H-Diplo | Robert Jervis International Security Studies Forum

Policy Roundtable III-4

The Practices and Politics of Cybersecurity Expertise

29 March 2024 |PDF: https://issforum.org/to/jprIII-4 | Website: rjissf.org | Twitter: @HDiplo

Editor: Diane Labrosse
Commissioning Editors: Rebecca Slayton and Lilly Muller
Production Editor: Christopher Ball
Pre-Production Copy Editor: Bethany Keenan

Contents

Introduction by Rebecca Slayton, Cornell University, and Lilly Muller, King's College London*

Experts play pervasive and multifarious roles in shaping the international order. A sophisticated body of literature within international relations explores how experts shape governmental and international policy—including their work in framing problems, gathering and interpreting ambiguous evidence, and proposing policy solutions.[1] Separately, scholars whose work is informed by the field of science and technology studies have examined how experts' technological work can enact international order more directly.[2] The essays in this forum focus on this latter mode of influence, showing how experts participate in politics by other means—specifically the making and breaking of the security of networked information systems.

By examining these practices, this forum calls attention to several key questions. What does cybersecurity mean to distinctive polities, and how do these meanings change over time? How do these different conceptions of cybersecurity shape what constitutes legitimate and authoritative expert practice? And to what extent can expert practices not only enact, but actively transform power relations? The seven research essays in this forum demonstrate substantial variation in how these questions are answered, and two concluding essays provide reflections on the significance of this variation for scholarship and policy.

This introductory essay frames this variation through a relational conception of expertise. We draw on scholarship which analyzes expertise not primarily as the possession of skills or knowledge, but rather as the enactment of relationships between experts, non-experts, and culturally valuable objects of expertise.[3] These relationships are enacted in ways that vary with time and place.

The first two essays in this forum analyze examples of how hackers attempt to use their skills to challenge dominant political structures.

Max Smeets analyzes the Cyber Partisans, a Belarusian hacking group that opposes the authoritarian government of Belarusian President Alexander Lukashenko. Smeets argues that the Cyber Partisans should not be understood as a proxy for state action, but rather as a social movement. He further suggests that integrating insights from social-movements theory into analyses of hacking may be a productive avenue for

* This forum is based in part on research supported by National Science Foundation CAREER award grant number 1553069, "Enacting Cybersecurity Expertise."

[1] The most prominent tradition in the field is the epistemic communities literature; see e.g., Peter Haas, "Introduction: Epistemic Communities and International Policy Coordination," *International Organization* 42:1 (1992): 1-35. However, other traditions have emerged in recent decades; for a relatively recent review, see Christian Bueger, "From Expert Communities to Epistemic Arrangements: Situating Expertise in International Relations," in *The Global Politics of Science and Technology-Vol. 1* (Springer, 2014): 39-54.

[2] This is the essence of "technopolitics." See, for example, Gabrielle Hecht, "Technology, Politics, and National Identity in France," in *Technologies of Power: Essays in Honor of Thomas Parke Hughes and Agatha Chipley Hughes*, ed. Michael Thad Allen and Gabrielle Hecht (Cambridge, MA: MIT Press, 2001): 253-294; and Hecht, ed., *Entangled Geographies: Empire and Technopolitics in the Global Cold War* (MIT Press, 2011). The emphasis on technologies as enacting political order is implicit in a small but growing body of work on the international politics of infrastructure; see, for example, Marieke de Goede and Carola Westermeier, "Infrastructural Geopolitics," *International Studies Quarterly* 66:3 (2022), https://doi.org/10.1093/isq/sqac033, https://doi.org/10.1093/isq/sqac033. However, these works still tend to give primary agency to national leaders rather than the more mundane day-to-day work of technologists.

[3] E. Summerson Carr, "Enactments of Expertise," *Annual Reviews of Anthropology* 39 (2010): 17-32.

international relations scholars. Such work might build productively on existing analyses of the politics of hacking.[4]

Matt Goerzen's essay examines the rise of an "anti-security" meme in the late 1990s underground. He argues that this meme was a response to the rise of a cybersecurity industry that threatened to co-opt members of the underground and eliminate the vulnerabilities that they wished to exploit. While there was no unified social movement, the tactics embraced under the banner of anti-security have resurfaced in prominent hacks on the surveillance industry in recent years—attacks which undermine both the technical and political viability of the corporations that support autocratic political regimes.

As Goerzen notes, hackers' participation in the underground sometimes served as a source of technical authority that enabled them to earn lucrative paychecks in the mainstream security industry. The alignment of hackers and corporations might seem to be another example of how hackers have worked the system to their advantage. Or, it might be interpreted as the co-option of hackers' expertise by the very establishment that they once opposed. Either way, however, these realignments came with new risks to the interests of both hackers and corporations.

In previous work, Ryan Ellis and Yuan Stevens show that bug bounty programs, which offer hackers financial rewards for information about vulnerabilities that can be exploited, cultivate precarity for hackers by turning them into gig workers.[5] In this forum, Ryan Ellis goes further by revealing the risks that such program pose for the mainstream security industry. Contrary to the many scholars who portray bug bounty programs as a means of correcting market failures that lead to poor security, Ellis shows that bug bounty programs can also serve as targets for exploitation by malicious actors, thereby creating new risks.

Ellis's essay demonstrates the risks that are associated with embedding expert knowledge and practice in the digital infrastructure. Andrew Dwyer further explores these risks, showing how globally-interconnected infrastructures for malware analysis and detection have shaped the possibilities for espionage and subversion. Thus, for example, the malware detection infrastructure of the Moscow-based endpoint detection vendor Kaspersky enabled it to download the US National Security Agency's tools—whether accidentally, as claimed by Kaspersky, or in the service of the Russian government, as claimed by the United States and its allies. Dwyer emphasizes that analysts' knowledge becomes embodied in malware detection networks, and trust in those networks creates expertise. By banning Kaspersky from US governmental networks, the US government effectively affirms its confidence in the technical authority of its analysts, but not in their moral authority.

In their article, Rebecca Slayton and Clare Stevens shift the focus from malware detection infrastructure to national critical infrastructure, which comprises the systems that are essential to the everyday safety and security of nation-states. They examine how experts construct and reconfigure boundaries around their authority and expertise, focusing in particular on the oft-cited boundary between security in information technologies (IT) and operational technology (OT). They argue that developing a bottom-up understanding

[4] See, for example, Gabriella Coleman, *Hacker, Hoaxer, Whistleblower, Spy: The Many Faces of Anonymous* (New York: Verso, 2014); Matt Goerzen and Coleman, *Wearing Many Hats: The Rise of the Professional Security Hacker*, Data and Society (2022), https://datasociety.net/wp-content/uploads/2022/03/WMH_final01062022Rev.pdf; Molly Sauter, *The Coming Swarm: DDOS Actions, Hacktivism, and Civil Disobedience on the Internet* (Bloomsbury Academic: New York, 2014).

[5] Ryan Ellis and Yuan Stevens, *Bounty Everything: Hackers and the Making of the Global Bug Marketplace*, Data & Society (January 2022), https://datasociety.net/library/bounty-everything-hackers-and-the-making-of-the-global-bug-marketplace/.

of how experts alternately produce, maintain, navigate, and transcend boundaries between such fields can improve the development of policies to manage boundary-spanning risks.

Jesse Sowell's essay examines the politics of maintaining reliable infrastructure, by showing how internet operators use "rough consensus" to establish credibility and authority within their technical community. While internet-operating organizations ground their authority in this technical and ostensibly apolitical mode of decisionmaking, Sowell describes their work as a kind of "low politics" that shapes everyday communications infrastructures. For example, Sowell opens with the example of the refusal of the European network operator organization to accommodate Ukraine's request to cut Russia off from the internet in the wake of the brutal Russian invasion of Ukraine.

In his essay, James Shires shifts the focus from maintaining digital connectivity, to efforts to exclude individuals from digital networks. In particular, he shows how the enactment of cybersecurity expertise in carceral conditions differs from its enactment in everyday commerce and government operations. Mainstream studies of cybersecurity emphasize the tradeoffs between usability and security: organizations want technologies that are easy to use, but only in authorized ways and by authorized individuals. By contrast, prisons and other carceral contexts aim to prohibit use and enforce digital exclusion for incarcerated individuals. Shires notes that expert knowledge and practice are necessary both for circumventing and enforcing such exclusion.

Two concluding essays synthesize some broader lessons from this forum. First, Jon Lindsay notes that these essays help us move beyond two common but contradictory narratives about cybersecurity expertise. One is that cybersecurity expertise is scarce, something held only by a few technical geniuses. The other is that cybersecurity expertise is widely available, as it is relatively easy to launch highly damaging cyberattacks. These essays complicate the first narrative by showing that cybersecurity expertise is diverse and distributed across social organizations, rather than something possessed by relatively rare hacker-geniuses, and they complicate the second by revealing the substantial labor needed to either maintain or compromise the security of complex socio-technical systems.

Aaron Gluck-Thaler concludes with a reflection on how and why some expert practices and institutions have become dominant, while others have been marginalized. As he notes, "the priorities of states and corporations do disproportionately shape the possible forms that cybersecurity expertise can take," as evidenced by the alliance of cybersecurity and national surveillance industries, and the marginalization and cooptation of the hacker underground. He argues for additional research that examines how expert practices are not only shaped by the interests of states, corporations, and civil society, but also help to co-produce them.

This observation brings us back to the relational conception of expertise. The expert authority of the individuals and communities which are discussed in this forum ultimately derives from their ability to persuade other powerful actors that they possess specialized knowledge and skills—whether through theatrical technical disruptions, the reliable provision of infrastructure, the enforcement of digital exclusion, or the identification of new exploits and the construction of means for preventing those exploits. It is through these relational practices that experts gain not only credibility with policymakers but also the ability to pursue politics by technical means. We hope that these essays will encourage other scholars to examine the daily work of technical experts as a locus of power in international politics.

Contributors:

Rebecca Slayton is Associate Professor in the Department of Science & Technology Studies and the Judith Reppy Institute for Peace and Conflict Studies, both at Cornell University. Her research examines how new fields of expertise become institutionalized and gain authority in the contexts of international security and cooperation. Her first book, *Arguments that Count: Physics, Computing, and Missile Defense, 1949–2012* (MIT Press, 2013), shows how the rise of a new field of expertise in computing reshaped public policies and perceptions about the risks of missile defense in the United States. In 2015, *Arguments that Count* won the Computer History Museum Prize. Slayton's current book project, *Shadowing Cybersecurity*, examines how expert knowledge and practice in cybersecurity have been shaped by conflicting notions of security, as well as the irreducible uncertainties associated with intelligent adversaries.

Lilly Pijnenburg Muller is a Research Associate in the War Studies Department at King's College London. She holds a non-resident fellowship at the Tech Policy Institute at Cornell University and the Norwegian Institute of International Affairs (NUPI). She is an interdisciplinary researcher in Critical Security Studies and Science and Technology Studies (STS) with interests in technology, the politics of (in)security, and power. During her Fulbright postdoctoral Fellowship in the Science and Technology Studies department at Cornell University (2022–2023), she co-edited this H-Diplo|RJISSF forum on cybersecurity expertise. Prior to joining Cornell, Lilly held research positions at the Oxford Martin School at the University of Oxford and the Norwegian Institute of International Affairs (NUPI). She received a PhD in War Studies from King's College London.

Andrew Dwyer is a Lecturer in Information Security at Royal Holloway, University of London. His interests lie at the intersection of understanding decisionmaking as it is mediated through computation, the role of cyber operations and capabilities, as well as "critical" approaches to the study of cybersecurity. Beyond Royal Holloway, he is the lead of the UK Offensive Cyber Working Group and has previously held research positions at Bristol and Durham universities after completing his DPhil at the University of Oxford in 2019.

Ryan Ellis is an Associate Professor of Communication Studies at Northeastern University. Ryan's research and teaching focuses on topics related to communication law and policy, infrastructure politics, and cybersecurity. He is the author of *Letters, Power Lines, and Other Dangerous Things: The Politics of Infrastructure Security* (MIT Press, 2020) and the editor (with Vivek Mohan) of *Rewired: Cybersecurity Governance* (Wiley, 2019). Prior to joining Northeastern, Ryan held fellowships at the Harvard Kennedy School's Belfer Center for Science and International Affairs and at Stanford University's Center for International Security and Cooperation (CISAC). He received a PhD in Communication from the University of California, San Diego. He is currently working on a book about hackers, precarious work, and bugs for MIT Press.

Aaron Gluck-Thaler is a PhD candidate in the Department of the History of Science at Harvard University, and an affiliate of the Berkman Klein Center for Internet & Society. Aaron studies the history of surveillance and its relationship to scientific practice. He is the 2023–2024 IEEE Life Members' Fellow in the History of Electrical and Computing Technology.

Matt Goerzen is a student in the History of Science department at Harvard University. This work compiles some loose research from the report "Wearing Many Hats," co-authored with Gabriella Coleman and published by Data & Society Research Institute with their financial support. Goerzen is interested in the history of computer security, with a particular emphasis on alternative historical imaginaries of computer security.

Jon R. Lindsay is an Associate Professor at the School of Cybersecurity and Privacy and the Sam Nunn School of International Affairs at the Georgia Institute of Technology. He is the author of *Information Technology and Military Affairs* (Cornell, 2020) and coauthor of *Elements of Deterrence: Strategy, Technology, and*

Complexity in Global Politics (Oxford, 2024). His latest book project is *Age of Deception: Cybersecurity and Secret Statecraft.*

James Shires is a Senior Research Fellow in Cyber Policy at Chatham House. He is a co-founder and trustee of the European Cyber Conflict Research Initiative (ECCRI), and a non-resident associate fellow with The Hague Program for International Cybersecurity. He speaks regularly and has published extensively on cybersecurity and global politics, including *The Politics of Cybersecurity in the Middle East* (Hurst/Oxford University Press, 2021).

Max Smeets is a Senior Researcher at the Center for Security Studies (CSS) at ETH Zurich, and Director of the European Cyber Conflict Research Initiative. He is the author of *No Shortcuts: Why States Struggle to Develop a Military Cyber-Force* (Oxford University Press & Hurst, 2022) and co-editor of *Deter, Disrupt or Deceive? Assessing Cyber Conflict as an Intelligence Contest* (Georgetown University Press, 2023), with Robert Chesney. He has published widely on cyber statecraft, strategy, and risk.

Jesse Sowell, PhD, is a Lecturer in Internet Governance and Policy at University College London's Department of Science, Technology, Engineering, and Public Policy (STEaPP), focusing on the operational institutions that ensure the Internet stays glued together in a secure and stable way. His work focuses on how governance and authority is constructed within these communities and institutions, how that authority differs from authority in conventional state-based institutions, and how to develop institutional and policy interfaces that can help bridge the gaps between the two. Prior to joining UCL, Jesse held positions as an Assistant Professor of International Affairs at Texas A&M University and as a Postdoctoral Cybersecurity Fellow at Stanford. Jesse holds a PhD in Technology, Management, and Policy from MIT.

Clare Stevens is a Teaching Fellow in International Security at the University of Portsmouth. Her research has looked at the controversies, politics and boundary work of defining "cybersecurity," including what it can teach us more broadly about security, secrecy and technologies in contemporary international security. She has recently co-authored a piece in *International Political Sociology* entitled "What Can a Critical Cybersecurity Do?" and an article piece on the contested politics of private cybersecurity expertise in *Contemporary Security Policy*, entitled "Assembling Cybersecurity."

"Collective Resistance in the Digital Domain: The Cyber Partisans as an Exemplar" by Max Smeets, ETH Zurich, Center for Security Studies

The Cyber Partisans, a Belarusian hacking group formed in September 2020, has claimed responsibility for several high-profile cyber operations, including an attack against the Belarusian railway system that reportedly halted Russian ground artillery and troop movement into Ukraine and allowed the group to access the complete database with personal information of those crossing the Belarusian borders.[1]

It might be tempting to describe the Cyber Partisans as a "cyber proxy,"[2] "mercenary,"[3] "semi-state actor" group,[4] or "intermediary."[5] But the Cyber Partisans do not fit these labels. The Cyber Partisans do not act as an intermediary for another government's interests, and have a history of independent operations against the government of Belarus. As it is a small group of closely linked individuals with a strong connection to Belarus, the Cyber Partisans also differ from other non-governmental "hacktivist" efforts, such as Anonymous.[6]

Instead, the Cyber Partisans are more akin to a digital resistance movement—a concept that is not yet well-described in the literature on cyber politics. A resistance movement is commonly defined 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."[7] The Cyber Partisans are specifically organized to resist—and ultimately overthrow—the Lukashenko regime in Belarus through the use of digital means. The Cyber Partisans' use of digital violence is what Michael Lipsky calls a "strategically deployed resource"—and not, for instance, a spontaneous eruption of hacktivist rage.[8] To this end, the group also works together with other resistance organizations that do apply kinetic force. We have not seen a violent digital resistance movement like this to date.

The activities of the Cyber Partisans highlight the need to broaden our theoretical perspectives in the study of non-state actors in cyberspace, beyond principle-agent models, mercantile analogies, or institutional design theories on delegation and orchestration. Cyber scholars should engage more with research on social movements, resource mobilization, and collective resistance.

[1] This essay draws on my previously published work. See Max Smeets and Brita Achberger, "Cyber Hactivists Are Busy Undermining Putin's Invasion," Monkey Cage, *Washington Post*, 13 May, 2022, https://www.washingtonpost.com/politics/2022/05/13/cyber-attack-hack-russia-putin-ukraine-belarus/. Also see Bryan Pietsch, "Hacking Group Claims Control of Belarusian Railroads in Move to "Disrupt" Russian Troops Heading near Ukraine," *Washington Post*, 25 January 2022, https://www.washingtonpost.com/world/2022/01/25/belarus-railway-hacktivist-russia-ukraine-cyberattack/.

[2] Erica Borghard and Shawn Lonegran, "Can States Calculate the Risks of Using Cyber Proxies?" *Orbis* 60:3 (2016): 396; Jamie Collier, "Proxy Actors in the Cyber Domain: Implications for State Strategy," *St. Antony's International Review* 13:1 (2017): 25-47.

[3] Tim Maurer, *Cyber Mercenaries: The State, Hackers, and Power*, (Cambridge: Cambridge University Press: 2017).

[4] Florian J. Egloff, *Cybersecurity and the Age of Privateering* (Oxford: Oxford University Press: 2022).

[5] Max Smeets, No Shortcuts: Why States Struggle to Develop a Military Cyber-Force (New York: Oxford University Press: 2022), 157-160.

[6] Gabriella Coleman, Hacker, Hoaxer, Whistleblower, Spy: The Many Faces of Anonymous (London: Verso, 2014).

[7] U.S. Joint Staff, "Joint Publication 1-02: Department of Defense Dictionary of Military and Associated Terms," Washington, D.C.: Government Printing Office, (November 2010, as amended through 15 June 2015), 206

[8] Michael Lipsky, "Protest as a Political Resource," *The American Political Science Review*, 62:4 (1968): 1144-1158.

Formation

The Cyber Partisans were formed in September 2020, following the elections in Belarus and the ensuing protests and brutal, repressive crackdown by the Lukashenko regime earlier in the year.[9] The group is one of three arms of the collective Suprativ, a larger resistance movement opposing the government.[10] The Suprativ movement includes two other groups: the Flying Storks, an activist network executing also kinetic operations;[11] and the PSS: People's Self-Defense Squad.[12] The latter group provides training for vigilante action against the government. This, for example, includes online videos on how to free yourself from zip ties when captured, how to make smoke bombs, or countering protest crackdown tactics by riot police.[13]

The Cyber Partisans' membership is said to include former IT sector professionals "who learned everything on the go. They're not hackers, none of them were hackers at any point."[14] The group has said on several occasions that it does not receive funding or support from Western governments, and reported that "most" of its members are Belarusian citizens located in Belarus, stressing its grass-roots origins and physical presence as protesters in Belarus.[15]

Unlike most other hacktivist groups, the Cyber Partisans have designated an official spokesperson.[16] Based in New York City, Yuliana Shemetovets gives interviews to explain the rationale behind their operations and the group's broader campaign goals.[17] She says that she does not know the identities of the Cyber Partisans but is given instructions through encrypted messaging.[18] "I don't know who they are, and I don't want to know," Shemetovets says. "Even if someone gets access to my phone…they are not going to find anything that can reveal any sensitive information."[19]

Numerous hacker groups have falsely claimed they were not affiliated to a government. Thus, we must ask: are the Cyber Partisans truly independent and not operating on behalf of a state? There is no definitive

[9] Whilst there is much public information to analyze about the Cyber Partisans' operations and organizational structure, there are still many unknowns and open questions. Not least, there is little information about where the members of the group outside of Belarus reside.

[10] The Belarusian regime proclaimed the group to be a "terrorist" movement. Šarūnas Černiauskas, "Belarus Hackers Declared Terrorists After Exposing Dubious Donation to Regime," Organized Crime and Corruption Reporting Project, 1 December 2021, https://www.occrp.org/en/daily/15590-belarus-hackers-declared-terrorists-after-exposing-dubious-donation-to-regime.

[11] busly_laciac, Telegram, https://t.me/busly_laciac.

[12] dns_main, Telegram, https://t.me/dns_main

[13] Буслы Ляцяць, "Совет от ДНС: Как освободиться от стяжки," 17 September 2021, https://www.youtube.com/watch?app=desktop&v=B9JAgT5gRzs; Буслы Ляцяць, "ДНС: КАК ПРОТИВОСТОЯТЬ ТАКТИКЕ КАРАТЕЛЕЙ," 22 December 2021, https://www.youtube.com/watch?v=s5KK2j_4iLI; dns_main, Telegram, 14 June 2021, https://t.me/dns_main/137.

[14] Ylenia Gostoli, "How I Became the Spokesperson for a Secretive Belarusian "Hacktivist" Group," *TRT World*, 10 February 2022, https://www.trtworld.com/magazine/how-i-became-the-spokesperson-for-a-secretive-belarusian-hacktivist-group-54617 .

[15]Радио 97, "ЭКСКЛЮЗИВ! Дмитрий Щигельский и Юлиана Шеметовец про движение Супраціў и Кибер-Партизан," 2 April 2022, https://www.youtube.com/watch?v=sqrXYvzGVz4.

[16] The Cyber Partisans have set up various communication channels to showcase their operational successes. They actively post and produce content for their official Cyber Partisans channels on Telegram, Twitter, and YouTube.

[17] Gabriella Coleman, "Cyber Partisans: An Insider's Interview on Truth, Terror, and Technology in the Lukashenko Regime," unknown date in 2021, hack_curio, https://hackcur.io/cyber-partisans-an-insiders-interview-on-truth-terror-and-technology-in-the-lukashenko-regime/.

[18] Gostoli, "How I Became the Spokesperson for a Secretive Belarusian "Hacktivist" Group."

[19] Gostoli, "How I Became the Spokesperson for a Secretive Belarusian "Hacktivist" Group."

evidence that the hacking collective is independent of state sponsorship. However, as Juan Andres Guerrero-Saade points out, the way in which the Cyber Partisans operate suggests they are an independent effort: "Most importantly, their limitations and tasking appear organic. They claim that in order to discover important government targets, they collaborate with a union of current and former Belarusian security officers (BYPOL) better acquainted with the inner workings of the government."[20]

This suggests that the Cyber Partisans is not a large, loosely connected group of international hackers, but rather a small, trusted group of individuals with a strong connection to Belarus. Current membership is said to be around 30 members.[21] According to the spokesperson of the group, four individuals are responsible for "ethical hacking" while the others provide support, analysis, and data processing.

The Cyber Activity of the Cyber Partisans

The Cyber Partisans have conducted a wide set of operations against the Belarusian regime since September 2020. They maintain a list of operations on the website of the Suprativ collective.[22] Early operational activity of the Cyber Partisans included Distributed Denial of Service (DDoS) attacks against government websites.[23] They also reportedly added Belarusian president Aleksandr Lukashenko's name to the Ministry of Internal Affairs' "Most Wanted" list.[24] The group later expanded their and disruptive as well as doxing activities, which involve the act of publicly releasing personal information.

The two largest sets of coordinated activities conducted by the Cyber Partisans are Operation Scorching Heat and Operation Inferno.[25] As part of Scorching Heat, the Cyber Partisans released the passport details of millions of Belarusians, obtained by hacking the government "Passport System" and traffic police database. According to the group, "The database contains all people who have a passport, residence permit or similar documents. We can't say for sure, because sometimes a person has several documents. But in a separate sample, we saw more than 11 million personal numbers."[26] Scorching Heat also released other government databases, including the 102 ambulance emergency call logs, the database of violations of the Department of

[20] Juan Andres Guerrero-Saade, "Hacktivism and State-Sponsored Knock-Offs | Attributing Deceptive Hack-and-Leak Operations," Sentinel Labs, 27 January 2022, (2022, January 27), https://www.sentinelone.com/labs/hacktivism-and-state-sponsored-knock-offs-attributing-deceptive-hack-and-leak-operations/; see also Andy Greenberg, "Why the Belarus Railways Hack Marks a First for Ransomware," *Wired*, 25 January 2022, https://www.wired.com/story/belarus-railways-ransomware-hack-cyber-partisans/.

[21] Gostoli, "How I Became the Spokesperson for a Secretive Belarusian "Hacktivist" Group."

[22] Suprativ, "Suprativ's Operations and Projects," 24 August 2021, https://telegra.ph/Suprativs-Operations-08-24.

[23] Cpartisans, Telegram, 26 October 202, https://t.me/cpartisans/48.

[24] Nexta Live, Telegram, 3 September 2020), https://t.me/nexta_live/10437.

[25] They can better be described as "campaigns" rather than "operations." Also see: Андрей Сошников, "Противостоящие Лукашенко "Киберпартизаны" получили паспортные данные и фото ВСЕХ белорусов. Фактчек Настоящего Времени и интервью с хакерами," Current Time TV, 30 July 2021, https://www.currenttime.tv/a/hakery-vzlomali-pasporta/31385554.html; "Cyber Partisans Hack Police Recordings with Brutal Orders," Belsat, 5 August 2021, https://belsat.eu/en/news/15-08-2021-cyber-partisans-hack-police-recordings-with-brutal-orders; On the concept of campaigns versus operations in the cyber context see: Richard Harknett and Max Smeets, "Cyber Campaigns and Strategic Outcomes," *Journal of Strategic Studies*, 45:4 (2022): 534-567.

[26] Андрей Сошников, "Противостоящие Лукашенко "Киберпартизаны" получили паспортные данные и фото ВСЕХ белорусов. Фактчек Настоящего Времени и интервью с хакерами," Current Time TV, 30 July 2021, https://www.currenttime.tv/a/hakery-vzlomali-pasporta/31385554.html.

Internal Security of the Interior Ministry, the video database of the Interior Ministry drones, and the video surveillance system of the detention center, among others.[27]

The second campaign, Inferno, ran from November to December 2021. It included three operations against organizations with ties to the Lukashenko regime. The first operation encrypted the workstations, databases, and servers of the Belarusian Academy of Public Administration. The second operation targeted Belaruskali, one of the largest state-owned companies producing potash fertilizers, and the third operation was aimed at the Mogilevtransmash, one of the largest vehicle manufacturing company in Belarus.[28]

The hack that has gained the most international attention, however, was not part of these two campaigns. In that attack, which targeted the Belarusian railway, the Cyber Partisans claimed to have put the train traffic control systems in the Belarusian cities of Minsk and Orsha into a "manual control" mode that would "significantly slow down the movement of trains" without creating "emergency situations."[29] The spokesperson explained that the attack aimed "to indirectly slow down Russian troops on the territory of Belarus, and show [that its] strategically most important infrastructure is overlooked by Lukashenko." Additionally, "Belarus is at the centre of Europe and a lot of other countries are using these systems. […] It's to show that Lukashenko is not only not safe for the people of Belarus, but also for its neighbours."[30]

The Cyber Partisans have also provided technical support to the Belarusian resistance movement. For example, during the 2020 protests against Lukashenko, the group shared three links to proxy servers via their Telegram account to the protesters marching in the streets.[31] In addition, the group developed new tools to provide secure channels of communication. They announced an encrypted SMS application to allow protesters to communicate securely, without an internet connection, and the development of a secure Telegram application (Partisan Telegram).[32]

The Cyber Partisans have developed a victory plan, called Momentum X, which consists of two phases. The first, Moment X, involves the launch of multiple actions aimed at eliminating the regime of Lukashenko. As stated on the group's website, "it is the beginning of an indefinite protest up to the moment of victory. The exact date will not be known until Moment X, which is set according to the necessary degree of readiness of the partisan organizations and the entire protesting community."[33] Second, there is Phase X. This is "a period of time during which Moment X can be declared at any point. The beginning of Phase X will be announced in advance."[34] During this phase, the group will also release their X-App to help paralyze the internal

[27] "Cyber Partisans Hack Police Recordings with Brutal Orders," Belsat, 15 August 2021, https://belsat.eu/en/news/15-08-2021-cyber-partisans-hack-police-recordings-with-brutal-orders.

[28] https://telegra.ph/Suprativs-Operations-08-24.

[29] Cpartisans, telegram, (2022, February 27), https://t.me/cpartisans/702; Reuters could not confirm the attacks against the railway's traffic system. It did note that the company's reservation website was down on Tuesday afternoon. Joel Schectman, Christopher Bing and James Pearson, "Ukrainian Cyber Resistance Group Targets Russian Power Grid, Railways," Reuters, 1 March 2022, https://www.reuters.com/technology/ukrainian-cyber-resistance-group-targets-russian-power-grid-railways-2022-03-01/.

[30] Gostoli, "How I Became the Spokesperson for a Secretive Belarusian "Hacktivist" Group."

[31] Cpartisans, telegram, 25 October 2020, https://t.me/cpartisans/43.

[32] Cpartisans, telegram, https://t.me/cpartisans_security.

[33] Suprativ, "Cyber-Partisan's Victory Plan," 24 August 2021, https://telegra.ph/Cyber-Partisans-victory-plan-08-24.

[34] Suprativ, "Cyber-Partisan's Victory Plan."

networks of the regime. It seeks to mobilize the community against "vulnerable points" of the regime through another application called the Vulnerability Points Map.[35]

Collaborations

The Cyber Partisans stress that their actions and collaborations are strictly intended to produce operative effects on Belarusian territory and infrastructure only. Thus, while the Ukrainian government has called on volunteers to join their IT Army in the fight against Russia, the Cyber Partisans do not participate in the IT Army's activities or execute operations outside of Belarus's borders.[36] The group is, however, willing to share best practices about the targeting of Russian forces.[37]

The Cyber Partisans owes much of its successful targeting, which is often a major issue for many other non-state groups, to its partnership with an organization of former Belarusian government officials, BYPOL. Launched in October 2020, BYPOL "unites hundreds of incumbent and former security officers looking to restore the rule of law and order in Belarus."[38] Their stated goal is democratic rule in Belarus, entailing new presidential and parliamentary elections, led by Svetlana Tikhanovskaya, the opposition presidential candidate in 2020. "The Cyber Partisans wrote to us to help them find a way to understand all the law enforcement and intelligence agencies," Aliaksandr Azarau, a former lieutenant colonel in Belarus's police force now working for BYPOL says. "They wanted to know how to penetrate inside these organizations to steal information. Because we work there, we know everything inside. We consulted with them on how to do this."[39] In exchange, BYPOL receives access to data from the Cyber Partisans to aid their investigations into the regime, which are subsequently published on BYPOL's Telegram channel.[40]

The group also frequently collaborates on projects with other arms of the Suprativ collective, a larger resistance movement opposing the government of Belarus. They help the Flying Storks, an activist network which also executes kinetic operations, by curating the Belarus Black Map, a database and comprehensive search system of identities and physical addresses of government officials, KGB officers, and anti-riot police to assist the work of opposition organizations.[41] This database can be searched to find the group's doxed profiles.

Furthermore, the Cyber Partisans have worked with news agencies and other journalism groups. It collaborated with CurrentTimeTV on the reporting of COVID-19 infections and deaths numbers for

[35] Suprativ, "Cyber-Partisan's Victory Plan."

[36] Burges, "Ukraine's Volunteer 'IT Army' Is Hacking in Uncharted Territory;" see also Stefan Soesanto, "The IT Army of Ukraine Structure, Tasking, and Ecosystem," *Cyberdefense Report*, June 2022, Center for Security Studies, https://css.ethz.ch/content/dam/ethz/special-interest/gess/cis/center-for-securities-studies/pdfs/Cyber-Reports-2022-06-IT-Army-of-Ukraine.pdf.

[37] Alex Pasternack, "How Hackers in Belarus Are Complicating Putin's Ukraine Invasion," Fast Company, 14 March 2022, https://www.fastcompany.com/90730789/cyber-partisans-hacking-belarus-putin-ukraine-invasion; Schectman, Bing, and Pearson, "Ukrainian Cyber Resistance Group Targets Russian Power Grid, Railways."

[38] "A Union of Belarusian Security Officers," BYPOL, unknown date, https://bypol.org/en/#rec316330668.

[39] Patrick Howell O'Neill, "Hackers are Trying to Topple Belarus's Dictator, with Help from the Inside," *MIT Technology Review* 26 Augusat 2021, https://www.technologyreview.com/2021/08/26/1033205/belarus-cyber-partisans-lukashenko-hack-opposition/.

[40] O'Neill, "Hackers Are Trying to Topple Belarus's Dictator."

[41] Flying Storks, "Belarus Black Map," unknown date, https://blackmap.org/.

Belarus.[42] Bellingcat has also used information released by the Cyber Partisans for its report on Wagnergate (an attempted Ukrainian sting operation)[43] and the uncovering of a Russian spy in Italy.[44]

Difficult Choices to Be Made

The Cyber Partisans are faced with a series of choices about their modus operandi. I discuss five choices here, which relate to location, scale, communication, engagement with international politics, and targeting.

Location: To ensure members of the Cyber Partisans are free from danger or threat, the most obvious policy would be for them to live outside of Belarus or Russia. This would significantly reduce the chances of their being found and detained. At the same time, the operations of the Cyber Partisans show that there is an inherent nexus between the conventional domain and cyberspace. It greatly helps a hacking group's operational effectiveness if people can provide physical access to the systems. For example, a hacking group can benefit from the work of insiders, who, for instance, can plug in a USB-stick to a device to spread malware. An insider might also be able to share information about the target environment—for example, telling the hacker group what type of (outdated) software is running on the workstations.

Scale: To scale up the operational efforts it would make sense for the Cyber Partisans to grow the group's membership (assuming it is true they only have a handful of operators and about 30 members). At the same time, bringing in more hackers makes operational security more challenging and introduces other security risks, such as insider threats that could pass on confidential information.

Communication: Collective resistance requires domestic and international support. Support inherently relies on effective communication. Having a spokesperson who can be interviewed, attend conferences, and speak at roundtables helps the Cyber Partisans in a number of ways. It not only makes it easier to engage with the group, it draws out the human element—and potentially human sacrifice - of their activity. Yet, it also leads to new vulnerabilities—not least to the relatives and friends of those who do not remain anonymous.

Engagement with international politics: The 2022 further invasion of Ukraine by Russia meant that the Cyber Partisans needed to be cognizant of the international political dynamics. On the one hand, the invasion has led to a growing international attention and interest in the region, including Lukashenko's close relationship with Russian President Vladimir Putin. This has helped the cause of the Cyber Partisans: they are part of a larger fight against the evils of authoritarianism. On the other hand, it has led the group's actions to be often folded into the larger events of Ukraine and misrepresented—as mischaracterizations of Cyber Partisans as a Ukrainian cyber proxy exemplifies.

Targeting: Conducting cyber attacks can help raise awareness of the Cyber Partisans' cause and existence. Yet, the more significant type of disruptive or destructive operations are hard to pull off and hardly weaken the

[42] Андрей Сошников, "Избыточная смертность—32 тысячи человек. Власти Беларуси многократно занижают статистику во время эпидемии коронавируса—данные утечки," Current Time TV, 9 August 2021, https://www.currenttime.tv/a/smertnost-v-belarusi/31401342.html.

[43] "Inside Wagnergate: Ukraine's Brazen Sting Operation to Snare Russian Mercenaries," Bellingcat, 17 November 2021, https://www.bellingcat.com/news/uk-and-europe/2021/11/17/inside-wagnergate-ukraines-brazen-sting-operation-to-snare-russian-mercenaries/.

[44] Christo Grozev, "Socialite, Widow, Jeweller, Spy: How a GRU Agent Charmed Her Way into NATO Circles in Italy," Bellingcat, 25 August 2022, https://www.bellingcat.com/news/2022/08/25/socialite-widow-jeweller-spy-how-a-gru-agent-charmed-her-way-into-nato-circles-in-italy/.

legitimacy of the targeted institution. From this perspective, doxing can be a more appealing option. Doxing is often easier to pull off - especially if it is combined with physical access—and it turns the spotlight on the victim organization, rather than attacking entity. In the case of the Cyber Partisans, the internal corruption that is revealed through the disclosure of the documents can help to erode support for the regime in Belarus.

Any resistance movement faces difficult organizational and operational choices. A fruitful area for further research is to draw on existing work on resource mobilization and more systematically compare how digital resistance movements differ from conventional resistance movements in their (violent) tactics, and address its implications.[45] It seems to be much easier to hit targets remotely when operating in cyberspace, but harder to cause major disruptions. Furthermore, cyber operations can more easily be misattributed to other groups, especially if they use common tactics, techniques, and procedures (TTPs).

It is likely also harder to attract and train talented individuals for cyber operations.[46] Yet, it is not clear whether digital resistance movements therefore have to organize themselves differently compared to conventional resistance movements or whether this means that conventional resistance movements are thus more or less likely to start adopting digital forms of violence. Existing scholarship on the macrostructural situations of social movements potentially provides a useful starting point for future research to address these questions and analyze the *conditions* in which digital resistance can emerge and be sustained.[47]

[45] See for example: John D. McCarthy and Mayer N. Zald, *The Trend of Social Movements in America: Professionalization and Resource Mobilization* (Morriston: General Learning Press: 1973); Charles Tilly, *From Mobilization to Revolution*, (Reading, MA: Addison-Wesley: 1978).

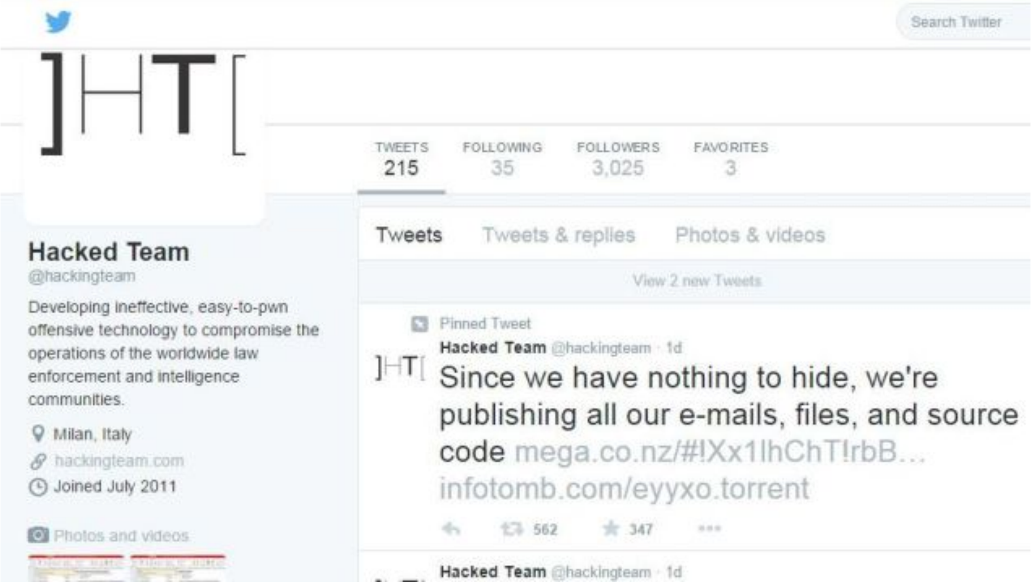
[46] See for example: Josh Lospinso, "Fish out of the Water: How the Military Is an Impossible Place for Hackers, And What To Do About It", War on the Rocks, 12 July 2018, https://warontherocks.com/2018/07/fish-out-of-water-how-the-military-is-an-impossible-place-for-hackers-and-what-to-do-about-it.

[47] Alberto Melucci, "An End to Social Movements? Introductory Paper to the Sessions on "New Social Movements and Change in Organizational Forms," *Social Science Information*, 24:4/5 (1984): 819-835.

"Visions of (In)Security: Anti-Security, Project Mayhem, and Unruly Expertise"
by Matt Goerzen, Harvard University

Introduction[1]

On 5 June 2015, the Twitter account of the cybersecurity company Hacking Team tweeted a surprise announcement: "Since we have nothing to hide, we're publishing all our e-mails, files, and source code." A link associated with the tweet provided access to virtually all of the company's corporate data, which largely confirmed suspicions that Hacking Team was supplying hacking and spyware tools to repressive regimes in Sudan, Saudi Arabia, Bahrain, and elsewhere.[2] Furthermore, the group's Twitter handle had changed from Hack*ing* Team to Hack*ed* Team. Hacking Team had been hacked.



Hacked Team
@hackingteam

Developing ineffective, easy-to-pwn offensive technology to compromise the operations of the worldwide law enforcement and intelligence communities.

Milan, Italy
hackingteam.com
Joined July 2011

Hacked Team @hackingteam · 1d

Since we have nothing to hide, we're publishing all our e-mails, files, and source code mega.co.nz/#!Xx1lhChT!rbB… infotomb.com/eyyxo.torrent

Hacked Team @hackingteam · 1d

Figure 1. Screenshot of Hacking Team's hacked twitter account. Note the updated description on the left.

They were not alone. Since 2011, anonymous hackers have targeted multiple companies and organizations suspected of supporting state oppression or otherwise curbing the possibility of political dissent. These hacks have often complemented the work of civil society groups who are interested in a broader notion of

[1] Thanks to Rebecca Slayton and Lilly Pijnenburg Muller for their generous editorial guidance. I would also like to acknowledge the support and feedback of everyone involved in the "Cybersecurity Expertise: Practice, Performance, Power" workshop held at Cornell University in June 2022, and Harvard University's "From One to Zero" workshop group. Thanks to Gabriella Coleman, Brian Friedberg, Sophia Reuss, and Stephen Watt for their feedback on later drafts. This essay draws on archival materials and oral histories gathered with the financial support of the Data & Society Research Institute.

[2] Lorenzo Franceschi-Bicchierai, "Hacker 'Phineas Fisher' Speaks on Camera for the First Time—Through a Puppet," *VICE Motherboard,* 20 July 2016, https://www.vice.com/en/article/78kwke/hacker-phineas-fisher-hacking-team-puppet; Lorenzo Franceschi-Bicchierai, "Hacking Team Has Lost Its License to Export Spyware," *VICE Motherboard,* 6 April 2016, https://www.vice.com/en/article/78k8dq/hacking-team-has-lost-its-license-to-export-spyware.

computer security, one aligned more with what the United Nations (UN) has called "human security"—an approach aimed at securing people's basic rights to food and political autonomy.[3]

Yet many of these hackers promote the notion of "anti-security," or "antisec." What does it mean to be "anti-security?" How do these hackers envision security in relation to the companies and nations that claim to provide it?

This essay traces the emergence of anti-security to structural changes near the turn of the millennium. With the mainstreaming of the World Wide Web, the growing viability of e-commerce, and increasing government commitments to secure network-connected infrastructure, some hackers began to decry the rise of what they called the "security industry." The specter of the security industry was invoked in Internet Relay Chat (IRC) channels, conference talks, hacker journals, and electronic textfiles—the underground's favored method of circulating small bundles of information.[4] The "security industry" referred loosely to a growing assortment of government-sponsored research shops, auditing firms, security service and tool vendors, consultancies, and an array of boutique start-ups run by ex-military and "white hat" hackers. Some hackers became convinced that this industry was exploiting the underground, both by villainizing hackers to justify their services, and by extracting hackers' hard-won knowledge about computer vulnerabilities. Industry activity led to the patching of vulnerabilities, shutting down the access that was prized by "black hat" hackers. Meanwhile, motivated by the prospect of legal safe harbor for research, mainstream legitimacy, and steady paychecks, white hat hackers were increasingly stepping up as this security industry's rank-and-file workers.

Amid speculation that the underground was dying,[5] an assortment of hackers pushed back, seeking to rebuke the security industry, punish "sell-outs" and "hacker pimps," and preserve the underground scene in an idealized state. In 2001, hackers launched the "Anti-Security Movement" as a concerted effort to denounce the full disclosure practices that had brought hackers and mainstream security figures together.[6] The following year, hackers initiated "Project Mayhem," a self-consciously black hat initiative to hack, dox, and otherwise shame and ridicule those white hat hackers and mainstream security experts seen to be contributing to the rise of the hated security industry.[7] Along the way, a handful of other actors and chaos agents entered the fray, offering their take on the industry and its worst excesses. While many of those involved articulated

[3] Ronald J. Deibert, director of the Citizen Lab at the University of Toronto, has explored human security in relation to cybersecurity. See Ronald J. Deibert, "Toward a Human-Centric Approach to Cybersecurity," *Ethics & International Affairs* 32:4 (2018): 411-424.

[4] Hackers favored plaintext .txt files for their small size, compatibility, and ease of dissemination. Textfiles were sometimes shared from a primary website or BBS, and at other times disseminated in a primarily ephemeral or peer-to-peer fashion—through FTPs, mirror sites, IRC file transfer, and, later, websites like pastebin. Many of the textfiles discussed here have been archived at textfiles.com, maintained by Jason Scott.

[5] These anxieties date to at least 1996, when *Phrack* editor Chris Goggans ("Erik Bloodaxe") stepped down from the role, calling the underground "terminally ill." "The community has degenerated. It has become a media-fueled farce. The act of intellectual discovery that hacking once represented has now been replaced by one of greed, self-aggrandization and misplaced post-adolescent angst," he declared. Unlike the hackers I focus on in this paper, this assessment led him to seek out employment in the nascent security industry. Erik Bloodaxe (Chris Goggans), "Phrack Editorial," *Phrack,* 7:48, File 2a of 18 (9 January 1996).

[6] Matt Goerzen and Gabriella Coleman, "Wearing Many Hats: The Rise of the Professional Security Hacker," Data & Society Research Institute, January 2022, https://datasociety.net/library/wearing-many-hats-the-rise-of-the-professional-security-hacker/.

[7] According to a textfile issued by the Phrack High Council (PHC), "A WHITEHAT IS ANYONE WHO HELPS THE SECURITY INDUSTRY (POSTING BUGS/INFO ETC),"
https://web.archive.org/web/20090207110001/http://dsr.segfault.es/stuff/website-mirrors/pHC/old/faq1.txt.

alternative visions for computer security, the escalating rhetorical denigrations and performative stunts of Project Mayhem largely overshadowed those visions—at least in the short term.

This essay will sketch these developments and reflect on their lasting significance. While many hackers ultimately judged both the original Anti-Security Movement and Project Mayhem as failures, they became powerful touchstones for later generations of politically-inclined hacker activists. The concept of anti-security supported a lasting vision of unruly, anti-establishment hacking—a mode of hacker expertise that is firmly autonomous from employment contracts, training certifications, or bug bounty leaderboards.

The Eclectic Origins of Anti-Security

In the early 1990s, prominent computer scientists and industry professionals spoke out against the idea of hiring hackers, arguing that it was like hiring burglars as bank guards, or arsons as fire marshals. Many hackers spent the better part of the 1990s fighting these perceptions.[8] The controversial "full disclosure movement" served as one crucial mechanism. Exemplified by the hacker-founded mailing list Bugtraq, full disclosure facilitated dialogue between members of the underground hacking scene and more mainstream technologists: university systems administrators, hobbyist and professional security researchers, and eventually even representatives of the companies whose products were being hacked. Participants openly shared knowledge of vulnerabilities and, often, even functioning exploit code. They developed a conception of hacking that was less tied to autonomy and freedom of access, and more oriented towards the discovery and documentation of security vulnerabilities as an intellectual and ultimately commercial pursuit. Meanwhile, this research enabled the exploitation of vulnerabilities in a "proof of concept" mode to shame vendors like Microsoft, shifting the burden of insecurity away from the hackers themselves and onto the negligence of corporations.

By the late 1990s, many organizations recognized the hacker underground as an incubator for computer security expertise. Casting aside earlier anxieties, they sought to hire from the underground. Hackers became military consultants, penetration testers on auditing teams, tool developers, and in-house experts at companies that were beginning to take security seriously after a half-decade of hacker-initiated bad press.

In 1999, security researcher Marcus J. Ranum articulated a new, structural reason not to hire experts from the underground: what might it mean for a burgeoning security industry to reward people whose prior activity (on "the dark side") was a chief reason for the security industry to exist in the first place? [9] What sort of perverse incentives might this create?[10]

"Instead of just having the 'bad guys' trying to find and exploit holes in systems, now we have the 'good guys' doing it too, or hiring 'ex-bad guys,' repackaging them as 'good guys,' and selling them to you for $400 an hour," he argued in a 1999 special security-themed issue of Usenix's *;login:* magazine. "When you read about a shocking new vulnerability found in something, research not only the vulnerability but the individual or

[8] Goerzen and Gabriella Coleman, "Wearing Many Hats."

[9] Marcus J. Ranum, "Selling Security: Fear Leads to ... the Dark Side," *;login:*, November 1999, https://static.usenix.org/publications/login/1999-11/features/darkside.html.

[10] Riffing off Chris Kelty's idea of a "recursive public" (more on this in the conclusion), we could conceptualize this as the "perversive private" sector—one that sustains itself through the production of the conditions that lead to perverse incentives. Chris Kelty, *Two Bits: The Cultural Significance of Free Software* (Durham, NC: Duke University Press, 2008).

organization that announced it. Ask yourself if they happen to sell a solution to the problem, and keep your skepticism in gear."[11]

Ranum soon found some strange bedfellows: a number of influential figures in the hacker underground began to amplify and adapt his message under the banner of the "Anti-Security Movement" (or "antisec," for short). In early 2001, the Iceland-based hacking group security.is established a webpage for the movement at anti.security.is.[12] A links page featured Ranum's talks and articles alongside the websites of supporting organizations, including the esteemed underground hacker group Association de Malfaiteurs (ADM).

The site's central manifesto called for an immediate end to the public, full disclosure of computer vulnerabilities. However, this anti-disclosure policy left a lot of room for debate and disagreement. Contrary to what anti-security might suggest, some participants argued anti-disclosure would serve as a net positive for the security of the internet. As the anonymous author of the manifesto put it:

> A digital holocaust occurs each time an exploit appears on Bugtraq, and kids across the world download it and target unprepared system administrators. Quite frankly, the integrity of systems worldwide will be ensured to a much greater extent when exploits are kept private, and not published.[13]

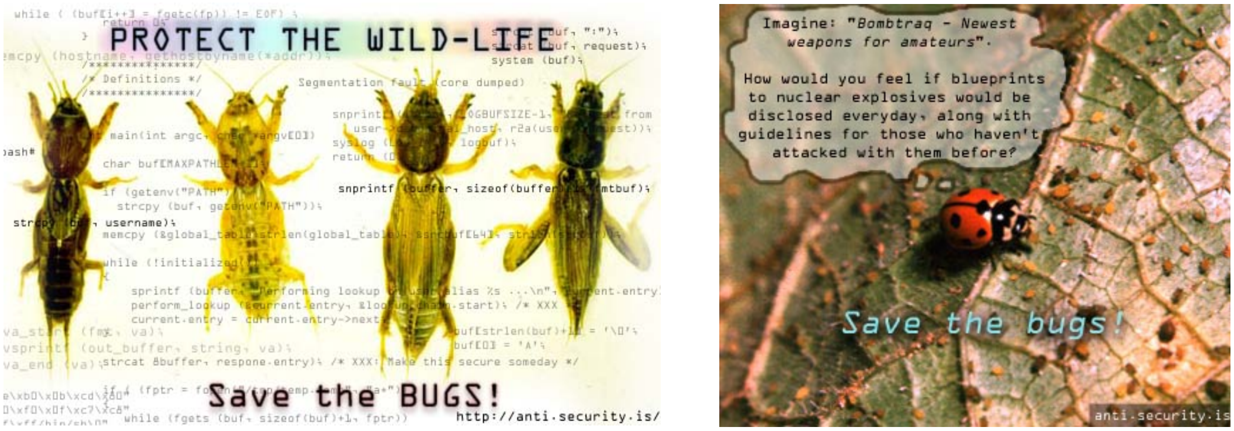

PROTECT THE WILD-LIFE
Save the BUGS!
http://anti.security.is/
Imagine: "Bombtraq - Newest weapons for amateurs".
How would you feel if blueprints to nuclear explosives would be disclosed everyday, along with guidelines for those who haven't attacked with them before?
Save the bugs!
anti.security.is

Figure 2. Two promotional images from the "Gallery" section of the anti.security.is website.

Many observers also noted a less altruistic motive for ending full disclosure: the hardcore underground researchers could continue to exploit the vulnerabilities they uncovered for their own use.[14]

[11] Ranum, "Selling Security."

[12] It is unclear exactly why they chose "anti-security" instead of "anti-disclosure" or "anti-security industry." One possibility is that it was an attempt to appropriate the idea of "anti-security" away from the script kiddies; one prolific website defacement crew operated under the banner of anti-security in the late 90s. Another is that it was chosen for its rhetorical punchiness.

[13] Anonymous, "Intro," Anti Security, 2001, https://web.archive.org/web/20010301215117/http://anti.security.is/.

[14] In a critical mode, user NightAxis parodied the movement's position: "Don't disclose, that way the few of us that have the knowledge to exploit can do so without a) making as big of a mess and b) can move in and out of systems

It quickly became clear that anti-disclosure could serve a range of other agendas, too. The site hosted a growing collection of often-incommensurate statements, alternate manifestos, and FAQs from supporters, and participants debated with one another on the site's message board.

Many critiqued the way full disclosure empowered the security industry's growth, along two dimensions. First, they argued that disclosure armed unskilled, wannabe-hacker "script kiddies" with exploits that would then be used for website defacements and other attention-grabbing stunts—thus justifying the criminalization of hacking and enhancing security industry sales pitches. Second, they argued that the employment prospects enabled by the security industry's growth incentivized hackers to disclose or sell underground knowledge as a form of resume building—effectively diminishing the power of the underground. As one participant broke it down:

> …commercial security services rely on the presence of full disclosure mechanisms to feed script kids in order for them [script kids] to reduce real world security and thus increase their [security industry] sales activity and raise their profits. While that happens, the underground takes the hard shot: hackers are tracked down, bugs are patched, backdoors are discovered, etc.[15]

Some rationalized anti-security as a way to maintain the possibility of meaningful resistance to corporate or state overreach. As one contributor put it:

> Imagine a secure internet...
>
> Content Filters, IDS [intrusion detection systems], censorship, law enforcement, ...
>
> i mean.. that's what security is also about…
>
> Is it really what you want? A secure internet would take away all your means to oppose... It would be the equivalent of a police state.[16]

But other participants were less roundly opposed to the prospect of getting paid to improve internet security. For some, the marketing of security services and anti-disclosure could co-exist. (Indeed, some members of ADM had founded a security firm called Qualys in 1999).[17] For others, limited forms of vulnerability

easier and c) people probably won't patch as frequently as less will be known/actively exploited." NightAxis, "RE: welcome!" Anti Security: General Discussions, 18 February 2001, https://web.archive.org/web/20011226050711/http://anti.security.is/chat/spjall_thradur.php?id=8&bordid=1&efni=General+discussions&msgcnt=8.

[15] ghkjfdhdf, "RE: Question your beliefs," Anti Security: General Discussions, 15 September 2001, https://web.archive.org/web/20011226042608/http://anti.security.is:80/chat/spjall_thradur.php?id=209&bordid=1&efni=General+discussions&msgcnt=8.

[16] justin case, "How do you justify full disclosure?" Anti Security: Texts, 28 January 2001, https://web.archive.org/web/20010425190554/http://anti.security.is/texts.php?file=article1.txt.

[17] That said, it's unclear whether everyone in ADM supported anti-security—some onlookers opined that the Anti-Security Movement proxied an internal power struggle between different factions within ADM. ADM members working for Qualys could participate in the market by selling scanning services for *known* vulnerabilities while simultaneously maintaining a principled opposition to the disclosure of private, 'zero-day' vulnerabilities. But as escalating events would make clear, some anti-security supporters resented other ADM members for selling knowledge as consultants to corporate and military clients.

disclosure were acceptable.[18] They criticized security industry *excesses* and the "pimping out" of "mediawhore" hackers. For these figures, vulnerability disclosure and professional hacking per se were not the problem. Hewing closer to Ranum's position, they instead expressed disgust with the growth of a profit-motivated, financialized security services industry that contributed to and benefited from the dissemination of fear, uncertainty, and doubt ("FUD"). Consider this entry from one of the site's FAQs:

> Q: Who profits from the infosec[19] war?
>
> A: Security companies do (this is their line of employment). They need to use scare tactics to motivate more people and companies into thinking their services are not only desirable, but necessary. It's simple Capitalism. These corporations make security popular and fashionable, and turn it into a consumer pastime. Why can't they carry out their jobs with less glamour?[20]

Still others framed it as a labor or intellectual property issue. Some argued that the exploit code appearing on mailing lists like Bugtraq was often posted without the consent of its authors—sometimes by people passing themselves off as the original author to accrue credit.[21] Some objected in a more general sense, suggesting that full disclosure was tantamount to giving away value. "By defending the virtues of fully disclosing vulnerabilities in public forums, one already contributes to the expansion of corporate empires, without receiving their fair share of the millions of dollars generated by the information technology industry."[22]

And yet others gestured vaguely towards visions of "real world security" or "people's security," where the internet could be made more secure for regular people without an escalation of an "attack/defense context."[23] The possibility of a public-interest approach to security would become a significant theme in subsequent discussions related to anti-security. But it struggled to secure attention in the midst of an escalating series of spectacular developments.

[18] Indeed, some applauded efforts by hackers like Elias Levy ("Aleph One"), Chris Wysopal ("Weld Pond"), and Jeff Forristal ("rain forest puppy") to engage with industry and government to develop standards for "selective" or "coordinated" modes of disclosure. This position might help us make sense of the involvement of underground stalwart ADM; some members were then growing their boutique company Qualys in France. Meanwhile other ADM members—as escalating events would make clear—had entered the security industry in more controversial ways, offering their skills as consultants for corporate and military clients.

[19] Infosec, short for information security, is a common term used by computer security practitioners to refer both to the field of computer security research, in general, and the subset of computer security research focused on controlling access to sensitive documents and communications, in particular.

[20] Anonymous, "Official antiSecurity FAQ," Anti Security: FAQ, 2001, https://web.archive.org/web/20010425132923/http://anti.security.is/FAQ.php?faq=official&lang=gr.

[21] This serves as a fascinating correlate to the dynamics explored by Hugh Gusterson in his chapter "The Death of the Authors of Death", whereby institutional restrictions against the publication of research by nuclear scientists precludes their capacity to accrue credit. Here, the subcultural norms of the underground black hat scene could be understood to have functioned in a similar manner. See Hugh Gusterson, "The Death of the Authors of Death" in Mario Biagioli and Peter Galison, eds. *Scientific Authorship: Credit and Intellectual Property in Science* (2003), 281-309.

[22] Anonymous, "Full Disclosure and Capitalism," Anti Security: Texts, 30 January 2001, https://web.archive.org/web/20010425132923/http://anti.security.is/texts.php?file=antisec.html.

[23] See ghkjfdhdf, "RE: Question your beliefs" for discussion of "attack/defense context" and "real world security"; and anonymous, "Full Disclosure and Capitalism" for discussion of "people's security."

Making Mayhem of Anti-Security

In early summer 2002, the relatively civil discourse found on anti.security.is was overshadowed by a parallel development. A hacker crew calling itself ~el8 began disseminating a textfile announcing "pr0jekt MAYHeM" (Project Mayhem).[24] Promising to "br1ng an end to the security community," the zine's authors posted a list of "missions"—exhortations to pollute the discourse on prominent security forums with bogus disclosures, and to hack both "media lmrz" (attention-seeking, resume-building white hat hackers) and prominent computer security figureheads, such as Purdue University computer scientist Eugene Spafford.[25] Subsequent issues showcased trophies from these hacking escapades—bash histories, home directory contents, and even email exchanges—intermixed with entertaining and outrageous editorial commentary. In one inflammatory example, a controversial ADM member's emails were leaked, suggesting he was avidly consulting for the US military.[26]

An associated group calling itself the Phrack High Council (PHC) appeared as well, targeting the esteemed underground journal *Phrack* for its purported complicity in the security industry's growth.[27] As they expressed it in a pithy diagnostic: "EVERY TECHNIQUE THAT IS RELEASED IN PHRACK IS NOW REALIZED BY THE SECURITY INDUSTRY. THE SEC INDUSTRY NOW SPENDS TIME TO THWART THESE TECHNIQUES."[28] PHC managed to wrest control of #phrack, a major node in the hacker IRC network on EFNET, away from the journal's editors. They then disseminated the unfinished version of *Phrack* 59, complete with spurious additions which made it seem like the journal's staff supported anti-security and were contrite about their supposed role in selling out the underground.

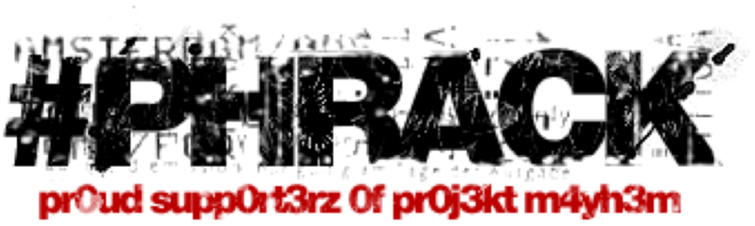


Figure 3. Circa 2002 advertisement for #phrack—under new moderation.

Others called for the boycott or detournement of DEF CON, the flagship annual Las Vegas-based hacker conference. They argued the conference's founders had perverted it into nothing more than a side show for their recently-established big ticket industry event Black Hat Briefings—a way to give Black Hat attendees a

[24] Anonymous, "~el8[2]," *~el8,* 2002, http://web.textfiles.com/ezines/EL8/el8.2.txt.

[25] Anonymous, "~el8[2]."

[26] Anonymous, "~el8[3]," *~el8,* 2002, http://web.textfiles.com/ezines/EL8/el8.3.txt.

[27] A mirror of the Phrack High Council website can be found at: https://web.archive.org/web/20020807201432/http://www.eurocompton.net/~fuk/phrack/index.html.

[28] Anonymous, "FAQ," 2002, https://web.archive.org/web/20090207110001/http://dsr.segfault.es/stuff/website-mirrors/pHC/old/faq1.txt.

glimpse of the hacker threat, and thus spread more industry-serving FUD. "You're participating to [sic] a scam to sell security to corporate black hat attendees. Defcon should be paying *you,"* reads a line on an anonymous flyer, probably produced by a hacker known as "Gweeds."[29]

This "hacker pimping" became a chief concern of Gweeds, who advanced the notion of a "Black Hat Bloc" as a more explicitly politicized correlate to Project Mayhem. Gweeds laid out his project at H2K2, the July 2002 edition of the Hackers on Planet Earth (HOPE) conference. His presentation, titled "Black Hat Bloc or How I Stopped Worrying About Corporations and Learned to Love the Hacker Class War,"[30] advanced an ardently anti-capitalist view of hacker security research. Gweeds contrasted a historical hacker conception of security, premised on open access and privacy, with the version of security he saw as at work in the security industry: "Security of their legitimacy to power, money, and public resources. The security of [corporation's] machines, of [corporation's] right to own the network that was built on public funds. The security to protect you from taking it back." The audience applauded loudly as he called out prominent members of the hacker scene for facilitating this shift. "They're making money, sure. But they're also increasing the reach of the police state at the expense of fellow hackers who will go to jail because of these crimes." Gweeds argued instead that the unabashedly black hat position should be valorized as a means to push back against the overreach of states and corporate entities, and secure the internet as a public good.

The talk was favorably reviewed by *The Register*, an influential tech publication. But the review itself attracted massive backlash.[31] Hackers who were moving into the security industry wrote in to the publication, undermining Gweeds' points and accusing the publication of rewarding anti-social elements of the scene with undeserved attention.[32]

By August 2002, many awaited the drama guaranteed to unfold at DEF CON 10. A much-speculated-about individual or group of known as "GOBBLES Security" was slated to speak. Though positioned outside both the antisec and security industry camps, they performatively pushed back on the respectability-seeking "white hat" trend in hacking. For instance, while GOBBLES discursively supported full disclosure, in early 2002 they hijacked the Anti-Security Movement website, mocking participant's ostensibly anodyne motives by poking fun at the idea that skilled black hat hackers would just "sit on [their] warez" (i.e., engage in research solely as an intellectual pursuit).[33] Moreover, when they did engage in full disclosure it was typically served to deride or humiliate other researchers. In this way, GOBBLES tacitly aligned with many aspects of the anti-security

[29] Anonymous, "TOP TEN REASONS NOT TO GIVE DEFCON YOUR FIFTY BUCKS," https://web.archive.org/web/20120509093625/http://lucifer.phiral.net/blackhatbloc/defcontoptenflyer.html.

See also: Anonymous, "REJECT HACKER EXPLOITATION—FIGHT BACK," https://web.archive.org/web/20120509093524/http://lucifer.phiral.net/blackhatbloc/defconrejectflyer.html.

[30] Gweeds, "Black Hat Bloc or How I Stopped Worrying About Corporations and Learned to Love the Hacker Class War," H2K2, 14 July 2002, https://infocondb.org/con/hope/h2k2/black-hat-bloc-or-how-i-stopped-worrying-about-corporations-and-learned-to-love-the-hacker-class-war.

[31] Thomas C. Greene, "Security Industry's Hacker-Pimping Slammed," *The Register*, 15 July 2002, https://www.theregister.com/2002/07/15/security_industrys_hackerpimping_slammed/.

[32] Greene, "Letters: Gweeds Gets Killed," *The Register,* 16 July 2002, https://www.theregister.com/2002/07/16/gweeds_gets_killed/.

[33] The defaced anti.security.is website was archived and can be found here: https://web.archive.org/web/20020127132025/http://defaced.alldas.de/mirror/2002/01/01/anti.security.is/.

position by performatively insisting on the continued viability of an unruly, non-commoditized form of hacker expertise.[34]

GOBBLES had become known for sophisticated advisories published to mailing lists like Full Disclosure, all wrapped in long-winded, tangential rhetoric that was coded in a cliched Eastern European ESL writing style punctuated with phoneticized turkey sounds. The schtick only made the unusual sophistication of GOBBLES' exploits all the more perplexing. Fascinated by the phenom, respected security researchers offered to foot the bill for a GOBBLES representative to attend DEF CON, according to a *WIRED* article titled "Hacker Humbles Security Experts."[35] Onlookers would make sport of attempting to figure out who, exactly, was behind the moniker for years to come.

The question of how some form of computer security could be pursued outside of the auspices of the security industry remained very much on the agenda. Speaking just before GOBBLES' much hyped talk, security researcher Steve Manzuik ("hellNbak") echoed some of the points Gweeds had made a month earlier during his presentation at H2K2.[36] Notably, Manzuik endorsed the idea of hackers engaging in security research and action to support non-profits and other civil society groups—even suggesting that proactive info-seeking hacking-and-leaking from companies like Enron would be a marked public good.

But the GOBBLES presentation, "The Wolves Among Us," quickly marked a return to spectacle.[37] The GOBBLES representative, "Nwonknu," was joined on stage by Stephen Watt and Silvio Cesare, hackers who had recently been employed with Qualys, the security firm founded by members of ADM.[38] Watt, who penned some of the early Anti-Security Movement texts under the handle "jim_jones," presented himself at DEF CON as "The Unix Terrorist." This name was drawn from the extensive, absurdist—and probably spurious—members roster found in the most recent edition of the ~el8 textfile.[39] The speakers proceeded with a sustained ironic tone, calling out white hat-type hackers in the room and passive-aggressively interrogating those who came up to the stage to defend themselves or attempt to dialogue. The talk later attained legendary status among resolutely underground hackers, and spurred on still more attempts in mailing lists, IRC channels, and conference hallways to figure out just what exactly was going on, who was GOBBLES, and whether the entire Anti-Security Movement was some strange form of proxy infight on the acceptable modes of professionalizing among those in ADM's orbit.

[34] In a 2001 *Phrack* report, GOBBLES also acknowledged that "some GOBBLES researchers are very loyal to anti.security.is philosophy." Phrack Staff, "SIGINT CONFIDENTIAL REPORT ON GOBBLES," *Phrack,* 11:58, File 3 of 15 (28 December 2001), http://phrack.org/issues/58/3.html.

[35] Brian McWilliams, "Hackers Humble Security Experts," *WIRED,* 16 January 2003 [references 2002 events], https://www.wired.com/2003/01/hackers-humble-security-experts/.

[36] hellNbak (Steve Manzuik), "Selling Out for Fun and Profit," DEF CON 10, 4 August 2002, https://infocondb.org/con/def-con/def-con-10/selling-out-for-fun-and-profit.

[37] GOBBLES Security, "Wolves Among Us," DEF CON 10, 4 August 2002, https://infocondb.org/con/def-con/def-con-10/wolves-among-us.

[38] This information comes from an interview with Stephen Watt conducted by me and Gabriella Coleman on 20 July 2019. The details are further substantiated in an unpublished manuscript authored by Watt.

[39] That is to say, it's unclear if Watt was involved with the production of ~el8, or whether he just adopted the name to align himself with the group and court attention. See anonymous, "el8[3]" for the membership roster in question.

The Rhetorical Escalation and Effectual Decline of Anti-Security Mayhem

All the while, Project Mayhem continued to roll on, with new missions and escalating rhetoric. The Anti-Security Movement's playful calls to "Protect the wild-life—Save the bugs!" via anti-disclosure were twisted into calls to "Save a bug, kill a white hat!" PHC spokesperson "gayh1tler" began calling for a "white hat holocaust."[40] In associated IRC channels, websites, and textfiles, casual racism, misogyny, and calls for violence became the rhetorical mode of the day.

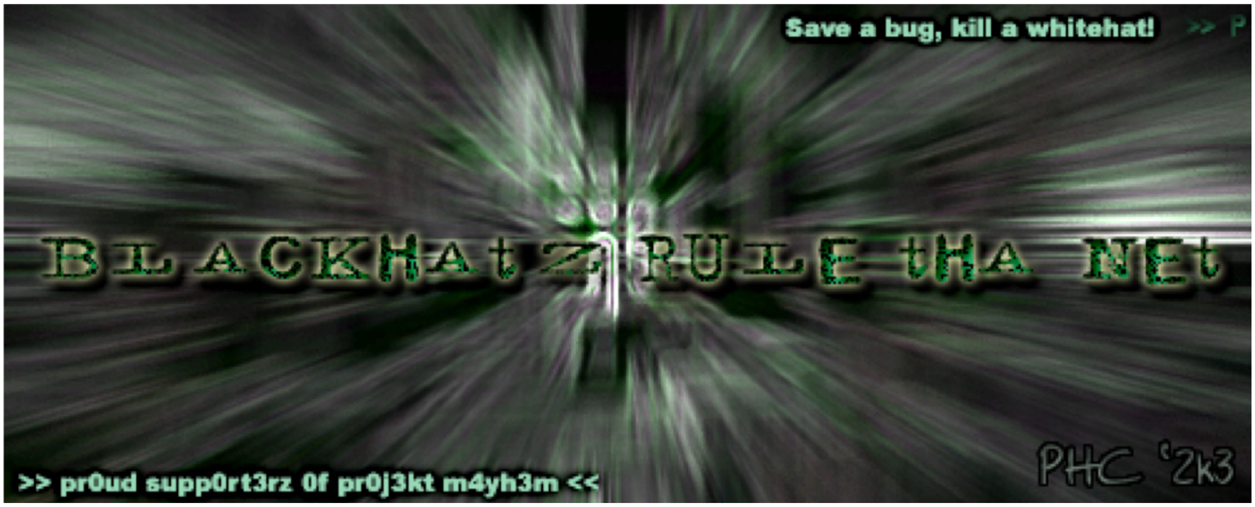

Save a bug, kill a whitehat! >> P
BLACKHAtZ RULE tHA NEt
>> pr0ud supp0rt3rz 0f pr0j3kt m4yh3m <<
PHC '2k3

Figure 4. Latter-day promotional image for Project Mayhem, with escalated rhetoric.

Supporters trolled mailing lists devoted to full disclosure, pointing to the growing list of hacked white hats and security industry figureheads as evidence that the black hat scene possessed greater expertise than the people institutionally tasked with promoting security.[41] Further evidence, they argued, of the "snake oil" being pitched by the blossoming security industry. "Who are the scriptkids now? You're outgunned and outclassed. Take a nap and retire, you pathetic leeches."[42]

Perhaps remarkably, the unfolding drama had until now remained under the radar of the mainstream technology press. This changed August 13, 2002 when *WIRED* published an article entitled "White-Hat Hate Crimes on the Rise." It described Project Mayhem as "a violent incarnation of the 'anti-sec' movement, a

[40] See archived website at: https://web.archive.org/web/20120509093303/http://lucifer.phiral.net/blackhatbloc/phrack/.

[41] Note that not all "white hats"—self-identified or otherwise—necessarily supported full disclosure. In the mid-1990s, full disclosure was seen by many as a necessary evil to motivate the computer industry to take security seriously. In this way, it was regarded by many as a "white hat" pursuit. But by 2001, many of the earlier supporters of full disclosure—including Elias Levy ("Aleph One"), the former moderator of BugTraq—were collaborating with an array of technology stakeholders to develop an effective "selective" or "coordinated" disclosure policy that favored direct disclosure of vulnerabilities to the responsible vendors, preferably with some time-delayed public disclosure to facilitate security research. Likewise, some "script kiddies" who benefited from full disclosure would have seen themselves as "black hats" and opposed the anti-disclosure mechanisms prioritized by the underground "black hat" elite.

[42] Anonymous, "A PHC PRODUCTION: THE REAL SCRIPTKIDDIES," *Full Disclosure,* 16 August 2002, https://seclists.org/fulldisclosure/2002/Aug/482.

campaign to persuade hackers not to publish information about the security bugs they uncover."[43] The performative excesses of Project Mayhem thus entirely eclipsed the critically reflexive strains that had, at least in part, motivated the broader movement.

Many observers roundly dismissed anti-security in moral terms. Some even mused that the whole event was a false flag designed to support the very security industry it ostensibly opposed. One onlooker speculated that "the entire shenanigan was designed to ensure job security for those who fear the economic trends in IT employment."[44] Another observed that "the fear PHC, ~el8 and such groups put into companies is actually helping sec.industry... This helps sell their service very well."[45]

A small group of diehards kept Project Mayhem and PHC alive in one form or another over the next few years.[46] But much of their power to provoke was lost. Indeed, even the prospect of hacking white hats lost its edge as it transformed into an odd honorific; so many skilled hackers had been compromised in one way or another that being targeted had almost become a credential, a badge of significance and expertise.

Looking back on these events in a *Phrack* "prophile" a year later, in 2003, security.is member "Digit" admitted a cynical motive for his involvement in the Anti-Security Movement:

> the true reasons behind antisec were not to create some greater security in the world or something like that which was mentioned in the FAQ and we took a lot of crap for. It was to keep security research where it belongs, with those that actually did it and at most a small tight knit group. That basically meant that people that found bugs, wrote exploits, and hacked wanted to keep their exploits/research private so that they had some nice private warez for some time ;>[47]

Whether all participants—across all the manifestations described above—were also cynically motivated is impossible to say.

Whatever the case, the notion of antisec, the specter of a "hacker class war," and the general notion that the very meaning of "security" could itself be contested, remained significant for a global cast of resolutely underground writers publishing in *Phrack* and other venues for years afterwards. Some saw antisec as an inspiration,[48] while others saw it as a misguided endeavor that had only accelerated the underground's

[43] Brian McWilliams, "White-Hat Hate Crimes on the Rise," *WIRED News,* 13 August 2002, https://web.archive.org/web/20020818095341/http://www.wired.com/news/culture/0,1284,54400,00.html.

[44] Chaos_Magician, "RE: A PHC PRODUCTION: THE REAL SCRIPTKIDDIES," Full Disclosure*,* 16 August 2002,
https://seclists.org/fulldisclosure/2002/Aug/485.

[45] Anonymous, "the sides of security(a 0day post)," Full Disclosure*,* 20 November 2002, https://seclists.org/fulldisclosure/2002/Nov/225.

[46] Notably, some diehards seem to have cross-pollinated with a new generation of hackers operating at the nexus between the underground hacking scene and the trolling subculture emerging on sites like Something Awful and 4chan. See: anonymous, "blackhat for life," wordpress, 2005–2006, https://antisec.wordpress.com/.

[47] Phrack Staff, "PROPHILE ON DIGIT," *Phrack,* 11:61, File 4 of 15 (13 August 2003), http://phrack.org/issues/61/5.html.

[48] See, for example: Anonymous, "The Indian Hacking Scene: Unofficial Memoirs of the Desi h4x0rs," *Phrack*14:67, File 16 of 16 (17 November 2010), http://phrack.org/issues/67/16.html; Anonymous, "Lines in the Sand: Which Side Are You on in the Hacker Class War," *Phrack,* 14:68, File 16 of 19 (14 April 2012), http://phrack.org/issues/68/16.html.

demise.[49] Either way, the specter of antisec ultimately went on to serve as an important touchstone for a next generation of hacker activists—establishing a vision of an unruly hacker expertise operating outside the auspices of a security mainstream.

The Politics and Permanence of Anti-Security

We can consider the underground of this period through the lens of what Christopher Kelty has called a "recursive public"—a public defined by its commitment to maintaining and defending the integrity of the technological tools and infrastructure it requires to exist.[50] While Kelty explored the concept in relation to free and open source software communities, it applies just as well here. In this case, the hacker underground relied on both an active member base and continued—privileged—access to vulnerabilities as the condition of their existence. But it also ultimately benefitted from a sustaining, animating legend—a testament that the underground was not quite dead yet.

Anti-security demonstrated that not everyone who possessed real, practical hacking expertise were primarily out for a paycheck. Many were motivated by values and conceptions of "security" that diverged from those dominant in the corporate and national security arenas. Furthermore, it demonstrated that black hat hacking and the hacker underground could remain viable subcultural enterprises—viable, at least, for anyone willing to accept, look beyond, or attempt to supplant the flippantly macho and toxic dynamics on offer. While hackers continued to warn of the underground's demise, texts generated by the Anti-Security Movement, ~el8, PHC, and the black hat bloc remained in circulation through website mirrors, textfile archives, and periodic porting-over to new web-based platforms. The specter of anti-security persisted, a touchstone frequently invoked by a next generation of hackers attempting to articulate, defend, or re-invigorate an underground sensibility and community both in their own textfiles and in publications like *Phrack*.

The lasting resonance of anti-security is most visible in the hacktivist endeavors that captured headlines throughout the 2010s. Most explicitly, in Anonymous" 2011 "Operation Antisec," which took aim at what participants called the "security intelligence complex"—an updated correlate to the "security industry" of yesteryear. Indeed, some of the targeted companies employed "white hat" hackers known in the days of anti-security 1.0. A few years later, a hacker (or group of hackers) known as "Phineas Fisher" began targeting technology companies they held responsible for the suppression of political activism across the globe, including Hacking Team, as discussed in the introduction. In 2016, Fisher exfiltrated data from the union of the Catalonian police force, before turning their attention to the Turkish Justice and Development Party in solidarity with Rojava and Bakur, two anti-capitalist autonomous regions in Kurdistan.[51]

In April 2016, Fisher disseminated a "DIY guide" to encourage others—prominently featuring a urinating ascii art character first found in an ~el8 textfile some 14 years earlier, with "#antisec" printed underneath.[52]

[49] See, for example: Anonymous, "The Underground Myth," *Phrack,* 12:65, File 13 of 15 (4 November 2008), http://phrack.org/issues/65/13.html; The Phrack Staff, "PHRACK PROPHILE ON Solar Designer," *Phrack,* 15:69, File 2 of 16 (6 May 2016), http://phrack.org/issues/69/2.html.

[50] Kelty, "Two Bits."

[51] Lorenzo Franceschi-Bicchierai, "Notorious Hacker 'Phineas Fisher' Says He Hacked The Turkish Government," *VICE Motherboard,* 20 July 2016, https://www.vice.com/en/article/yp3n55/phineas-fisher-turkish-government-hack.

[52] Phineas Fisher, "Hack Back! A DIY Guide," Pastebin, 17 April 2016, https://web.archive.org/web/20160417201517/http://pastebin.com/0SNSvyjJ; Phineas Fisher released a previous DIY guide on 8 August 2014. Phineas Fisher, "Hack Back! A DIY Guide for those without the patience to wait for

In 2019, they appropriated the tactics of what they called the "infosec industry," announcing a "Hacktivist Bug Hunting Program" that aimed to incentivize hackers to secure "material of public interest" from banks, private prison operators, and other targets with the promise of monetary reward. As Fisher explained it, "this program is my attempt to make it possible for good hackers to earn an honest living uncovering material in the public interest, rather than having to sell their labor to the cybersecurity, cybercrime, or cyberwar industries" [translated from the original Spanish].[53]

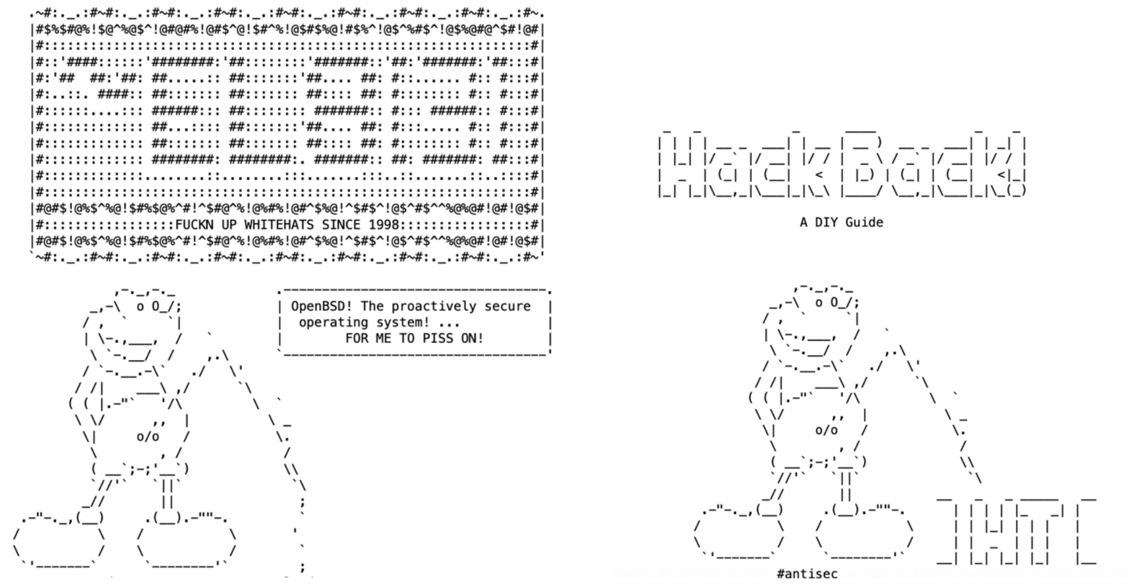

FUCKN UP WHITEHATS SINCE 1998
OpenBSD! The proactively secure operating system! ... FOR ME TO PISS ON!
HackBack
A DIY Guide
#antisec

Figure 5. Ascii art from ~el8 issue 3 from August 2002 (left) and ascii art from Phineas Fisher's "hack back" DIY guide from April 2016 (right).

Fisher and likeminded hackers thus continue the tendencies of the earlier anti-security epoch: the spectacular, irreverent hack-and-leak logic exemplified by Project Mayhem, the unruly expertise typified by figures like GOBBLES, and the nascent envisioning of a public-interest mode of hacking. Only now, as mentioned in the introduction, their activities often align with the work of non-profit organizations like Citizen Lab—whose founder, Ron Deibert, has called for a "human-centric approach to cybersecurity."[54] Indeed, Citizen Lab's investigations often hinge on the same cast of actors, demonstrating how the tools produced by companies

whistleblowers," Pastebin, 8 August 2014, https://web.archive.org/web/20140808212532/http://pastebin.com/cRYvK4jb. Interestingly, Fisher claims to have used many publicly-disclosed vulnerabilities in their attacks—effectively making them a "script kiddy" in the old elitist view espoused by the antisec 1.0-era black hats.

[53] Subcowmandante Marcos, "Hack Back! A DIY Guide to Robbing Banks," Distributed Denial of Secrets, Archived 25 November 2019. https://web.archive.org/web/20191125194808/https://data.ddosecrets.com/file/Sherwood/HackBack_EN.txt

[54] Deibert, "Toward a Human-Centric Approach to Cybersecurity."

like Hacking Team, Gamma Group, and NSO Group are used to target journalists, activists, and dissidents around the globe.[55]

From this perspective, it is possible to imagine backwards what some of the original anti-security supporters may have been imagining forward in their visions of "people's security" delivered by a "black hat bloc" of unruly computer security experts in opposition to a corporatized, militaristic security industry. Or at least, we can recognize that the inchoate rhetoric of anti-security has since inspired such imaginings in others.

[55] In the 2019 edition of "Hack Back!" Fisher acknowledges the influence of Citizen Lab directly, citing the organization's collected investigations of Hacking Team (https://citizenlab.ca/tag/hacking-team/) and FinFisher (https://citizenlab.ca/tag/finfisher/).

"Hacking The Hackers: When Bug Bounty Programs Are the Target." by Ryan Ellis, Northeastern University

In the past decade, bug bounty programs have become a common way to manage the identification and reporting of previously unknown and undisclosed software flaws.[1] Bounty programs, in their most simple form, pay hackers who find and disclose new bugs.[2] These programs have become a familiar security tool: Google, Facebook, the Department of Defense, Tesla, the retailer Lululemon, and hundreds of others now rely on bug bounties to help improve their security.[3] Bounties provide a way to organize hacker expertise and labor: enrolling them in a market that pays for the successful uncovering and reporting a new flaw.[4]

In a recent report, *Bounty Everything: Hackers and the Making of the Global Bug Market Place*,[5] Yuan Stevens and I show how the bounty model of crowdsourced security creates new hazards for the hackers/workers who participate in bounty programs. While bounty programs offer a number of possible benefits, they also mimic other forms of gig work and saddle precarious workers with new risks.

In this short essay, I further argue that bounty programs serve as *targets*—sites of potential malicious exploitation and attack—that create previously unacknowledged risks. In reconceiving bounty programs as targets, I draw on recent publicly reported incidents that (1) provide a window into the security of bug bounty programs and platforms, and (2) demonstrate the perceived value of coopting or repurposing key elements of exploits and attacks. Ultimately, this reframing calls for prioritizing the security of bounty programs.

Hacking the Hackers: Subverting Bounty Platforms

Bounty platforms organize and manage bounty programs on behalf of clients, and have been hacked in recent years. Documented incidents provide a window into bounty security and indicate the value that an adversary might find in subverting a bounty program.

In 2019, HackerOne, a bounty platform that hosts several hundred bounty programs and counts over one million registered hackers, reported an unauthorized breach.[6] A hacker working under the handle haxta4ok00

[1] This material is based upon work supported by the National Science Foundation under Grant Nos. 1915815 and 2203175. Any opinions, findings, and conclusions or recommendations expressed in this material are those of the author(s) and do not necessarily reflect the views of the National Science Foundation.

[2] For an overview of bounty programs and their history, see: Ryan Ellis and Yuan Stevens, *Bounty Everything: Hackers and the Making of the Global Bug Marketplace,* Data & Society, 2022. https://datasociety.net/library/bounty-everything-hackers-and-the-making-of-the-global-bug-marketplace/. For an examination of the evolution and expansion of bounty programs to address sociotechnical harms, see: Josh Kenway. Camille François, Sasha Costanza-Chock, Inioluwa Deborah Raji, and Joy Buolamwini, *Bug Bounties for Algorithmic Harms?*, Algorithmic Justice League, 2022. https://www.ajl.org/bugs.

[3] Ellis and Stevens, *Bounty Everything.* For a partial list of programs, see: "Public Bug Bounty Programs." https://github.com/projectdiscovery/public-bugbounty-programs.

[4] For a detailed history of the shifting landscape of hacker expertise, see: Matt Goerzen and Gabriella Coleman, *Wearing Many Hats: The Rise of the Professional Security Hacker,* Data & Society, 2022. https://datasociety.net/library/wearing-many-hats-the-rise-of-the-professional-security-hacker/.

[5] Ellis and Stevens, *Bounty Everything.*

[6] For a detailed account of the case, see the HackerOne incident report. HackerOne, "Incident Report | 2019-11-24 Account Takeover via Disclosed Session Cookie," 11 November 2019, https://hackerone.com/reports/745324.

found an unauthorized way to access a HackerOne security analyst's account (an expired session cookie was inadvertently disclosed during an earlier interaction).[7] The hacker reported the bug to HackerOne and the company resolved the problem by revoking the session cookie, thus removing the path to unauthorized access. Haxta4ok00 received a $20,000 bounty for reporting the issue.[8] The story attracted attention in the popular press. The irony was clear: a company devoted to harnessing the power of hackers for good had been hacked.[9]

Access to the session cookie indicated the potential value of bounty programs and platforms as a target. Gaining access to the HackerOne security analyst's account would allow an attacker to access otherwise unknown and unfixed bugs across a selection of HackerOne's programs.[10] HackerOne provides a range of services for the companies and organizations that have bounty programs on their platform, including review and triage of incoming reports.[11] This moderation is one of the core services that bounty platforms provide: public bounty programs regularly receive a flood of invalid reports. Triage service, whether outsourced to a bounty platform or performed by in-house staff, is essential to identifying new bugs within a steady stream of invalid or irrelevant submissions.[12] The hacker haxta4ok00 demonstrated how useful targeting a bounty program or platform might be. In accessing the employee account, they instantly gained visibility into bugs that had been submitted but not yet reviewed or remediated by the vulnerable host organization. HackerOne's incident report was plain: with this somewhat simple flaw an intruder would have been able to access all of the programs and all of the reports associated with the analysts' account; they would have been able to access both the metadata associated with the report and the contents of the report.[13]

While credential theft or spoofing provides one way to take advantage of the pool of unfixed bugs, insider threats present a similar risk. In June, 2022 HackerOne disclosed another security incident. An unidentified employee who worked in triage improperly leveraged their access to submitted bugs. Rather than pushing the bugs through the triage and mitigation pipeline, they attempted to steal bugs for personal gain.[14] They created a fake HackerOne account and submitted these bugs to a number of different bounty programs in the hopes of claiming payment for novel bugs.[15]

These two incidents are striking in just how ordinary they are. Credential theft enabled by weak security practices, and insiders accessing systems for unauthorized purposes, are not unique or surprising vectors of compromise. These sorts of incidents regularly happen across organizations. What makes these cases relevant or sobering is the potential impact that the subversion of a bounty platform might cause. Bounty programs, by their very nature, gather and host sensitive data that, if it fell into the wrong hands, would allow malicious actors to generate novel exploit and attacks. Identifying and developing novel attacks requires a specific type

For details concerning HackerOne's size, scope, and operation see: HackerOne, *The 2021 Hacker Report: Understanding Hacker Motivations, Development and Outlook,* 2021. 2; "Public Bug Bounty Programs."

[7] HackerOne, "Incident Report | 2019-11-24 Account Takeover via Disclosed Session Cookie."

[8] HackerOne, "Incident Report | 2019-11-24 Account Takeover via Disclosed Session Cookie."

[9] "HackerOne Pays $20,000 Bug Bounty After 'Sloppy Breach,'" BBC, 5 December 2019, https://www.bbc.com/news/technology-50670433.

[10] HackerOne, "Incident Report | 2019-11-24 Account Takeover via Disclosed Session Cookie."

[11] For a discussion of the role of bounty platforms, see Ellis and Stevens, *Bounty Everything.*

[12] See Ellis and Stevens, *Bounty Everything.*

[13] Security analysts have segmented access, they can only access reports for programs to which they assigned rather than the entire customer base. HackerOne reported that this particular account allowed for access to "less than 5%" of the programs on their platform. HackerOne, "Incident Report | 2019-11-24 Account Takeover via Disclosed Session Cookie."; "HackerOne Pays $20,000 Bug Bounty After 'Sloppy Breach,'" BBC.

[14] HackerOne, "June 2022 Incident Report," 1 July 2022. https://hackerone.com/reports/1622449.

[15] HackerOne, "June 2022 Incident Report."

of expertise, including the discovery of vulnerabilities on endpoints that can be exploited. Rather than cultivating this expertise, some actors might find it easier or more advantageous to simply steal the capability. Indeed, as the next set of cases show, that is precisely what has happened.

Reuse/Recycle/Redeploy: Coopting Expertise

As Ben Buchanan observes in his book, *The Hacker and the State: Cyber Attacks and the New Normal of Geopolitics*, hacking is one of the central ways that states try to shape geopolitics.[16] The use of cyber operations to spy, disrupt, and destabilize requires the development of novel exploits and attacks and, by extension, the cultivation of expertise.[17] This cultivation takes many different forms, including bureaucratic reorganization, formal investment in training and tool development, contracting with third-party proxies, and other techniques. Additionally, a number of recent cases point toward a different approach to cultivating expertise: co-option. As states race to compete and create new capabilities, the theft and redeployment of an adversary's (or, as the case below suggests, an ally's) capabilities is an attractive option. Rather than investing resources in developing native tools or components (such as finding a novel flaw that can be used as a basis for an exploit or attack), stealing another's hard work is a useful shortcut.

Examples of co-opting are readily available. The United States reportedly "piggybacked" on South Korean capabilities to gain visibility into North Korean computer networks during a period when its own access was otherwise limited.[18] Reports in *Der Spiegel* tied to documents leaked by former National Security Agency (NSA) contractor Edward Snowden, describe this hacking of hackers as a common practice christened as "fourth-party" collection.[19] Other examples of this approach—states (or other actors) hacking into other ongoing hacking attempts—point to the widespread prevalence of fourth-party hacking by other parties.[20] These accounts make it clear that the value of this approach is two-fold. In addition to gaining capabilities that were otherwise unavailable or out of reach, cooption provides an added layer of deniability or obfuscation.

Perhaps the most high-profile example of coopting is the case of ETERNALBLUE. In 2016, a group known as "the Shadow Brokers" claimed to have successfully pilfered a collection of secret NSA tools and documents. In time, they released these powerful capabilities, including ETERNALBLUE, a powerful exploit that targeted Windows machines. This tool was repurposed, first by malicious actors linked to North Korea as part of the ransomware attack known as "WannaCry" and then, a few months later by the Russian-backed destructive attack, "NotPetya." While WannaCry was broadly disruptive, the damage associated with

[16] See Ben Buchanan, The Hacker and the State: Cyber Attacks and the New Normal of Geopolitics (Cambridge: Harvard UP, 2020), 7.

[17] See Buchanan, *The Hacker and the State*. For a discussion of how states harness hackers for particular ends, see Luca Follis and Adam Fish, *Hacker States* (Cambridge: MIT Press, 2020).

[18] Choe Sang-Hun, "North Korean Hackers Stole U.S.-South Korean Military Plans, Lawmakers Say," *New York Times*, 10 October 2017.

[19] This practice is not isolated to a particular target or particular hacking campaign. Joseph Cox, "The Murky World of Spies Hacking Other Spies," *The Daily Beast,* 4 October 2017, https://www.thedailybeast.com/the-murky-world-of-spies-hacking-other-spies. See also Jacob Appelbaum, Aaron Gibson, Claudio Guarnieri, Andy Müller-Maguhn, Laura Poitras, Marcel Rosenbach, Leif Ryge, Hilmar Schmundt und Michael Sontheimer, "NSA Preps America for Future Battle," *Der Spiegel,* 1 January 2015, https://www.spiegel.de/international/world/new-snowden-docs-indicate-scope-of-nsa-preparations-for-cyber-battle-a-1013409.html.

[20] Cox, "The Murky World of Spies Hacking Other Spies."

NotPetya was unprecedented, totaling an estimated $10 billion.[21] The most destructive cyberattack in history was, in part, a story of the effectiveness of repurposing or coopting expertise through theft.

Expertise is not only something to be developed; sometimes it can be stolen.

Conclusion: Bug Bounty Programs as Target

Reconsidering bounty programs as targets reveals an important point: the utility of bounty programs can be compromised by weak security practices. Bounty programs are designed and intended to improve security, but poor controls or protection of the bug pipeline can undermine any presumed or hoped-for security gains. Recasting bounty programs as targets highlights the importance of prioritizing security in handling the submission, review, and mitigation of bugs. Bugs are valuable and they should be protected like other valuable assets. Here, as in other domains, increasing security raises the costs for attackers.

There are, however, larger lessons as well. Bounty programs or other caches of unpatched bugs are always going to be attractive targets. Reducing their value is, however, possible. Developing well-resourced and integrated bounty programs can reduce "time-to-fix," the period between submission and mitigation deployment. Ensuring that bugs are, in effect, soon-to-spoil goods reduces their value. For an attacker, the value of unknown and unpatched bugs starts to erode once mitigations are developed and deployed.[22] Capturing a collection of soon-to-be-fixed bugs is less valuable than capturing long-lasting and durable vulnerabilities. Shrinking the window between initial disclosure and mitigation requires not just quickly reviewing or validating a bug, it also rests on successfully integrating an organization's bounty program with in-house staff that will develop and deploy the eventual fix.[23] Making sure an effective patch or update is available and adopted is vital to limiting the utility of a bug and shrinking the value of bounty programs as targets.

These simple solutions face an uncertain future. For the past decade, bounty programs have been adopted, in part, as a strategy for lowering the costs associated with security work.[24] The development of bounty platforms that mirror other forms of gig work rest on this basic premise.[25] These developments create cross-winds that can complicate matters. Bug bounty programs are increasingly run by bounty platforms. These companies are styled as *lean platforms*—they thrive by increasing scale, adding more bounty programs, more hackers, more submissions, while keeping costs and directly employed workers to a minimum.[26] This model might be difficult to reconcile with costly investments in security (as the above hacking of bounty platforms might hint). At the same time, the business model of bounty platforms requires signing up an increasing number of companies to offer bug bounty programs. Not all companies may be ready to respond to the rush

[21] Andy Greenberg, "The Untold Story of Not Petya, the Most Devastating Cyberattack in History," *Wired,* 22 August 2018, https://www.wired.com/story/notpetya-cyberattack-ukraine-russia-code-crashed-the-world/.

[22] Erode, but not shrink to zero. Known bugs are still useful for hacking campaigns. For example, see Nicholas Weaver, "0-Days, N-Days, iPhones, and Android," *Lawfare Blog,* 20 August 2016, https://www.lawfareblog.com/0-days-n-days-iphones-and-android.

[23] Bruce Schneier describes this period—between the identification of a vulnerability and the adoption of patch—as the "the window of exposure." Schneier, "Full Disclosure," *Crypto-Gram,* 15 November 2001. https://www.schneier.com/crypto-gram/archives/2001/1115.html.

[24] Ellis and Stevens, *Bounty Everything.*

[25] Ellis and Stevens, *Bounty Everything.*

[26] On lean platforms, see Nick Srnicek, *Platform Capitalism* (Malden: Polity, 2017).

of incoming reports.[27] "Time to fix" may well suffer as overwhelmed organizations sputter in the face of a flood of new bugs. In such instances, bug bounty programs can create new risks. This state of affairs is not a technical failing, and not an inevitability, but a question of how expertise is to be organized and valued.

[27] Security expert, and one of the key figures in the development of bug bounties, Katie Moussouris emphasizes that adopting a bug bounty programs without first developing organizational maturity—the ability to process and handle bugs—is a recipe for disaster. See Andrew Marino, "How the Commercialization of Bug Bounties is Creating More Vulnerabilities—An Interview with the CEO of Luta Security, Katie Moussouris," 7 July 2020, https://www.theverge.com/2020/7/7/21315870/cybersecurity-bug-bounties-commercialization-katie-moussouris-interview-vergecast-podcast.

"The Recursive, Geopolitical, and Infrastructural Expertise of Malware Analysis and Detection." by Andrew Dwyer, Royal Holloway, University of London

The computational device used to read this essay, and the network providing access to this forum, are both likely to be protected by an often-ignored infrastructure to analyse and detect malicious software. Most evident in anti-virus and endpoint detection technologies, an infrastructure is silently at work on the background of our digital devices, dependent on a largely hidden network of people and computers that span the globe.

In this essay, I argue that this infrastructure is dependent on the recursive folding of the expertise of people who analyse and write detections for malware—malware analysts—alongside the exploitation of greater computer automation and reasoning. Endpoint detection vendors generate detections by engaging in iterative feedback loops and recursive practices to limit and contour the cyber operations of states and cyber-criminals as much as the infrastructure itself is exploited for geopolitical advantage by states. The capacity to shape geopolitical action is dependent on, and sustained by, complex techno-human hybrids of expertise between malware analysts and computation, extending from an analyst's hand-written detection to machine learning algorithms that construct new features to detect "suspicious" malware attributes.[1] Techno-human expertise is likewise facilitated by the sharing and analysis of big data as well as novel organisational practices that embed recursive, non-linear feedback loops into an infrastructure of malware analysis and detection. Together, techno-human expertise and a recursive infrastructure support one another, enabling endpoint detection vendors to purportedly claim that they can identify suspicious activity quickly, pre-emptively, and at scale. Notwithstanding the extent to which this claim is true, without such a recursive infrastructure of malware detection, the Internet and today's geopolitical landscape, would look considerably—if not radically—different.

The geopolitical importance of the combination of a recursive malware analysis and detection infrastructure and techno-human expertise is made most explicit in the removal of the Moscow-based endpoint detection vendor Kaspersky from government networks in various countries. In 2015, Israeli intelligence operators who were exploiting Kaspersky's computer network discovered a trove of hacking tools from the US National Security Agency (NSA). The operatives tipped off their counterparts at the NSA, who concluded that the Russian government had used Kaspersky's infrastructure to gain access to those tools. As first reported in the *New York Times* in 2017, this resulted in Kaspersky and its infrastructure becoming a "Google search for sensitive information."[2] In the same year, the United States had prohibited the federal government from using the services or products of the Russian-based endpoint detection provider, arguing that Kaspersky threatened the integrity and confidentiality of government information.[3] By 2022, the US Federal Communications Commission (FCC) had added Kaspersky to a list of firms that could not be paid by the FCC's Universal

[1] For more, see Andrew Dwyer, "Malware Ecologies: A Politics of Cybersecurity," PhD diss. (University of Oxford, 2019), https://ora.ox.ac.uk/objects/uuid:a81dcaae-585b-4d5b-922f-8c972b371ec8/; Andrew C Dwyer, "Cybersecurity's Grammars: A More-than-Human Geopolitics of Computation," *Area* 55:1 (2023): 10–17, https://doi.org/10.1111/area.12728.

[2] Nicole Perlroth and Scott Shane, "How Israel Caught Russian Hackers Scouring the World for U.S. Secrets," *New York Times*, 10 October 2017, https://www.nytimes.com/2017/10/10/technology/kaspersky-lab-israel-russia-hacking.html.

[3] Office of the Press Secretary, "DHS Statement on the Issuance of Binding Operational Directive 17-01," US Department of Homeland Security, 13 September 2017, http://web.archive.org/web/20220901082417/https://www.dhs.gov/news/2017/09/13/dhs-statement-issuance-binding-operational-directive-17-01.

Service Fund.[4] Similar calls and guidance to limit the use of Kaspersky have been made across Europe and elsewhere.[5]

Yet Kaspersky argued that the NSA tools were gathered inadvertently, as part of its routine automated "proactive protection technologies."[6] To understand why such an argument is even plausible, and how it links to contemporary endpoint protection's use of techno-human expertise, we must examine the recursive infrastructure of malware analysis and detection. As I shall argue, despite the particularities of the case, Kaspersky's infrastructural capacities for analysis and detection are not unique, but rather are a feature of most endpoint detection products.

For the remainder of this essay, I explore how the techno-human expertise of malware analysts and computer automation and reasoning fuse into a geopolitically important infrastructure, where Kaspersky becomes a threat to national security as much as it may limit the exploitation of computer vulnerabilities by state cyber operations and cyber-criminals. I first outline how endpoint detection operates. Second, I discuss how three concurrent strategies of malware detection—the static, contingent, and recursive—employ different techno-human variants of expertise and how they fuse in a recursive infrastructure. I then return, third, to the case of Kaspersky to demonstrate how recursive malware detection infrastructures intersect with geopolitics and outline how malware analyst expertise can shape the capacity of other geopolitical actors as it becomes articulated through a recursive, techno-human infrastructure of international reach and depth.

Endpoint Detection

At its most simple and common, malware detection is conducted by "engines" on millions, if not billions, of endpoints, such as personal computers. Detections are distributed to engines by endpoint detection vendors and data returned to them, forming a massive, networked infrastructure. Engines operate in the background, monitoring and assessing an endpoint's environment, paying particular attention to any new introduced software and code like a new saved document from an email. When first developed in the late 1980s, engines used simple "matching" techniques on the few known malware variants at the time. Since then, the growing sophistication, volume, and scale of malware has driven the development of new, increasingly automated, techniques to reduce the burden on analyst labour. Malware analysts have thus developed a wide range of expertise in techniques that use computer automation and reasoning, including behavioural monitoring, utilising big data from malware repositories such as Alphabet-owned VirusTotal and from customer telemetry (data collected from engines on endpoints), as well as developing cloud-based technologies for scalable and remote analysis. Malware analysts also work with data scientists and engineers to integrate machine learning algorithms into endpoint detection products.

[4] Federal Communications Commission, "FCC Expands List of Equipment and Services That Pose Security Threat," US Federal Communications Commission, 25 March 2022, http://web.archive.org/web/20220928110917/https://www.fcc.gov/document/fcc-expands-list-equipment-and-services-pose-security-threat.

[5] For example, in the UK see Ian Levy, "Use of Russian Technology Products and Services Following the Invasion of Ukraine," UK National Cyber Security Centre (blog), 29 March 2022, http://web.archive.org/web/20221107180345/https://www.ncsc.gov.uk/blog-post/use-of-russian-technology-products-services-following-invasion-ukraine.

[6] Kaspersky, "Preliminary Results of the Internal Investigation into Alleged Incidents Reported by US Media (Updated with New Findings)," Kaspersky Daily (blog), 25 October 2017, https://www.kaspersky.com/blog/internal-investigation-preliminary-results/19894/.

At the greatest extent with machine learning algorithms, increasing automation, and engagement with big data practices have enabled supposedly predictive forms of malware detection. As contemporary endpoint detection integrates machine learning algorithms, it has become more dependent on computer reasoning to develop new insights into what is suspicious through "unseen" clusters, patterns, and relations between data. By integrating computer reasoning that can construct new features upon which suspicious activity can be identified, the infrastructure transforms the terrain upon which actors—both state and non-state[7]—must negotiate.[8] This has come with the benefits of scale and speed of detection to limit and shape the activity of various actors but has often brought with it a higher false positive rate.

Despite endpoint detection being ever-more reliant on contemporary machine learning techniques, at its core remains the expertise of malware analysts. They remain fundamental to the decision of what is, and is not, malicious. They are essential to reduce false positive rates and ensure that computer reasoning is closely optimised to what malware analysts define as malicious. Therefore, even as analysts are supported by a range of engineers, teams ensuring strong quality assurance, and are part of techno-human hybrids, they provide what is the "ground truth" to maliciousness that is quintessential for computer automation and reasoning on malware.

Hybrid Expertise

Based on the outskirts of the quiet English town of Abingdon-on-Thames, the headquarters of Sophos sits as one node in an international recursive infrastructure of malware detection connected by big data practices. In 2017, I spent seven months conducting an autoethnography at Sophos' malware analysis and detection laboratory, SophosLabs. Whilst there, I sought to understand how software is identified as malicious. However, I also found how the expertise of malware analysts, built up over lengthy periods of sitting behind a computer screen, experimenting with deeply embodied notions of what is malicious and not, has become tied to computer automation and reasoning in recursive feedback loops. Techno-human hybrid expertise is shared between malware analysts, computers, and others, and complicates *where* expertise is located and *who* and *what* can claim to be assessing what is malicious. Yet, this requires an infrastructure to support such activity. In what follows, I describe how relationships between analysts and computers have evolved through three strategies of malware detection: the static, contingent, and recursive.

Static Strategies

The deep analysis of malware is central to an analyst's training and is a method core to the static strategy of detection. That is, suspicious software and code are isolated, extracted, and examined on an analyst's computer, which functions as a metaphorical microscope where the specimen is sliced, spliced, and rendered visible to the analyst to inquire about its maliciousness. Historically following an often slow and methodical process to determine whether software is malicious, analysts develop a "signature" that specifies the malware's unique attributes for the endpoint detection engine to use. These malware detection signatures become the "ground truth" to what is considered malicious, which is based on the social and embodied relations that malware analysts acquire over time. This includes learning a range of technical skills, in conversations between analysts on malware techniques, in peer review of detection signatures, as well as through the practical experimentation of analysis. Conventionally, this meant that a software's designation as

[7] Florian J Egloff, *Semi-State Actors in Cybersecurity* (Oxford: Oxford University Press, 2022).

[8] David Beer, "The Problem of Researching a Recursive Society: Algorithms, Data Coils and the Looping of the Social," *Big Data & Society* 9: 2 (2022): 20539517221104996, https://doi.org/10.1177/20539517221104997.

malicious could be linked to the skills and labour of individual malware analysts. Whilst the static strategy is no longer the only method for detecting malware, signatures are still commonly produced to improve the overall accuracy of detection and have become essential to big data sharing and machine learning algorithms. Yet, this is not a unidirectional process. Computer automation and reasoning have recursively, through the infrastructure, transformed labour practices, as shall be explored in greater detail below.

Contingent Strategies

As the volume of malware grew rapidly in the 1990s, malware analysts developed a range of analytic strategies that identified common behaviours rather than individual malware attributes. These contingent strategies monitor software and code in execution as well as computational environments (e.g., systems, networks) in real-time, using a set of pre-defined rules and algorithms to detect and prevent any suspected malicious behaviour. This automates some of a malware analyst's expertise, thereby increasing the speed and scale of malware detection and prevention. However, contingent detection will only catch behaviours that are already deemed suspicious by malware analysts, such as the use of common entry points or the exploitation of already-known vulnerabilities. This can be useful for ransomware detection, using encryption of numerous files over a short timeframe as an indicator of suspicious behaviour. However, it cannot identify as yet-unseen behaviours and techniques. The contingent strategy increases the speed and scale of application of an analyst's expertise rather than an explicit re-working through computer reasoning as in recursive strategies for analysis and detection.

Recursive Strategies

In recent years, the production and sharing of big data have facilitated recursive logics for detection, thereby re-shaping the expert labour of malware analysts. Machine learning algorithms employ feedback loops, such as in "deep" convolutional neural network algorithms,[9] enabling computer reasoning to be used to detect malware in new ways.

The data used to train machine learning algorithms to identify "malicious" attributes derive from the vast number of detection signatures that have been written by malware analysts working in various endpoint detection vendors over many years. These are collectively shared with repositories including VirusTotal. Thus, collectives of analysts past and present create a foundation for machine learning algorithms to identify attributes of malicious behaviour. The logic of this process may be incomprehensible to the analysts who provided the learning data itself. This is due to the capacity of machine-learning algorithms to develop new abstractions and thus create features to detect malicious attributes based on "unseen" connections that people may not be able to recognise easily, if at all. The resulting detections by machine-learning algorithms are not a linear extension of the malware analyst, but integrate the capacity for computation to reason. This is a deeply techno-human hybrid expertise—neither wholly human nor computational alone.

Today, static strategies have also been reworked by contingent and recursive strategies in this recursive infrastructure. Analyst labour is now directed by detection data produced by computer automation and reasoning. When I was sitting at my computer at SophosLabs, I was writing detection signatures according to what detection data from contingent and recursive strategies suggested was being "missed" by analysts. Therefore, the whole infrastructure is engaged in recursive logics that combine the expertise of analysts past

[9] These are a type of algorithm that take image-based inputs and develop features to identify attributes, for more in malware analysis and detection, see Joshua Saxe and Hillary Sanders, *Malware Data Science: Attack Detection and Attribution* (San Francisco: No Starch Press, 2018).

and present to sustain the promise of an anticipatory, pre-emptive malware detection. Recursive strategies of malware detection sustain endpoint detection organisational feedback loops to prioritise and order the limited labour of analysts. Loop upon loop! Still, as much as machine learning algorithms enable new forms of reasoning to construct new features to detect malicious attributes, analysts crucially provide a collective re-grounding to a "truth" of what is malicious through their detections, that in turn is used in contingent and recursive strategies of analysis and detection. Computer reasoning in machine learning algorithms may be able to optimise to the collective output of analyst detections in repositories and customer telemetry, but computers can never mathematically assess maliciousness in the same way that an analyst can.

Recursive malware detection infrastructures—through relying on new forms techno-human expertise in recursive strategies—shape the terrain for cyber operations as much as for criminal groups, in part because they have a capacity to detect as-yet "unknown" malware and limit certain behaviours. Highly sophisticated actors must work to avoid such infrastructures and sometimes explicitly exploit them. Internationally, through big data sharing and use, malware detection infrastructures are productive of techno-human hybrids beyond individual endpoint detection vendors. Collectively, multiple vendors now overlap and shape the detection on many computers and networks we all rely on, making computation more difficult to exploit. This does not necessarily prevent widespread exploitation, as was seen in two now-infamous events in 2017, WannaCry and NotPetya, in part because not all of the recursive and contingent strategies developed by endpoint detection vendors are used consistently. Additionally, recursive strategies do not necessarily "work" better, because computer reasoning is reliant on optimisation, leading to higher volumes of false positives. Nonetheless, by focusing on how the terrain is modified by a collective techno-human hybrid expertise, we gain a new perspective on why states may be increasingly employing subversive practices to conduct cyber operations.[10]

Recursive Geopolitics

Malware analysis and detection is today an integrated part of a contemporary geopolitical and recursive infrastructure that draws together malware analysts and exploits computer reasoning at scale through techno-human hybrids of expertise, which is distributed via endpoint detection engines. These endpoints transform the terrain of billions of digital devices across the world, thereby exerting geopolitical effects on the types of cyber operations states may conduct. This infrastructure no longer operates primarily in a responsive and linear mode. Rather, it attempts to identify software as malicious proactively and pre-emptively before it is executed on computers and networks through recursive, big data practices. This in turn permits a techno-human expertise that is reliant on big data and feedback loops of the recursive infrastructure, as much as that infrastructure exists to sustain and support that expertise. It is not possible to detach the capacity of the recursive infrastructure from techno-human expertise that is developed in endpoint detection. Without the expertise, the recursive infrastructure would not "work", and without the infrastructure, there would be very different forms of techno-human expertise less dependent on recursive practices.

Returning to the case of Kaspersky, one can understand how the removal of endpoint detection engines from government networks is at least in part explainable by the recursive infrastructure of malware detection and its reliance on big data to generate collective techno-human hybrid expertise. In Kaspersky's case, a NSA contractor enabled Kaspersky's endpoint detection engine to scan documents and send this as customer

[10] Lennart Maschmeyer, "The Subversive Trilemma: Why Cyber Operations Fall Short of Expectations," *International Security* 46:2 (2021): 51–90, https://doi.org/10.1162/isec_a_00418.

telemetry back to the endpoint detection provider, leading to the upload of confidential documentation.[11] It is not possible to independently assess whether Kaspersky was used by the Russian government to deliberately target the NSA, if it was part of a broader campaign, or simply a benign result of its recursive infrastructure. However, it is precisely contemporary malware analysis and detection infrastructures which make this possible. The practice of monitoring environments and collecting customer telemetry is not a unique feature of Kaspersky. Many endpoint detection vendors remotely upload "unseen" and "suspicious" software and code to analyse, often using cloud computing. Most endpoint detection engines are likely to flag at least some of the NSA's tools as "suspicious." Without such practices, endpoint detection vendors would be unable to offer the promise of identifying new malware quickly and at scale. This undoubtedly offers greater security for many everyday threats, but also has geopolitical implications. Kaspersky's infrastructure is more a norm than exception.

As computers and geopolitics are more deeply and widely interconnected than ever before, the recursive infrastructure of contemporary malware analysis and detection is attaining greater geopolitical potential. The capacity of such a recursive infrastructure and the techno-human expertise required may be imperceptible to most people, but it is core to the provision of cybersecurity for many. When a cyber campaign cannot cause an effect on a computer due to a detection written by an analyst based on previously observed behaviour, or a machine learning algorithm detects a cyber operation due to the identification of attributes that fit patterns and relations on previously identified malicious features, this shapes the terrain of engagement for states. For sophisticated actors, this is evidently not insurmountable, as much as the infrastructure itself is exploitable, but the expertise of malware analysts, as they become techno-human and tied into recursive infrastructures, is of quintessential geopolitical interest.

[11] Alex Hern, "NSA Contractor Leaked US Hacking Tools by Mistake, Kaspersky Says," *The Guardian*, 26 October 2017, http://web.archive.org/web/20221008105906/https://www.theguardian.com/technology/2017/oct/26/kaspersky-russia-nsa-contractor-leaked-us-hacking-tools-by-mistake-pirating-microsoft-office.

"Expertise at the Boundaries: Understanding Critical Infrastructure Cybersecurity" by Rebecca Slayton, Cornell University, and Clare Stevens, University of Portsmouth

Critical infrastructure organizations around the world are increasingly connecting the operational technology (OT) that controls physical devices, processes, and events with the information technologies (IT) that comprise cyberspace. Many organizations seek the integration of OT and IT in order "to gain competitive advantage, become more efficient, profitable and reliable."[1] However, this revolution also comes with new challenges and risks. Increased connectivity increases the complexity of large technical systems and the corresponding potential for "normal accidents."[2] It also increases the risk of cyber-attacks. For example, in both 2015 and 2016, Russian hackers successfully shut down sections of Ukraine's electric power grid. Though these attacks have been overshadowed by the physical attacks associated with Russia's invasion of Ukraine, nations around the world are devoting growing attention and resources to the challenges of security that come with what the World Economic Forum has described as the "Fourth Industrial Revolution."[3]

Historically, critical infrastructure organizations have secured operational technology primarily through physical isolation. Through most of the twentieth century, the computers controlling operational technology ran behind locked doors and tall fences, and thus needed few if any additional security controls. In fact, the lack of computer passwords or other computer security controls could be understood as a safety feature in the context of operational technology, because it ensured that operators would not be locked out in an emergency.[4] But as organizations have begun to connect these computers to broader networks that enable remote access, cybersecurity has become increasingly important. Unfortunately, implementing cybersecurity in legacy systems incurs production downtime and capital improvement costs that are expensive—sometimes prohibitively so.

Decisions about how to manage the integration of IT and OT thus entail trade-offs between different kinds of public goods and risks, including security, safety, reliability, and economy. These trade-offs are not merely arcane technical issues, but policy challenges that cross boundaries between governments, sectors, and fields of expertise. In this essay, we first outline the challenges of managing risks that span these traditional boundaries. We then focus on a challenge that underlies risk management across all of these boundaries: the challenge of creating credible expertise in a still-emerging area of technology that creates new opportunities for emerging threats. While policy often seeks to manage boundary-spanning risks from the top-down—with governmental policies shaping cooperation between public and private organizations, which in turn coordinate and structure the work of different kinds of experts—we argue for a bottom-up approach that examines how the cultural practices of experts alternately produce, maintain, navigate, and transcend boundaries.

[1] Rob Hayes, "Managing the Successful Convergence of IT and OT What I Wish I'd Known," Deloitte, 2020, 1, https://www2.deloitte.com/content/dam/Deloitte/global/Documents/Risk/gx-deloitte-managing-the-successful-convergence-of-it-and-ot.pdf.

[2] Charles Perrow, "Organizing to Reduce the Vulnerabilities of Complexity," *Journal of Contingencies and Crisis Management* 7:3 (1999): 150-155, https://doi.org/10.1111/1468-5973.00108.

[3] Madeline Carr and Feja Lesniewska, "Internet of Things, Cybersecurity and Governing Wicked Problems: Learning from Climate Change Governance," *International Relations*, 34:3 (2020): 391–412. https://doi.org/10.1177/0047117820948247.

[4] Joe Weiss, *Protecting Industrial Control Systems from Electronic Threats* (New York, NY: Momentum Press, 2010).

Three Kinds of Boundaries: Governmental, Organizational, and Professional

The physical infrastructures associated with operational technologies often cross boundaries between and within different nations, creating interdependencies and complicating questions of responsibility. For example, the North American electrical grid is a giant system of systems spanning Canada, the United States, and Mexico. Similarly, in Europe the bulk transmission of electric power requires the cooperation of operators spanning from Eastern Russia to the Republic of Ireland, and this does not include the many thousands of distribution operators across the continent. Many critical infrastructures are also indirectly connected to more distant nations that produce critical components and flows of oil, gas, coal, and other materials. As a case in point, the European Natural Gas Network constitutes more than 200,000 km of transmission pipelines, over 2 million kilometers of distribution network stretching across the continent, and is operated by a complex combination of large private corporations and European government agencies.[5]

Regional and international organizations have created regulations that establish minimum levels of security. For example, in 2016 the European Union issued a Directive on the Security of Networks and Information Systems (NIS) that required member states to appoint national authorities to serve as a single point of contact for coordinating cross-border issues and to develop policy frameworks to ensure that critical infrastructure operators are implementing security safeguards that are proportional to risk.[6] Similarly, the United States has delegated authority to establish cybersecurity standards to an industry group, the North American Electric Reliability Corporation, which operates across the United States, Canada, and a small portion of Mexico.

These policies leave considerable ambiguity surrounding what constitutes an adequate response to cyber risk. Furthermore, many organizations fall out of the scope of regulation due to jurisdictional issues. For example, the United States regulates electricity production and transmission through the Federal Energy Regulatory Commission (FERC), which has the authority to establish both reliability and security standards. In the United States, however, electricity distribution is regulated by state and local agencies that have traditionally focused primarily on the economic and reliability needs of ratepayers, not national security.[7] Responsibility for cybersecurity and many other aspects of critical infrastructure thus remains diffuse, with unclear lines of authority.

A second set of boundaries lies in the split between the private organizations that often provide critical infrastructure, and the governmental organizations that are responsible for national security. This divide is particularly problematic in nations with strong traditions of privatization. For example, the majority of electrical power in the United States is produced by investor-owned utilities whose primary goal is to turn a profit for shareholders, not provide national security.

Nations in North America and Western Europe have attempted to align public goods such as cybersecurity with corporate interests through public-private partnerships, a term used to describe a broad range of organizational arrangements. Scholars warn that public-private partnerships are "no silver bullet," with

[5] European Union Agency for the Cooperation of Energy Regulators, "Gas Factsheet," ACER, 2021, https://www.acer.europa.eu/gas-factsheet#:~:text=The%20EU%20gas%20network%20is,compressor%20and%20pressure%20reduction%20stations.

[6] "Directive (EU) 2022/2555 of the European Parliament and of the Council," 14 December 2022, European Union, https://eur-lex.europa.eu/legal-content/EN/TXT/PDF/?uri=CELEX:32022L2555&from=EN

[7] Aaron Clark-Ginsberg and Rebecca Slayton, "Regulating Risks within Complex Sociotechnical Systems: Evidence from Critical Infrastructure Cybersecurity Standards," *Science and Public Policy* 46: 3 (2019): 339-346. Slayton, "Performing Cybersecurity Expertise: Challenges for Public Utility Commissions," *Berkeley Technology & Law Journal*, 35: 3 (2021): 757-792.

continuing tensions and disagreement about appropriate arrangements for sharing information and delegating authority and responsibility for security.[8] Governments that are committed to free market principles have tried to make private organizations responsible for the security of their own networks, but this strategy becomes problematic when the threats to those networks are other state actors with a potential impact on national security. Indeed, most private organizations want to pass responsibility for protection from nation-state threats to the federal government.[9]

Because different kinds of threats use similar tactics, it is impossible to simply split responsibility for cybersecurity along these lines—delegating responsibility for defense against nation-states to the government, and responsibility for defense against criminals and less resourced threats to private organizations. Indeed, the lines between these different kinds of threats are themselves quite blurry; as Max Smeets notes in his essay for this forum, states sometimes act through or tacitly allow hacking by criminal organizations. Practical decisions about what security measures to implement, and at what cost, must be oriented towards defending against multiple kinds of threats, not just a few. Thus many governments have attempted to establish regulations to ensure that critical infrastructure organizations are managing security in a manner that is commensurate with risks not only to organizational goals, but also to national security. The diversity and complexity of critical infrastructure defies any one-size-fits all security solution. Furthermore, regulators lack the expertise and local knowledge needed to establish regulations that are can ensure security in the complex and variable contexts of critical infrastructure. As a result, regulations leave considerable discretion to private actors who must weigh tradeoffs between cost, reliability, and security.

This leads to the third boundary-spanning challenge: how can credible expertise be created in the newly emerging field of OT cybersecurity? Many industry observers note a gap between expertise in OT and IT—that is, between the practices of those who work with physical control systems and those who work with office-environment computers. The UK's National Cyber Security Centre (NCSC) explains: "Where cyber security for IT has traditionally been concerned with information confidentiality, integrity and availability, OT priorities are often safety, reliability and availability, as there are clearly physical dangers associated with OT failure or malfunction."[10] These different priorities require different rhythms and practices. Physical infrastructure changes gradually in order to maintain high levels of reliability and safety, but information infrastructure changes more rapidly.[11] Industrial control systems are expected to last for 25-125 years, while most information technology products are expected to last 3-5 years.

The fast-moving pace of information technology is both a vulnerability and a strength. Economies of scope accrue to information technology companies that capture an early market share, giving them a strong incentive to "ship it Tuesday, get it right by version three," i.e. to ship insecure products.[12] Without large incentives to develop more secure products, information technology companies rely instead on rapidly patching vulnerabilities as they are discovered. Patching can produce unexpected interactions with high consequences in industrial control systems, such as a loss of control over hazardous equipment. Organizations have traditionally scheduled industrial control systems maintenance months or even years in

[8] Myriam Dunn-Cavelty and Manuel Suter, "Public–Private Partnerships Are No Silver Bullet: An Expanded Governance Model for Critical Infrastructure Protection," *International Journal of Critical Infrastructure Protection* 2:4 (2009): 179-187.

[9] Dunn-Cavelty and Suter, "Public–Private Partnerships are no Silver Bullet"; Madeline Carr, "Public–Private Partnerships in National Cyber-Security Strategies," *International Affairs* 92:1 (2016): 43-62.

[10] NCSC, "Operational Technologies," n.p.

[11] Weiss, *Protecting Industrial Control Systems from Electronic Threats*.

[12] Ross Anderson, "The Economics of Information Security," *Science* 314 (2006): 610-613.

advance to ensure safe and reliable operation, but the need for frequent updating poses challenges to these practices.[13]

Unbounding Cybersecurity Expertise

In summary, IT communities have traditionally focused on security, while OT communities have focused more on safety, but policymakers are pushing for the integration of safety and security.[14] Over the past twenty-five years, engineers and regulators have tried to develop shared practices and standards to overcome these differences, but significant tensions remain.[15] The field of OT cybersecurity is still emerging at the point of new socio-technical developments. As Raj Badiani, the "Head of Digital" at Raytheon UK notes, "the OT cyber security maturity remains comparatively under-developed."[16]

Many discussions of the gap between OT and IT suggest a top-down approach to bridging these different areas of expert practice.[17] On the UK's National Cyber Security Centre blog, one "senior security architect" argues:

> …operators should ensure that both OT and IT systems are equally and consistently accounted for in their overall approach to risk management. Not to do so could result in differences and deficiencies in the way cyber security policies are applied and risks are managed across an operator's combined IT and OT estate. […] The most effective operators are those where any friction between OT and IT teams has been reduced and where the overall approach to risk management is applied consistently in both IT and OT environments. [18]

The need to manage cultural differences, or "friction," between IT and OT communities is increasingly a feature of official and regulatory discourse. Industry observers note that "IT and OT exhibit widely-differing cultural values across several dimensions," suggesting that collaboration and coordination problems are inevitable.[19] Regulators often describe culture as a tool for achieving policy objectives. For example, the UK's

[13] Weiss, *Protecting Industrial Control Systems from Electronic Threats*.

[14] NCSC, "CNI System Design: Secure Remote Access." Ola Michalec, Sveta Milyaeva, and Awais Rashid, "When the Future Meets the Past: Can Safety and Cyber Security Coexist in Modern Critical Infrastructures?" *Big Data & Society* 9:1 (2022): 2, https://doi.org/10.1177/20539517221108369.

[15] Rebecca Slayton and Aaron Clark-Ginsberg, "Beyond Regulatory Capture: Coproducing Expertise for Critical Infrastructure Protection," *Regulation & Governance* 12:1 (2018): 115-130.

[16] Raj Badiani, "Contemporary Cyber Security and CNI: Converging IT and OT Cyber Security," TechUK Industry Views, 2021, https://www.techuk.org/resource/contemporary-cyber-security-and-cni-converging-it-and-ot-cyber-security.html.

[17] Ascentor, "The OT and IT Debate _ Make UK," Make UK: The Manufacturers Organisation, 2020.

[18] NCSC, "CNI System Design: Secure Remote Access," National Cyber Security Centre, 2020, https://www.ncsc.gov.uk/blog-post/cni-system-design-secure-remote-access.

[19] G. Murray, M. N. Johnstone, and C. Valli, "The Convergence of IT and OT in Critical Infrastructure," Australian Information Security Management Conference Proceeding (2017) 151, https://doi.org/10.4225/75/5a84f7b595b4e; Thams, cited in Joe Pettit, "Ask the Experts: How IT and OT Can Collaborate in the Name of ICS Security," *TripWire*, 2019.

Centre for the Protection of National Infrastructure argues that "getting security culture right will help develop a security conscious workforce, and promote the desired security behaviours you want from staff."[20]

Anthropologists and sociologists argue that culture is emergent, and thus cannot be used as an instrument to achieve engineering goals. Too often, talk of security or safety culture ignores competing interests and power differences within organizations.[21] This discourse frames humans as the weakest link, and obfuscates more fundamental structural problems and responsibilities for creating the problems that individuals must resolve.[22] Policies that focus only on essentialized differences or cultural stereotypes will likely produce efforts at top-down *coordination* that do not necessarily lead to effective on-the-ground *collaboration*.[23]

We argue for the need to study the situated practices that alternately produce, cross, and transcend these boundaries.[24] Sociologists of science have conceptualized boundaries between fields of expertise not as natural barriers to be managed, overcome, or erased, but as social constructs to be studied. For example, scientists often engage in "boundary-work:" rhetorical efforts to distinguish their work from that of non-scientists.[25] A key finding of this work is that socially-constructed boundaries shift with time and place. Others have examined how different fields of work coordinate their work through boundary objects: artifacts that take on distinctive meanings in different fields of activity, yet retain sufficient stability to enable coordination.

These studies suggest new research questions and strategies that go beyond efforts to engineer cultural cooperation from the top-down. How do experts rhetorically construct or challenge boundaries around their field of work, excluding or including insiders? What technologies and concepts have proven useful in coordinating fields of work that have traditionally been disparate? What interests are at stake in efforts to maintain boundaries between distinctive fields, and what interests are driving efforts to merge or overcome such boundaries? How do these interests and practices vary across national contexts, and how can an understanding of these distinctive interests inform regulatory guidelines and practices? Rather than attempting to engineer culture from the top-down, we need to better understand how experts construct and navigate boundaries around their areas of expertise.

Conclusion

Critical infrastructure cybersecurity is a "wicked problem" of coordination among multiple nations, sectors, organizations, occupations with highly complex interrelationships.[26] As other scholars have argued, "[a]ny understanding of resilience, the dynamics that produce safety, on a societal level needs to be based on study

[20] CPNI, "SeCuRE 4: Assessing Security Culture," Centre for the Protection of National Infrastructure, 2021, https://www.cpni.gov.uk/secure-4-assessing-security-culture, emphasis added.

[21] Susan S Silbey, "Taming Prometheus: Talk about Safety and Culture," *Annual Review of Sociology* 35 (2009): 341-69.

[22] Benoît Bernard, "Regulating Nuclear Safety through Safety Culture," *Journal of Safety Science and Resilience* 2:3 (September 2021): 172–178, https://doi.org/10.1016/j.jnlssr.2021.08.001.

[23] Petter Almklov, Stian Antonsen, Rolf Bye, and Anita Øren. "Organizational Culture and Societal Safety: Collaborating across Boundaries," *Safety Science* 110 (2018): 89-99.

[24] James Shires, "Enacting Expertise: Ritual and Risk in Cybersecurity," *Politics and Governance* 6:2 (2018): 31–40, https://doi.org/10.17645/pag.v6i2.1329.

[25] Thomas Gieryn, "Boundary-Work and the Demarcation of Science from Non-Science: Strains and Interests in Professional Ideologies of Scientists," *American Sociological Review* 48:6 (1983): 781-795.

[26] Carr and Lesniewska, "Internet of Things, Cybersecurity and Governing Wicked Problems."

of work practices that cross organizational boundaries."[27] We argue for research that examines how the expertise is constructed as workers negotiate these boundaries. How do these experts generate credibility and authority in the workplace, and to decision makers in the private sector, and to policymakers? In contrast to managerial or structural approaches that attempt to coordinate culture from the top down with formalized frameworks and protocols, a focus on how workers generate authority through and across different boundaries can help us better understand the processes that go into managing resilience.

[27] Almklov, Antonsen, Bye, and Øren, "Organizational Culture and Societal Safety: Collaborating across Boundaries," 3.

"Expertise, Authority, and Rulemaking in the Internet's Infrastructure"

by Jesse Sowell, University College London

Introduction

As the conflict between Russia and Ukraine escalated in March 2022, Ukraine asked technical communities that coordinate critical Internet resources to effectively disconnect Russia from the Internet, ostensibly to limit Russian propaganda and disinformation campaigns.[1] Both the Internet Corporation for Assigned Names and Numbers (ICANN, the organization maintaining top-level domain name services) and the Regional Internet Registries (RIRs, organizations delegating IP addresses necessary for Internet communication) declined to intervene.[2] Perhaps anticipating the request, the Executive Board of the Réseaux Internet Protocol Européens Network Coordination Centre (the RIPE NCC: the RIR that serves Europe, the Middle East, and Russia) had just issued a resolution reaffirming longstanding norms that the "means to communicate should not be affected by domestic political disputes, international conflicts or war."[3] The *Resolution* further highlights that "the RIPE NCC can be trusted as authoritative and free from bias or political influence" precisely because it "guarantees equal treatment for all those responsible for providing Internet services…across a diverse geographical and political region."[4]

Although triggered by the "high politics" of international conflict, the norms and assertions of authority at play in the *Resolution* have been historically relegated to the "low politics" of technical coordination.[5] Despite RIRs central role in Internet governance and security, scholars have devoted far more attention to more prominently visible organizations such as ICANN and the Internet Governance Forum (IGF), with the result that many scholars and government officials treat these organizations as the familiar, de facto Internet governance bodies.[6] Michel J.G. Van Eeten and Milton Mueller critique this focus as "lamppost science," challenging scholars to find Internet governance among the "heterogeneous organizational forms" and

[1] Mykhailo Fedorov, "Letter from the Vice Prime Minister of Ukraine to RIPE NCC," E-mail, March 2, 2022, https://www.ripe.net/publications/news/announcements/request-from-ukrainian-government.pdf; Jon Brodkin, "Ukraine Asks ICANN to Revoke Russian Domains and Shut down DNS Root Servers" (Ars Technica, 3 February 2022), https://arstechnica.com/tech-policy/2022/03/ukraine-wants-russia-cut-off-from-core-internet-systems-experts-say-its-a-bad-idea/.

[2] For ICANN's response, see Göran Marby, "ICANN Response to Request from Ukraine," E-mail, 2 March 2022, https://www.icann.org/en/system/files/correspondence/marby-to-fedorov-02mar22-en.pdf; for the RIPE NCC's response, see Hans Petter Holen, "RIPE NCC Response to Request from Ukrainian Government" (RIPE Network Coordination Centre, 10 March 12022), https://www.ripe.net/publications/news/announcements/ripe-ncc-response-to-request-from-ukrainian-government.

[3] RIPE NCC, "RIPE NCC Executive Board Resolution on Provision of Critical Services" (RIPE Network Coordination Centre, 1 March 2022), https://www.ripe.net/publications/news/announcements/ripe-ncc-executive-board-resolution-on-provision-of-critical-services.

[4] RIPE NCC.

[5] Nazli Choucri and David D. Clark, "Who Controls Cyberspace?" *Bulletin of the Atomic Scientists* 69:5 (2013): 21–31, https://doi.org/10.1177/0096340213501370. Jeanette Hofmann, Christian Katzenbach, and Kirsten Gollatz, "Between Coordination and Regulation: Finding the Governance in Internet Governance," *New Media & Society* 19:9 (2017): 1413–1414, https://doi.org/10.1177/1461444816639975.

[6] Michel J.G. van Eeten and Milton Mueller, "Where Is the Governance in Internet Governance?" *New Media & Society* 15: 5 (2013): 729–32, https://doi.org/10.1177/1461444812462850.

"massively distributed authority and decision-making power" that contribute to managing a complex, decentralized, and distinctly global Internet.[7]

This essay contributes to this challenge by analyzing how technical organizations such as RIRs generate what is described here as *operational epistemic authority*: the ability to produce, evaluate, and sustain credible and legitimate knowledge about the on-the-ground, day-to-day operation of a complex system by those actors who contribute to the independent functions of that system.

This article argues that the RIPE NCC and similar organizations facilitate creating and sustaining operational epistemic authorities among operator communities, in particular through what has been called "rough consensus." Rough consensus is a process wherein technologically knowledgeable experts openly debate solutions (manifest here as rules such as resource policies in the RIRs and standards in the IETF) intended to contribute to, improve, or sustain overall system function. Discussion and debate iteratively refine solutions until they have the support of most (hence the "rough") of those who participate in this process. Solutions are considered authoritative *within* these technical communities precisely because they are these actors' best effort at evaluating, articulating, and documenting their collective knowledge and experience related to the function of the complex system at hand, in this case Internet infrastructure operations.

To ensure the integrity of the rough consensus process, and consequently the credibility of solutions created by this process, organizations such as the RIPE NCC are often delegated administrative authority to ensure the knowledge creation process itself is not biased. Part of this administrative authority, as illustrated by the *Resolution*, is to ensure that rules are based on operational experience, not political motivations considered outside the core function of the system. Thus, in this context, administrative authority is about ensuring the integrity of the knowledge creation and rulemaking process. Rulemaking authority, which is exercised by the community through rough consensus, is about *evaluating* the knowledge and evidence contributing to the substance of rules that shape the system.

While rough consensus is sufficient for generating authority within the functional organizations that must work together to keep the internet running, it does not necessarily generate authority for outsiders who lack the technical literacy to participate in the consensus-building process. Rather, the origins of this form of authority in the early Internet emerged from choices to encourage decisionmaking among technical communities, then took on a life of its own as private authorities began formalizing processes in the absence of public guidance. As public authorities take more interest in how this critical infrastructure functions, understanding the differences between operational epistemic authorities and traditional rule-making processes will be critical to reconciling two different, but arguably complementary modes of governance.

This essay explains how rough consensus works, and how operational communities might work more effectively with the public policy stakeholders who must ultimately grapple with these decisions' effects on broader policy issues. Empirically, this work is based on a combination of archival analysis of the documents that are maintained by these communities (email lists, RIR policy documents), and, importantly, fieldwork in these communities that included passive participant observation and interviews. The interview subjects include leadership and participants in RIRs and network operator communities in all five internet registry regions, and the interviews focused on operational infrastructure governance (including variants of rough consensus) and the institutional economics of transnational cooperation within and across these communities.

In what follows, the concepts of operational epistemic authority and its relation to similar concepts in international relations are described and evaluated. Given these foundations, the historical origins of "rough

[7] Van Eeten and Mueller, 729, 730.

consensus" as a means of generating and sustaining this kind of authority, and how it functions within the operational community, is described. The conclusions discuss the need to find better ways for operational communities to engage with public policymakers, arguing for the notion of an operational policy interface akin to the well-known science policy interface in environmental studies.

Operational Epistemic Authority

The notion of operational epistemic authorities is similar to Peter Haas's "epistemic communities" in that they comprise transnational communities of experts who share "normative and principled beliefs," "causal beliefs," and "notions of validity" that shape their domain of knowledge.[8] Operational epistemic authorities are also similar to Michael Zürn's "pure epistemic authorities," in that they derive their authority from their reputation for engagement with transnational issue areas (such as human rights), monitoring capabilities, and, taken together, their credible assessment of knowledge of their respective issue domains.[9]

However, unlike Haas's and Zürn's concepts (respectively), operational epistemic authorities do not derive their authority from their influence advising governments and shaping the policies of nation-states.[10] Rather, their authority derives from the application of their collective knowledge to create, update, and sustain the norms, best practices, standards, and rules for coordinating and managing complex engineering systems "in the wild."[11] These footnotes should be merged into one.

As the stewards of a transnational network of largely private, independent networks, operational epistemic authorities must work in a coordinated fashion in order to effectively sustain a globally connected Internet.

[8] Peter M. Haas, "Introduction: Epistemic Communities and International Policy Coordination," *International Organization* 46: 1 (1992): 4, https://doi.org/10.1017/S0020818300001442.

[9] Michael Zürn, *A Theory of Global Governance: Authority, Legitimacy, and Contestation* (Oxford: Oxford University Press, 2018), 52-53, https://doi.org/10.1093/oso/9780198819974.001.0001.

[10] Zürn, 52–53.

[11] This notion of tacit knowledge is rooted in learning by doing, i.e., through experience in the field. D. C. North, *Institutions, Institutional Change, and Economic Performance*, Political Economy of Institutions and Decisions (Cambridge: Cambridge University Press, 1990) provides an excellent analogy: "[o]ne cannot learn to play a good game of tennis solely from a book, and even with practice there is an immense difference between players,"(74). Similarly, rough consensus is one means of sharing knowledge that improves everyone's "game." See also Richard R. Nelson and Sidney G. Winter, *An Evolutionary Theory of Economic Change* (Cambridge, MA: Belknap Press of Harvard University Press, 1982) for a discussion of tacit knowledge in ongoing engagements among actors and Michael Polanyi, *The Tacit Dimension*, 1st ed., Terry Lectures 1962 (Garden City, N.Y: Doubleday, 1966) for the original articulation of the concept. The application of experience over theory is rooted in both the management of common resource systems and early industrial research. Elinor Ostrom and Edella Schlager, "The Formation of Property Rights," in Susan Hanna, Carl Folke, and Karl-Goran Mäler, eds., *Rights to Nature: Ecological, Cultural, and Political Principles of Institutions for the Environment* (Washington, DC: Island Press, 1996), 127-156, highlights that "rules devised by resource users are based on years, decades, and sometimes centuries of experience…[s]uch information is gleaned while engaging in everyday…activities," (143). From the history of industrial research, operational actors share what Edwin T. Layton, in "Scientific Technology, 1845-1900: The Hydraulic Turbine and the Origins of American Industrial Research," [*Technology and Culture* 20, no. 1 (January 1979): 64–89, https://doi.org/10.2307/3103112], characterized as hydraulic engineer's "antipathy" for limited "tractable" models, focusing on experimentation to develop pragmatic responses that are not accounted for in existing models of the system (79). While the focus here is operational epistemic authorities managing shared resources critical to the function of the Internet's infrastructure, the principles are arguably applicable to the management of other network infrastructures. In particular, there are similarities to the coproduction of expertise and self-regulatory processes evaluated by Rebecca Slayton and Aaron Clark-Ginsberg, "Beyond Regulatory Capture: Coproducing Expertise for Critical Infrastructure Protection," *Regulation & Governance* 12:1 (2018): 115–130, https://doi.org/10.1111/rego.12168.

That is, they must achieve some level of internal consensus about how they interoperate. Additionally, they must constantly update their knowledge as they modify internet technologies and topologies. If operational epistemic authorities become ineffective at managing critical resources, they lose credibility both within the community and beyond it.

Operational epistemic authorities function as credible knowledge assessors, effectively monitoring and responding to changes in their knowledge domain, updating norms, rules, and best practices apace with those changes. Although limiting their decisions to system function, their decisions can (and often do) have distributional consequences that can impact issue areas that rely on Internet communication. These "downstream" implications include issue areas such as economic development, and as illustrated by the Resolution, international conflicts.

A perennial challenge facing operational epistemic authorities is defining the scope of their decision-making authority. Participants in the IETF and operational communities distinguish between the authority to make the rules shaping day-to-day activities (which is held by the broader community) and what is referred to here as administrative authority, such as the obligation to *facilitate*, but not interfere in, the rule-making by these communities themselves. In the cases below, administrative authority includes developing and maintaining organizations that *facilitate* rough consensus, *manage resource rights* based on rules created via rough consensus, and *ensure the integrity* of these governance processes.

Historically, operator communities have eschewed overstepping the boundary between their operational epistemic authority to sustain core Internet functions and the more conventional authority of "public" policymakers. The sections below present a very brief history of rough consensus to illustrate how operational epistemic authority is created and sustained within these communities before turning to the challenge of engaging and coordinating with public policymakers.

The Origins of "Rough Consensus" in the IETF

The US Defense Department funded the early development of the research network that would ultimately become the Internet, but left much of the development and decision-making to the researchers and engineers who developed early protocol standards, implementations, and network deployments.[12] As Janet Abbate has noted:

> …managers preferred to take the informal approach whenever possible. Having been researchers themselves, they subscribed to the view that the best way to get results in basic research was to find talented people and give them room to work as they saw fit. *They also tended to believe that differences of opinion could be debated rationally by the parties involved and decided on their technical merits...*[13]

This mode of management laid the foundations for what would later become "rough" consensus. By the early 1990's, although the IETF had emerged as the organization coordinating the development of TCP/IP (Transport Control Protocol/Internet Protocol), it was facing competition from the International Standards

[12] For detailed histories, see Abbate, *Inventing the Internet*, 2000 and Katie Hafner and Matthew Lyon, *Where Wizards Stay up Late: The Origins Of The Internet* (New Yor: Simon & Schuster, 1999).

[13] Abbate, 55, emphasis added. See also Yates and Murphy, *Engineering Rules*, in particular chapter 7 for their analysis of Internet standards development.

Organization (ISO) and its Open Systems Interconnection (OSI) protocol.[14] The "protocol wars" were driven by both technical and cultural differences. Andrew Russell suggests that the ISO's organizational culture "resembled contemporary democratic bodies insofar as it featured voting, partisan compromises, and rule-making behavior designed to protect financial interests."[15] The IETF considered its mode of decision-making to be better suited for technical decision-making. Russell notes that the IETF's distaste for ISO and the OSI model "stemmed from their frustration with the technical aspects of OSI as well as with ISO *as a bureaucratic entity*. Where TCP/IP was developed through *continual experimentation* in a *fluid* organizational setting, Internet engineers viewed OSI committees as *overly bureaucratic* and *out of touch* with existing networks and computers."[16]

In 1992, a subcommittee of the IETF, the Internet Architecture Board (IAB), proposed replacing IP addresses with addresses from the OSI model. Other members of the IETF objected not only on technical grounds but on procedural grounds. The IAB's proposal had not gone through the typical process of developing consensus. The ensuing "palace revolt" helped cement rough consensus as *the* mode of community-based decision-making in the IETF. Engineer David Clark was partly responding to the IAB's misstep at the IETF's annual meeting later that year when he famously declared: "We reject: kings, presidents, and voting. We believe in: rough consensus and running code."[17]

Clark's notion of "rough consensus" refers to agreement among most but not necessarily all members of the IETF actively participating in deliberation about standards or operating technologies. Such informal, consensus-based decisionmaking among scientists and engineers was not new when Clark articulated it. Rough consensus has been quietly contributing to the development of complex engineering systems, in domestic, international, and transnational contexts, since at least the late 1800s.[18] Clark's simple yet compelling articulation of this process was embraced by the IETF[19] and related operational communities that have since emerged, here in particular the RIRs and numbers communities.

Rough Consensus and the Obligations of the RIRs

Around the same time that Clark articulated the notion of rough consensus, the internet was being transformed from a research network managed by the United States' Defense Advanced Projects Agency and the National Science Foundation (NSF) into a commercial network whose physical infrastructure was largely deployed by corporations and whose core protocol development and coordination resources were governed by non-profit organizations.[20] As organizations and operational communities emerged to manage and

[14] TCP/IP is the combination of the Transmission Control Protocol and the Internet Protocol, both longstanding standards that provide the foundations for Internet communication.

[15] Andrew L. Russell, "'Rough Consensus and Running Code" and the Internet-OSI Standards War," *IEEE Annals of the History of Computing* 28:3 (2006): 48–61, https://doi.org/10.1109/MAHC.2006.42, 53.

[16] Russell, 53, emphasis added.

[17] Clark, "A Cloudy Crystal Ball," Slide 19 (Plenary Presentation presented at the 24th Meeting of the Internet Engineering Task Force, Cambridge, MA, 13-17 July 1992, https://groups.csail.mit.edu/ana/People/DDC/future_ietf_92.pdf).

[18] JoAnne Yates and Craig N. Murphy, *Engineering Rules: Global Standard Setting Since 1880* (Johns Hopkins University Press, 2019), 4, chapter 1.

[19] The IETF's form of rough consensus is itself codified (through the rough consensus process) in RFC 7282 *On Consensus and Humming in the IETF* (Pete Resnick, "On Consensus and Humming in the IETF," RFC (Fremont, CA, USA: Internet Engineering Task Force (IETF); RFC Editor; RFC 7282 (Informational), June 2014), https://doi.org/10.17487/RFC7282).

[20] Abbate, Inventing the Internet, 195-204.

coordinate functional domains previously managed by the NSF or documented in the IETF, they extended and adapted the traditions of rough consensus as a means rule-making via collective credible knowledge assessment.

At the same time as the modern commercial Internet grew and evolved, the IETF's focus shifted to protocols and standards, with more operational interests coalescing in network operator communities that focus on on address management, routing, and interconnection. As these function-specific communities emerged, they carried with them the ethos and decision-making processes established in the IETF. In 1994, one of the largest operator communities, the North American Network Operators' Group (NANOG) spun off from the Merit Corporation, which three universities formed in 1966 to facilitate academic networking, and which won a 1987 contract to work on the NSFNET. According to NANOG's own history, it emerged as "a forum to exchange technical information and discuss implementation issues among network service providers."[21] Similar network operator groups (communities) emerged in other regions as the Internet expanded and/or regional networks were connected to the Internet. In 1992, the European network operator community, the RIPE community, established the RIPE NCC in Amsterdam as the first of the modern RIRs.[22] Shortly thereafter, the Asia-Pacific Network Coordination Centre (APNIC) was formed in Australia to serve the Asia Pacific region.[23]

As community experience with managing Internet registries that delegate IP addresses grew, operators came together in the IETF to propose formalizing the criteria for establishing RIRs in what became RFC 2050.[24] These criteria, which were evaluated and refined via rough consensus, became the constitutional norms of the RIRs.[25] The criteria for establishing an RIR highlights bottom-up, epistemic governance: networking authorities (i.e. experts among those who manage network infrastructures and interconnection between these, who are typically not government actors) in the region must legitimize the organization as credible and capable to perform the functions of an Internet registry. For instance, RFC 1366 highlights that "the organization will commit appropriate resources to provide stable, timely, and reliable service to the geographic region."[26]

The remaining three of the five modern RIRs were formed after RFC 2050 was published in 1996. In 1997, the American Registry for Internet Numbers (ARIN) was formed in the United States to coordinate number resources for the United States, Canada, and some Caribbean and North Atlantic islands.[27] In 2002, the Latin American and Caribbean Internet Addresses Registry (LACNIC) was established in Uruguay to coordinate

[21] "A Community Built on Openness, Authenticity, and Collaboration," North American Network Operators' Group, 2023, https://www.nanog.org/about/our-story/.

[22] The Number Resource Organization NRO, "Regional Internet Registries" The Number Resource Organization, 22 April 2021, https://nro.net/about/rirs/.

[23] NRO.

[24] V. G. Cerf, "IAB Recommended Policy on Distributing Internet Identifier Assignment and IAB Recommended Policy Change to Internet "Connected" Status," RFC (Fremont, CA, USA: IETF (Internet Engineering Task Force); RFC Editor; RFC 1174 (Informational), August 1990), https://doi.org/10.17487/RFC1174. E. Gerich, "Guidelines for Management of IP Address Space," Request for Comments, Internet Request for Comments (Fremont, CA, USA: IETF (Internet Engineering Task Force); RFC Editor; RFC 1466 (Informational), May 1993), https://doi.org/10.17487/RFC1466.

[25] For revisions, see K. Hubbard et al., "Internet Registry IP Allocation Guidelines," Request for Comments (Fremont, CA, USA: IETF (Internet Engineering Task Force); RFC Editor; RFC 2050 (Historic), November 1996, https://doi.org/10.17487/RFC2050. Also R. Housley et al., "The Internet Numbers Registry System," RFC (Fremont, CA, USA: RFC Editor; RFC Editor; RFC 7020 (Informational), August 2013, https://doi.org/10.17487/RFC7020.

[26] Gerich, "Guidelines for Management of IP Address Space," 1–2.

[27] NRO.

number resources for Latin America and some Caribbean islands.[28] In 2005, the African Network Coordination Centre (AFRINIC) was established in Mauritius to coordinate number resources for Africa.[29] By the time the more well-known ICANN was founded in 1998,[30] and the IGF later in 2006,[31] the norms of "rough consensus" were well established among the operator groups coordinating core infrastructure functions.

While the network operator community agrees on broad goals such as providing prompt and reliable service to given regions, it also recognizes that, in practice, the core goals of an Internet registry "may sometimes conflict with each other or with the interests of individual end users, Internet service providers, or other number resource consumers." Because rules cannot be completely or perfectly specified in advance, "[c]areful analysis, judgment, and cooperation among registry system providers and consumers at all levels via community-developed policies are necessary to find appropriate compromises to facilitate Internet operations."[32] The IETF's rough consensus process was not just used to document IR norms. RIRs and the attendant communities adopted and adapted rough consensus as the foundation for their own policy development processes for the development of IP address delegation policies. The RIRs themselves served as the administrative authorities ensuring the integrity of those processes.

Reconciling conflicts among core goals and concerns about fairness to consumers and non-technical communities requires not only credible knowledge assessment of operational realities, but also careful deliberation of the scope of these decisions. Rough consensus' strength in expert evaluation is also a weakness: rough consensus is not always amenable to participation by non-technical stakeholders. The section below describes and evaluates the process of rough consensus within technical and operational communities before finally turning to the challenges of engaging members of more traditional public authorities.

Coming to Rough Consensus

Rough consensus' rejection of majoritarian voting, and the politics of vote trading that comes along with this model of decision-making, gives it credibility and legitimacy within technical communities.[33] Debates in the rough consensus process are rigorous and contentious, requiring deep knowledge of the topic at hand. These debates take place in public fora, in a combination of regular in-person meetings and on public e-mail lists in between in-person meetings.

Both the IETF and the RIRs use rough consensus to determine what constitutes credible contributions to standards and resource policy development (respectively). To participate in the consensus process, a participant must propose their contribution at an in-person meeting or via email lists that are established for collecting contributions. All proposed contributions brought to these fora must be addressed, regardless of

[28] NRO.

[29] NRO.

[30] ICANN, "ICANN's Early Days," ICANN History Project, 2023, https://www.icann.org/en/history/early-days.

[31] United Nations, "UN General Assembly Resolution 60-252," ITU, 27 April 2006, https://www.itu.int:443/en/wtisd/Pages/res60-252.aspx.

[32] R. Housley et al., "The Internet Numbers Registry System," RFC (Fremont, CA, USA: RFC Editor; RFC Editor; RFC 7020 (Informational), August 2013), https://doi.org/10.17487/RFC7020, 2.

[33] Arend Lijphart, Patterns of Democracy: Government Forms and Performance in Thirty-Six Countries (New Haven: Yale University Press, 1999).

whether the actor presenting that contribution is a large multinational corporation or a small network, a large group or an individual. If most of the participants in the discussion deem the contribution to be credible and that it does not disproportionately privilege (or disadvantage) a set of operational actors or technologies, the community standards documents are updated to reflect that contribution. This process iterates as contributions are introduced and debated until there are no more contributions (or contestations), and contributions are largely agreed upon by the participants.

The criterion for what it means to come to "rough" consensus is also critical to the credibility of the outcome and the community's long-term buy-in to the process, especially among those that may not agree with particular revisions that have contributed to an outcome. In contrast with majoritarian voting, consensus is not achieved by the first group that can rally a simple 51-49 majority. Votes are not counted in the majoritarian sense of picking winners, and intrinsically, losers. Participants cannot simply vote "no." Contestations in the negotiation of a rule (such as a particular standard in the IETF or a resource policy in an RIR) must be justified and evaluated in terms of community knowledge, experience, and operational implications. Contestations must be debated and, ideally, reconciled among those who participate in the process. Reconciling credible contestations means integrating the rationale for that contestation, such as correcting a deficiency in the current solution identified by participants or accounting for a corner condition, into the current solution, then presenting this integrated solution to the community for further scrutiny. To reach "rough" consensus, this iterative process of contributions, contestations, and revisions continues until all of the contributions have been considered and all of the contributions and contestations have either been evaluated and integrated, or dismissed.

If a participant (or group of participants) cannot justify the rationale for contesting a solution in reference to some combination of (1) existing community knowledge of the system, (2) existing best practices, and/or (3) evidence accepted as credible by the community, their contestations are dismissed. If a participant makes a regular habit of making these kinds of (incredible) contestations, their status as a credible contributor, and, subsequently, their status as an authority, diminishes. Former prestige as an authority and former accomplishments do foster some tolerance, but if these actors engage in invalid or incredible contestations, those contestations will be dismissed as well. Moreover, not all of these processes reach rough consensus. In some cases, the process may iterate for years. In other cases, if contestations cannot be reconciled, and there is little to no activity in in-person sessions or email lists on the issue, that particular consensus process may be abandoned with no resolution.

In these distinctly transnational communities, absent a hegemon ("we reject: kings, presidents"), the perceived integrity of the process by participants is exceptionally important. A number of participants interviewed indicated that even when some of their contributions are not used (either a recommended contribution or a contestation), they accept the result of the rough consensus process because they believe that their peers have heard and evaluated their ideas based on shared operations values rooted in technical efficiency, efficacy, and interoperability. Decisions based on rough consensus are ultimately considered by the community as authoritative not just because of their technical merits, but because the participants are normatively committed to rough consensus as a form of credible knowledge assessment.

When the community generates rough consensus, it transforms diffuse community knowledge into authoritative documentation of that knowledge that is manifest in standards, protocols, and/or resource policy. Haas's first three characteristics of epistemic communities are at play here. The community applies *shared* norms to critically evaluate new knowledge, even if it is initially presented by a minority. Principles and causal beliefs are significantly informed by operational experience. Finally, notions of validity are based on communal evaluation and confirmation through testing and experimentation in the wild ("running code"). Taken together, these confer credibility (or not) to claims of knowledge used to develop authoritative rules and create administrative obligations related to critical Internet resource management.

A consequence of this process is that epistemic authority is not hierarchically delegated. Rather, it is fluid, and may be accrued or eroded based on the communal evaluation of individuals' and organizations' credibility in their contributions to these debates. Individuals and organizations *gain* credibility, and consequently authority, within the community when they make constructive contributions that are deemed credible by the community, adding to or updating shared knowledge through accepted practices such as rough consensus. Individuals and organizations *lose* credibility when they consistently attempt to inject incredible contributions, and/or attempt to subvert or sidestep the rough consensus process.

Developing an Operational Policy Interface

Members of RIR communities are quick to highlight their position that "we do not make *public* policy, we make *[Internet] resource* policy."[34] These communities argue that their rules are strictly about operations and Internet resources. Decisions made through rough consensus do have distributional consequences, though. As governmental authorities increasingly look to influence the management of online communications and platforms, they are discovering Internet standards and operations as a potential instrument to achieve their goals. An emerging challenge is the development of an operations-policy interface that facilitates engagement between these distinct spheres of authority that promotes public goods without compromising the principles and practices that have sustained the governance of a distinctly liberal, transnational Internet.

An early strategy for developing an operations-policy interface was to invite public policy and regulatory actors directly into the consensus process. While collaborative in spirit, the technical and tacit knowledge necessary for effective and credible engagement in the rough consensus process is a distinct barrier to entry by non-experts. Many conventional policy makers find the rough consensus process at best frustrating, and, at worst, can perceive it as protectionist and actively alienating.

Within these operational communities, a number of actors in leadership, alongside longstanding participants, recognize the important role operational communities (as a collective) can play as "honest brokers,"[35] contributing technical knowledge and assessments in support of state actors developing evidence-based public policy. Although not labeled as such by these communities, there have been early efforts at developing operations-policy interfaces. Some RIRs have established formal and informal communication channels with law enforcement, supporting these actors learning how to use public facing infrastructure services (such as the registry of IP addresses) to more effectively investigate cybercrime. As a form of broader engagement, the RIPE NCC sustains an ongoing Round Tables forum for engaging with government representatives and regulators from its region on issues related to "the governance and operation of the Internet."[36] Similarly, APNIC has established the Cooperation Special Interest Group (SIG), whose charter indicates that it was established to "act as a forum to develop and clarify the APNIC community's position on issues of relevance to the public sector, or on matters for which a community position has been sought" and that it "should focus on information sharing, outreach, capacity building, and other activities that will advance APNIC's vision for a global, open, stable, and secure Internet."[37] At a recent IETF meeting in London, IETF and ISOC leadership, along with technical community members and public policy actors (each in their capacity as

[34] This language, and similar utterances with minor differences in phrasing, was used consistently across a number of interviews.

[35] Roger A. Pielke, Jr., *The Honest Broker: Making Sense of Science in Policy and Politics* (Cambridge, UK: Cambridge University Press, 2007).

[36] Réseaux IP Européens Network Coordination Centre, "Roundtable Meetings," RIPE Network Coordination Centre, 2022, https://www.ripe.net/participate/meetings/roundtable/roundtable-meetings.

[37] "Cooperation SIG," APNIC, 2022, https://www.apnic.net/community/participate/sigs/cooperation-sig/.

Internet community members, not as representatives of their respective organizations) actively discussed the challenges of more effective engagement between the technical community and state-based actors, with the objective of actively developing a more effective model of engagement.

While these are valuable steps that engage with the broader global governance system, the informal character of these relationships makes these relationships tenuous. Although there are certainly some actors in these communities that still eschew government engagement, policy entrepreneurs among leadership and the broader communities recognize the benefits of engaging as honest brokers for both the community and public policy making. Here, these communities' self-imposed aversion to "making public policy" can be turned to diplomatic benefit. Many of these actors do want to provide credible advice to state and international actors who are wrestling with distinguishing between where the operational impact of technical standards and resource policies end, and where the distributional consequences of these decisions and the role of public policy begins. Committing further analysis to developing an operations-policy interface between these spheres of authority, while remaining cognizant of the differences between these distinct, yet arguably complementary, modes of decision-making and governance can substantively contribute to greater integration of these valuable sources of knowledge, capabilities, and capacities into the global governance system.

"Enforcing Exclusion: Cybersecurity Expertise in Carceral Conditions" by James Shires, Chatham House

> Technology changes quickly, so imagine being in a coma for 10 years and waking up to use the newest technology. This is the reality for prisoners who have never seen or touched smartphones, tablets, or even laptops or computers.[1]
>
> "Access to a cell(phone) was a non-issue," said an inmate… "You see the guys out in the common areas... with their phones, and nobody bothers (them)… There are unused phones in walls, ceilings, buried outside, stashed in food, EVERYWHERE".[2]

Carceral conditions are places of deliberate exclusion from society, including its digital aspects. However, carceral conditions are also - as infrastructures, organizations, government agencies, and actors in capitalist markets—swept along by and participants in successive waves of digitalization. While prisons are the archetypal carceral condition, there are many others: juvenile "correctional" facilities, secure hospitals, immigration detention centers, and so on. Carceral conditions are also not the exclusive preserve of the state. Many are operated by private companies, while non-state actors, such as criminal gangs and terrorist organizations, also operate carceral sites. All these sites display the same tension between exclusion and digitalization evident in the two quotations above: the deliberate deprivation of liberty includes exile from our contemporary digital world, yet this exclusion is regularly circumvented by both the unauthorized and entirely legitimate use of information communications technologies. Carceral conditions are, in the words of Carolyn McKay, highly permeable.[3]

This essay argues that cybersecurity expertise operates differently in the context of incarceration than it does in other types of organizational contexts. Under the standard logic, cybersecurity expertise manages a trade-off between benefits and risks that is ultimately aimed at digital adoption, but in carceral conditions it becomes a means to facilitate or enforce digital exclusion.

This argument proceeds in four parts. It first lays out the standard logic of cybersecurity and suggests how carceral conditions challenge this logic. Second, it identifies four ways in which carceral conditions are increasingly digitalized: through surveillance technologies, contraband devices, managed access, and broader carceral infrastructures. Third, it uses examples of cybersecurity incidents to investigate the distinct role of cybersecurity expertise in carceral conditions. It concludes by reflecting on the relationship between carceral cybersecurity and its more familiar cousin.

Cybersecurity can be broadly defined as the prevention and mitigation of malicious interference with digital devices and networks. Such definitions have been criticized (including by this author) for their technological myopia, fetishizing digital objects above those that are affected by their function or lack thereof.[4] However,

[1] Reaz Ahmed et al., "Cons and Pros: Prison Education through the Eyes of the Prison Educated," *Review of Communication* 19:1 (2019): 69–76, https://doi.org/10.1080/15358593.2018.1555645.

[2] Meg Kinnard, "Inmate: Prison Officers Complicit in Cellphone Problem," AP NEWS, 24 April 2018, https://apnews.com/article/829196f22fb24cdc8336e3af83a402d2.

[3] Carolyn McKay, "Video Links from Prison: Permeability and the Carceral World," *International Journal for Crime, Justice and Social Democracy* 5:1 (March 1, 2016): 21–37, https://doi.org/10.5204/ijcjsd.v5i1.283.

[4] James Shires, *The Politics of Cybersecurity in the Middle East* (London, UK: Hurst, 2021).

while alternative visions and philosophies of cybersecurity—whether "human-centric," feminist, or other[5]—shift the ultimate rationale for cybersecurity towards the rights and freedoms of people rather than technologies, they often still seek to prevent and mitigate malicious interference by acting on the digital devices and networks themselves: i.e., securing them.

The idealized model of digital communication is one of seamless interconnection, where individuals work and live together across borders and despite geographic distance. Cybersecurity measures introduce friction for individual and organizational goals, including new layers of complexity, extra checks, and limits to functionality.[6] Making subversion more difficult thus also means making legitimate activity slower and less convenient. Even where cybersecurity measures do not directly introduce friction for network users, they do so at an organizational level. Diverting scarce resources to cybersecurity precludes investment in other areas. Of course, there are longer-term arguments that cybersecurity measures are economically beneficial, both for the organization and for a state or society, but these arguments do not deny that cybersecurity causes friction: they claim that the trade-off is worth it.[7]

The extreme case of cybersecurity-induced friction is disconnection. During or immediately after a cyber-attack, disconnection can be a rational short-term strategy. However, such actions are a last resort, precisely because the impact of disconnection is itself highly negative. Sometimes, disconnection is not an urgent response to a specific incident, but a broader security strategy. Infamously, highly sensitive digital networks are "air-gapped," meaning that there *should be* no link between them and the wider internet.[8] Many security and intelligence agencies—and increasingly, private cybersecurity companies and even non-governmental organizations (NGO)s—operate Sensitive Compartmented Information Facilities (SCIFs), which are designed to prevent accidental connections to exterior networks and wireless electronic surveillance.[9] In very different contexts, spies, criminals, and terrorists seeking to counter electronic surveillance institute similar disciplines of disconnection: leaving devices off with the battery removed; or turning them on only briefly and then driving quickly elsewhere (e.g. to prevent a drone strike based on geo-location).

Crucially, the decision to disconnect is the result of the same trade-off between benefits and risks. In this way, the fundamental logic of cybersecurity is preserved even in settings where individuals or organizations deliberately choose to forgo the benefits of digital technologies. There is an old adage in cybersecurity that the only way to be 100 percent secure is to turn off your computer and lock it away. This is funny because it is unthinkable, turning the logic of cybersecurity on its head and leading to what—for most inhabitants of the digital world and certainly most cybersecurity experts—is a clearly absurd conclusion.

Carceral conditions are, stereotypically, sites of *intentional* digital exclusion. The other main form of intentional digital exclusion is the internet shutdown, which usually occurs when a government sees internet access as a

[5] Ronald J. Deibert, "Toward a Human-Centric Approach to Cybersecurity," *Ethics & International Affairs* 32:4 (ed 2018): 411–424; Julia Slupska, "Safe at Home: Towards a Feminist Critique of Cybersecurity," *St. Anthony's International Review* 15:1 (2019):83-100, https://papers.ssrn.com/abstract=3429851.

[6]For a more in-depth discussion of the concept, see Anna Lowenhaupt Tsing, *Friction* (Princeton: Princeton University Press, 2004).

[7]Jon Randall Lindsay, "Restrained by Design: The Political Economy of Cybersecurity," *Digital Policy, Regulation and Governance* 19:6 (2017): 493–514, https://doi.org/10.1108/DPRG-05-2017-0023.

[8] Notoriously, this included the Iranian uranium enrichment facility in Natanz. Kim Zetter and Huib Modderkolk, "Revealed: How a Secret Dutch Mole Aided the U.S.-Israeli Stuxnet Cyberattack on Iran," Yahoo! News, 2 September 2019, https://perma.cc/3AB6-AX8T.

[9] James Shires, "Cyber-Noir: Cybersecurity and Popular Culture," *Contemporary Security Policy* 41:1 (2020): 82–107, https://doi.org/10.1080/13523260.2019.1670006; Ronald J. Deibert, *Reset: Reclaiming the Internet for Civil Society* (Canada: House of Anansi Press Ltd, 2020).

facilitator for protest, unrest, or oppositional organizing. There are important differences between internet shutdowns and carceral conditions. First, it is unhelpful to conceive of country-level shutdowns as a single "site," as they require coordination between multiple internet service providers, whereas carceral conditions are more tightly bounded and clearly bordered geographic locations.[10] More importantly, internet shutdowns follow the friction-based logic of cybersecurity above, albeit with a different conception of who is protected from what. Internet shutdowns are usually temporary and reversed as soon as possible, with "essential" functioning preserved.[11]

Of course, there are many other places where the internet is unavailable or access is intermittent. In 2022, there were five billion internet users worldwide, meaning that 37 percent of the global population does not have regular access to the internet.[12] One can trace the reasons for this "digital divide" both at a global level and in individual countries to uneven economic flows and structural intersectional discrimination, especially the tendencies of market-based capitalism to ignore hard-to-reach areas and marginalized communities. But such digital exclusion is largely a result of structural factors, rather than an outcome intended by specific individuals or organizations, and so is not considered further here.

Despite their exclusionary stereotype, the permeability of carceral conditions can be categorized into four main forms (Table 1).

Table 1: the permeability of carceral conditions to digital technologies

Permission/User	Authorized	Unauthorized
Incarcerated individuals	*1) Managed digital access*	*3) Contraband digital devices*
Guards and staff	*2) Digital infrastructure*	*4) Leakage of surveillance data*

The simplest form of permeability is 1) the provision of managed digital access for incarcerated individuals. Such provision follows longer histories of communication between carceral sites and the outside world, from letter writing to prison phone calls.[13] Managed access to the internet can take the form of internet rooms, kiosks, tablets or computers installed in cells, or even a "PrisonCloud:" a single internet infrastructure enabling access to legal services, monitoring, and recreational internet use.[14] Here, the risk of circumvention of security controls is connected to several thorny moral debates, where the demands of cybersecurity rub against the public perception of incarceration as punishment, its integration into the broader justice system, and its potential as a tool for rehabilitation and reform.

The second form of permeability is the adoption of digital technologies by carceral operators. Especially when privatized, they seek to reduce costs and achieve economies of scale by streamlining and digitalizing

[10] Indeed, carceral conditions often constitute national borders. Louise Amoore, "Biometric Borders: Governing Mobilities in the War on Terror," *Political Geography* 25:3 (2006): 336–351.

[11] Larry Greenemeier, "How Was Egypt's Internet Access Shut Off?," *Scientific American*, 28 January 2011, https://perma.cc/T3LM-E88P.

[12] "Internet usage worldwide - Statistics & Facts," statista, https://www.statista.com/topics/1145/internet-usage-worldwide/.

[13] Steven J. Jackson, "Ex-Communication: Competition and Collusion in the U.S. Prison Telephone Industry," *Critical Studies in Media Communication* 22:4 (2005): 263–280, https://doi.org/10.1080/07393180500288329.

[14] Aysha Kerr and Matthew Willis, "Prisoner Use of Information and Communications Technology," *Trends and Issues in Crime and Criminal Justice* (Canberra: Australian Institute of Criminology, 2018), 11.

everything from food provision to laundry via digital management.[15] Various terms for this have emerged, from Carolyn McKay's "carceral automaton," to Constantine Gidaris' "techno-carcerality."[16] Here, digital technologies form part of the background infrastructure that is necessary for carceral sites to operate in contemporary society—and, through individual tracking and monitoring devices, to transgress their geographical boundaries. The logic of cybersecurity thus operates as it does everywhere else, with the benefits of digital adoption weighed against its risks, and cybersecurity measures undertaken based on their cost, perceived effectiveness, and regulatory or public pressure: in short, on their friction.

Both 1) and 2) are permitted forms of permeability, in that they are accepted—even embraced—by the operators of carceral sites. While there are some distinct characteristics, in general cybersecurity logics operate here as elsewhere in society. This, I argue, is not the case for 3) and 4), the two unauthorized forms of permeability to which I now turn. The examples discussed below are taken from carceral conditions worldwide—in the US, Syria, Iran, Russia, Egypt, and China—and it is important to recognize that there are significant differences between these sites.[17]

Regarding contraband digital devices, discourses on prison violence often revolve around calls for better detection and prevention of contraband cellphone communication. In April 2018, an episode of severe violence, which was described as the "worst US prison riot in the last 25 years," took place at Lee Correctional Institution, a maximum-security prison in South Carolina.[18] Cellphones were a contributing factor in two ways: first, because the cellphone contraband market became an object of dispute between rival gangs (costing up to $1,000 dollars each); and second, because cellphones enabled the continuation and intensification of outside gang competition within prisons.[19] This was the most violent but not the only incidence of contraband cellphones being connected to violent acts, including extortion, scams, and murders.[20]

Various forms of cybersecurity expertise can be deployed to counter the rise in contraband devices. South Carolina's preferred technological solution was a selective jamming apparatus manufactured by the Maryland

[15] Ruha Benjamin, ed., *Captivating Technology: Race, Carceral Technoscience, and Liberatory Imagination in Everyday Life* (Durham: Duke University Press Books, 2019).

[16] Carolyn McKay, "The Carceral Automaton: Digital Prisons and Technologies of Detention," *International Journal for Crime, Justice and Social Democracy* 11:1 (March 1, 2022): 100–119, https://doi.org/10.5204/ijcjsd.2137; Constantine Gidaris, "Rethinking Confinement through Canada's Alternatives to Detention Program," *Incarceration* 1:1 (July 1, 2020), https://doi.org/10.1177/2632666320936436. See also https://www.carceral.tech/.

[17] Prisons or detention facilities in these states operate under different rules and norms, and the kinds of violence inflicted routinely on inmates—and, occasionally, guards and other staff—also vary significantly. Most simply, the conditions in these carceral sites range from overcrowding, lack of hygiene, and brutal violations of human rights to more indirectly violent forms of separation, community, and gang rivalry. Carceral conditions in these states also have different genealogies, with contemporary manifestations and consequences, such as racialized over-incarceration of Black men in the US, and differently colonial origins in Iran and the Middle East.

[18] Victoria McKenzie, "Are Cellphones Really to Blame for Spike in S.C. Prison Violence?," CBS News, 19 April 2018, https://www.cbsnews.com/news/south-carolina-prison-riot-are-cellphones-really-to-blame-for-spike-in-s-c-prison-violence/.

[19] The previous year, around 7000 cellphones or cellphone parts had been confiscated from carceral facilities in the state (up from around 2000 in 2008). Seanna Adcox, "Supplying Cellphones to South Carolina Inmates Could Bring a $10,000 Fine, plus 10 Years in Prison," *South Carolina Post and Courier*, 30 January 2018, https://www.postandcourier.com/politics/supplying-cellphones-to-south-carolina-inmates-could-bring-a-10-000-fine-plus-10-years/article_845d8102-05fb-11e8-956a-d7b6a75f5e7f.html.

[20] Kirk Brown, "Wireless Companies Are Hampering the Call-Blocking System at an SC Prison, Director Says," *The Greenville News*, 11 May 2019, https://www.greenvilleonline.com/story/news/local/2019/05/11/wireless-companies-hamper-call-blocking-sc-prison-director-says/1157792001/.

company Tecore, costing $500,000 a year.[21] In 2021, the Federal Communications Commission (with a chequered history in conflicts of interest regarding prison technology companies) agreed to expand federal permissions for prisons to use jamming technologies; this time in coordination with cellphone companies for specific phones, rather than a complete block.[22] Elsewhere, a former Lee prison official turned cybersecurity consultant argued there is an "urgent need for [cybersecurity company Cellebrite's] digital intelligence" software to analyze confiscated cellphones, suggesting that they provide valuable intelligence for solving crime and preventing violence.[23] These cybersecurity measures are not designed to smooth digital adoption or preserve individuals' or organizations' digital access, but to collect intelligence and enforce exclusion.

Turning to the second form of unauthorized permeability, carceral surveillance leaks are an increasingly frequent occurrence. The state's urge to document, classify and record is evident even in the most shocking of cases, such as the "Caesar" photos of over 50,000 tortured and murdered civilian detainees taken in military intelligence prisons in Damascus from 2011-2013, at the start of the Syrian civil war. These photos were taken by a military photographer tasked with documenting prison deaths, who transferred files to USB sticks, shared them with a collaborator and backed them up online, and eventually left Syria to provide testimony against the regime.[24] Even when the state does not have this urge, individuals often do; videos by guards of migrants rioting in Libyan migration centers circulate on social media, and Russian and Egyptian prison guards intimidate detainees and their families by recording instances of rape and abuse.[25]

In some cases, these leaks result from a cybersecurity compromise. In August 2021, hackers leaked images and video from the notorious Evin prison in northern Tehran, including CCTV videos of mistreatment, beatings, and displays of callous ignorance.[26] In February 2022, there was another hack-and-leak of the largest state prison in Iran. As well as posting video of the prison control room, these hackers also released documents with the charges of around 2,000 inmates, including many imprisoned for attending protests,

[21] In 2019, this solution notably failed to prevent Lee inmates from streaming on Facebook Live from the prison; a failure officials put down to either "tampering" or a lack of communication regarding frequencies between mobile providers and Tecore, who failed to adjust their detection accordingly.

[22] Meg Kinnard, "FCC to Consider Cellphone Blocking Options for State Prisons across U.S.," Associated Press, 10 July 2021, https://www.theitem.com/stories/fcc-to-consider-cellphone-blocking-options-for-state-prisons-across-us,366931.

[23] Bolchoz, "How Data Found on Contraband Cell Phones Is Being Used to Thwart Criminal Activities Inside and Outside of Corrections Facilities - Cellebrite," Cellebrite Blog, 3 March 2022, https://cellebrite.com/en/how-data-found-on-contraband-cell-phones-is-being-used-to-thwart-criminal-activities-inside-and-outside-of-corrections-facilities/.

[24] Garance le Caisne, "'They Were Torturing to Kill': Inside Syria's Death Machine," *The Guardian*, 1 October 2015, https://www.theguardian.com/world/2015/oct/01/they-were-torturing-to-kill-inside-syrias-death-machine-caesar.

[25] Ian Urbina, "The Secretive Prisons That Keep Migrants Out of Europe," *The New Yorker,* 23 November 2021, https://www.newyorker.com/magazine/2021/12/06/the-secretive-libyan-prisons-that-keep-migrants-out-of-europe; Staff Report, "Ex-Inmates Reveal Details of Russia Prison Rape Scandal," BBC News, 9 August 2022, https://www.bbc.com/news/world-europe-62465043; Orla Guerin, "The Shadow over Egypt," BBC News, 23 February 2018, https://perma.cc/B5UW-PZKE.

[26]Jon Gambrell, "Leaked Footage Shows Grim Conditions in Iran's Evin Prison," AP NEWS, 23 August 2021, https://apnews.com/article/technology-health-religion-iran-prisons-01dfade61d7a706d630bf83d30d8cb02. Many political prisoners are reportedly held in Evin, and conditions are notoriously bad—many wings are run by the Iranian intelligence agencies and Revolutionary Guards, and its feared reputation has developed since it was built in 1971, under the British- and US-trained secret police of the Shah. The Evin control room was reportedly running Windows 7, which is no longer supported by Microsoft and so is extremely insecure.

other minor political activities, or even on suspicion alone.[27] In China, documents detailing violent carceral practices against the Uighur minority in Xinjiang were reportedly obtained by someone who hacked police computer servers in the Xinjiang region and sent the decrypted documents to a US-based China scholar before releasing them to the media in May 2022.[28] Notably, after previous leaks, the Chinese government issued a directive for police departments to improve their cybersecurity, securing databases and restricting access to information. Cybersecurity expertise thus not only ensures that prison systems run, but also seeks to prevent leaks from these systems. The less secure the surveillance systems, the harder it is for states to keep prisons far from scrutiny.

In sum, in carceral conditions, cybersecurity facilitates exclusion by underpinning the digital transition of prison infrastructure, limiting access to the internet for those incarcerated, and preventing leaks in and out of digital devices, communications, and databases. Cybersecurity expertise in these settings uniquely seeks to deny participation in digital society to those who are incarcerated, rather than minimize the risks of digital adoption. It follows a logic of exclusion, rather than one of participation.

At least since Michel Foucault's field-changing study of prison punishment and discipline, criminology has been attentive to the relationship between carceral conditions and societal structures more broadly.[29] On one level, this relationship is metaphorical: we can analyze or look at society *as if it were* a prison.[30] On another level, carceral governance transforms states and societies by reshaping social relations between those who are incarcerated and their families and networks, as well as providing expertise and technologies that are then transferred to schools, hospitals, care homes, corporations, and other institutions.[31] The carceral state, in this sense, is not just an analogy, nor even a political-economic term highlighting the dependence of national economies on incarcerated labor. It identifies a logic of borders, confines, surveillance—and resistance—that is far from unique to carceral conditions.

How far should we take this argument? It has clear analytical potential: if carceral logics extend beyond carceral sites, then cybersecurity expertise aimed at preventing or denying digital access, rather than facilitating or maximizing it, might exist in many other fields and social areas. However, extending arguments about carceral conditions into a more general point about limits and constraints in broader society, as Foucauldian analyses sometimes do, risks diluting the critique of the specific kinds of violence involved in and caused by incarceration.

[27] Tzvi Joffre, "Hacker Group Leaks Footage, Files from Iranian Prisons," *The Jerusalem Post*, 8 February 2022, https://www.jpost.com/middle-east/article-695837.

[28] John Sudworth, "Xinjiang Police Files: Inside a Chinese Internment Camp" *BBC News*, 24 May 2022, https://www.bbc.co.uk/news/resources/idt-8df450b3-5d6d-4ed8-bdcc-bd99137eadc3.

[29] Michel Foucault, *Discipline and Punish: The Birth of the Prison* (London: Allen Lane, 1977).

[30] For some societies, this is rapidly becoming a literal, rather than metaphorical condition; for a detailed discussion of the carceral quality of the occupied Palestinian Territories, see Amahl A. Bishara, *Crossing a Line: Laws, Violence, and Roadblocks to Palestinian Political Expression* (Stanford: Stanford University Press, 2022).

[31] These links include the transfer of surveillance practices and technologies between these locations; for discussion, see Madison van Oort, "Employing the Carceral Imaginary: An Ethnography of Worker Surveillance in the Retail Industry," in Ruha Benjamin, ed., *Captivating Technology: Race, Carceral Technoscience, and Liberatory Imagination in Everyday Life* (Durham: Duke University Press Books, 2019), 209-223. Notoriously, Amazon's current Director of Learning and Development for its warehouse employees started their career in private prisons. Jules Roscoe, "Amazon's Newest Training Exec Used to be a Private Prison Manager," *Vice,* 22 September 2022, https://www.vice.com/en/article/4ax4en/amazons-newest-training-exec-used-to-be-a-private-prison-manager. From the perspective of cybersecurity expertise, internal digital surveillance is an accepted part of the field due to the prevalent concept of "insider" threat.

Whatever direction this argument takes in future, the exclusionary logic of carceral cybersecurity demonstrated here contains two lessons for cybersecurity studies in general, and scholars of cybersecurity expertise in particular. First, it reminds us of the adjunct place that cybersecurity occupies as an enabler and follower of a broader digital transformation, preventing us from losing sight of the trade-offs between security, friction, and even disconnection that underpin most cybersecurity practices in pursuit of an unattainable ideal of complete cybersecurity. Second, it underlines that there are significant omissions in this progressive account, with considerable and often violent consequences for incarcerated individuals worldwide. Although exclusionary logics of cybersecurity expertise function in less visible, often deliberately minimalized or overlooked sites, understanding such logics is nonetheless vitally important.

"Varieties of Cybersecurity Expertise"

By Jon R. Lindsay, Georgia Institute of Technology

We often hear two very different narratives about expertise in cybersecurity. One is that it is a rare quality possessed by hackers and other technical wizards. The other is that low barriers to entry make cyber expertise almost irrelevant. These two views are contradictory: how can expertise be both precious and easy? Another problem is that they do not stand up empirically. If expertise is rare, then it should be hard to acquire or hire. But the cybersecurity industry is booming, employing over five million professionals worldwide, many of whom perform rather mundane tasks on a day-to-day basis. If expertise is unnecessary, then we should see a lot of disruptive cyber attacks. But after decades of warnings of catastrophic cyberwar, we see many low-level intrusions but few sophisticated cyber-physical attacks. Both narratives are misleading, as the essays in this collection reveal.

Pushing back on the first conventional wisdom (about hacker genius), the essays in this forum demonstrate there are actually many varieties of relevant expertise in this domain, and they tend to be distributed across social networks and enabling infrastructures. While there are many talented individuals in the cybersecurity business, genius is insufficient for expertise, which often depends in practice on a lot of mundane maintenance and repair work. Expertise is a feature of social organizations more than lone individuals, and as a feature of social organizations, expertise is often expressed through institutional tussles, political controversy, and conceptual misunderstandings across heterogenous groupings of experts.

Illustrating these points, Matt Goerzen highlights the fraught relationship between countercultural ideologies and the emerging security industry in hacker groups. James Shires gives us a refreshingly different take on cybersecurity expertise by exploring its manifestation in prisons, where the problem is more about information getting out than hackers getting in. Jesse Sowell explores the "operational epistemic authority" enacted by networks of professionals who manage internet resources on a day-to-day basis, relying on a "rough consensus" to establish internet standards and protocols. And Rebecca Slayton and Clare Stevens highlight frictions and boundary-negotiation work between experts on the security of information technologies (IT) that coordinate software interactions, and experts on the security operational technology (OT) that control manufacturing hardware and industrial processes. The guiding values and content of these very different varieties of expertise in these essays are not easily reduced to one master archetype, nor should they be.

Pushing back on the second conventional wisdom (about low barriers of entry), substantial expertise is necessary for doing anything at all in cybersecurity, both on the offense and defense. The barriers to entry on both sides of the cyber contest are often higher than is widely appreciated, because operations depend on the acquisition and maintenance of technical capabilities and the coordination of institutional work. Previous work by Slayton and Max Smeets, two contributors to this roundtable, has persuasively made this point by detailing the hidden work of offensive cyber operations.[1]

Illustrating these points in this roundtable, Andrew Dwyer focuses on the evolution of malware analysis work in three phases as analysts have become more dependent on automated infrastructure to make sense of threat activity. Smeets explores the role of social movements, as contrasted with state spy agencies, in coordinating independent cyber campaigns. And Ryan Ellis demonstrates that infrastructural reliance complicates the

[1] Rebecca Slayton, "What Is the Cyber Offense-Defense Balance? Conceptions, Causes, and Assessment," *International Security* 41:3 (2017): 72–109; Max Smeets, "Cyber Arms Transfer: Meaning, Limits and Implications," *Security Studies* 31:1 (2022): 65–91.

simple distinction between offense and defense, since hackers can coopt bug-bounty programs intended to enhance security.

Another form of cybersecurity expertise is on display in this roundtable: the expertise of scholars who turn their gaze to cybersecurity. They make different choices to foreground some phenomena and methods while backgrounding others. Here we thus have a heterogeneity of perspectives on and approaches to studying a domain that itself is distinguished by considerable institutional and infrastructural heterogeneity. For simplicity, I will group them into two camps of interpretive and policy approaches, although the boundaries between them are naturally blurry, as befits the subject. One group is most interested in helping us to understand how to understand the world while the other is most interested in how those understandings shape practical outcomes. One test for membership is the sort of expertise one needs to engage with each essay, as the interpretive approaches tend to assume more familiarity with critical scholarship while the policy approaches seem to be written with a broader audience in mind.

Leading off the interpretive group, Goerzen takes a historical approach to the emergence and development of the "antisec" community, providing a rich internal history of this motley crew, but attending less to their broader context, relevance, or impact. Dwyer takes a more ethnographic approach that draws on his own experience as a practitioner in malware analysis, previewing what one hopes will be a more detailed study of the dimly understood field of commercial threat intelligence. Shires, in his fascinating essay on prison cybersecurity, offers an analytic typology of the permeability of carceral conditions, developing a 2x2 matrix of actors (prisoners and guards) and uses (authorized or unauthorized). These three essays all ask us to look at unfamiliar communities to unsettle familiar assumptions, while they vary in the degree to which they stick closely to a descriptive narrative (Goerzen) or venture a more analytical framing (Shires).

Shading into the policy group, Slayton and Stevens show how the boundary work of expert communities has important implications for critical infrastructure regulatory policy. Ellis focuses on a specific problem of bug-bounty hacking, highlighting a general problem in cybersecurity where defensive resources and intentions are coopted to undermine cybersecurity, highlighting the risks of unintended consequences in cybersecurity. Sowell examines the complex process of internet governance at the operational level that establish working norms, a process quite different from that envisioned in high-level policy discussions of internet norms. Smeets, finally, delves into the work of the so-called Cyber Partisans in the context of the ongoing war in Ukraine, highlighting the role of grassroots organizations in an ostensibly state-centric conflict, and gesturing toward promising engagement with the literature on social movements. These four essays problematize distinctions that sit at the heart of different policy conversations, i.e., about the relationship between IT and OT, offense and defense, norms and operations, and, respectively, threat actors and movements. This adds important nuance to policy choices that shape the domain.

Expertise is both complicated and necessary. Indeed, the necessity of multiple forms of expertise for navigating the problems of cybersecurity is making it ever more complicated in practice. Cyberspace is a domain of interconnection fragmented by barriers, which means it inherently connects different expert communities and boundary management becomes inherently contested. Attackers can exploit these contested seams between communities, and defenders can try to sew them up, but in the process the social fabric becomes more entangled.

"De-essentializing Cybersecurity Expertise"
by Aaron Gluck-Thaler, Harvard University

What constitutes cybersecurity expertise? The essays presented here address this question not by offering rigid definitions but by showcasing the heterogeneous ways in which cybersecurity is enacted. On topics ranging from anti-statist hackers to the technologists who enforce carceral conditions, the essays demonstrate how cybersecurity expertise is compatible with distinct political projects. They similarly push on conventional narratives of how one becomes a cybersecurity expert, for example by completing a certificate program. The authors instead study how cybersecurity expertise has been cultivated in diverse ways, whether through delegating decision making to malware detection algorithms or through stealing bug-bounty reports. The practices considered in this forum invite us to rethink top-down theoretical and regulatory approaches to cybersecurity. However, as I will argue in conclusion, an emphasis on the heterogeneity of cybersecurity practices should not obscure how they are linked to a shared history. As some of the essays suggest, this is not a history of peaceful co-existence, but one that involves certain forms of expertise, particularly those amenable to the priorities of states and corporations, appropriating other forms of expertise.

The conception of cybersecurity expertise that emerges across these essays is untethered to a fixed set of practices, politics, or sites. Matt Goerzen's essay brings us directly into this framing by charting how an "unruly" form of expertise coalesced from the 1990s through the early 2000s in groups associated with the anti-security hacking movement and the hacker underground. These hackers were unified less by what they were for and more by what they were against: the "security industry."[1] Their anti-establishment rhetorical stance accommodated dissimilar political positions and was shared both by hackers concerned with security research becoming complicit with a "police state" and by those who believed that financial incentives would corrupt the values of security professionals. Goerzen examines how these hackers resisted the commercial appropriation of their access to and knowledge of computer vulnerabilities. Sensationalist hacks and irreverent proclamations that self-consciously eschewed respectability became ways to safeguard the character of a distinctive type of cybersecurity expert. Goerzen's account shows how these practices solidified into an enduring sensibility that was later taken up by others who sought to challenge corporate or national security-oriented views of security.

Why is cybersecurity expertise so fluid and capable of being bought or appropriated? Ryan Ellis's essay suggests how this might have something to do with the form of knowledge that is produced by cybersecurity experts, as well as with how that knowledge has been managed commercially. Building on earlier research with Yuan Stevens,[2] Ellis considers how the bug-bounty industry is increasingly amassing the principal fruits of cybersecurity research in the form of reports of exploitable and yet to be published vulnerabilities. Ellis shows that bug-bounty platforms have become an attractive target for hackers because they afford the opportunity to bypass traditional and more time-intensive paths for acquiring knowledge of vulnerabilities. Bug reports that codify existing knowledge, Ellis argues, enable this practice of "coopting expertise."

Max Smeets's essay defines the Cyber Partisans, a Belarusian hacking group, in terms of how it organizes and acquires know-how. To Smeets, the question of how members of the Cyber Partisans acquire their expertise should be central to how scholars theorize them. Smeets provides evidence that the group is a closely linked,

[1] For a complementary study, see also Matt Goerzen and Gabriella Coleman, "Wearing Many Hats: The Rise of the Professional Security Hacker," Data & Society, 14 January 2022, https://datasociety.net/library/wearing-many-hats-the-rise-of-the-professional-security-hacker/.

[2] Ryan Ellis and Yuan Stevens. "Bounty Everything: Hackers and the Making of the Global Bug Marketplace." Data & Society, 12 January 2022, https://datasociety.net/library/bounty-everything-hackers-and-the-making-of-the-global-bug-marketplace/.

grassroots collective of non-state actors with varying levels of technical expertise who collaborate with other local groups. As a result, Smeets argues that the collective does not easily fit into labels of "cyber proxies" or "intermediaries," and is similarly insufficiently captured by existing theoretical understandings of cybersecurity politics that privilege the state. To better understand such groups, Smeets argues that cybersecurity scholars can benefit from deepening their engagements with literature on resistance movements.

While Smeets suggests the theoretical dangers of neglecting how technical actors acquire expertise, Andrew Dwyer's account foregrounds the geopolitical stakes. Dwyer studies a hybridization of expertise between malware analysts and the machine-learning algorithms they employ, which identify malicious behaviour based on the often-opaque recognition of patterns within datasets of past signatures and decisions by other analysts. In Dwyer's reading, the infrastructure that enabled Kaspersky Lab's endpoint detection technology to scan and upload the National Security Agency's hacking tools was not exceptional. Regardless of whether the Russian government motivated or used Kaspersky to access these tools, Dwyer argues that this pervasive infrastructure can facilitate hacking operations between states and therefore requires a new geopolitical calculus.

The essay by Rebecca Slayton and Clare Stevens focuses directly on how top-down oriented policy and theoretical analytics inadequately account for the local specificity of cybersecurity practices. Studying approaches to managing the risks of operational technology (OT) and information technology (IT), Slayton and Stevens argue that attempts to coordinate between these two areas often assume an unchanging boundary between what distinguishes an OT from an IT expert. Slayton and Stevens instead argue that such boundaries are precisely the thing to be explained, namely, how and why experts construct and traverse different boundaries. The authors emphasize that this approach can shift attention away from formalized protocols and towards how cybersecurity expertise and authority are produced from the bottom-up.

Jesse Sowell's essay further underscores how technical communities can produce their authority through on-the-ground practices. Sowell studies the conventions of Regional Internet Registries (RIRs) and other organizations that govern internet infrastructure. Sowell shows how governance by these communities is carried out through a process of achieving "rough consensus" rather than obtaining a majority number of votes. Solutions become credible within RIRs when they are seen as reflecting the operational experience of those within the community. Technical expertise in this reading is inseparable from the cultural norms that are internal to a community of practice. These norms, Sowell is careful to note, have a history in other engineering communities. Attempts to bypass those norms, for example by introducing considerations that community members see as reflecting "political influence," in turn decrease a community member's credibility.

James Shires's contribution highlights a pitfall of essentializing cybersecurity expertise. Shires argues that cybersecurity expertise is driven by a specific logic in carceral environments that distinguishes it from other organizational contexts. Rather than being directed towards maximizing participation in digital environments while minimizing risk, Shires shows how this form of expertise enforces the exclusion of individuals who are incarcerated. Through practices like locating contraband devices and preventing leaks of compromising prison surveillance footage, these cybersecurity experts maintain the boundaries of prisons. Progressive accounts of cybersecurity as normatively desirable—as that which will enable seamless digital adoption—risk obscuring how cybersecurity can also enable violent practices.

The essays collectively present a forceful argument against fixed views of what it means to be a cybersecurity expert. In closing, I want to note that an emphasis on the heterogeneity and locality of expert practices should not come at the expense of studying how some practices have assimilated others. As Goerzen's essay indicates, antisec hackers were operating in a historical context where corporate and national-security oriented forms of expertise were creeping into and coopting the politics of different technical communities. This is not to say that militarized or commercial cybersecurity practices have extinguished all others. The legacy of antisec and the practices of the Cyber Partisans demonstrate ongoing and distinct traditions, as do experts who approach cybersecurity as an enabler of human and civil rights. Nonetheless, the priorities of states and corporations do disproportionately shape the possible forms that cybersecurity expertise can take. This is evidenced by the gargantuan corporate and military cybersecurity sectors; the increasingly brazen attempts by governments to tailor internet infrastructure policy to their needs, as noted by Sowell; and the bug-bounty programs, professional malware analysts, and carceral logics considered by Ellis, Dwyer, and Shires, respectively.

One way to account for why certain forms of cybersecurity expertise have become dominant is to acknowledge that expert practices are interrelated. The experts examined across these essays are part of a shared, and still understudied, history of conflict over the meaning of cybersecurity and how it should be practiced. In researching this broader history, it would be a mistake to assume fixed corporate or governmental priorities. Such a study should instead consider how cybersecurity practices have been co-extensive with the production of those priorities. How have disparate approaches to cybersecurity become flattened, coopted, or homogenized? How have different experts reinforced or challenged these practices of appropriation? How has cybersecurity policy to date encouraged or restricted the ascendancy of different forms of expertise? This forum provides much needed leverage to these questions by foregrounding the diversity of cybersecurity practices.