attribution of several Chinese cyber actions (White House 2021). The same can be said about the recent Russian cyber campaigns against Ukraine (Australia 2022; Canada 2022; United Kingdom 2022; United States 2022; European Council 2022). Even the broader call to action issued in March 2022 by Canada, as the chair of the Freedom Online Coalition against state-sponsored disinformation targeting Ukraine, did not have the endorsement of any African or Latin American country (Global Affairs Canada 2022).

NATO and EU countries led these initiatives, with other *like-minded* countries, including Australia, Japan, and New Zealand. However, no African or Latin American country joined the collaborative attribution efforts. One might argue that the effort to include countries from Africa and Latin America requires establishing clear points of contact in different countries, which might gather stakeholders from law enforcement agencies, intelligence agencies, and military organizations. This is undoubtedly laborious in light of the difference in institutional maturity in these regions but reaps benefits as it creates a network among them. Egloff and Smeets (2021) present a framework for public attribution that helps guide efforts for broader collaborative attributions. The process would necessarily include sharing information regarding cyber campaigns to obtain countries' support for the attribution. There is no need for an all-or-nothing approach, as countries might refrain from disclosing the targeting state. Examples include collaborative attributions for the ransomware WannaCry and the 2020 collaborative attribution regarding 2019 cyber operations against Georgia. In the latter, the EU and some countries attributed the action to Russia; others refrained from doing so and only deplored the incidents, while still others (France and Germany) remained silent (European Council 2020; Roguski 2020). In May 2022, the European Council, supported by Turkey, North Macedonia, Montenegro, Albania, Bosnia and Herzegovina, Iceland, Ukraine, Moldova, and Georgia, condemned the Viasat incident, stating that 'such behaviour is contrary to the expectations set by all UN Member States, including the Russian Federation, of responsible State behaviour and the intentions of States in cyberspace' (European Council 2022).

There is debate over the consequences of attribution, but that topic is beyond the scope of this paper. In any case, most scholars agree that attribution remains an essential tool for geopolitical reasons. The expansion in the number and geographical representation of countries supporting attribution is a net gain for those leading the initiative, regardless if they wish to 'signal unaccepted behaviour' or 'shape international norms' (Bateman 2022).

The sharing of technical information prior to collaborative attribution would benefit African and Latin American countries, as it would enhance their capacity to detect threats. In this case, information exchange could become a two-way street. For NATO and the EU, this could mean increasing the telemetry – by receiving raw data on malicious behaviour detected in Latin America and Africa – to assess or confirm behaviour from sophisticated threat actors.

But the most relevant initiative would be to provide tangible elements to raise the bar on cyber defence among nations. This point has been agreed upon on multiple occasions in forums such as the UN GGE and the OEWG, though it has rarely been put into practice. Information-sharing would help Latin America and Africa better defend themselves against threats. Further, collaborative attribution efforts would solidify a more transparent framework of responsible state behaviour in cyberspace.

When engaging with the Global South, the relevant NATO-affiliated COEs are a suitable platform for capacity-building and collaboration, enabling an exchange of ideas. It is important for NATO and EU states to provide engagement across various sectors and multiple organizations and institutions across those sectors to maximize the return on investment. To alleviate concerns of foreign digital colonialism, we propose that the focus of the engagement be placed on academia, who can advocate change from inside the country through their established stakeholder engagements. However, many in the Global South face the problem of access to EU and NATO countries, particularly involving lengthy visa processes. A possible initiative to enable engagement and collaboration is for EU members to proactively allow established cybersecurity researchers from the Global South to seek long-term visas to provide the necessary flexibility to attend events in the relevant countries. A focus on Africa will provide the best return on investment, as Africa contains more countries and is less aligned with NATO and the EU than Latin America is (as illustrated by Figures 1 and 2). Recruiting cyber policy advisors from the major economies in the regions, and based in their respective countries, will provide additional analysis and engagement to identify areas for cooperation and engagement.

4. SURVEILLANCE AND DOMESTIC TARGETING

Voting patterns in the UN indicate that Africa and Latin America lean towards Russia and China on some issues. The collaborative attribution experience shows opportunities for information-sharing among Africa and Latin America on one side and NATO and the EU on the other. But for a coherent values-based cyber policy, more attention needs to be given to human rights online, especially with the recent publication of the European Digital Rights and Principles (European Commission 2022a).

The use of OCCs is changing from a taboo in international relations to an action compatible with responsible state behaviour. However, the number of countries developing and deploying OCCs is uncertain, especially given the ‘lack of agreement about the realities of cyber proliferation’ (Smeets 2022). The estimates range from more than 30 to well over 100 nations, depending on the data source. A recent study suggests that 29 countries used OCCs, while a total of 86 nations acquired them from private vendors (Izycki 2022a). In both Africa and Latin America, 12 countries in each region purchased OCCs, with five countries each having multiple private providers as of 2020. Furthermore, the private vendors frequently selling OCCs are headquartered in NATO or EU countries such as Germany, Italy, Canada, the United Kingdom, the United States, and France (Izycki 2022b).

It is worth noting that other countries are also responsible for commercializing OCCs to the Global South. For example, the Australian Strategic Policy Institute extensively mapped Chinese companies selling artificial intelligence (AI) and surveillance technologies globally (Cave et al. 2019), while the Canadian research institution Citizen Lab reported that Israeli companies such as Circles, the NSO Group, and Candiru are selling OCCs to autocratic governments (Marczak et al. 2018; Marczak, Scott-Railton, Berdan et al. 2021). Citizen Lab also provided evidence of a North Macedonian company providing an offensive cyber solution called Predator; the company was part of a group self-described as 'EU-based and regulated, with six sites and R&D labs throughout Europe' (Marczak, Scott-Railton, Razzak, et al. 2021).

Commercialization of OCCs to the Global South could encourage 'digital colonialism' (Coleman 2019), which might potentially push Latin America and Africa closer to Russia and China. Our objective is not to condemn this commercialization but rather to indicate that responsible state behaviour in cyberspace must comply with digital human rights. That discussion must not be confined to inter-state rivalry; it must also include cases of digital oppression by states (Deibert and Pauly 2019).

We will briefly present the cases of Honduras, Mexico, and Panama, which illustrate how Latin American countries irresponsibly deploy OCCs against domestic targets. All three countries show a 0.5 correlation with the EU and NATO in voting patterns. Similarly, the African nations of Ethiopia and Togo can be considered. Both countries have a closer alignment with Chinese and Russian voting patterns, as they both scored -0.63246 in correlation towards the EU and NATO.

The Honduran case was included in the Citizen Lab report of an extensive investigation into the Bulgarian company Circles, closely associated with the Francisco Partners, a company that also managed the NSO Group. The report documents IP addresses associated with the Honduran National Directorate of Investigation and Intelligence. In addition, the same report presented evidence that Chile, Ecuador, El Salvador, Guatemala, Mexico, and Peru also acquired the software that exploits the SS7 routing protocol for mobile phones (Marczak et al. 2020).

In Panama, the abuse of OCCs is closely related to former President Martinelli, who is under prosecution for the unlawful use of hacking tools against political opponents, business leaders, and union leaders. The revelations began with the help of Italian provider Hacking Team (WikiLeaks 2015) and include the use of Pegasus spyware (the NSO Group) under similar conditions (Marczak et al. 2018). Martinelli and his sons are currently charged with money laundering (Reuters 2022).

Mexico has, on several occasions, used OCCs against civilian targets, including some researchers on soft-drink consumption (Scott-Railton, Marczak, Razzak, et al. 2017; Scott-Railton, Marczak, Guarnieri, et al. 2017; Scott-Railton 2017). In addition, recent leaks by the Guacamaya Group provided evidence of continuous abusive practices, which were confirmed by President Obrador (ElHacker 2022).

The Ethiopian example is one of the oldest reported cases of OCC abuse. Citizen Lab reported several instances of the use of Hacking Team and Cyberbit (an Israeli company part of Elbit Systems) against domestic targets and the Ethiopian diaspora in the US (Marczak et al. 2014; Marczak et al. 2015; Marczak et al. 2017). In addition, Citizen Lab reported the targeting of religious leaders and opposition parties in Togo during the nationwide protests for political reform. Again, the NSO Group's Pegasus spyware was used (Scott-Railton et al. 2020).

While several countries in Africa and Latin America have national security interests that are compatible with OCCs, the cases presented above illustrate situations in which domestic targeting had a clear political motivation. This is part of a larger trend that has been ongoing since the early 2010s (Izycki 2022a). The UN is already concerned with that issue and has pledged to ensure the protection of human rights online through a couple of reports from the Secretary-General and the High Commissioner for Human Rights on the right to privacy online (UNGA 2020; OHCHR 2022). In both cases, the reports address the need for considering that 'even if legitimate goals are being pursued, such as national security objectives or the protection of the rights of others, the assessment of the necessity and proportionality of the use of spyware severely limits the scenarios in which spyware would be permissible' (OHCHR 2022).

The fact that Global South countries acquired OCCs from companies subject to NATO or EU countries' legislation for surveillance based on gender, ethnic, and political grounds should be addressed. A modest start was the US Department of Commerce's blacklisting of four companies (Bureau of Industry and Security 2021) and EU organizations' ongoing investigations regarding the abuse of NSO Pegasus spyware in Greece, Hungary, Poland, and Spain (Marzocchi and Mazzini 2022).

In a study that focused on countering OCC proliferation, DeSombre et al. (2021) point to some policy recommendations to shape and limit proliferation. These include developing know-your-vendor regulations, blacklisting countries that use OCCs to infringe upon human rights, and eventually pursuing legal action against providers contravening agreed rules. In addition to those policy recommendations, we propose the creation of a digital ombudsman to investigate complaints against OCC companies subject to EU countries' legislation. It is worth noting that the abuses in surveillance and domestic targeting usually occur through governmental organizations from the purchasing countries. Therefore, it is unlikely that a domestic investigation in those countries would uncover domestic abuses. As an example, Mexico arrested just one individual for the use of NSO Pegasus (Reuters 2021).

The intention of the ombudsman is to assess compliance from the vendor with EU legislation on human rights online rather than promote legal action against a sovereign nation. If violations were to be found, the EU could enforce the termination of the contract on the grounds of unlawful use, similar to the application of General Data Protection Regulation (GDPR) rules and trade restrictions based on environmental concerns. In a recent example of vendor's liability, the US Supreme Court ruled that the NSO Group was not protected by the Foreign Sovereign Immunity Act because it only applies to sovereign countries (US Supreme Court 2023). Thus, the trial on the violation of Meta's terms and services will proceed.

The ombudsman reinforces NATO and EU countries' commitment to an open internet and is compliant with the EU standardization strategy (European Commission 2022b). It should also empower private vendors to terminate contracts with foreign governments that have committed human rights violations, based on the vendors' legal obligations to EU or NATO countries' jurisdictions.

The Global South's use of OCCs to exert information control and domestic political surveillance is also being enabled by NATO- and EU-based companies. This legitimizes the Russian and Chinese model of a state-controlled internet, as it levels the ground with the argument that 'everyone hacks'. A values-based cyber policy should distinguish the use of OCCs for national security purposes (i.e. combating terrorism and organized crime) from human rights abuses against domestic targets as part of responsible state behaviour in cyberspace.

5. CONCLUSION

The invasion of Ukraine marks the start of a new multipolar era, as Western nations struggle to receive a clear commitment from emerging countries to sanctions against Russia. The EU and NATO still expect the Global South to eventually condemn Russia's unlawful actions against Ukraine, implying that the Global South will have more say in cyberspace policy issues where new concepts, norms, and perceptions are being constructed through UN negotiations and state practice.

Russia and China's growing cyber policy influence is illustrated through the voting patterns (on four Russian proposals), technology investment projects, and advocacy of digital sovereignty. Some of their cyber policy positions have found echoes in Africa and Latin America; however, it is not yet an ideological identity. With the inclusion of the 2022 Programme of Action vote, a voting pattern emerges that implies that inclusivity and capacity-building opportunities within the proposal may be what attracts support for the vote. The EU's existing cyber diplomacy towards the Global South regions has shown limited success. Therefore, NATO and EU countries must engage in a straightforward values-based approach that includes capacity-building and information-sharing with the Global South.

The fact that collaborative attributions have not included countries from Africa and Latin America creates further distance between the 'Western Bloc' and the Global South. This is an excellent opportunity for NATO countries to create a broader coalition of countries supporting the political goals inherent in collaborative attribution. This would help bridge the gap created by the 'democracies vs. autocracies' frame widely used by Western leaders regarding the Russian invasion of Ukraine, to focus instead on responsible state behaviour more akin to international law.

This initiative contains an underlying tension. The possibility that NATO countries are conducting similar cyber operations targeting the new partners in the *like-minded* group could create friction. As an example, during the Cold War, the US sponsored and/or endorsed several coups d'état against Latin American countries. If a NATO member was discovered to be sponsoring a cyber operation attempting to interfere in an election, it would almost certainly revive political ruses among Latin American countries.

Moreover, a values-based cyber policy should take into consideration the cases where African and Latin American potential partners are behaving as threat actors themselves. The acquisition of OCCs is an inevitable trend that can harm human rights online, but this is an essential part of EU rhetoric. Getting the right balance between the use of OCCs as a legitimate raison d'état and human rights requires effort from EU countries. Given the need to update European legislation, this would qualify as a worthy EU effort.

By creating mechanisms for civil society, researchers, NGOs, or individuals from targeted audiences to formalize complaints before the country's technology providers, EU countries can set in motion an alternative mechanism to curb a digital autocratic trend. It would not be an assault on the sovereign use of these capabilities but a domestic enforcement of laws and regulations to providers that sell them abroad. In addition, by engaging with established academic researchers in the Global South, EU countries gain a platform with which to alter perspectives and practices inside nations that are resistant or sceptical of current cyber diplomacy initiatives.

The recommendations presented in this paper aim to counterbalance the economic influence that China currently holds on Africa and Latin America. In addition, the effort to build a values-based cyber policy that engages NATO and the EU with the countries from Africa and Latin America can strengthen historical and cultural ties between the Global North and Global South. This is particularly relevant in the current multipolar era, where cyberspace has become a pertinent dimension for states, companies, and individuals.

REFERENCES

Acharya, Amitav. 2018. *The End of American World Order*, 2nd ed. Cambridge, UK; Malden, MA: Polity Press.

Africa Defence Forum. 2023. ‘Zimbabwe Turns to Chinese Technology to Expand Surveillance of Citizens’. *ADF Magazine*, 17 January 2023. Accessed 17 March 2023.https://adf-magazine.com/2023/01/zimbabwe-turns-to-chinese-technology-to-expand-surveillance-of-citizens/.

Australia. 2022. ‘Attribution to Russia of Malicious Cyber Activity against Ukraine’. Australian Government Department of Foreign Affairs and Trade. 2022. Accessed December 2022. https://www.foreignminister.gov.au/minister/marise-payne/media-release/attribution-russia-malicious-cyber-activity-against-ukraine.

Bateman, Jon. 2022. ‘The Purposes of U.S. Government Public Cyber’. Carnegie Endowment for International Peace. 28 March 2022. Accessed December 2022. https://carnegieendowment.org/2022/03/28/purposes-of-u.s.-government-public-cyber-attribution-pub-86696.

Borrel, Josep. 2020. ‘Cyber diplomacy and shifting geopolitical landscapes’. EU Cyber Forum. 14 September 2020. Accessed 22 February 2023. https://www.eeas.europa.eu/eeas/cyber-diplomacy-and-shifting-geopolitical-landscapes_en.

Bureau of Industry and Security. 2021. ‘Addition of Certain Entities to the Entity List’. Federal Register. 11 April 2021. Accessed December 2022. https://www.federalregister.gov/documents/2021/11/04/2021-24123/addition-of-certain-entities-to-the-entity-list.

Canabarro, Diego, and Thiago Borne. 2015. ‘The Brazilian Reactions to the Snowden Affairs: Implications for the Study of International Relations in an Interconnected World’. *Conjuntura Austral*, Porto Alegre 6(30): 50–74.

Canada. 2022. ‘Statement on Russia’s Malicious Cyber Activity Affecting Europe and Ukraine’. *Canada Global Affairs*, 10 May 2022. Accessed December 2022. https://www.canada.ca/en/global-affairs/news/2022/05/statement-on-russias-malicious-cyber-activity-affecting-europe-and-ukraine.html.

Cave, Danielle, Fergus Ryan, and Vicky Xiuzhong Xu. 2019. ‘Mapping More of China’s Tech Giants: AI and Surveillance’. Australian Strategic Policy Institute. 28 November 2019. Accessed December 2022. https://www.aspi.org.au/report/mapping-more-chinas-tech-giants.

CFR. 2018. ‘NotPetya’. Council on Foreign Relations. April 2018. Accessed March 2021. https://www.cfr.org/cyber-operations/notpetya.

Cohen, Eliot A. 2022. ‘The Return of Statecraft: Back to Basics in the Post-American World’. *Foreign Affairs*, 19 April 2022. https://www.foreignaffairs.com/articles/world/2022-04-19/return-statecraft.

Cohen, Michael, Henry Farrell, and Martha Finnemore. 2013. ‘Hypocrisy Hype: Can Washington Still Walk and Talk Differently?’ *Foreign Affairs*, 1 November 2013. https://www.foreignaffairs.com/articles/united-states/2013-10-15/end-hypocrisy.

Coleman, Danielle. 2019. ‘Digital Colonialism: The 21st Century Scramble for Africa through the Extraction and Control of User Data and the Limitations of Data Protection Laws’. *Michigan Journal of Race & Law* 24(2): 417.

Collett, R. 2021 ‘Cyber Diplomacy: A New Way to Visualise UN Voting Records’. Cybercapacity.org. 2021. Accessed 29 September 2022. https://cybercapacity.org/new-way-to-visualise-un-cyber-diplomacy-voting-records/.

CyberPeace Institute. 2022. ‘Workshop on Advancing the Cyber Programme of Action (PoA)’. CyberPeace Institute. 7 July 2022. https://cyberpeaceinstitute.org/news/cyber-programme-of-action/.

Deibert, Ronald J., and Louis W. Pauly. 2019. ‘Mutual Entanglement and Complex Sovereignty in Cyberspace’. In *Data Politics*, edited by Didier Bigo, Engin Isin, and Evelyn Ruppert, 81–99. London: Routledge, 2019.

DeSombre, Winnona, James Shires, J. D. Work, Robert Morgus, Patrick Howell O’Neill, Luca Allodi, and Trey Herr. 2021. ‘Countering Cyber Proliferation: Zeroing in on Access-as-a-Service’. *Atlantic Council*, 1 March 2021. Accessed December 2022. https://www.atlanticcouncil.org/in-depth-research-reports/report/countering-cyber-proliferation-zeroing-in-on-access-as-a-service/#policyrecommendations.

DHS. 2016. ‘Joint Statement from the Department Of Homeland Security and Office of the Director of National Intelligence on Election Security’. DHS Press Office. 17 October 2016. Accessed December 2022. https://www.dhs.gov/news/2016/10/07/joint-statement-department-homeland-security-and-office-director-national.

Dietrich, C., and P. Pawlak. 2022. ‘Tracking UN Voting Patterns on Cybercrime’. *Directions Blog*, 1 February 2022. Accessed 29 September 2022. https://directionsblog.eu/tracking-un-voting-patterns-on-cybercrime/.

Egloff, Florian J., and Max Smeets. 2021. ‘Publicly Attributing Cyber Attacks: A Framework’. *Journal of Strategic Studies*, 10 March. https://doi.org/10.1080/01402390.2021.1895117.

ElHacker. 2022. ‘Grupo Guacamaya hackeó SEDENA (Secretaría de la Defensa Nacional de México) y filtró 6 TB información’. *ElHacker*, 3 October 2022. Accessed December 2022. https://blog.elhacker.net/2022/10/grupo-guacamaya-hackea-SEDENA-mexico-filtran-6TB-datos.html.

European Commission. 2020. ‘The EU’s Cybersecurity Strategy for the Digital Decade’. JOIN(2020) 18 final, Brussels, 16 December 2020. Accessed March 2023. https://ec.europa.eu/newsroom/dae/redirection/document/72164.

European Commission. 2022a. ‘European Declaration on Digital Rights and Principles’. European Commission. 15 December 2022. Accessed 28 December 2022. https://digital-strategy.ec.europa.eu/en/library/european-declaration-digital-rights-and-principles.

European Commission. 2022b. ‘An EU Strategy on Standardisation – Setting Global Standards in Support of a Resilient, Green and Digital EU Single Market’. COM(2022) 31 final, Brussels, 2 February 2022. Accessed March 2023. https://ec.europa.eu/docsroom/documents/48598.

European Commission. 2023. ‘Global Gateway: EU, Latin America and Caribbean partners launch in Colombia the EU-LAC Digital Alliance’. European Commission, 14 March 2023. Accessed 17 March 2023.https://ec.europa.eu/commission/presscorner/detail/en/ip_23_1598.

European Council. 2020. ‘Declaration by the High Representative on Behalf of the European Union – Call to Promote and Conduct Responsible Behaviour in Cyberspace’. Council of the European Union. 21 February 2020. Accessed December 2022. https://www.consilium.europa.eu/en/press/press-releases/2020/02/21/declaration-by-the-high-representative-on-behalf-of-the-european-union-call-to-promote-and-conduct-responsible-behaviour-in-cyberspace.

European Council. 2022. ‘Russian cyber operations against Ukraine: Declaration by the High Representative on behalf of the European Union’. Council of the European Union. 10 May 2022. https://www.consilium.europa.eu/en/press/press-releases/2022/05/10/russian-cyber-operations-against-ukraine-declaration-by-the-high-representative-on-behalf-of-the-european-union/.

Global Affairs Canada. 2022. ‘Statement on behalf of the Chair of the Freedom Online Coalition: A Call to Action on State-Sponsored Disinformation in Ukraine’. *Global Affairs Canada*, 2 March 2022. Accessed December 2022. https://www.canada.ca/en/global-affairs/news/2022/03/statement-on-behalf-of-the-chair-of-the-freedom-online-coalition-a-call-to-action-on-state-sponsored-disinformation-in-ukraine.html.

Hicks, Jacqueline. 2022. ‘Export of Digital Surveillance Technologies from China to Developing Countries’. Institute of Development Studies. 1 August 2022. Accessed 29 December 2022. https://opendocs.ids.ac.uk/opendocs/handle/20.500.12413/17644.

Hurel, Louise Marie. 2022. ‘Beyond the Great Powers: Challenges for Understanding Cyber Operations in Latin America’. *Global Security Review* 2, Article 7.

Ifeanyi-Ajufo, Nnenna. 2022. ‘International Cooperation and Cybersecurity in Africa’. Directionsblog. 16 December 2022. Accessed 28 December 2022. https://directionsblog.eu/international-cooperation-and-cybersecurity-in-africa/.

Izycki, Eduardo. 2022a. ‘Cyber Power Diffusion: Global, Regional and Local Implications’. In *Modelling Nation-State Information Warfare and Cyber-operations*, edited by B. Van Niekerk, T. Ramluckan, and N. Kushwaha, 79–110. London: Academic Conferences and Publishing International Limited.

Izycki, Eduardo. 2022b. *Nation-State Cyber Offensive Capabilities: An In-depth Look into a Multipolar Dimension*. São Paulo: Dialética.

Lațici, Tania. 2020. ‘Understanding the EU’s Approach to Cyber Diplomacy and Cyber Defence’. European Parliamentary Research Service. 28 May 2020. https://www.europarl.europa.eu/thinktank/en/document/EPRS_BRI(2020)651937.

Layne, Christopher. 2018. ‘The US–Chinese Power Shift and the End of the Pax Americana’. *International Affairs* 94(1): 89–111.

LetsTalkCyber. 2020. ‘Six Paris Call Working Groups Announced’. LetsTalkCyber. 16 November 2020. Accessed 22 February 2023. https://letstalkcyber.org/news/six-paris-call-working-groups-announced.

Malena, Jorge. 2021. ‘The Extension of the Digital Silk Road to Latin America: Advantages and Potential Risks’. Council of Foreign Relations. 19 January 2021. Accessed 17 March 2023. https://www.cfr.org/blog/extension-digital-silk-road-latin-america-advantages-and-potential-risks.

Marczak, Bill, Claudio Guarnieri, Morgan Marquis-Boire, and John Scott-Railton. 2014. ‘Hacking Team and the Targeting of Ethiopian Journalists’. Citizen Lab. 12 February 2014. Accessed 2019. https://citizenlab.org/2014/02/hacking-team-targeting-ethiopian-journalists/.

Marczak, Bill, Geoffrey Alexander, Sarah McKune, John Scott-Railton, and Ron Deibert. 2017. ‘Champing at the Cyberbit: Ethiopian Dissidents Targeted with New Commercial Spyware’. Citizen Lab. 6 December 2017. Accessed December 2022. https://citizenlab.ca/2017/12/champing-cyberbit-ethiopian-dissidents-targeted-commercial-spyware/.

Marczak, Bill, John Scott-Railton, and Sarah McKune. 2015. ‘Hacking Team Reloaded? US-Based Ethiopian Journalists Again Targeted with Spyware’. Citizen Lab. 9 March 2015. Accessed December 2022. https://citizenlab.ca/2015/03/hacking-team-reloaded-us-based-ethiopian-journalists-targeted-spyware/.

Marczak, Bill, John Scott-Railton, Bahr Abdul Razzak, Noura Al-Jizawi, Siena Anstis, Kristin Berdan, and Ron Deibert. 2021. ‘Pegasus vs. Predator: Dissident’s Doubly-Infected iPhone Reveals Cytrox Mercenary Spyware’. Citizen Lab. 16 December 2021. Accessed December 2022. https://citizenlab.ca/2021/12/pegasus-vs-predator-dissidents-doubly-infected-iphone-reveals-cytrox-mercenary-spyware/.

Marczak, Bill, John Scott-Railton, Kristin Berdan, Bahr Abdul Razzak, and Ron Deibert. 2021. ‘Hooking Candiru: Another Mercenary Spyware Vendor Comes into Focus’. Citizen Lab. 15 July 2021. Accessed December 2022. https://citizenlab.ca/2021/07/hooking-candiru-another-mercenary-spyware-vendor-comes-into-focus/.

Marczak, Bill, John Scott-Railton, Sarah McKune, Bahr Abdul Razzak, and Ron Deibert. 2018. ‘Hide and Seek: Tracking NSO Group’s Pegasus Spyware to Operations in 45 Countries’. Citizen Lab. 18 September 2018. Accessed December 2022. https://citizenlab.ca/2018/09/hide-and-seek-tracking-nso-groups-pegasus-spyware-to-operations-in-45-countries/.

Marczak, Bill, John Scott-Railton, Siddharth Prakash Rao, Siena Anstis, and Ron Deibert. 2020. ‘Running in Circles’. Citizen Lab. 1 December 2020. Accessed December 2022. https://citizenlab.ca/2020/12/running-in-circles-uncovering-the-clients-of-cyberespionage-firm-circles/.

Martin, Alexander. 2022. ‘The Future of the Internet Is up for Vote at the U.N.’. *Record*. 28 September 2022. Accessed 29 September 2022. https://therecord.media/the-future-of-the-internet-is-up-for-vote-at-the-u-n/.

Marzocchi, Ottavio, and Martina Mazzini. 2022. ‘Pegasus and Surveillance Spyware’. European Parliament. May 2022. Accessed December 2022. https://www.europarl.europa.eu/RegData/etudes/IDAN/2022/732268/IPOL_IDA(2022)732268_EN.pdf.

Moreno, Jaime. 2022. ‘China Seen Backing “Digital Authoritarianism” in Latin America’. VOA News. 14 January 2022. Accessed 17 March 2023. https://www.voanews.com/a/china-seen-backing-digital-authoritarianism-in-latin-america-/6398072.html.

NATO. 2016. ‘Cyber Defence’. North Atlantic Treaty Organization. Accessed 22 February 2023. https://www.nato.int/nato_static_fl2014/assets/pdf/pdf_2016_07/20160627_1607-factsheet-cyber-defence-en.pdf.

NATO. 2022. ‘Strategic Concepts’. North Atlantic Treaty Organization. 18 July 2022. Accessed 22 February 2023. https://www.nato.int/cps/en/natohq/topics_56626.htm.

Nye, Joseph S. 2010. ‘The Future of American Power: Dominance and Decline in Perspective’. *Foreign Affairs* (November 1): 2–12. https://www.foreignaffairs.com/united-states/future-american-power.

OHCHR. 2022. ‘The Right to Privacy in the Digital Age’. Human Rights Council. 12 September 2022. Accessed December 2022. https://documents-dds-ny.un.org/doc/UNDOC/GEN/G22/442/29/PDF/G2244229.pdf?OpenElement.

Paris Call. 2022. ‘South-South Cooperation in Cyber Capacity Building and Good Practices’. Paris Peace Forum. 11 November 2022. https://parispeaceforum.org/en/events/south-south-cooperation-in-cyber-capacity-building-and-good-practices/.

Regissahui, Magby Henri Joel. 2019. ‘Overview on the China-Africa Trade Relationship’. *Open Journal of Social Sciences* 7(7): 381–403.

Reuters. 2021. ‘Mexico Detains Man Implicated in Pegasus Spy Plot against Journalist’. Reuters. 8 November 2021. Accessed March 2023. https://www.reuters.com/world/americas/mexico-detains-man-implicated-pegasus-spy-plot-against-journalist-2021-11-08/.

Reuters. 2022 ‘Panama Ex-president Martinelli to Stand Trial on Money Laundering Charge’. Reuters. 9 December 2022. Accessed December 2022. https://www.reuters.com/world/americas/panama-ex-president-martinelli-stand-trial-money-laundering-charge-2022-12-10/.

Roguski, Przemysław. 2020. ‘Russian Cyber Attacks Against Georgia: Public Attributions and Sovereignty in Cyberspace’. *Just Security*, 6 March 2020. Accessed March 2021. https://www.justsecurity.org/69019/russian-cyber-attacks-against-georgia-public-attributions-and-sovereignty-in-cyberspace/.

Roy, Diana. 2022. ‘China’s Growing Influence in Latin America’. Council on Foreign Relations. 12 April 2022. Accessed December 2022. https://www.cfr.org/backgrounder/china-influence-latin-america-argentina-brazil-venezuela-security-energy-bri.

Scott, David. 2008. *The Chinese Century?: The Challenge to Global Order*. Basingstoke, UK; New York: Palgrave MacMillan.

Scott-Railton, John. 2017. ‘Reckless Exploit: Mexican Journalists, Lawyers, and a Child Targeted with NSO Spyware’. Citizen Lab. 19 June 2017. Accessed 20 November 2018. https://citizenlab.ca/2017/06/reckless-exploit-mexico-nso/.

Scott-Railton, John, Bill Marczak, Bahr Abdul Razzak, Masashi Crete-Nishihata, and Ron Deibert. 2017. 'Reckless III – Investigation Into Mexican Mass Disappearance Targeted with NSO Spyware'. Citizen Lab. 10 July 2017. Accessed 2019. https://citizenlab.ca/2017/07/mexico-disappearances-nso/.

Scott-Railton, John, Bill Marczak, Claudio Guarnieri, and Masashi Crete-Nishihata. 2017. 'Bitter Sweet: Supporters of Mexico's Soda Tax Targeted With NSO Exploit Links'. Citizen Lab. 2 November 2017. Accessed 2018. https://citizenlab.org/2017/02/bittersweet-nso-mexico-spyware/.

Scott-Railton, John, Siena Anstis, Sharly Chan, Bill Marczak, and Ron Deibert. 2020. 'Nothing Sacred – Religious and Secular Voices for Reform in Togo Targeted with NSO Spyware'. Citizen Lab. 3 August 2020. Accessed 2020. https://citizenlab.ca/2020/08/nothing-sacred-nso-sypware-in-togo/.

Smeets, Max. 2022. *No Shortcuts: Why States Struggle to Develop a Military Cyber-Force*. Oxford: Oxford University Press.

Trubowitz, Peter, and Peter Harris. 2019. 'The End of the American Century? Slow Erosion of the Domestic Sources of Usable Power'. *International Affairs* 95(3): 619–39. https://doi.org/10.1093/ia/iiz055.

Tugendhat, Henry, and Voo, Julia. 2021. 'China's Digital Silk Road in Africa and the Future of Internet Governance'. China-Africa Research Initiative, Policy Brief No 60. Accessed 29 December 2022. https://saiia.org.za/wp-content/uploads/2021/01/CARI_PB60_TugendhatVooChinaDigitalSilkRoadAfrica.pdf.

UNGA. 2020. 'Road Map for Digital Cooperation: Implementation of the Recommendations of the High-Level Panel on Digital Cooperation'. United Nations General Assembly. 29 May 2020. Accessed December 2022. https://documents-dds-ny.un.org/doc/UNDOC/GEN/G22/442/29/PDF/G2244229.pdf?OpenElement.

United Kingdom. 2022. 'Russia Behind Cyber-Attack with Europe-wide Impact an Hour Before Ukraine Invasion'. GOV.UK. 10 May 2022. Accessed December 2022. https://www.gov.uk/government/news/russia-behind-cyber-attack-with-europe-wide-impact-an-hour-before-ukraine-invasion.

United States. 2022. 'Attribution of Russia's Malicious Cyber Activity Against Ukraine'. United States Department of State. 10 May 2022. Accessed December 2022. https://www.state.gov/attribution-of-russias-malicious-cyber-activity-against-ukraine/.

US Supreme Court. 2023. NSO Group Technologies Limited, et al., Petitioners v. WhatsApp Inc., et al. No. 21-1338. January 2023. Accessed 2023. https://www.supremecourt.gov/search.aspx?filename=/docket/docketfiles/html/public/21-1338.html.

Weber, Valentin. 2022. 'How to Strengthen the Program of Action for Advancing Responsible State Behavior in Cyberspace'. *Just Security*, 19 February 2022. Accessed 23 December 2022. https://www.justsecurity.org/80137/how-to-strengthen-the-programme-of-action-for-advancing-responsible-state-behavior-in-cyberspace/.

White House. 2021. 'Background Press Call by Senior Administration Officials on Malicious Cyber Activity Attributable to the People's Republic of China'. White House. 19 July 2021. Accessed December 2022. https://www.whitehouse.gov/briefing-room/press-briefings/2021/07/19/background-press-call-by-senior-administration-officials-on-malicious-cyber-activity-attributable-to-the-peoples-republic-of-china/.

WikiLeaks. 2015. 'Hacking Team'. Wikileaks. 8 July 2015. Accessed 20 November 2018. https://wikileaks.org/hackingteam/emails/.