doi:10.1017/aju.2019.32

SYMPOSIUM ON CYBER ATTRIBUTION

THE BUMPY ROAD TO A MEANINGFUL INTERNATIONAL LAW OF CYBER ATTRIBUTION

*William C. Banks**

Attributing computer network intrusions has grown in importance as cyber penetrations across sovereign borders have become commonplace. Although advances in technology and forensics have made machine attribution easier in recent years, identifying states or others responsible for cyber intrusions remains challenging. This essay provides an overview of the attribution problem and its international legal dimensions and argues that states must develop accountable attribution mechanisms for international law to have practical value in this sphere.

Scope and Nature of the Attribution Problem

When it comes to cyber activity, we live in what Lucas Kello calls a state of "unpeace"—"mid-spectrum rivalry lying below the physically destructive threshold of interstate violence, but whose harmful effects far surpass the tolerable level of peacetime competition."[1] More than thirty nations are now capable of effectively using cyber tools as weapons. Moreover, as the devices and techniques for deploying them become more advanced, identifying the boundary between cyber effects that do not rise to the level of armed force and more damaging cyber weapons that do remains contentious and difficult.

Some states and nonstate actors use cyber tools to strike with impunity, knowing (or at least strongly suspecting) that their digital attacks will not prompt a response, certainly not a kinetic response. The exfiltration of data and intellectual property has been going on for some time, at great cost to governments and private industry. More recent and increasingly sophisticated forms of offensive hacking are capable of causing more significant harm, even catastrophic damage, such as shutting down financial systems, sabotaging critical infrastructure, and scrambling communications. These activities all place a premium on knowing the source of the cyber intrusion, so that states can respond accordingly.

The inability to identify the source potentially increases the risks of confusion and escalation. For example, when the United States released an unclassified summary of its Department of Defense Cyber Strategy in September 2018, attention focused on its commitment to "defend forward to disrupt or halt malicious cyber activity at its source, including activity that falls below the level of armed conflict."[2] In other words, U.S. Cyber Command will interdict cyber threats outside combat zones. Yet some see the "defending forward" idea as putting the U.S. military on an offensive, rather than defensive, footing. The recent shift in U.S. cyber policy deepens a cyber variant on a classic security dilemma between the United States and other states: As one state takes steps to defend itself in

* *Board of Advisers Distinguished Professor, Professor of Law and Professor of Public Administration & International Affairs emeritus, Syracuse University.*

[1] Lucas Kello, The Virtual Weapon and International Order 78 (2018).

[2] U.S. Dep't of Def., Summary Cyber Strategy 1 (2018).

cyberspace, it inadvertently threatens other states with what appears to be offensive action. In practice, "defending forward" can look like attacking forward to those experiencing an intrusion. One implication is an increased tendency to escalate conflicts.[3] In an environment where escalation fears are on the rise, the possibility that cyber intrusions could spike destructive and even destabilizing conflicts between states places a premium on confident attribution of cyber intrusions.

Technical and Practical Challenges

Attribution is defined as "identifying the agent responsible for the action."[4] Attribution of cyber intrusions can be challenging. States often use technical tools to hide their responsibility and/or rely on nonstate proxies to carry out cyber activities for them.[5] However, significant technological strides in attributing cyber events in the last decade have made the task "more nuanced, more common, and more political than has typically been acknowledged."[6] The nuance involves combining experienced and disciplined technical operators with the intuition and judgment of intelligence professionals. The political aspect includes assessing what is at stake in making the attribution judgment, starting with the damage incurred, whether physical, financial, or reputational.[7] As such, attribution is often expressed in degrees of certainty, and it requires input from a range of actors and sources, including technical forensics, human intelligence, signals intelligence, history, and diplomatic relations.[8]

Understanding the components of attribution is essential for shaping a legal and policy strategy to deter harmful cyber intrusions in the future.[9] As cyber intrusions have proliferated in recent years, states have invested in doing attribution well and, as a result, deterring or at least discouraging states and other cyber intruders.[10] When attribution is done badly or not at all, states lose credibility and likely effectiveness in dealing with those who would harm the state and its citizens. These risks hold for state-on-state interactions across the spectrum of cyber operations—from espionage to destructive attacks on infrastructure.

Although identifying the machines responsible for a cyber intrusion is no longer a difficult task for the most advanced states,[11] identifying the persons, organizations, or states that are legally responsible for the cyberattack remains challenging.[12] The problems derive from technical means of deception and anonymity, but they are also due to the vagaries of the process of fixing responsibility for cyberattacks within the international community and the malleability and open-endedness of the few attribution rules that currently exist in international law.[13]

[3] Ben Buchanan & Robert D. Williams, *A Deepening U.S.-China Cybersecurity Dilemma*, LAWFARE (Oct. 24, 2018); Robert Chesney, *An American Perspective on a Chinese Perspective on the Defense Department's Cyber Strategy and 'Defending Forward'*, LAWFARE (Oct. 23, 2018).

[4] David D. Clark & Susan Landau, *Untangling Attribution*, 2 HARV. NAT'L SEC. J. 531, 531 (2011).

[5] *See* John P. Carlin, *Detect, Disrupt, Deter: A Whole-of-Government Approach to National Security Cyber Threats*, 7 HARV. NAT'L SEC. J. 391, 409 (2016); JEFFREY CARR, INSIDE CYBER WARFARE 29, 139–40 (2010).

[6] Thomas Rid & Ben Buchanan, *Attributing Cyber Attacks*, 38 J. STRATEGIC STUD. 4 (2015).

[7] *Id.* at 7 ("[A]ttribution is an art as much as a science.").

[8] *See* Carlin, *supra* note 5, at 396–97 (discussing the expertise required for the complex attribution analysis).

[9] *Id.*

[10] Rid & Buchanan, *supra* note 6, at 4.

[11] *See, e.g.*, Carlin, *supra* note 5, at 416; Herbert Lin, *Attribution of Malicious Cyber Incidents: From Soup to Nuts*, 70 J. INT'L AFFAIRS 75, 82–83 (2017).

[12] Carlin, *supra* note 5, at 409; Lin, *supra* note 11, at 84.

[13] *See* William Banks, *State Responsibility and Attribution of Cyber Intrusions After Tallinn 2.0*, 95 TEX. L. REV. 1487, 1494–97 (2017).

https://doi.org/10.1017/aju.2019.32 Published online by Cambridge University Press

Legal Challenges in Attributing an Unlawful Cyber Intrusion

The law of attribution for cyber intrusions aims to identify and place responsibility for internationally wrongful acts. The customary law of state responsibility for violations of international law, and with it the law of attribution, is drawn largely from the long-term work of the International Law Commission (ILC) and its *Draft Articles on Responsibility of States for Internationally Wrongful Acts*.[14] The ILC rules were commended to the member states by the UN General Assembly in 2001 and have become the authoritative guidepost for public international cyber law. Drawing from the ILC rules, the starting point in the cyber context is that "a State bears international responsibility for a cyber-related act that is attributable to the State and that constitutes a breach of an international legal obligation."[15] In essence, states are responsible for the wrongful cyber-related acts of their own officials, agents, contractors, nonstate actors, and other states, to the extent they actually control the operations.[16] States do not escape legal responsibility for internationally wrongful acts by perpetrating them through proxies.[17]

Outside an armed conflict, international law forbids cyber intrusions that violate the prohibition on intervention.[18] Based on the international law principle of sovereignty, the principle forbids coercive intervention by cyber means.[19] The consensus among experts is that state-on-state cyber intrusions that are not coercive but are "detrimental, objectionable, or otherwise unfriendly"[20] are not international legal violations. As confirmed by the ICJ in the *Nicaragua* judgment, "the element of coercion … forms the very essence of … prohibited intervention."[21]

Yet international law has never had a precise definition of "coercion." According to a consensus among the cyber experts who contributed to *Tallinn 2.0*, "coercion is not limited to physical force, but rather refers to an affirmative act designed to deprive another State of its freedom of choice … to force that State to act in an involuntary manner or involuntarily refrain from acting in a particular way."[22] A state compels another state by, for example, providing cyber training or supplying malware to a terrorist group operating in the compelled state.[23] Yet defining the full range of cyber conduct that qualifies as "coercion" has been more difficult. The International Group of Experts (IGE) that provided the analysis in *Tallinn 2.0* could only agree on the anodyne statement "that as a general matter, States must act as reasonable States would in the same or similar circumstances when considering responses to them."[24] The IGE elaborated:

> Reasonableness is always context dependent. It depends on such factors as, *inter alia*, the reliability, quantum, directness, nature (e.g., technical data, human intelligence), and specificity of the relevant available information when considered in light of the attendant circumstances and the importance of the right

[14] UN Int'l Law Comm'n, Report of the International Law Commission, Draft Articles of State Responsibility, UN GAOR, 53rd Sess., Supp. No. 10, UN Doc. A/56/10 (2001) [hereinafter Articles on State Responsibility].

[15] Tallinn Manual 2.0 on the International Law Applicable to Cyber Operations 3 (Michael N. Schmitt ed., 2017) [hereinafter Tallinn Manual 2.0].

[16] *Id.* at 87–92 (rule 15).

[17] *Id.* at 94–95 (rule 17).

[18] *Id.* at 312 (rule 66(1)).

[19] *Id.* at 312–13.

[20] *Id.* at 85 (rule 15(7)).

[21] Military and Paramilitary Activities in and against Nicaragua (Nicar. v. U.S.), Merits, 1986 ICJ Rep. 14, para. 205 (June 27).

[22] Tallinn Manual 2.0, *supra* note 15, at 317 (rule 66(18)).

[23] Michael N. Schmitt & Liis Vihul, *Proxy Wars in Cyberspace: The Evolving International Law of Attribution*, I Fletcher Sec. Rev. 55, 60 (Spr. 2014).

[24] Tallinn Manual 2.0, *supra* note 15, at 81.

https://doi.org/10.1017/aju.2019.32 Published online by Cambridge University Press

> involved. These factors must be considered together. Importantly in the cyber context, deficiencies in technical intelligence may be compensated by, for example, the existence of highly reliable human intelligence.[25]

Recognizing that customary international law has not developed a set of understandings or recognized state practice on what level of attribution is acceptable or necessary for establishing state responsibility for cyber actions, the IGE concluded that "States may agree between themselves to a rule of responsibility specific to a cyber act or practice."[26] The result would be lex specialis to the extent the rule conflicts directly with general principles of state responsibility.[27] Similar uncertainty surrounds the degree of certainty on attribution a state should achieve before asserting the legal right to take self-help measures in response that would otherwise violate international law.

Deciding How and When a "Victim" State May Respond

If a state is victimized by an internationally wrongful act below the use of force threshold, the "State may be entitled to take countermeasures, whether cyber in nature or not."[28] Countermeasures are responses that otherwise would violate international law, designed to prevent a responsible state from continuing its unlawful cyber intervention.[29] Countermeasures require prior notice to the offending state, and they must have as their purpose inducing compliance with international law.[30] Punitive countermeasures are forbidden.[31] Short of countermeasures, states may respond to cyber intrusions through retorsions, acts that are "unfriendly" but lawful.[32] Examples include protests, denying access to state resources, and economic sanctions.[33]

The *Tallinn 2.0* experts agreed that

> in the context of unilateral self-help measures, the reality is that States must make *ex ante* determinations with respect to attribution of a cyber operation to another State before responding. … [T]he State may be faced with a situation to which it may have to respond in an extremely short time frame, without recourse to the full range of information that might be available in the non-cyber context.[34]

The IGE opined that "as a general matter the graver the underlying breach … , the greater the confidence ought to be in the evidence relied upon by a State considering a response[35] … because the robustness of permissible self-help responses … grows commensurately with the seriousness of the breach."[36] However, the severity of the cyber intrusion directed at an injured state is also relevant, so that a state confronted with "low-level cyber operations

[25] *Id.* at 81–82.

[26] *Id.* at 80.

[27] *Id.* at 81.

[28] *Id.* at 111 (rule 20).

[29] *Id.*

[30] *Id.* at 111–12.

[31] *Id.* at 124.

[32] *Id.* at 112; Articles on State Responsibility, *supra* note 14, chapeau to ch. II of pt. 3, para. 3 of commentary.

[33] Thomas Giegerich, *Retorsion*, Max Planck Encyclopedia of Public International Law (Rüdiger Wolfrum ed., 2017).

[34] Tallinn Manual 2.0, *supra* note 15, at 81.

[35] In support of its position, the IGE cited Oil Platforms (Iran v. U.S.), 2003 ICJ Rep. 161, para. 33 (Nov. 6) (separate opinion of Judge Higgins); Corfu Channel Case, (UK v. Alb.), 1949 ICJ Rep. 4, para. 17 (Apr. 9); Application of Convention on Prevention and Punishment of Crime of Genocide (Bosn. & Herz. v. Serb. & Montenegro), 2007 ICJ Rep. 108, paras. 209–10 (Feb. 26); and Application of Convention on Prevention and Punishment of Crime of Genocide (Croat. v. Serb.), 2015 ICJ Rep. 3, para. 178 (Feb. 3).

[36] Tallinn Manual 2.0, *supra* note 15, at 82.

https://doi.org/10.1017/aju.2019.32 Published online by Cambridge University Press

that are merely disruptive" may be expected to amass more evidence for attribution than a state victimized by "devastating cyber operations and needing to respond immediately to terminate them."[37]

In a similar vein, the time it takes to produce a high confidence attribution judgment can limit the lawful responses to cyber operations. Mistaken attribution can lead to an unlawful response even if the state made a reasonable attribution judgment and implemented countermeasures.[38] If a state victimized by an internationally wrongful cyber intrusion engages in countermeasures and ends up being wrong about state attribution, the victimized state has committed an internationally wrongful act.[39] On the other hand, if the victim state waits until it has high confidence in its attribution of a state's responsibility for the intrusion, any countermeasures may be construed as punishment, a form of reprisal forbidden under international law.[40] As a result, cyber deterrence may be undermined because the legally less risky but weak self-help retorsion responses to an intrusion are unlikely to deter similar cyber intrusions in the future.

Prospects for International Legal Regulation

The short-term prospects for international legal regulation of cyber attribution are not encouraging. The failure of a state to provide persuasive proof of attribution is not itself an internationally wrongful act. Nor are there burdens of proof or additional legal criteria for establishing attribution. The 2015 United Nations Group of Governmental Experts (GGE) report noted that accusations of wrongful acts by states "should be substantiated,"[41] but the GGE gave no indication of which or how much evidence would suffice or even count. The U.S. view, articulated by State Department Legal Adviser Brian Egan in November 2016, is that "a State acts as its own judge of the facts and may make a unilateral determination with respect to attribution of a cyber operation to another State. … [T]here is no international legal obligation to reveal evidence on which attribution is based prior to taking appropriate action."[42]

States are likewise not obligated to provide evidence of attribution when responding to another state's cyber intrusions.[43] While the IGE acknowledged the value in such a disclosure requirement, it found insufficient state practice and opinio juris to recognize "an established basis under international law for such an obligation."[44] The IGE noted that the highly classified nature of such attribution assessments is the primary reason for the absence of customary international law on this important point.[45]

Although attribution is necessarily probabilistic, the process serves its purpose if it convinces the responsible state (and victim state's citizens) that a response to the cyber intrusion is called for.[46] The fact that attribution judgments draw on many different sources of information has one major temporal implication—early judgments made with less information are generally less believable than later judgments made with more information. Continuing investigation may reveal additional useful information, which may (or may not) reinforce attribution

[37] *Id.*

[38] *Id.* at 82–83.

[39] *Id.* at 118–20.

[40] *Id.* at 116.

[41] Group of Governmental Experts on Developments in the Field of Information and Telecommunications in the Context of International Security, Report, UN Doc. A/70/174, para. 24 (July 22, 2015).

[42] *See* Brian J. Egan, Remarks on International Law and Stability in Cyberspace at Berkeley Law School (Nov. 10, 2016).

[43] Tallinn Manual 2.0, *supra* note 15, at 83.

[44] *Id.*

[45] *Id.*

[46] Clement Guitton, Inside the Enemy's Computer: Identifying Cyber Attackers 66 (2017).

https://doi.org/10.1017/aju.2019.32 Published online by Cambridge University Press

judgments made earlier. Over time, an international consensus may develop on the minimum level of involvement needed to declare that a state is legally responsible for a cyber operation.

States are better able to attribute cyber intrusions than they were a decade ago. Yet the technical environment is so dynamic that new tools constantly both improve and occlude attribution capabilities. Spoofing and other challenges can greatly complicate attribution and a response when an immediate response is required, particularly when a state is the suspected perpetrator. As cyber international relations now stand, a few states benefit from the absence of express cyber norms on what suffices to attribute state responsibility for cyber exploitation because they have the most offensive cyber capabilities. However, in general, those states are also the most vulnerable to cyber intrusions. Meanwhile, the disparity between states that are strong and weak at attribution results in the equivalent of an arms race between advances in detection versus detection evasion. Evasion is getting easier faster, so states that do not have advanced attribution capabilities can reliably invest in hiding themselves.[47]

As the most advanced cyber states recognize the risks of cyber escalation, those states have good reason to become more transparent about attribution in service of the mutual restraint that could be gained by sharing attribution information. But to date, state concerns about revealing intelligence sources and methods counsel against transparency.[48]

[47] Bruce Schneier, Click Here to Kill Everybody: Security and Survival in a Hyper-Connected World 54–55 (2018).

[48] *Id.* at 54.

https://doi.org/10.1017/aju.2019.32 Published online by Cambridge University Press