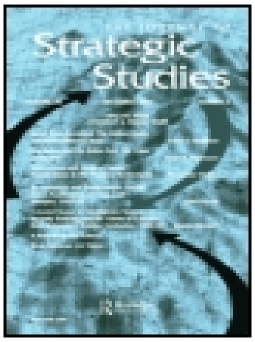

Journal of Strategic Studies

Routledge
Taylor & Francis Group

ISSN: (Print) (Online) Journal homepage: https://www.tandfonline.com/loi/fjss20

Publicly attributing cyber attacks: a framework

Florian J. Egloff & Max Smeets

To cite this article: Florian J. Egloff & Max Smeets (2021): Publicly attributing cyber attacks: a framework, Journal of Strategic Studies, DOI: 10.1080/01402390.2021.1895117

To link to this article: https://doi.org/10.1080/01402390.2021.1895117

Published online: 10 Mar 2021.

JOURNAL OF STRATEGIC STUDIES
https://doi.org/10.1080/01402390.2021.1895117

ARTICLE

OPEN ACCESS

Check for updates

Publicly attributing cyber attacks: a framework

Florian J. Egloff and Max Smeets

Center for Security Studies (CSS), ETH, Zürich, Switzerland

ABSTRACT
When should states *publicly* attribute cyber intrusions? Whilst this is a question governments increasingly grapple with, academia has hardly helped in providing answers. This article describes the stages of public attribution and provides a *Public Attribution Framework* designed to explain, guide, and improve decision making of public attribution by states. Our general argument is that public attribution is a highly complex process which requires trade-offs of multiple considerations. Effective public attribution not only necessitates a clear understanding of the attributed cyber operation and the cyber threat actor, but also the broader geopolitical environment, allied positions and activities, and the legal context. This also implies that more public attribution is *not* always better. Public attribution carries significant risks, which are often badly understood. We propose the decision maker's attitude towards public attribution should be one of 'strategic, coordinated pragmatism'. Public attribution – as part of a strategy – can only be successful if there is a *consistent* goal, whilst the avenues for potential negative counter effects are assessed on a *case-by-case* basis.

KEYWORDS Public attribution; attribution; cyber operations; cyber intrusions; deterrence; norms; intelligence

Introduction

When should states *publicly* attribute cyber attacks? This question has become increasingly important for decision making. The growing relevance of this question is partially due to the fact that states have become better at attributing cyber operations.[1] Attribution is – and remains – a tedious process.[2] But, contrary to conventional wisdom, a lot of the significant cyber activity has been attributed. It

CONTACT Florian J. Egloff florianegloff@ethz.ch Center for Security Studies (CSS), ETH Zürich, IFW, Haldeneggsteig 4, Zürich 8092, Switzerland

[1] This is often done in collaboration with the private sector. Dimitri Alperovitch, 'Stopping the Next Cyber Conflict', *The Cipher Brief*, (28 January 2018). https://www.thecipherbrief.com/column_article/stopping-next-cyber-conflict.

[2] For an excellent overview see: Thomas Rid and Ben Buchanan, 'Attributing Cyber Attacks', *The Journal of Strategic Studies*, 38/1–2 (2015), 4–37; David D. Clark and Susan Landau, 'Untangling attribution', in *Committee on Deterring Cyberattacks* (ed.), Proceedings of a Workshop on Deterring Cyberattacks: Informing Strategies and Developing Options for US Policy, (Washington, DC: The National Academies Press 2010), 25–40; and Earl Boebert, 'A survey of challenges in attribution', in: *Committee on Deterring Cyberattacks* (ed.), Proceedings of a Workshop on Deterring Cyberattacks: Informing Strategies and Developing Options for US Policy, (Washington, DC: The National Academies Press: 2010), 41–52.

is equally driven by governments' ongoing desire to shape the political and normative environment of cyber operations, vis-a-vis the realization that few measures have worked in the past. Public attribution is believed to be an important measure to help create a more stable cyberspace.

The Netherlands, for example, has identified public attribution as a cornerstone issue in their latest Cyber Defense Strategy.[3] As stated in the document; '[t]he increasing cyber threat requires a strong international response based on international agreements. That is still insufficient. The government wants to more frequently approach cyber attack perpetrators (publicly) about their behavior. [. . .] An active political attribution policy contributes to the deterrent ability and making the Netherlands less attractive as a target of cyber attacks. A state actor who (publicly) is held accountable for his actions will make a different assessment than an attacker who can operate in complete anonymity. The Netherlands thus contributes to combating impunity in the digital domain.'[4]

But does public attribution truly lead to a better deterrence posture and make countries a less attractive target? Is more public attribution always better? And what are the potential unintended consequences of public attribution?

Political science and security studies have addressed some elements of public attribution.[5] Some of have turned to game-theory to derive insights.[6] Others use historical examples from research on covert action to derive their claims. For example, we know that introduction of information into the public domain by itself does not have to lead to political effects. Rather, effects stem from a collective recognition that something is exposed and the politicization thereof.[7] States will also regularly shirk responsibility despite implausible deniability.[8] The political effect also depends on how resolved the attacking state is – exposure of a more resolved type being more likely to induce

[3] Ministerie van Defensie, 'Defensie Cyber Strategie 2018: Investeren in digitale slagkracht voor Nederland' (2018).

[4] Authors' translation.

[5] For general discussions on public attribution see: Florian J. Egloff, 'Contested public attributions of cyber incidents and the role of academia', *Contemporary Security Policy* 1 (2020), 55–81; Timo Steffens, *Attribution of Advanced Persistent Threats* (Springer: 2020); Florian J. Egloff, 'Public Attribution of Cyber Intrusions', *Journal of Cybersecurity*, 6/1 (2020), 1–12; Gil Baram, and Uri Sommer. 'Covert or Not Covert: National Strategies During Cyber Conflict', Paper presented at the 2019 11th International Conference on Cyber Conflict (CyCon), 28–31 May 2019. and Clement Guitton, 'Achieving attribution', *PhD Thesis*, (London: King's College London 2014).

[6] Benjamin Edwards, Alexander Furnas, Stephanie Forrest, and Robert Axelrod. 'Strategic Aspects of Cyberattack, Attribution, and Blame', *Proceedings of the National Academy of Sciences* 114/11 (2017), 2825–30; and Sandeep Baliga, Ethan Bueno De Mesquita, and Alexander Wolitzky. 'Deterrence with Imperfect Attribution', *American Political Science Review* 114/4 (2020), 1155–78.

[7] Lisa Stampnitzky, 'Truth and Consequences? Reconceptualizing the Politics of Exposure,, *Security Dialogue* 51/6 (2020), 597–613; and Thomas Eason, Oliver Daddow, and Rory Cormac. 'From Secrecy to Accountability: The Politics of Exposure in the Belgrano Affair', *The British Journal of Politics and International Relations* 22/3 (2020), 542–60.

[8] Rory Cormac and Richard J. Aldrich. 'Grey Is the New Black: Covert Action and Implausible Deniability', *International Affairs* 94/3 (2018), 477–94.

escalation.[9] Even after exposure, attackers and other interested parties continue to contest and shape the narrative around the public attribution claim, with some attackers also claiming responsibility themselves.[10] This raises significant challenges, particularly in democracies.[11] Thereby, deliberate non-acknowledgement of exposure can be used for escalation control.[12] Indeed, one of us theorized public attribution of cyber intrusions to represent a strategic activity to shape the operational space, with a particular aim to set the 'rules of the game'.[13] What all of this literature has in common is that it focuses on the goals and political outcomes of public attribution, but does not characterize the decision considerations in-depth nor provide a framework for public attribution. The dynamics of public attribution are still poorly understood. We aim to remedy that here.

As a first step to uncover these dynamics, we start by asking: what is required to make disciplined and high-level decisions regarding the public attribution of cyber operations by other actors? We define public attribution as the act to publicly disclose information about the malicious cyber activity to a machine, specific perpetrator, and/or ultimately responsible adversary.[14] We argue that public attribution is a highly complex process which requires trade-offs of multiple considerations. Effective public attribution not only necessitates a clear understanding of the attributed cyber operation and the cyber threat actor, but also the broader geopolitical environment, allied positions and activities, and the legal context. This also implies that more public attribution is *not* always better. Public attribution carries significant risks which are often badly understood. We propose decision makers' attitude towards public attribution is one of 'strategic, coordinated pragmatism'. Public attribution – as part of a strategy – can only be successful if there is a *consistent* goal, whilst the avenues for potential negative counter effects are assessed on a *case-by-case* basis.

[9] Jacob Otto and William Spaniel. 'Doubling Down: The Danger of Disclosing Secret Action', *International Studies Quarterly* (2020), 1–12 10.1093/isq/sqaa081

[10] Egloff, 'Contested Public Attributions of Cyber Incidents and the Role of Academia'; Florian J. Egloff, 'Cybersecurity and Non-State Actors: A Historical Analogy with Mercantile Companies, Privateers, and Pirates', DPhil Thesis (University of Oxford 2018), 144–168, 187–192; and Michael Poznansky, and Evan Perkoski. 'Rethinking Secrecy in Cyberspace: The Politics of Voluntary Attribution', *Journal of Global Security Studies* 3/4 (2018), 402–16.

[11] Marcus Schulzke, 'The Politics of Attributing Blame for Cyberattacks and the Costs of Uncertainty', *Perspectives on Politics* 16/4 (2018), 954–68. Not 'just' in democracies, as James Shires aptly demonstrates, see James Shires, 'Hack-and-Leak Operations: Intrusion and Influence in the Gulf', *Journal of Cyber Policy* 4/2 (2019), 235–56; James Shires, 'The Simulation of Scandal: Hack-and-Leak Operations, the Gulf States, and U.S. Politics', *Texas National Security Review* 3/4 (2020), 10–29.

[12] Austin Carson, 'Facing Off and Saving Face: Covert Intervention and Escalation Management in the Korean War', *International Organization* 70/1 (2016), 103–31; and Austin Carson, *Secret Wars: Covert Conflict in International Politics* (Princeton: Princeton University Press 2018).

[13] Egloff, 'Public Attribution of Cyber Intrusions'.

[14] Definition based on Lin's three levels of attribution. Herbert S. Lin, 'Attribution of Malicious Cyber Incidents: From Soup to Nuts', *Journal of International Affairs* 70/1 (2016).

We provide a *Public Attribution Framework* designed to explain, guide, and improve decision making of public attribution by states. The *Public Attribution Framework* distinguishes between goals and four categories – intelligence, incident severity, geopolitical context, and post-attribution actions – that act as enablers or constraints to pursue those goals. The combination of those enable careful decisions about whether to disseminate information about an adversary's actions to the public, to privately tell an adversary, or to restrict the knowledge of the intrusion to the government and potentially other partners.[15]

The remainder of this article is outlined as follows. Section II introduces the *Public Attribution Framework*, describing the different goals and enabling or constraining factors which governments should take into consideration when they decide to publicly attribute. To demonstrate the relevance and workings of the framework, Section III in turn provides an illustrative case study of the Dutch public attribution of the Russian attempted OPCW hack. The final section concludes and discusses avenues for future research.

A framework for public attribution

This section describes the *Public Attribution Framework*. We distinguish between the goals an actor is pursuing and four categories – intelligence, incident severity, geopolitical context, and post-attribution actions – that act as enablers or constraints upon these goals. This is also the order we follow in our discussion: we first highlight the diversity of goals that public attribution could serve and then discuss the enabling/constraining categories. We explain that each of these categories include a number of subcategories that need to be balanced to make careful decisions surrounding public attribution. In our discussion of the individual categories, we use the *ceteris paribus* assumption, that is all 'other things held constant'.

Thereby, we acknowledge that when responding to a particular intrusion, a government is likely to work through the enabling or constraining categories to prepare the response options serving the goals. Nevertheless, when considering where to place public attribution as a 'means', we deem it to be useful to reflect on the goals public attribution may serve *before an intrusion occurs*. This is to ensure that the maximum strategic value can be gained by the use of the means, rather than to be pushed into a responsive stance by adversarial action. Accordingly, we also first discuss the goals here.

The Public Attribution Framework is summarized in Figure 1.

[15] This article assumes that decision makers do not *intentionally* want to misattribute or misidentify adversaries.

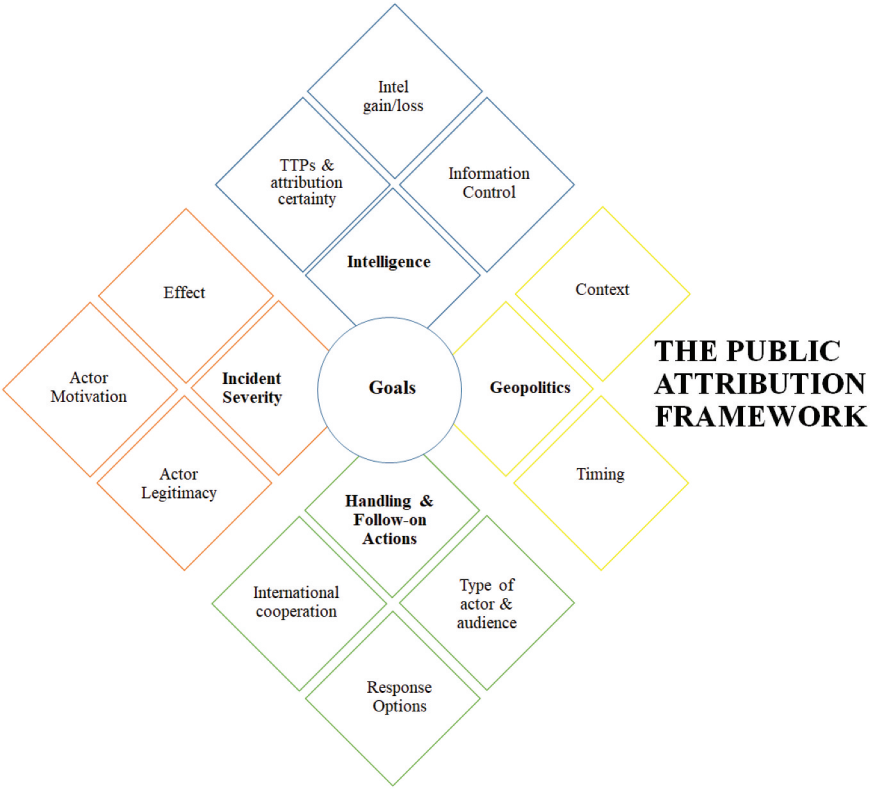

Figure 1. The public attribution framework.

Goals

Public attribution success can be defined as achieving the desired goals of a government. Public attribution is a means towards an end. Hence, by definition, there is no absolute measure of success, as it inherently depends on the goals set by the respective government (for an overview in table format, see Table A1 in the Appendix).

One objective could be *norm setting*,[16] that is clarifying and enforcing a set of standards for the appropriate behavior of states and other actors in cyberspace.[17] In less formal terms, it is about establishing the 'rules of the road' in cyberspace and beyond.[18] The Carnegie Endowment has created a repository of cyber norms

[16] Finnemore and Hollis differentiate between enforcement, constitution, defense, and deterrence, and note that this is not an exhaustive list of goals. We expand on their work here. Martha Finnemore, and Duncan B. Hollis. 'Beyond Naming and Shaming: Accusations and International Law in Cybersecurity', *European Journal of International Law* (2020), 969–1003.

[17] Definition based on Martha Finnemore and Kathryn Sikkink, 'International Norm Dynamics and Political Change', *International Organization* 52/4 (1998), 887–917.

[18] For an overview see: Joseph S. Nye, 'The Regime Complex for Managing Global Cyber Activities', *Global Commission on Internet Governance Paper Series*, 1 (2014), https://www.cigionline.org/sites/default/files/gcig_paper_no1.pdf

agreements since 2007.[19] It shows a wide range of mostly declaratory statements about what cyber norms states should adhere to from various government and non-government actors. Indeed, if anything, the repository shows there is no shortage of cyber norms agreements and initiatives.[20] But norms building can only succeed if norm violations are acknowledged. The idea is that public attribution may help to demarcate what is deemed to be appropriate behavior and can help ensure the adversary conforms to it.

Another principle objective of public attribution could be *coercion*, that is to deter or to compel.[21] Deterrence is conventionally defined as 'dissuading an adversary from doing something by threatening him with unacceptable punishment if he does it.'[22] In turn, compellence refers to one of two objectives: to get an adversary to do something (s)he has not yet, or to stop an activity undertaken by an adversary.[23] Coercion is one of the topics which has received the most attention in the cyber conflict.[24] Yet, most scholars are critical about the potential to deter or compel adversarial cyber activity.[25] Still, one could argue that public attribution could support coercive efforts both directly or indirectly. The very act of public coercion can change the (rational) cost-benefit calculus of the adversary through, for example, delegitimization and shame.[26] Or the disclosure of malicious activity may enable follow-on activity, such as the enforcement of sanctions, which in turn influences the incentive-structure of the adversary.

[19] Carnegie Endowment for International Peace, 'International Cyber Norms', https://carnegieendowment.org/specialprojects/cybernorms?lang=en

[20] A similar observation was made in: Martha Finnemore and Duncan B. Hollis, 'Constructing Norms for Global Cybersecurity', *The American Journal of International Law* 110/3 (2016), 425–479.

[21] Norm breaking or calling out can be linked to coercion, if there is an inherent cost to breaking the norm.

[22] Robert J. Art, 'To What Ends Military Power?' *International Security* 4/4 (1980), 3–35.

[23] Ibid. Also see: Schelling, Thomas C. *Arms and Influence* (Yale University Press 1966), 69–91.

[24] A systematic review of the literature shows that main topics of analysis in the literature between 1995–2019 were 'cyberwar', 'coercion, and 'norms'. Over 20 articles on coercion were published in the top Political Science journals since the mid 1990s. See: Max Smeets and Robert Gorwa, 'Cyber Conflict in Political Science: A Review of Methods and Literature', *2019 ISA Annual Convention* (Toronto, 2019, March). 10.31235/osf.io/fc6sg; For an overview of the cyber deterrence debate see: Stefan Soesanto and Max Smeets, 'Cyber Deterrence: The Past, Present, and Future', in Frans Osinga and Tim Sweijs (eds)., *Netherlands Annual Review of Military Studies 2020*, (2021), 385–400, 10.1007/978-94-6265-419-8_20

[25] Erica D. Borghard and Shawn W. Lonegran, 'The Logic of Coercion in Cyberspace', *Security Studies* 26/3 (2017), 452–8; Adam P. Liff, 'Cyberwar: A New "Absolute Weapon"? The Proliferation of Cyberwarfare Capabilities and Interstate War', *Journal of Strategic Studies* 35/3 (2012), 401–28; Timothy J. Junio, 'How Probable is Cyber War? Bringing IR Theory Back In to the Cyber Conflict Debate', *Journal of Strategic Studies* 36/1 (2013), 125–33; Adam P. Liff, 'The Proliferation of Cyberwarfare Capabilities and Interstate War, Redux: Liff Responds to Junio', *Journal of Strategic Studies* 134–38; Erik Gartzke, 'The Myth of Cyberwar: Bringing War in Cyberspace Back Down to Earth', *International Security*, 38/2 (2013), 41–73; Whilst often conflated, this discussion is distinct from literature looking at whether cyber operations can compel or deter. See: Max Smeets and Herbert Lin, 'Offensive cyber capabilities: To What Ends?' CyCon X: Maximising Effects, T. Minárik, R. Jakschis, L. Lindström (Eds.), (Tallinn: NATO CCD COE Publications 2018); and Max Smeets, 'The Strategic Promise of Offensive Cyber Operations', *Strategic Studies Quarterly*, 12/3 (2018), 90–113.

[26] For a more in-depth discussion on the possible link between norm and deterrence see: Joseph S. Nye Jr., 'Deterrence and Dissuasion in Cyberspace', *International Security* 41/3 (2017), 44–71.

Public attribution can also serve to cause friction, that is a *counter-threat* objective. Thus, one aims to ensure that the adversary has to spend valuable time and resources on capability-development and potentially follow the grueling counterintelligence leads.[27] An example is the case of the Swiss government disclosing detailed technical information about a specific operation by a threat actor aliased Turla, which potentially forced developers to go back to the drawing table and reconsider their use of certain tools and infrastructure. Interestingly, a research group at Columbia University examined the impact of disclosures on nine APT groups.[28] One of their main conclusions was that, 'If the disclosures discussed in the case studies were intended to deter similar future behavior, they unequivocally failed. [. . .] Were the intent to cause friction, then the hiatus seen in APT operations would suggest that disclosures were successful.'[29]

A closely related objective could be *prevention and defence*.[30] Spreading information about potential threats can be a means for such prevention and could create an enhanced defensive posture. As Anne Neuberger, then Director of the NSA's Cybersecurity Directorate, states about the NSA openly attributing the exploitation of a vulnerability to the Russian military intelligence service GRU, 'we chose to do it because we see that it makes targeted network owners more quickly patched and secure and build the resilience of their systems. Network administrators have way more vulnerabilities to address than they have time for, or frankly, money for and way more alerts than they can act on. So if we can say, "This particular vulnerability is being used by a nation-state intelligence service", we see network administrators moving quickly and addressing it. And that's our fundamental goal: our fundamental goal is improving cybersecurity.'[31]

Often overlooked, the public disclosure of malicious cyber activity by a state can also serve to create or enhance *community building* amongst

[27] Indeed, for this reason, activity such as uploading malware samples on Virus Total, a malware database now owned by Google, has been seen as an important instrument of the US new strategy of Persistent Engagement and Defend Forward. See: Michael Fischerkeller and Richard J. Harknett, 'Persistent Engagement and Cost Imposition: Distinguishing Between Cause and Effect', *Lawfare*, (6 February 2020), https://www.lawfareblog.com/persistent-engagement-and-cost-imposition-distinguishing-between-cause-and-effect; Michael Fischerkeller, 'The Fait Accompli and Persistent Engagement in Cyberspace,' *War on the Rocks*, (24 June 2020), https://warontherocks.com/2020/06/the-fait-accompli-and-persistent-engagement-in-cyberspace/

[28] This is one of the few existing studies which has actually tried to assess the impact of public disclosures of malicious cyber activity.

[29] Matthew Armelli, Stuart Caudill, John Patrick Dees, Max Eager, Jennifer Keltz, Ian Pelekis, John Sakellariadis, Virpratap Vikram Singh, Katherine von Ofenheim, and Neal Pollard, 'Named but Hardly Shamed: The Impact of Information Disclosures on APT Operations', *SIPA Capstone Project*, (Spring 2020), 94.

[30] This point is particularly well-addressed in the terrorism literature. See for example: Cynthia Lum, Lesly W. Kennedy, and Alison J. Sherley, 'The Effectiveness of Counter-Terrorism Strategies', *A Campbell Systematic Review*, (January 2006). For the roles of the state in cybersecurity, see: Myriam Dunn Cavelty and Florian J. Egloff. 'The Politics of Cybersecurity: Balancing Different Roles of the State', *St Antony's International Review* 15/1 (2019), 37–57.

[31] CBS News, 'NSA Cybersecurity Directorate's Anne Neuberger on protecting the elections', (19 August 2020), https://www.cbsnews.com/news/nsa-cybersecurity-directorates-anne-neuberger-on-protecting-the-elections/

relevant attribution actors.[32] Sharing sensitive information with other states and then acting together publicly is a political signal of shared threat perception. This may serve as a starting point for building more extensive response options beyond public attribution.

Finally, public attribution can serve *to enhance the domestic and/or international legitimacy or credibility* of actors involved in the attribution process. This is particularly important, as attribution capabilities are not observable to outside observers. A high-profile attribution case, receiving widespread attention in the media, can shed a positive light on the role of intelligence and security organizations. Government bureaucracies often compete for scarce resources. Public attribution might help to justify and secure budgets. It can also be part of a strategic effort to expand or protect the turf of the organization responsible for the disclosure.[33]

Four enablers and constraints for public attribution

This subsection discusses the four main enabling or constraining factors of public attribution: intelligence, incident severity, geopolitics, and response (i.e. handling and follow-on actions). For an overview and list of relevant questions pertaining to each factor see Table A2 in the Appendix.

Intelligence

The first set of factors of the *Public Attribution Framework* focus on intelligence, particularly the ability to collect and process information about foreign countries and their agents. We can distinguish between four factors within this category of intelligence: the degree of attribution certainty; the potential intelligence gains and losses; the tactics, techniques and procedures used by the intruder; and the ability to control relevant attribution information.

First, the degree of certainty about the intruder's identity and responsibility plays an important role in evaluating whether or not to publicly disclose information about a malicious act. Knowing who sat in front of the keyboard to conduct an intrusion is one thing. Knowing who is responsible for the cyber intrusion is another.[34] Healey has developed a spectrum assigning ten categories of state responsibility for a particular intrusion, shown in Table 1.[35]

[32] Egloff, 'Public Attribution of Cyber Intrusions'

[33] For a more theoretical discussion on bureaucratic competition see: Todd Kunioka and Lawrence S. Rothenberg, 'The Politics of Bureaucratic Competition: The Case of Natural Resource Policy', *Journal of Policy Analysis and Management* 12/4 (1993), 700–725.

[34] Lin, Attribution of Malicious Cyber Incidents.

[35] We adapted Healey's language of cyber attacks to cyber operations. The framework has been subsequently adopted and extended in Maurer. Jason Healey, 'Beyond Attribution: Seeking National Responsibility for Cyber Attacks', Atlantic Council: Cyber Statecraft Initiative, Issue Brief (2011), http://www.atlanticcouncil.org/publications/issue-briefs/beyond-attribution-seeking-national-responsibility-

Table 1. Spectrum of state responsibility conducting cyber operations.

Spectrum	Description
State-integrated	The national government conducts operations using integrated third-party proxies and government cyber forces.
State-executed	The national government conducts the operation using cyber forces under their direct control
State-rogue-conducted	Out-of-control elements of cyber forces of the national government conduct the operation.
State-ordered	The national government directs third-party proxies to conduct the operation on its behalf.
State-coordinated	The national government coordinates third-party attackers such as by 'suggesting' operational details.
State-shaped	Third parties control and conduct the operation, but the state provides some support.
State-encouraged	Third parties control and conduct the operation, but the national government encourages them as a matter of policy.
State-ignored	The national government knows about the third-party intrusions but is unwilling to take any official action
State-prohibited-but inadequate	The national government is cooperative but unable to stop the third-party operation.
State-prohibited	The national government will help stop the third-party operation.

According to Healey, '[t]he global national security community needs to shift resources from the technical attribution problem to solving the responsibility problem. This re-establishes state-to-state symmetry and enables a wider range of options open to sovereign nations: diplomatic, intelligence, military, and economic responses.'[36]

As one slides down the scale of state responsibility, it becomes harder to publicly attribute offensive cyber operations to a state, for at least two reasons. First, the state is less likely to have control over the outcome of the operation. Second, it becomes easier for the adversarial government to deny involvement outright. Equally, when a government has limited evidence to link an operation or campaign to a state it reduces the incentives for public attribution.

Second, a decisionmaker needs to balance the intelligence gains and losses following the public disclosure of malicious cyber activity. 'Whatever the context, disclosing what you know to an adversary always has downsides', as Aitel and Tait note.[37] Public attribution might provide insights into a government's attribution sources and methods; it possibly tells the adversary what you are capable of seeing. Entry points for intelligence collection might be lost following public attribution. Indeed, it is possible one agency might be engaging in information collection through tracking the threat actor, whilst another allied agency is calling out the actor. An actor, now

incyberspace; and Tim Maurer, '"Proxies" and Cyberspace', *Journal of Conflict and Security Law* 21/3 (2016).

[36] Healey, Beyond Attribution, 7.

[37] Dave Aitel and Matt Tait, 'Everything You Know About the Vulnerability Equities Process is Wrong', *Lawfare*, (18 August 2016), https://www.lawfareblog.com/everything-you-know-about-vulnerability-equities-process-wrong

knowing it is observed, might change its operations, stay silent for a period of time, or even completely abandon its activity.

The logical causal chain of operational disruption is this: the more technical details a government discloses about an operation, the more potential victims can a) spot the past actions of the actor's operations and b) protect against the specific means of operating disclosed. The actual disruption depends on the defensive uptake of the information, and the ease with which the operational tactics, techniques, and procedures can be adapted to evade such defensive practices.

The loss of intelligence – particularly for high-value targets – significantly reduces the incentives to attribute a cyber operation or campaign. However, the reverse can also happen: watching an adversary adapt to a non-expected public disclosure can lead to an intelligence gain, and if carefully managed, to more visibility into the adversary's actions as they struggle to clean up operations and find out how they were detected.

Furthermore, selectively revealing certain tactics, techniques, and procedures (TTPs) used in the cyber operation may positively affect visibility. TTPs is a frequently used concept in cybersecurity, describing the behavior of a threat actor.[38] Tactics refers to the tactical objective behind a certain set of activities by the threat actor. Techniques is about how the threat actor achieves a tactical objective. Procedures is about how the threat actor implements the technique to achieve an objective. A government may ask: where do we have most visibility? Where would we like to shift an actor's operations to? This would speak to deliberately revealing the TTPs where one has least visibility into, so as to incentivise shifting an actor's behaviour into a more visible space. This is an advanced counterintelligence tradecraft and assumes integration and steering of one's overall sensors and response toolkit with regard to a particular actor and beyond. Other actors will watch and learn from the disclosed information. Thus, to optimally profit from such an intelligence gain, the teams analyzing actors using similar operational techniques, or, where knowledge about other actors relies heavily on similar investigative techniques, ought to be pre-briefed about the imminent disclosure.

Third, a government is not always able to control what information about a cyber intrusion ends up in the public domain. Simultaneous tracking of (adversarial) cyber operations by different actors is not uncommon.[39] For

[38] MITRE, 'Tactics, Techniques, and Procedures', (12 September 2019), https://security.radware.com/ddos-experts-insider/hackers-corner/tactics-techniques-procedures/

[39] The risks of mutual tracking is higher than one may think, as intelligence agencies have a tendency and incentive to target and track the same entities. For an excellent discussion see Kaspersky, 'Spy Wars: How nation-state backed threat actors steal from and copy each other', (4 October 2017), https://www.kaspersky.com/about/press-releases/2017_spy-wars-how-nation-state-backed-threat-actors-steal-from-and-copy-each-other; For a discussion on the implications of national strategy of states also see: Max Smeets, 'US cyber strategy of persistent engagement & defend forward: implications for the alliance and intelligence collection', *Intelligence and National Security* 35/3 (2020), 444–453.

example, we know that both the Canadian intelligence agency (CSEC) as well as Kaspersky Lab were both tracking the espionage activities of the French General Directorate of External Security (also known as 'SnowGlobe' or 'Animal Farm').[40] Private sector actors – or academic research groups like the CitizenLab – might also disclose information about threat actors before a government wants to do this. Thereby, in contrast to their underreporting on civil society as targets, the private sector often discloses that government actors were targeted, even if they often remain abstract enough to not identify the particular victim department.[41] The key judgement is whether the government thinks it has traction on who is tracking the campaign and to what degree it plans to use outside publicity as a precursor or substantiator of its own attribution claims. Thereby, the secrecy of one's own victim status will often be honored by security companies, but not by the media. Media pressure can rise quickly and narrow the choices with regard to how and when to publicly attribute. Thus, having a plan of how to change one's communication strategy, should the fact of one's own victim status leak to the media, is essential. As campaigns often involve multiple victims across different countries, the likelihood of someone else having awareness of a campaign can be quite high. Overall, if the government has low confidence in their ability to control the flow of information about the intrusion, then a more proactive public information strategy is needed.

Fourth, there is another reason why it is important to consider the tactics, techniques, and procedures (TTPs) used by the intruder. As a general rule, there is more to gain by 'burning' previously unknown exploits and tools part of an advanced operation. If a government discloses a multi-million dollar operation using previously publicly unknown advanced tooling, there is a significant loss to the intruder when it comes to follow on operations.[42] It may have to go back to the drawing table and rebuild its platform. In turn, if an actor only uses publicly available tools and commonly used techniques, little is lost.

Incident severity

The second set of factors relates to the severity of the malicious activity. We consider three main aspects: the legitimacy of the responsible actor; the attacker's motivation; and the effect of the operation.

[40] INFOSEC, 'Animal Farm APT and the Shadow of French Intelligence', (8 July 2015), https://resources.infosecinstitute.com/animal-farm-apt-and-the-shadow-of-france-intelligence/#gref; and Kaspersky Lab, 'Animals in the APT Farm', (6 March 2015), https://securelist.com/animals-in-the-apt-farm/69114/.

[41] Maschmeyer, Lennart, Ronald J. Deibert, and Jon R. Lindsay. 'A Tale of Two Cybers – How Threat Reporting by Cybersecurity Firms Systematically Underrepresents Threats to Civil Society', *Journal of Information Technology & Politics* 18/1 (2021), 1–20.

[42] It could also be that multiple platforms used by an actor are connected. In that case, the public disclosure of one platform can be particularly damaging as it may subsequently disclose and burn related capability.

The goals of cyber operations range from espionage, subversion, to outright destruction.[43] Within these classes of activity, there are some activities that are deemed more legitimate, even legal under international law, as most states engage in similar activities. For example, espionage is illegal in every country in the world, but international law is largely silent on the permissibility of peacetime espionage. This does not mean states need to accept espionage, but it does mean that if a government intends to use public attribution as a means to explain their understanding of (il)legitimate actions in cyberspace, espionage cases may not be the best set of activity to start with.

There are two additional considerations when it comes to disclosure and the impact of cyber operations. First, when it comes to effect operations, there is often a disjuncture between intended effect and actual effect. When Robert Morris released one of the first computer worms over the internet in 1988, he did not intend to take down so many computer systems.[44] The worm's purpose was to measure the size of the Internet but a critical bug in the spreading mechanism transformed it into a highly disruptive attack.[45] Not only early, non-state cyber activity had unintended outcomes. Even the most advanced operations often have unintended effects. Stuxnet, for example, infected numerous computer systems in the United States and the rest of the world – whilst it did not trigger the Stuxnet payload, companies still spend millions of dollars to clean-up the malware in their systems.[46]

Second, most of the cyber activity takes place below the threshold of armed attack. Specific cyber operations are often linked into broader multi-year campaigns.[47] Individually, a certain operation may not seem geopolitically relevant enough to attribute. Yet, cumulatively, a set of connected operations may have significant strategic impact. This makes it harder to consider public attribution on a case-by-case basis. Doing this, one would be more likely to focus only on the highly disruptive events, and miss out on the long term set of activities.

[43] For a similar distinction see: Ben Buchanan, *The Hacker and the State: Cyber Attacks and the New Normal of Geopolitics* (Harvard University Press 2020)

[44] For more detailed discussion on intended effect and how to control it: Raymond, David, Gregory Conti, Tom Cross and Robert Fanelli, 'A Control Measure Framework to Limit Collateral Damage and Propagation of Cyber Weapons', in: K. Podins, J. Stinissen, M. Maybaum (Eds.), *5th International Conference on Cyber Conflict*, (Tallinn: NATO CCD COE Publications 2013),' 4; Bob Page, 'A Report of the Internet Worm', (7 November 1988). http://www.ee.ryerson.ca/~elf/hack/iworm.html; and Ted Eisenberg, David Gries, Juris Hartmanis, Don Holcomb, M. Stuart Lynn, Thomas Santoro, 'The Cornell Commission: On Morris and the Worm', *Communications of the ACM* 32/6 (1989), 706–709

[45] At least, this was stated by Robert Morris in court.

[46] Geoff McDonald, Liam O Murchu, Stephen Doherty, Eric Chien, 'Stuxnet 0.5: The Missing Link', *Symantec*, (26 February 2013), http://www.symantec.com/content/en/us/enterprise/media/security_response/whitepapers/stuxnet_0_5_the_missing_link.pdf; Ralph Langner, 'To Kill a Centrifuge: A Technical Analysis of What Stuxnet's Creators Tried to Achieve', The Langner Group, (November 2013), http://www.langner.com/en/wp-content/uploads/2013/11/To-kill-a-centrifuge.pdf; and Jon R. Lindsay, 'Stuxnet and the Limits of Cyber Warfare', *Security Studie*s 22/3 (2013), 365–404.

[47] Richard J. Harknett and Max Smeets, 'Cyber campaigns and strategic outcomes', *Journal of Strategic Studies*, (2020), 10.1080/01402390.2020.1732354

The legitimacy and motivation of the intruder – and potentially responsible party – matters too. Some actors can be named, but not shamed. Consider the hypothetical case of ISIS conducting a distributed denial of service attack against government websites of the United Kingdom. The malicious activity temporary overwhelms the website with traffic and shuts down from a lack of bandwidth. Some people cannot access the website. The attackers did not gain access to any government information. ISIS will not be 'shamed' if the U.K. government would release a statement or report about the malicious activity. Publicly calling out ISIS for their DDoS attack may only call more attention to their organization, and help to terrorize: the public may not immediately recognize that this was a relatively minor activity and it may spur fears that ISIS is capable of conducting much more disruptive or destructive cyber operations against national infrastructure.

Indeed, some cyber operations may be conducted with the strategic aim of being publicly attributed by the victim.[48] If this is the case, the government may want to consider an over-/under-reaction to thwart the adversary's expectations (see e.g. Salisbury response).

Overall, if the disclosed information contradicts an actor's identity constructions, it is likely to be perceived as negative, as the actor's reputation of its chosen identity suffers. If the disclosed information, however, supports an actor's chosen identity construction, then it is likely to be perceived as positive, as the actor's self-image is strengthened. If an actor wants to construct an identity as a state that can easily flout international norms, then violating a strong norm may be more effectful, and useful for the state. By contrast, a state that wants to construct an identity as a state that adheres to a shared normative framework would suffer a reputational hit from a public attribution showing its violation of the international norm.

Geopolitical situation

Third, the geopolitical situation inevitably plays a role in public attribution decision-making. We previously discussed the 'spectrum of responsibility' and argued that the level of state responsibility matters. We also argued that the legitimacy of the attackers is important for public attribution. From a geopolitical perspective, what is of course at least as important is the relationship between the attributing government and the intruder. After all, allied governments hack into each other's networks all the time. Most of the time this is for espionage purposes.[49] A common

[48] For a discussion on actors' considerations in revealing their *own* cyber operations see: Michael Poznansky and Evan Perkoski, 'Rethinking Secrecy in Cyberspace: The Politics of Voluntary Attribution', *Journal of Global Security Studies* 3/4 (2018), 402–416.

[49] For a longer discussion on whether cyber conflict should therefore be seen as an intelligence contest see: Joshua Rovner, 'Cyber War as an Intelligence Contest', *War on the Rocks*, (16 September 2019), https://warontherocks.com/2019/09/cyber-war-as-an-intelligence-contest/; and Robert Chesney and Max Smeets, 'Policy Roundtable: Cyber Conflict as an Intelligence Contest', *Texas National Security Review*

practice is third party collection: you spy on a strategic organization to use its data collection on your potential targets to your advantage.[50]

In cases when the intruder is not a close ally, not all considerations are intuitively accessible to decision makers. Some aspects will need in-depth country specific knowledge, for example, how an accusation enters the domestic politics of the accused country. How to deal with disagreements and accusations may also have a culturally specific connotation.

The history of interaction with the state in question sets the context for the current action.[51] Thus, a public attribution that was preceded by private attributions may be read by the attacker as an escalation.[52] Public attribution can also entrench relations of enmity and distrust, potentially leading to a tying hands effect.[53]

Second, timing matters greatly for the attribution decision, both operationally and politically. Operationally, one likely only wants to go public after having derived the maximum intelligence value out of observing an adversary. Politically, the question has to be asked whether there are other time-sensitive political agendas that attribution can support or be detrimental for? Attribution could act both as a leverage builder or as a spoiler for diplomatic discussions. For example, the United States government used public attribution as a leverage builder in the instance of the PLA indictment of 2014, where

(17 September 2020), https://tnsr.org/roundtable/policy-roundtable-cyber-conflict-as-an-intelligence-contest/

[50] There is even the potential practice of fourth party collection. As Guerrero-Saade and Raiu explain, 'Different agencies, be they friend or foe, have a purview over different geographical areas and desirable sectors. Their analysts are likely to have a greater acquaintance with desirable targeting and the context in which to interpret the information received from their collection. This presents a valuable opportunity: to co-opt the collection methods of a foreign intelligence service to receive the same raw information being collected on targets of interest to the latter – or ideally both – intelligence services; this practice is known as fourth-party collection.' Juan Andrés Guerrero-Saade and Costin Raiu, 'Walking in your enemy's shadow: When fourth-party collection becomes attribution hell', *Virus Bulletin Conference*, (2017, October), https://media.kasperskycontenthub.com/wp-content/uploads/sites/43/2018/03/07170728/Guerrero-Saade-Raiu-VB2017.pdf.

[51] Steffens, *Attribution of Advanced Persistent Threats.*

[52] A detailed discussion of potential escalation mechanisms in cyberspace goes beyond the focus of this article. For more research on this see: Ben Buchanan, *The Cybersecurity Dilemma: Hacking, Trust and Fear Between Nations* (Oxford: Oxford University Press 2017); Erica D. Borghard and Shawn W. Lonergan, 'Cyber Operations as Imperfect Tools of Escalation', *Strategic Studies Quarterly* (Fall 2019), 122–145, https://www.airuniversity.af.edu/Portals/10/SSQ/documents/Volume-13_Issue-3/Borghard.pdf; Sarah Kreps and Jacquelyn Schneider, 'Escalation Firebreaks in the Cyber, Conventional, and Nuclear Domains: Moving beyond Effects-Based Logics', *Journal of Cybersecurity* 5/1 (2019), 1–11, 10.1093/cybsec/tyz007; Jason Healey and Robert Jervis, 'The Escalation Inversion and Other Oddities of Situational Cyber Stability', T*exas National Security Review* 3/4 (2020); and Ben Buchanan and Fiona Cunningham, 'Preparing the Cyber Battlefield: Assessing A Novel Escalation Risk in a U.S.-China Crisis', *Texas National Security Review* 3/4 (2020).

[53] James D. Fearon, 'Signalling Foreign Policy Interests: Tying Hands Versus Sinking Costs', *The Journal of Conflict Resolution* 41/1 (1997), 68–90. On how public attributions can affect the underlying knowledge creation processes, see Florian J. Egloff, and Myriam Dunn Cavelty. 'Attribution and Knowledge Creation Assemblages in Cybersecurity Politics', *Journal of Cybersecurity* (2021, forthcoming), http://doi.org/10.1093/cybsec/tyab002.

the indictment was unsealed during the preparations for the presidential US-China summit.[54] It sped up the discussions around cybersecurity issues and gave it presidential priority.

Attribution can also influence the time-sensitive agenda of other policies and events. For example, the rule of law requires that the investigations are carried out without a prejudgment as to who perpetrated a crime. Hence, if a trial in court is envisaged, public attribution may be a hindrance to it. Potentially, this even affects basic human rights law. How can the intruder get the right to a fair trial when the state has already decided that you are guilty, e.g. through the cyber sanctions regime? A basic principle in law (*nulla poena sine lege*) dictates that you cannot be punished without law that criminalizes your behaviour. Thus, a government needs to ask whether the punishment itself is illegal (i.e. cyber sanctions) under general human rights law.

Handling and follow-on actions

Lastly, the determination *whether* to publicly attribute, depends on *how* an actor can publicly attribute and what set of actions can be taken following the disclosure. The first consideration for governments is whether public attribution enables potential follow-on policy or military responses. Of course, public attribution itself can be seen as a policy response. For example, consider Actor A conducting cyber operations against Actor B. Actor B might find out about the operation and the identity of the attacker. In private, Actor B could privately discuss the matter with Actor A and explain that this behavior cannot be tolerated. If this private dialogue does not have the intended effect, Actor B could publicly disclose the details of the intrusion, sending a signal to Actor A about the importance of the issue and willingness to raise the stakes.

Whilst not all policy responses require public attribution, it certainly helps for a great deal of cases. For instance, without public attribution it is difficult to call out/condemn state actors at international fora, such as the United Nations. Public attribution might also make it easier to rally support from allies for military or other responsive measures. A public attribution might also enable counter response; in other words, a policy response from the adversary.

Furthermore, the attributing actor needs to consider whether creating an environment, where public attribution becomes a regularity, is desirable for

[54] The United States Department of Justice, 'U.S. Charges Five Chinese Military Hackers for Cyber Espionage Against U.S. Corporations and a Labor Organization for Commercial Advantage: First Time Criminal Charges Are Filed Against Known State Actors for Hacking', (19 May 2014), https://www.justice.gov/opa/pr/us-charges-five-chinese-military-hackers-cyber-espionage-against-us-corporations-and-labor.

the long-term interests of the actor. By publicly attributing one legitimizes its use as a tool of statecraft. This consideration should shape both the decision of whether to publicly attribute and how exactly to do so. Thereby, the impact of non-attribution and non-public objection to activity observed by others should also be considered. Tolerating behavior over time can be read as acknowledging its legitimacy, which makes public objection all the more important.

Second, the larger international audience should be considered. We already discussed that malicious actors likely respond differently to the public disclosure of their cyber operations. But the actual attribution audience is bigger. Public attribution is not merely a dyadic activity: it goes beyond the attributing and attributed party, with potentially negative or positive consequences for international stability. For example, some have argued that, one of the reasons why Iran has ramped up their information operations in recent years is, because of the public media attention given to the Russian influence operations in the United States. The idea is that calling out of Russian activity created emulation effects: other countries realize there is a strategic space that might be exploited. Public attribution, in this respect, can reveal to an international audience what a government finds strategically important – creating a focal point for malicious activity.

A third factor is international cooperation. Coordinated multi-state public attribution can have benefits from both a norm-setting and coercive perspective, especially for smaller and middle powers. This type of international coordination, however, has an important temporal dimension: it can delay public attribution as significant diplomatic lead time and preparation is needed. Some countries have streamlined such joint public attribution processes – for example the Five Eyes – which allows them for that lead time to be shortened.

Illustrative Case Study: The attempted hack of the Organisation for the Prohibition of Chemical Weapons (2018)

The previous section laid out a framework for public attribution, distinguishing between the goals of an actor and four enabling or constraining factors. We provided short examples to explain each dimension of the *Public Attribution Framework*. The purpose of this section is to further highlight the possible interaction between these different dimensions through an empirical analysis of an 'illustrative case'.[55] Illustrative cases are descriptive in nature and 'designed to shed light on a particular situation, set of circumstances, and the social relations and processes that are embedded in them.'[56] This is in line

[55] On the role of illustrative cases compared to other case studies see: Jack Levy, 'Case Studies: Types, Designs, and Logics of Inference', *Conflict Management and Peace Science* 25/1 (2008), 1–18.

[56] Ashley Crossman, 'Conducting Case Study Research in Sociology', (23 June 2019), https://www.thoughtco.com/case-study-definition-3026125.

with the objectives of this article: the *Public Attribution Framework* is not utilized as a 'model' (if this, then that), but rather serves to show the complexity of relevant considerations.[57] For this study, we selected the public attribution case of the Dutch expelling four Russian officers after allegedly trying to hack into the Organisation for the Prohibition of Chemical Weapons (OPCW).[58] To reconstruct the public attribution considerations of the attempted hack on the chemical weapons watchdog, we primarily rely on newspaper reporting, think tank/NGO publications, and the press releases and statements provided by the respective governments.[59]

On 4 October 2018, the Dutch Ministry of Defence publicly announced that it had disrupted an attempted cyber operation against the OPCW. The Dutch government explained that four officers from Russia's military intelligence organization GRU were caught in a car, parked close to the OPCW headquarters in The Hague, with electronic equipment 'installed for the purpose of infiltrating the OPCW's network'.[60] The Russian officers were detained and later expelled.[61] A Financial Times' reporter described the Dutch government's decision to hold a press conference about the attempted hack of the GRU as 'an unusual step'.[62] Russia rejected the allegations.[63]

The *Public Attribution Framework* can usefully help us analyze this case of public attribution. Let us start by examining the potential goals of Dutch authorities' decision to go public. In a joint statement the Dutch Prime Minister, Mark Rutte, and the British Prime Minister, Theresa May, said: 'This attempt to access the secure systems of an international organisation working to rid the world of chemical weapons demonstrates the GRU's disregard for the global values and rules that keep us all safe.'[64] U.K. Foreign Secretary Jeremy Hunt equally stated that 'The GRU's actions are reckless and indiscriminate. [. . .] This pattern of behavior demonstrates their desire to operate without regard to international

[57] We are grateful for reviewer 1 in providing the language to clarify the purposes of our study.

[58] Pippa Crerar, Jon Henley and Patrick Wintour, 'Russia accused of cyber-attack on chemical weapons watchdog', *The Guardian*, (4 October 2018), https://www.theguardian.com/world/2018/oct/04/netherlands-halted-russian-cyber-attack-on-chemical-weapons-body.

[59] A timeline of how the Dutch security service caught the Russian GRU officers is provided by the Financial Times: Mark Odell, 'How Dutch security service caught alleged Russian spies', *Financial Times*, (4 October 2018), https://www.ft.com/content/b1fb5240-c7db-11e8-ba8f-ee390057b8c9

[60] Dutch Government, 'Netherlands Defence Intelligence and Security Service disrupts Russian cyber operation targeting OPCW', (4 October 2018), https://www.government.nl/latest/news/2018/10/04/netherlands-defence-intelligence-and-security-service-disrupts-russian-cyber-operation-targeting-opcw

[61] Alicia Sanders-Zakre, 'Russia Charged With OPCW Hacking Attempt', Arms Control Association, (2018, November), https://www.armscontrol.org/act/2018-11/news/russia-charged-opcw-hacking-attempt

[62] Odell, 'How Dutch security service caught alleged Russian spies'.

[63] Sanders-Zakre, 'Russia Charged With OPCW Hacking Attempt'.

[64] Government of the Netherlands, 'Joint statement by Prime Minister May and Prime Minister Rutte on cyber activities of the Russian military intelligence service, the GRU', (4 October 2018), https://www.government.nl/latest/news/2018/10/04/joint-statement-by-prime-minister-may-and-prime-minister-rutte-on-cyber-activities-of-the-russian-military-intelligence-service-the-gru

law or established norms and to do so with a feeling of impunity and without consequences.'[65]

The attempted OPCW hack violated international law about the integrity of diplomatic institutions. The Vienna Convention on Diplomatic Relations specifies that the privileges of a diplomatic mission enable diplomats to perform their duties without threat of influence.[66] At the same time, much political espionage violates this international legal principle. For example, the NSA is known to have gained deep access into the networks of the United Nations.[67] It also did not violate norms described in any document proposed at the UN Government Group of Experts (GGE), the Global Commission on the Stability of Cyberspace, or other (semi-formal) international institutions.[68] In other words, if public attribution is primarily used as a vehicle to *set norms*, this might not be the best case to pick for the Dutch and British government. One would be more adept in publicly calling out acts that demarcate the type of activity one would not be interested in engaging in oneself or its close allies, as for example acts of industrial espionage, election hacking, or more disruptive intrusions against critical infrastructure during peacetime.

However, if the principal goal of the Dutch disclosure was a *counter-threat activity*, one might argue the public attribution case made more sense. In reality, it may have been less about Russia violating a specific norm, and more about a broader diplomatic confrontation between the Kremlin and several Western countries.[69]

To improve our understanding of this decision and objective of Dutch attribution, we have to look at the enabling and constraining factors of the *Public Attribution Framework*. Considering the *intelligence* dimension, it is significant that the Dutch security service caught the GRU officers red-handed. On April 10, the four Russian men arrived at Amsterdam's Schiphol airport on a flight from Moscow on diplomatic passports. A day later, when the officers rented a car to begin initial reconnaissance around the OPCW headquarters, the Dutch intelligence received information from their British counterparts that Russians are potentially attempting a 'close access hack' of the OPCW computer networks.[70]

[65] National Cyber Security Centre, 'Reckless campaign of cyber attacks by Russian military intelligence service exposed', (3 October 2019), https://www.ncsc.gov.uk/news/reckless-campaign-cyber-attacks-russian-military-intelligence-service-exposed&xid=17259,1500000,15700023,15700124,15700149,15700186,15700191,15700201,15700214

[66] United Nations, 'Vienna Convention on Diplomatic Relations', (18 April 1961), https://legal.un.org/ilc/texts/instruments/english/conventions/9_1_1961.pdf

[67] Buchanan, *The Hacker and the State.*

[68] For an overview of the UN GGE initiatives see: Anders Henriksen, 'The end of the road for the UN GGE process: The future regulation of cyberspace', *Journal of Cybersecurity* 5/1 (2019), 1–9.

[69] Senior governments have also alluded to this. 'Our exposure of this Russian operation is intended as an unambiguous message that the Russian Federation must refrain from such actions,' said Dutch Defense Minister Ank Bijleveld. See: Sanders-Zakre, 'Russia Charged With OPCW Hacking Attempt'.

[70] see note 62 above.

On April 13, the Dutch counter-intelligence officers apprehended the Russian officers when they parked their vehicle close to the OPCW headquarters, confiscating all their electronic equipment, including mobile phones, Wi-Fi antenna, a computer, a transformer and specialist hacking equipment. At a presentation, the Dutch authorities presented copies of the men's passports and photographs of the equipment as well as a detail overview of their movements. They even disclosed the receipts of the taxi ride from the GRU barracks to Moscow airport. Despite these high levels of attribution certainty, there was little – if any – intelligence loss for the UK or Dutch intelligence services.[71]

Overall, the type of intelligence gathered made it much easier for the Dutch government to disclose publicly Russia's activities through a press conference. If either there was less evidence that GRU officers were behind the attempted intrusion or if the information was obtained in a different manner (for example, the Dutch being in the computer systems of the GRU and having gathered information in this manner), a similar public display would have been much harder. Furthermore, the Dutch caught the GRU officers red-handed in mid-April, and managed to control the information until the public release in October. As there was no leak or other information in the public realm about this case, the Dutch government was able to better time their public disclosure – and coordinate with other international actors, as discussed below – to maximize the impact.[72]

Turning to the second dimension of the framework, *incident severity*, one should observe that the intended effect behind this hack was not as severe relative to other (Russian) operations.[73] The four men did not intend to take down critical infrastructure or release an explosive worm into the wild. One could argue that the low-severity would be a constraining factor to publicly disclose, particularly as it may lead to less willingness to also publicly attribute by other actors.

Yet, the broader *geopolitical context and timing* reveal the significance of the operation. The public attribution decision took place over the background of the investigations into the downing of MH-17 and the investigations into the poisoning of former Russian military intelligence officer Sergei Skripal and his daughter in Salisbury, United Kingdom. As a reporter notes, 'OPCW's scientists were testing samples taken from Salisbury that would be

[71] Also, it prevented other activity from the officers as one of the laptops shoed that they had purchased train tickets onwards to Bern and printed out maps of Russian diplomatic facilities in the area.

[72] Bellingcat and The Insider website later publicly released information of their online investigation and found that Aleksei Morenets, Evgenii Serebriakov, Oleg Sotnikov and Alexey Minin were real identities rather than aliases. Bellingcat, '305 Car Registrations May Point to Massive GRU Security Breach', (4 October 2018), https://www.bellingcat.com/news/2018/10/04/305-car-registrations-may-point-massive-gru-security-breach/

[73] For an overview see: Andy Greenberg, *Sandworm: A New Era of Cyberwar and the Hunt for the Kremlin's Most Dangerous Hackers* (Penguin Random House 2019).

verified as Russian-made novichok within days. [. . .] The GRU had to act quickly.'[74] The GRU had already tried and failed remote cyber intrusions on the OPCW, UK Foreign Office and laboratory at Porton Down, near Salisbury.[75] This close-access operation was the last viable option.[76]

This further explains why the United Kingdom and the Netherlands were motivated to push back against the Russian military intelligence service GRU. OPCW provided proof of a much broader attempt to undermine the investigations of the international organisation looking into the Salisbury poisoning, which further contributed to a picture of Russian culpability. Furthermore, the audience extends beyond Russia in this attribution case. As Jeremy Wright told Sky news: 'One of the things that we've tried to get across, in response to what happened in Salisbury, is that this is a problem the whole world needs to wake up to.'[77]

Geopolitically, thus, the case was ideal not just as an opportunity for the Dutch to internationalize push-back against Russia in the context of its activities against the MH-17 investigation, but also, as contributing to their close ally's aims of internationalizing the response to the GRU's actions in Salisbury, Europe, and beyond. For example, in the joint press conference by the Dutch and the British, UK Ambassador Peter Wilson attributed not only the OPCW and Salisbury incidents, but also activities against the MH-17 investigation in Malaysia to one of the GRU officers involved, Yevgeniy Serebriakov.[78]

This also informs the last dimension of the *Public Attribution Framework: handling and follow-on actions*. The case shows how important the time to cue up follow-on actions after the decision to go public is. The Dutch public disclosure helped to build a strong international case against Russian brazen foreign influence campaigns, and particularly the Skripal poisoning. In fact, the Dutch used the six months between the incident and the public attribution not only to decide on whether and how to go public, but also joined the UK (and US) in a significant diplomatic campaign to motivate other states to also issue public statements. On the same day as the Dutch and British went public, a federal grand jury in Pennsylvania indicted seven Russian military intelligence officers – including the four men expelled from the Netherlands – accusing them of 'hacking into U.S. and international anti-doping agencies

[74] Lizzie Dearden, 'Russia hack: Taxi receipts to lager cans – the trail of evidence left by spies who tried to attack the chemical weapons watchdog', *Independent*, (5 October 2018), https://www.independent.co.uk/news/world/europe/russia-hack-spies-chemical-weapons-doping-us-evidence-opcw-hague-gru-a8569326.html

[75] Ibid.

[76] See note 75 above.

[77] Philip Whiteside, 'Jeremy Wright says attempts to hack OPCW "show Russia responsible for Salisbury poisonings"', Sky News, (7 October 2018), https://news.sky.com/story/jeremy-wright-says-attempts-to-hack-opcw-show-russia-responsible-for-salisbury-poisonings-11520515

[78] UK Foreign and Commonwealth Office, 'Minister for Europe statement: attempted hacking of the OPCW by Russian military intelligence', (4 October 2018).

and sports federations and of accessing data related to 250 athletes from about 30 countries.'[79] The Justice Department also said that the agents' targets included 'Westinghouse Electric Corporation, the Organization for the Prohibition of Chemical Weapons, and a Swiss lab that was testing for an exotic poison used in the attempted assassinations of former KGB agent Sergei Skripal and his daughter.'[80] Within the next day, in total seventeen other states and two international organizations attributed, supported the attributions, or otherwise diplomatically supported the NL/UK/US action.[81] The international coordination and collaboration undoubtedly influenced the Dutch decision to publicly attribute and its understanding of the goals it could pursue with this act.[82]

Conclusion

'The good news is that attribution – identifying who is responsible – is now largely a solved problem,' according to Dmitri Alperovitch, co-founder of Crowdstrike.[83] Yet, the bad news is that public attribution – what to do when you think you know who is responsible – remains an unresolved issue. We provided a framework that can be used by decision makers to decide whether to publicly attribute cyber operations to adversaries. It helps to make policymakers aware of the factors or equities that need to be balanced to make careful decisions surrounding public attribution. The key guiding questions for thinking through each element of the *Public Attribution Framework* are provided in the Appendix.

Public attribution inherently requires coordination and collaboration between different agencies within government, sometimes with competing missions and viewpoints. In this article, we were less concerned about the nature of this interagency process. Given that the implementation will inherently depend on the setup of authorities and responsibilities in different countries, the purpose was *not* to provide a decision-making chart for implementation on how to go about publicly attributing cyber operations. Instead, we sought to discuss the set of factors that require careful deliberation for *any* government when they

[79] Bill Chappell and Carrie Johnson, 'U.S. Charges 7 Russian Intelligence Officers With Hacking 40 Sports And Doping Groups', *National Public Radio*, (4 October 2018), https://www.npr.org/2018/10/04/654306774/russian-cyber-unit-accused-of-attacking-opcw-chemical-weapons-watchdog?t=1611604462119

[80] Ibid.

[81] The final tally was: the Netherlands, United Kingdom, United States, Canada, Australia, New Zealand, Czech Republic, Denmark, Estonia, Finland, France, Germany, Japan, Latvia, Norway, Poland, Romania, Slovakia, Sweden, Ukraine, EU, NATO.

[82] On October 15, EU foreign ministers also adopted a new regime of restrictive measures at a meeting in Luxembourg against 'a new regime of restrictive measures against those who use or develop chemical weapons or those who assist to do so, regardless of nationality.' Sanders-Zakre, 'Russia Charged With OPCW Hacking Attempt'.

[83] Alperovitch, 'Stopping the Next Cyber Conflict'.

consider public attribution.[84] Our general argument is that public attribution is a highly complex process which requires trade-offs of multiple considerations.

There is no linear relationship between the different equities. Indeed, this article both simplifies and complexifies. It simplifies in that it tries to address a complex issue into just a few factors to consider for a government. But it also complexifies in that it argues that public attribution is not a binary issue, there are no neat boundaries between the different factors, and the trade-offs are incredibly difficult to consider concurrently. It also suggests the difficult *individual* decisions each state has to make. Two governments will never be exactly the same in their factors to consider. They may require different response options, access to information about the threat actor, or be situated in a very different geographical context.

Like the Vulnerability Equity Process (VEP) established by several countries about the disclosure of unknown vulnerabilities, 'there are no hard and fast rules' for public attribution.[85] Flexibility is required on an ad hoc basis. Hence, we provided merely a 'framework' and not a step-by-step 'instruction manual' for public attribution.[86] However, *unlike* the VEP, we do *not* expect decision makers to provide complete transparency on their own guiding principles for public attribution. In other words, we hope decisionmakers internalize this framework, adjust it to their national context and priorities, but not necessarily publish it.

Also, there is tension between the various factors we discuss in the article. Our goal was not to be normative about these trade-offs. For example, we did not intend to make claims when it is acceptable to suffer intelligence losses for reaching a certain goal, such as norm-setting. Instead, we assess how each factor may be evaluated and how governments could combine this information for their decision-making.[87] In that sense, we proposed the state's attitude towards public attribution be one of 'strategic, coordinated pragmatism'.[88] Public attribution requires *consistent* goals, whilst the negative effects are assessed on a *case-by-case* basis.

[84] Private sector actors may also publicly attribute. As these actors will likely have a different set of considerations, we only consider state actors.

[85] Michael Daniel, 'Heartbleed: Understanding When We Disclose Cyber Vulnerabilities', *The White House*, (28 April 2014), https://obamawhitehouse.archives.gov/blog/2014/04/28/heartbleed-understanding-when-we-disclose-cyber-vulnerabilities; also see: Aitel and Tait, 'Everything You Know About the Vulnerability Equities Process Is Wrong'.

[86] Also see: Tristan Caulfield, Christos Ioannidis, and David Pym, 'The U.S. Vulnerabilities Equities Process: An Economic Perspective', (15 November 2017), http://www0.cs.ucl.ac.uk/staff/D.Pym/VEP.pdf

[87] Yet, there is no expectation that one can come up with a repeatable scoring system for public attribution, which some have attempted to do for the VEP. See Sasha Romanosky, 'Developing an Objective, Repeatable Scoring System for a Vulnerability Equity Process', *Lawfare*, (4 February 2019), https://www.lawfareblog.com/developing-objective-repeatable-scoring-system-vulnerability-equities-process

[88] This phrase is inspired by studies on human rights which frequently focus on the dual aspects of coordinated and systematic work vis-a-vis sensitivity and pragmatism from the leadership. See for example: Kartin Kinzelbach and Julian Lehmann, 'Can Shaming Promote Human Rights? Publicity in Human Rights Foreign Policy', GPPI, (14 December 2015), https://www.gppi.net/2015/12/14/can-

There are several avenues for future research. More research would be welcome not just on why states publicly attribute, but also *how* to publicly attribute. There is also still little research looking at the actual effectiveness of different forms of disclosure. Both empirical research and scenario-based analysis would help in making sense of this complex process.

Acknowledgements

We gratefully acknowledge the support by our research group at the Center for Security Studies at ETH Zürich, as well as the Grugq, Silas Cutler, and Timo Steffens. Special thanks go to two anonymous reviewers whose comments helped us sharpen our argument further.

Disclosure statement

No potential conflict of interest was reported by the author(s).

Notes on contributors

Florian J. Egloff is a Senior Researcher in Cybersecurity at the Center for Security Studies (CSS) at ETH Zurich. His current research projects focus on the politics of public attribution, the role of non-and semi-state actors in cyber security, and the use of cyber intrusions for political purposes.

Max Smeets is a senior researcher at the Center for Security Studies (CSS) at ETH Zurich. He also serves as the director of the European Cyber Conflict Research Initiative (ECCRI.eu), an organization promoting the interdisciplinary study of cyber conflict and statecraft in Europe and beyond.

ORCID

Florian J. Egloff http://orcid.org/0000-0002-0290-667X
Max Smeets http://orcid.org/0000-0003-4057-6445

Bibliography

Aitel, Dave and Matt Tait, 'Everything You Know about the Vulnerability Equities Process Is Wrong', *Lawfare*, 18 Aug., 2016. https://www.lawfareblog.com/everything-you-know-about-vulnerability-equities-process-wrong

Alperovitch, Dimitri, 'Stopping the Next Cyber Conflict', *The Cipher Brief* 28 Jan., 2018. https://www.thecipherbrief.com/column_article/stopping-next-cyber-conflict

shaming-promote-human-rights; and Alice Bah Kuhnke, 'Introduction to the Human Rights Strategy', Government of Sweden: Ministry for Culture and Democracy, (31 March 2017) https://www.government.se/4ab459/contentassets/08bcf332d33e40908f918f0cd29a13ae/a-strategy-for-national-efforts-with-human-rights

Armelli, Matthew, Stuart Caudill, John Patrick Dees, Max Eager, Jennifer Keltz, Ian Pelekis, John Sakellariadis, Virpratap Vikram Singh, von Ofenheim Katherine, and Neal Pollard, 'Named but Hardly Shamed: The Impact of Information Disclosures on APT Operations', *SIPA Capstone Project*, (Spring 2020).

Art, Robert J., 'To What Ends Military Power?', *International Security* 4/4 (1980), 3–35. doi:10.2307/2626666.

Bah Kuhnke, Alice, 'Introduction to the Human Rights Strategy,' Government of Sweden, 31 Mar., 2017 (Ministry for Culture and Democracy). https://www.government.se/4ab459/contentassets/08bcf332d33e40908f918f0cd29a13ae/a-strategy-for-national-efforts-with-human-rights

Baliga, Sandeep, Ethan Bueno De Mesquita, and Alexander Wolitzky, 'Deterrence with Imperfect Attribution', *American Political Science Review* 114/4 (2020), 1155–78. doi:10.1017/S0003055420000362.

Baram, Gil and Uri Sommer. 'Covert or Not Covert: National Strategies during Cyber Conflict,' Paper presented at the 2019 11th International Conference on Cyber Conflict (CyCon), 28-31 May 2019.

Bellingcat, '305 Car Registrations May Point to Massive GRU Security Breach,' 4 Oct., 2018. https://www.bellingcat.com/news/2018/10/04/305-car-registrations-may-point-massive-gru-security-breach/

Boebert, Earl,'A Survey of Challenges in Attribution,' in: *Committee on Deterring Cyberattacks* (ed.), Proceedings of a Workshop on Deterring Cyberattacks: Informing Strategies and Developing Options for US Policy (Washington, DC: The National Academies Press 2010), 41–52

Borghard, Erica D. and Shawn W. Lonegran, 'The Logic of Coercion in Cyberspace', *Security Studies* 26/3 (2017), 452–58. doi:10.1080/09636412.2017.1306396.

Borghard, Erica D. and Shawn W. Lonergan, 'Cyber Operations as Imperfect Tools of Escalation,' *Strategic Studies Quarterly* (Fall 2019), 122–45. https://www.airuniversity.af.edu/Portals/10/SSQ/documents/Volume-13_Issue-3/Borghard.pdf

Buchanan, Ben, *The Cybersecurity Dilemma: Hacking, Trust and Fear between Nations* (Oxford: Oxford University Press 2017).

Buchanan, Ben, *The Hacker and the State: Cyber Attacks and the New Normal of Geopolitics* (Cambridge, MA: Harvard University Press 2020).

Buchanan, Ben and Fiona Cunningham, 'Preparing the Cyber Battlefield: Assessing A Novel Escalation Risk in A U.S.-China Crisis', *Texas National Security Review* 3/4 (2020), 54–81.

Carnegie Endowment for International Peace, 'International Cyber Norms,' 19 Feb., 2021. https://carnegieendowment.org/specialprojects/cybernorms?lang=en

Carson, Austin, 'Facing off and Saving Face: Covert Intervention and Escalation Management in the Korean War', *International Organization* 70/1 (2016), 103–31. doi:10.1017/S0020818315000284.

Carson, Austin, *Secret Wars: Covert Conflict in International Politics* (Princeton: Princeton University Press 2018).

Caulfield, Tristan, Christos Ioannidis, and David Pym, 'The U.S. Vulnerabilities Equities Process: An Economic Perspective,' 15 Nov., 2017. http://www0.cs.ucl.ac.uk/staff/D.Pym/VEP.pdf

CBS News, 'NSA Cybersecurity Directorate's Anne Neuberger on Protecting the Elections,' 19 Aug., 2020. https://www.cbsnews.com/news/nsa-cybersecurity-directorates-anne-neuberger-on-protecting-the-elections/

Chappell, Bill and Carrie Johnson, 'U.S. Charges 7 Russian Intelligence Officers With Hacking 40 Sports And Doping Groups', *National Public Radio*, 4 Oct., 2018. https://

www.npr.org/2018/10/04/654306774/russian-cyber-unit-accused-of-attacking-opcw-chemical-weapons-watchdog?t=1611604462119

Chesney, Robert and Max Smeets, 'Policy Roundtable: Cyber Conflict as an Intelligence Contest,' *Texas National Security Review*, 17 Sep., 2020. https://tnsr.org/roundtable/policy-roundtable-cyber-conflict-as-an-intelligence-contest/

Clark, David D. and Susan Landau, 'Untangling Attribution,' in: *Committee on Deterring Cyberattacks* (ed.), Proceedings of a Workshop on Deterring Cyberattacks: Informing Strategies and Developing Options for US Policy, (Washington, DC: The National Academies Press 2010), 25–40.

Cormac, Rory and Richard J. Aldrich, 'Grey Is the New Black: Covert Action and Implausible Deniability', *International Affairs* 94/3 (2018), 477–94. doi:10.1093/ia/iiy067.

Crerar, Pippa, Jon Henley, and Patrick Wintour, 'Russia Accused of Cyber-attack on Chemical Weapons Watchdog', *The Guardian*, 4 Oct. 2018. https://www.theguardian.com/world/2018/oct/04/netherlands-halted-russian-cyber-attack-on-chemical-weapons-body

Crossman, Ashley, 'Conducting Case Study Research in Sociology,' 23 Jun. 2019. https://www.thoughtco.com/case-study-definition-3026125

Daniel, Michael, 'Heartbleed: Understanding When We Disclose Cyber Vulnerabilities', *The White House*, 28 Apr. 2014. https://obamawhitehouse.archives.gov/blog/2014/04/28/heartbleed-understanding-when-we-disclose-cyber-vulnerabilities

Dearden, Lizzie, 'Russia Hack: Taxi Receipts to Lager Cans – The Trail of Evidence Left by Spies Who Tried to Attack the Chemical Weapons Watchdog', *Independent*, 5 Oct. 2018. https://www.independent.co.uk/news/world/europe/russia-hack-spies-chemical-weapons-doping-us-evidence-opcw-hague-gru-a8569326.html

Dunn Cavelty, Myriam and Florian J. Egloff, 'The Politics of Cybersecurity: Balancing Different Roles of the State', *St Antony's International Review* 15/1 (2019), 37–57.

Dutch Government, 'Netherlands Defence Intelligence and Security Service Disrupts Russian Cyber Operation Targeting OPCW', 4 Oct. 2018. https://www.government.nl/latest/news/2018/10/04/netherlands-defence-intelligence-and-security-service-disrupts-russian-cyber-operation-targeting-opcw

Eason, Thomas, Oliver Daddow, and Rory Cormac, 'From Secrecy to Accountability: The Politics of Exposure in the Belgrano Affair', *The British Journal of Politics and International Relations* 22/3 (2020), 542–60. doi:10.1177/1369148120930588.

Edwards, Benjamin, Alexander Furnas, Stephanie Forrest, and Robert Axelrod, 'Strategic Aspects of Cyberattack, Attribution, and Blame', *Proceedings of the National Academy of Sciences* 114/11 (2017), 2825–30. doi:10.1073/pnas.1700442114.

Egloff, Florian J. 'Cybersecurity and Non-State Actors: A Historical Analogy with Mercantile Companies, Privateers, and Pirates', DPhil Thesis, (University of Oxford 2018).

Egloff, Florian J., 'Contested Public Attributions of Cyber Incidents and the Role of Academia', *Contemporary Security Policy* 41/1 (2020), 55–81. doi:10.1080/13523260.2019.1677324.

Egloff, Florian J., 'Public Attribution of Cyber Intrusions', *Journal of Cybersecurity* 6/1 (2020), 1–12. doi:10.1093/cybsec/tyaa012.

Egloff, Florian J. and Myriam Dunn Cavelty, 'Attribution and Knowledge Creation Assemblages in Cybersecurity Politics', *Journal of Cybersecurity* (2021 *forthcoming*). doi:10.1093/cybsec/tyab002.

Eisenberg, Ted, David Gries, Juris Hartmanis, M. Don Holcomb, Stuart Lynn, and Thomas Santoro, 'The Cornell Commission: On Morris and the Worm', *Communications of the ACM* 32/6 (1989), 706–09. doi:10.1145/63526.63530.

Fearon, James D., 'Signalling Foreign Policy Interests: Tying Hands Versus Sinking Costs', *The Journal of Conflict Resolution* 41/1 (1997), 68–90. doi:10.1177/0022002797041001004.

Finnemore, Martha and Duncan B. Hollis, 'Constructing Norms for Global Cybersecurity', *The American Journal of International Law* 110/3 (2016), 425–79. doi:10.1017/S0002930000016894.

Finnemore, Martha and Duncan B. Hollis, 'Beyond Naming and Shaming: Accusations and International Law in Cybersecurity', *European Journal of International Law* 31/3 (2020), 969–1003. doi:10.1093/ejil/chaa056.

Finnemore, Martha and Kathryn Sikkink, 'International Norm Dynamics and Political Change', *International Organization* 52/4 (1998), 887–917. doi:10.1162/002081898550789.

Fischerkeller, Michael, 'The Fait Accompli and Persistent Engagement in Cyberspace', *War on the Rocks*, 24 Jun. 2020. https://warontherocks.com/2020/06/the-fait-accompli-and-persistent-engagement-in-cyberspace/

Fischerkeller, Michael and Richard J. Harknett, 'Persistent Engagement and Cost Imposition: Distinguishing between Cause and Effect', *Lawfare*, 6 Feb. 2020. https://www.lawfareblog.com/persistent-engagement-and-cost-imposition-distinguishing-between-cause-and-effect

Gartzke, Erik, 'The Myth of Cyberwar: Bringing War in Cyberspace Back down to Earth', *International Security* 38/2 (2013), 41–73. doi:10.1162/ISEC_a_00136.

Gorwa, Robert and Max Smeets, 'Cyber Conflict in Political Science: A Review of Methods and Literature,' *2019 ISA Annual Convention* (Toronto 2019 March). doi:10.31235/osf.io/fc6sg.

Government of the Netherlands, 'Joint Statement by Prime Minister May and Prime Minister Rutte on Cyber Activities of the Russian Military Intelligence Service, the GRU', 4 Oct. 2018. https://www.government.nl/latest/news/2018/10/04/joint-statement-by-prime-minister-may-and-prime-minister-rutte-on-cyber-activities-of-the-russian-military-intelligence-service-the-gru

Greenberg, Andy, *Sandworm: A New Era of Cyberwar and the Hunt for the Kremlin's Most Dangerous Hackers* (New York: Penguin Random House 2019).

Guerrero-Saade, Juan Andrés and Costin Raiu, 'Walking in Your Enemy's Shadow: When Fourth-party Collection Becomes Attribution Hell', *Virus Bulletin Conference*, Oct. 2017. https://media.kasperskycontenthub.com/wp-content/uploads/sites/43/2018/03/07170728/Guerrero-Saade-Raiu-VB2017.pdf

Guitton, Clement, 'Achieving Attribution', *PhD Thesis*, (London: King's College London 2014)

Harknett, Richard J. and Max Smeets, 'Cyber Campaigns and Strategic Outcomes', *Journal of Strategic Studies* (2020), 1–34. doi:10.1080/01402390.2020.1732354.

Healey, Jason, 'Beyond Attribution: Seeking National Responsibility for Cyber Attacks', Atlantic Council: Cyber Statecraft Initiative, Issue Brief (2011). http://www.atlanticcouncil.org/publications/issue-briefs/beyond-attribution-seeking-national-responsibility-incyberspace

Healey, Jason and Robert Jervis, 'The Escalation Inversion and Other Oddities of Situational Cyber Stability', *Texas National Security Review* 3/4 (2020), 30–53.

Henriksen, Anders, 'The End of the Road for the UN GGE Process: The Future Regulation of Cyberspace', *Journal of Cybersecurity* 5/1 (2019), 1–9. doi:10.1093/cybsec/tyy009.

INFOSEC, 'Animal Farm APT and the Shadow of French Intelligence', 8 Jul. 2015. https://resources.infosecinstitute.com/animal-farm-apt-and-the-shadow-of-france-intelligence/#gref

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

Junio, Timothy J., 'How Probable Is Cyber War? Bringing IR Theory Back in to the Cyber Conflict Debate', *Journal of Strategic Studies* 36/1 (2013), 125–33. doi:10.1080/01402390.2012.739561.

Kinzelbach, Kartin and Julian Lehmann, 'Can Shaming Promote Human Rights? Publicity in Human Rights Foreign Policy', GPPI, 14 Dec. 2015. https://www.gppi.net/2015/12/14/can-shaming-promote-human-rights

Kreps, Sarah and Jacquelyn Schneider, 'Escalation Firebreaks in the Cyber, Conventional, and Nuclear Domains: Moving beyond Effects-Based Logics', *Journal of Cybersecurity* 5/1 (2019), 1–11. doi:10.1093/cybsec/tyz007.

Kunioka, Todd and Lawrence S. Rothenberg, 'The Politics of Bureaucratic Competition: The Case of Natural Resource Policy', *Journal of Policy Analysis and Management* 12/4 (1993), 700–25. doi:10.2307/3325347.

Lab, Kaspersky, 'Animals in the APT Farm,' 6 Mar. 2015. https://securelist.com/animals-in-the-apt-farm/69114/

Lab, Kaspersky, 'Spy Wars: How Nation-state Backed Threat Actors Steal from and Copy Each Other', 4 Oct. 2017. https://www.kaspersky.com/about/press-releases/2017_spy-wars-how-nation-state-backed-threat-actors-steal-from-and-copy-each-other#:~:text=In%20November%202014%2C%20Kaspersky%20Lab,language)%2C%20as%20well%20as%20Animal

Langner, Ralph 'To Kill A Centrifuge: A Technical Analysis of What Stuxnet's Creators Tried to Achieve', The Langner Group, Nov. 2013. http://www.langner.com/en/wp-content/uploads/2013/11/To-kill-a-centrifuge.pdf

Levy, Jack, 'Case Studies: Types, Designs, and Logics of Inference', *Conflict Management and Peace Science* 25/1 (2008), 1–18. doi:10.1080/07388940701860318.

Liff, Adam P., 'Cyberwar: A New 'Absolute Weapon'? The Proliferation of Cyberwarfare Capabilities and Interstate War', *Journal of Strategic Studies* 35/3 (2012), 401–28. doi:10.1080/01402390.2012.663252.

Liff, Adam P., 'The Proliferation of Cyberwarfare Capabilities and Interstate War, Redux: Liff Responds to Junio', *Journal of Strategic Studies*, 36/1 (2013), 134–38.

Lin, Herbert S., 'Attribution of Malicious Cyber Incidents: From Soup to Nuts', *Journal of International Affairs* 70/1 (2016). https://jia.sipa.columbia.edu/attribution-malicious-cyber-incidents

Lindsay, Jon R., 'Stuxnet and the Limits of Cyber Warfare', *Security Studies* 22/3 (2013), 365–404. doi:10.1080/09636412.2013.816122.

Lum, Cynthia, Lesly W. Kennedy, and Alison J. Sherley, 'The Effectiveness of Counter-Terrorism Strategies', *A Campbell Systematic Review* 2/1 (January 2006), 1–50. doi:10.4073/csr.2006.2.

Maschmeyer, Lennart, Ronald J. Deibert, and Jon R. Lindsay, 'A Tale of Two Cybers - How Threat Reporting by Cybersecurity Firms Systematically Underrepresents Threats to Civil Society', *Journal of Information Technology & Politics* 18/1 (2021), 1–20. doi:10.1080/19331681.2020.1776658.

Maurer, Tim, 'Proxies' and Cyberspace', *Journal of Conflict and Security Law* 21/3 (2016), 383–403. doi:10.1093/jcsl/krw015.

McDonald, Geoff, Liam O Murchu, Stephen Doherty, and Eric Chien, 'Stuxnet 0.5: The Missing Link', *Symantec*, 26 Feb. 2013. http://www.symantec.com/content/en/us/enterprise/media/security_response/whitepapers/stuxnet_0_5_the_missing_link.pdf

Ministerie, van Defensie, Defensie Cyber Strategie 2018: Investeren in digitale slagkracht voor Nederland 2018.

MITRE, 'Tactics, Techniques, and Procedures', 12 Sept. 2019. https://security.radware.com/ddos-experts-insider/hackers-corner/tactics-techniques-procedures/

National Cyber Security Centre, 'Reckless Campaign of Cyber Attacks by Russian Military Intelligence Service Exposed', 3 Oct. 2019. https://www.ncsc.gov.uk/news/reckless-campaign-cyber-attacks-russian-military-intelligence-service-exposed&xid=17259,1500000,15700023,15700124,15700149,15700186,15700191,15700201,15700214

Nye, Joseph S., Jr., 'The Regime Complex for Managing Global Cyber Activities', *Global Commission on Internet Governance Paper Series*, 1 (2014), https://www.cigionline.org/sites/default/files/gcig_paper_no1.pdf

Nye, Joseph S., Jr., 'Deterrence and Dissuasion in Cyberspace', *International Security* 41/3 (2017), 44–71. doi:10.1162/ISEC_a_00266.

Odell, Mark, 'How Dutch Security Service Caught Alleged Russian Spies', *Financial Times*, 4 Oct. 2018. https://www.ft.com/content/b1fb5240-c7db-11e8-ba8f-ee390057b8c9

Otto, Jacob and William Spaniel, 'Doubling Down: The Danger of Disclosing Secret Action', *International Studies Quarterly* (2020), 1–12. doi:10.1093/isq/sqaa081.

Page, Bob, 'A Report of the Internet Worm', 7 Nov. 1988. http://www.ee.ryerson.ca/~elf/hack/iworm.html

Poznansky, Michael and Evan Perkoski, 'Rethinking Secrecy in Cyberspace: The Politics of Voluntary Attribution', *Journal of Global Security Studies* 3/4 (2018), 402–16. doi:10.1093/jogss/ogy022.

Raymond, David, Gregory Conti, Tom Cross, and Robert Fanelli, 'A Control Measure Framework to Limit Collateral Damage and Propagation of Cyber Weapons,' in K. Podins, J. Stinissen, and M. Maybaum (Eds.), *5th International Conference on Cyber Conflict*, (Tallinn: NATO CCD COE Publications 2013).

Rid, Thomas and Ben Buchanan, 'Attributing Cyber Attacks', *The Journal of Strategic Studies* 38/1–2 (2015), 4–37. doi:10.1080/01402390.2014.977382.

Romanosky, Sasha, 'Developing an Objective, Repeatable Scoring System for a Vulnerability Equity Process', *Lawfare*, 4 Feb. 2019. https://www.lawfareblog.com/developing-objective-repeatable-scoring-system-vulnerability-equities-process

Sanders-Zakre, Alicia, 'Russia Charged With OPCW Hacking Attempt', Arms Control Association, Nov. 2018. https://www.armscontrol.org/act/2018-11/news/russia-charged-opcw-hacking-attempt

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

Schulzke, Marcus, 'The Politics of Attributing Blame for Cyberattacks and the Costs of Uncertainty', *Perspectives on Politics* 16/4 (2018), 954–68. doi:10.1017/S153759271800110X.

Shires, James, 'Hack-and-Leak Operations: Intrusion and Influence in the Gulf', *Journal of Cyber Policy* 4/2 (2019), 235–56. doi:10.1080/23738871.2019.1636108.

Shires, James, ‘The Simulation of Scandal: Hack-and-Leak Operations, the Gulf States, and U.S. Politics’, *Texas National Security Review* 3/4 (2020), 10–29.

Smeets, Max, ‘The Strategic Promise of Offensive Cyber Operations’, *Strategic Studies Quarterly* 12/3 (2018), 90–113.

Smeets, Max, ‘US Cyber Strategy of Persistent Engagement & Defend Forward: Implications for the Alliance and Intelligence Collection’, *Intelligence and National Security* 35/3 (2020), 444–53. doi:10.1080/02684527.2020.1729316.

Smeets, Max and Herbert Lin, ‘Offensive Cyber Capabilities: To What Ends?’, in T. Minárik, R. Jakschis, and L. Lindström (eds.), *CyCon X: Maximising Effects* (Tallinn: NATO CCD COE Publications 2018), 55–72.

Soesanto, Stefan and Max Smeets, ‘Cyber Deterrence: The Past, Present, and Future’, in Frans Osinga and Tim Sweijs (eds.), *Netherlands Annual Review of Military Studies 2020* (2021), 385–400. Breda, The Netherlands. doi:10.1007/978-94-6265-419-8_20.

Stampnitzky, Lisa, ‘Truth and Consequences? Reconceptualizing the Politics of Exposure’, *Security Dialogue* 51/6 (2020), 597–613. doi:10.1177/0967010620904576.

Steffens, Timo, *Attribution of Advanced Persistent Threats* (Berlin/Heidelberg: Springer 2020).

UK Foreign and Commonwealth Office, Minister for Europe statement: attempted hacking of the OPCW by Russian military intelligence, 4 Oct. 2018.

United Nations, ‘Vienna Convention on Diplomatic Relations’, 18 Apr, 1961. https://legal.un.org/ilc/texts/instruments/english/conventions/9_1_1961.pdf

The United States Department of Justice, ‘U.S. Charges Five Chinese Military Hackers for Cyber Espionage against U.S. Corporations and a Labor Organization for Commercial Advantage: First Time Criminal Charges are Filed against Known State Actors for Hacking’, 19 May 2014. https://www.justice.gov/opa/pr/us-charges-five-chinese-military-hackers-cyber-espionage-against-us-corporations-and-labor

Whiteside, Philip, ‘Jeremy Wright Says Attempts to Hack OPCW ‘Show Russia Responsible for Salisbury Poisonings’, Sky News, 7 Oct. 2018. https://news.sky.com/story/jeremy-wright-says-attempts-to-hack-opcw-show-russia-responsible-for-salisbury-poisonings-11520515

Appendix

Table A1. Overview questions related to the goals pursued through public attribution.

Category	Factor	Questions
Goals	Norm-setting	● Does it promote norm-setting? ● Is the aim to humiliate and shame the nation-state supporting the group?
	Coercion	● Does it dissuade or compel adversaries?
	Counter-threat	● To what degree does the public attribution cause friction in the actor's operational activity?
	Prevention and defence	● Does the public attribution lead to an enhanced defensive posture?
	Community building	● How does the public attribution contribute to defensive cybersecurity? ● Does it help to create a community of network defenders and a coalition of attribution states?
	Legitimacy and reputation building	● Does it enhance domestic and/or international legitimacy of involved actors? ● What is the influence of the public attribution on public relations?

Table A2. Overview questions enablers and constraints for public attribution.

Category	Factor	Main Question	Sub-questions
Intelligence	Certainty	What is the level of certainty we have about the actor's (type of) responsibility of the cyber operation or campaign?	● What is the actor's relationship to the state? ● What is the chance the actor has acted against the state's interest?
	Intel gain-loss	What is the information gain and loss following public attribution?	● Which entry points for intelligence collection will likely be lost following public attribution? ● How will the adversary likely adapt to a public disclosure? ● Will the adversary likely find out how they were detected?
	Information control	How much is it able to control the flow of information about the threat actor it considers to publicly attribute?	● Who's expected to simultaneously track the intruder? ● (How) will other actors publish attribution information about the intruder? ● What media pressure is expected related to the incident?
	Types of TTPs	What are the tactics, techniques and procedures (TTPs) used by the intruder?	● What are the resources the intruder has likely invested in the operation? ● Are publicly available tools used? ● Will public disclosure 'burn' previously unknown exploits and tools part of an advanced operation?
Incident Severity	Operational Effect	What's the (intended) effect of the cyber operation?	● Is the operation geopolitical relevant enough to publicly attribute? ● Is the operation part of a broader campaign? ● What are the short- and long-term effects of the operation? ● Was there likely collateral damage?
	Actor Legitimacy	What is the impact of public attribution on the intruder's legitimacy?	● What is the impact on the legitimacy and status of the attacker? Can the actor be 'shamed'? ● Is the cyber operations conducted with the strategic aim of being publicly attributed by the victim?
	Actor Motivation	What is the motivation behind the intruder's cyber operation?	● Does it strengthen the intruder's self-image and legitimacy? Does the disclosed information contradict an actor's identity constructions?

(*Continued*)

Table A2. (Continued).

Category	Factor	Main Question	Sub-questions
Geopolitical Situation	Context	What is the geopolitical context of the public attribution act?	● How does the disclosure influence the domestic politics of the accused country? ● What is the historical relationship with the state? ● In which networks was activity observed? ● Was the intrusion previously discussed in private? ● Is public attribution expected to lead to a ‘tying the hands’ effect?
	Timing	What is the geopolitical timing of the public attribution act?	● Has the intelligence value of observing the intruder been maximized? ● What impact does public attribution have on the right to a fair trial?
Handling and follow-on actions	Type of Actor & Audience	How does it affect the actors’ behavior and legitimacy when their cyber activity is publicly attributed?	● How are the actor’s operational practices influences by public attribution? ● How does public attribution legitimize its use as a tool of statecraft?
	Response options	Does public attribution enable potential follow-on policy or military responses?	● Is public attribution legally necessary to implement a follow-on response? ● What are the tools available to respond? ● Will public attribution make it easier to rally support from allies for military or other responsive measure? ● Does public attribution create a focal point for malicious activity? ● Is it expected to lead to a counter response?
	International Cooperation	How does public attribution influence international cooperation?	● Is there a joint public attribution procedure in place? ● How much does international coordination delay the attribution process?