2019

The attribution problem and cyber armed attacks

Lorraine Finlay

Christian Payne

This article was originally published as:
Finlay, L., & Payne, C. (2019). The attribution problem and cyber armed attacks. *American Journal of International Law, 113*, 202-206.

Original article available here:
https://doi.org/10.1017/aju.2019.35

This article has been published in the *American Journal of International Law.*
Published by Cambridge University Press and the American Society of International Law.
Available at: https://doi.org/10.1017/aju.2019.35

Finlay, L., and Payne, C. (2019) The attribution problem and cyber armed attacks. *American Journal of International Law, 113*. doi: 10.1017/aju.2019.35

doi:10.1017/aju.2019.35

SYMPOSIUM ON CYBER ATTRIBUTION

THE ATTRIBUTION PROBLEM AND CYBER ARMED ATTACKS

Lorraine Finlay & Christian Payne***

In late 2018, the U.S. Secretary of Homeland Security suggested that "cyber-attacks now exceed the risk of physical attacks."[1] Yet the law has not kept pace with this reality. In particular, identifying who is responsible for a cyberattack makes it difficult to regulate this conduct. A state often cannot practically respond to a threat unless it knows from where the threat emanates and potentially who is responsible. Attribution of cyber conduct is critical from a legal perspective because the unlawful act must be attributable to another state for state responsibility to be engaged.

This essay provides an overview of this attribution problem in the context of cyberattacks that might qualify as armed attacks for purposes of the *jus ad bellum*. These attacks are especially grave and, as such, ones that states will most want to respond to decisively. Understanding the availability and limits of a lawful response is therefore critical. The essay begins by briefly reviewing the current state of international law concerning armed attacks and self-defense, and whether and how cyberattacks fit within this framework. It then examines the attribution problem, noting that there are both technical and legal hurdles to overcome when attempting to reconcile conventional approaches to attribution with the unconventional characteristics of a cyberattack. Finally, it proposes a more contextually appropriate model for attribution that states could use in the case of cyber armed attacks to address the attribution problem.

Cyber Armed Attacks and Self-Defense

There is ostensibly general agreement that the commonly accepted rules of public international law surrounding the use of force also apply to cyberattacks. The starting point here is the prohibition on the use of cross-border force,[2] absent authorization by the Security Council[3] or circumstances of self-defense.[4] However, while the right to self-defense potentially applies to cyberattacks, it is available only in response to an "armed attack."[5] Therefore, whether a cyberattack can be classified as a "use of force" or an "armed attack" is critical in understanding the lawful responses that are available to a state.

* *Law Lecturer, Murdoch University; Adjunct Senior Lecturer, University of Notre Dame, Sydney.*

** *Lecturer, Murdoch University.*

[1] Kirstjen M. Nielsen, *Rethinking Homeland Security in an Age of Disruption*, U.S. Dep't of Homeland Security (Sept. 5, 2018).

[2] UN Charter art. 2(4).

[3] *Id.* art. 39.

[4] *Id.* art. 51.

[5] Tallinn Manual 2.0 on the International Law Applicable to Cyber Operations 339–48 (Michael N. Schmitt ed., 2d ed. 2017) [hereinafter Tallinn Manual].

In theory, there is no obstacle to a cyberattack reaching the threshold of both a "use of force" and an "armed attack." A cyberattack will constitute a use of force "when its scale and effects are comparable to non-cyber operations rising to the level of a use of force."[6] In order to objectively assess whether a cyberattack rises to this level, states likely will consider the severity, immediacy, directness, invasiveness, measurability of effects, military character, state involvement, and presumptive legality of the operation.[7]

A precise definition of what constitutes an "armed attack" is more elusive. It is well-established that "it will be necessary to distinguish the most grave forms of the use of force (those constituting an armed attack) from other less grave forms" with "armed attacks" being different in their "scale and effects" from lesser uses of force.[8] However, the exact threshold is not clear. The *Tallinn Manual* suggests that "acts of cyber intelligence gathering and cyber theft, as well as cyber operations that involve brief or periodic interruption of non-essential cyber services" will not meet the threshold of an "armed attack."[9] To date, few if any of the known cyberattacks appear to meet the threshold for a "use of force," let alone for an "armed attack."[10]

Further, while the inherent right of self-defense in theory may extend to cyber armed attacks, there are a number of challenges that arise in this context. These challenges, discussed below, suggest that the traditional concept of self-defense is of limited value in the context of cyber armed attacks.

The Attribution Problem

There are, broadly speaking, two aspects to the attribution problem. The first is the technical problem of how to identify the true origin of a particular attack and the identity of those who carried it out. The second is the legal question of whether and if so, when, factual attribution allows a state to be held responsible for the cyberattack under international law.

The technical consequences of the unique characteristics of cyberattacks have been widely acknowledged. Dan Efrony and Yuval Shany have highlighted two key differences between cyber and other kinds of attacks: the boundlessness and anonymity of the cyber domain.[11] As cyberspace is not limited by physical or geographical borders, questions of jurisdiction and enforcement become more complex. Perpetrators use various techniques to obscure their true location, and isolating the origin of a cyberattack is extraordinarily difficult when attacks are routed through multiple machines in multiple locations across the world. These techniques make the task of retrospectively establishing a forensic link between an attacker and an incident extremely difficult. They allow cyberattacks to be carried out with little risk to those involved, with perpetrators able to mask their identity and potentially even misattribute blame.

Thus, cyberattacks often take more time to detect and evaluate than traditional kinetic attacks. This analysis might require using classified intelligence and technological capabilities that a state would prefer not to reveal. All of these factors make it challenging to attribute cyberattacks.

These technical difficulties have two main consequences for attribution, both of which are critical in relation to cyber armed attacks. The first is that the risk of misattribution is heightened, leading to the potential for serious

[6] *Id.* at 300.

[7] *Id.* at 334–37.

[8] Military and Paramilitary Activities in and Against Nicaragua (Nicar. v. U.S.), Merits, 1986 ICJ Rep. 14, para. 191 (June 27) [hereinafter Nicaragua].

[9] Tallinn Manual, *supra* note 5, at 341.

[10] Dan Efrony & Yuval Shany, *A Rule Book on the Shelf? Tallinn Manual 2.0 on Cyberoperations and Subsequent State Practice*, 112 AJIL 583 (2018).

[11] Dan Efrony, The Cyber Domain, Cyber Security and What About the International Law? (The Federmann Cyber Security Center).

conflict escalation if a state mistakenly targets an innocent third party in its self-defense response. The second is that the length of time it will realistically take to correctly identify the perpetrator will very likely mean that it will be significantly harder for a state to satisfy the factors of immediacy and necessity required to lawfully exercise its right of self-defense.

For example, the famous formulation in the *Caroline* case described the necessity requirement for the use of force in self-defense as one in which the response is immediate, "leaving no choice of means and no moment for deliberation."[12] The recent WannaCry ransomware attack highlights the difficulties here. This attack by a ransomware cryptoworm began on May 12, 2017 and affected more than 230,000 computers across 150 countries within a single day. In attributing the attack to North Korea in December 2017, the U.S. Homeland Security Advisor acknowledged that the U.S. government needed "a lot of time to look through classified, sensitive information" and to "put it together in a way that allowed us to make a confident attribution," but that "the most important thing is to do it right and not to do it fast."[13] However, given the time that had elapsed in resolving the attribution issue, the traditional necessity requirement might at that point have precluded any right to respond in self-defense (assuming all of the other requirements for self-defense had been met). The highly controversial[14] intention element to self-defense referred to in the *Oil Platforms* case[15] is similarly problematic, as discerning the intention of an attacker is likely to take time and may entail questions of identity. Therefore, delayed attribution suggests self-defense was unlikely to be available as a lawful response to the attack, even assuming it met the "armed attack" threshold.

Turning to the legal problems of attribution, a state will not be held responsible for an internationally wrongful act unless that act is attributable to the state under international law,[16] with the conventional principles of state responsibility applying to cyberattacks.[17] The conduct of a nonstate actor may be attributed to a state where the nonstate actor is "acting on the instructions of, or under the direction or control of" the state, or where the state acknowledges and adopts the conduct as its own.[18]

The principles of attribution relating to the actions of nonstate actors are particularly relevant to cyberattacks, given that "the relative availability and cheapness of the technology necessary to mount a cyber-attack makes significant attacks launched by private individuals or corporations far more feasible compared with other types of warfare."[19] Attribution here is also critical given the prevailing view that cyberattacks conducted by nonstate actors that are not attributable to states are not a use of force for the purposes of the UN Charter regime.[20]

Traditionally there has been a high threshold for attributing the actions of nonstate actors to states, requiring the state to have had "effective control" over each specific operation for which attribution is sought.[21] The International

[12] Letter from Secretary of State Daniel Webster to British Minister to the United States Lord Alexander Baring Ashburton (Aug. 6, 1842).

[13] Thomas P. Bossert, *Press Briefing on the Attribution of the WannaCry Malware Attack to North Korea*, White House (Dec. 19, 2017).

[14] William H. Taft, *Self-Defense and the Oil Platforms Decision*, 29 Yale J. Int'l L. 295, 299 (2004).

[15] Oil Platforms (Iran v. U.S.), 2003 ICJ Rep. 161, para. 64 (Nov. 3).

[16] G.A. Res. 56/83, Annex, Responsibility of States for Internationally Wrongful Acts, UN Doc. A/RES/56/83, art. 2 (Jan. 28, 2002) [hereinafter ARSIWA].

[17] Tallinn Manual, *supra* note 5, at 84–87.

[18] *Id.* at 94–100.

[19] Christian Payne & Lorraine Finlay, *Addressing Obstacles to Cyber-Attribution: A Model Based on State Response to Cyber-Attack*, 49 Geo. Wash. Int'l L. Rev. 535, 536 (2018).

[20] Tallinn Manual, *supra* note 5, at 174–76. The legal framework surrounding possible state responses to cyberattacks conducted by nonstate actors is beyond the scope of this essay, but is an issue of considerable importance.

[21] Nicaragua, *supra* note 8, at para. 115.

Criminal Tribunal for the Former Yugoslavia adopted a broader standard of "overall control" in *Prosecutor v. Tadić*.[22] Although lowering the threshold, *Tadić*'s "overall control" standard applied to an organized and hierarchically structured group. This quasi-military context is easily distinguishable from the more decentralized groupings usually associated with cyberattacks. Further, the ICJ has rejected the use of the "overall control" test.[23]

In any event, regardless of whether "effective control" or "overall control" ultimately applies, the attribution threshold remains extremely high. Mere encouragement or support by a state through financing and equipping groups will be insufficient for attribution.[24] Simply financing a cyberattack or providing a safe haven to nonstate perpetrators would not appear to meet the threshold for the state itself to be held responsible for a cyberattack.

One possible response to these challenges is to lower the attribution threshold in cases of cyberattacks.[25] If the existing legal framework creates a threshold so high that states have no prospect of attributing responsibility for a cyber armed attack, states will be effectively precluded from any lawful response, including self-defense. This creates a dangerous practical lacuna, and emboldens perpetrators by allowing them to operate with impunity. It is unrealistic to assume that a state in these circumstances would simply not respond to a cyber armed attack. Instead, states will likely respond outside of the existing legal framework. An excessively high attribution threshold ultimately leaves states with no lawful option for self-defense and undermines the rules-based international order by incentivizing covert retaliatory cyber operations.

On the other hand, lowering the attribution threshold also carries potential dangers. As discussed above, there is a serious risk of misattribution, which has particularly grave consequences in relation to cyber armed attacks, where a state may retaliate in self-defense. Failing to maintain strict attribution requirements gives rise to a significant risk of escalation and conflict.[26]

Ultimately, when dealing with cyber armed attacks, attribution, and self-defense, the bar is set so high that it will rarely be met practice. The key questions are: has there been an armed attack; have the requirements of self-defense been established; and can responsibility be attributed to a state? For each question there are significant difficulties with applying conventional principles to cyber armed attacks.

A New Model for Attribution

It is critical to establish an attribution standard that strikes the right balance. As we have previously observed, there is the risk of escalating tension and conflict from misdirected retaliation due to flawed attribution or, alternatively, precluding law-abiding states from lawfully responding to cyberattacks because the legal prerequisite of attribution is practically impossible to establish.[27]

Therefore, approaching cases of cyber armed attack using the conventional principles of attribution is unlikely to be helpful, given that the threshold questions cannot realistically be answered. Instead, a more nuanced model that applies variable attribution standards based upon the remedy the victim state chooses to pursue may offer a more

[22] Prosecutor v. Tadić, Case No. IT-94–1-A, Judgment, para. 120 (Int'l Crim. Trib. for the Former Yugoslavia, Appeals Chamber July 15, 1999) [hereinafter Tadić].

[23] Application of the Convention on the Prevention and Punishment of the Crime of Genocide (Bosn. & Herz. v. Serb. & Montenegro), Merits, 2007 ICJ Rep. 43, para. 406 (Feb. 26).

[24] Tadić, *supra* note 22, at 120.

[25] A second possibility is to extend the right to self-defence to cyber armed attacks by nonstate actors, which raises many of the same issues discussed in recent years about the use of self-defence in the context of terrorism. This issue is beyond the scope of this essay, but is worth further exploration.

[26] Payne & Finlay, *supra* note 19, at 564.

[27] *Id.* at 566.

useful path forward. In cases where a state wishes to respond to a cyber armed attack using force, a strict test of attribution should continue to apply. This minimizes the otherwise grave risk of misdirected retaliation due to the technical obstacles to attribution. If, however, a state seeks redress through nonforcible legal or diplomatic channels instead, these concerns are less significant and the attribution threshold can be lowered.

One option would be a strict liability model that holds host states indirectly responsible for breaches occurring within their territory. Vincent-Joël Proulx proposed such a model in the context of terrorism.[28] Allowing it to authorize acts of self-defense in relation to cyberattacks is, in our view, dangerously flawed due to the risk of misattribution. However, a strict liability approach holds appeal when states are instead pursuing negotiated outcomes through legal means, as the adverse effects of misattribution are significantly reduced. Under this approach an injured state may, for example, be able to use countermeasures against the state from which the cyberattack allegedly originated (assuming the other legal requirements for countermeasures are met),[29] even if there was no evidence that the host state had any involvement in the attack beyond its territory being used. A strict liability model also encourages state cooperation by creating potential liability for reparations. When we consider that some estimates have put the potential costs from the previously mentioned WannaCry attack at US$4 billion,[30] the strong incentive that this model establishes for states to actively cooperate to prevent cyberattacks is readily apparent.

The lowered threshold in this context incentivizes host states to cooperate with injured states, particularly if assisting an injured state in collecting evidence and preventing future attacks would help to discharge a host state's obligations. This strict liability model is distinct from the due diligence obligation under Rule 6 of the *Tallinn Manual 2.0,* as it does not establish a separate wrongful act on the part of the host state. Rather, this approach would apply strict liability to questions of attribution where a state has suffered a prohibited use of cyber force but chooses to pursue remedies through nonforcible legal processes rather than forceful retaliation. Used in this way, strict liability avoids the scenario where the injured party could neither protect itself before the fact, nor receive justice after it.

Conclusion

The conventional principles of self-defense and attribution are poorly suited to dealing with cyber armed attacks. The technical problems of attribution in cyberspace may make it practically impossible for victim states to meet the legal requirements of both attribution and self-defense. Under the current legal framework, states are effectively prevented from utilizing the right to self-defense in response to a cyber armed attack.

Of course, self-defense is not the only legal option available to states.[31] In particular, in many scenarios, countermeasures may be a useful form of nonforcible self-help. Similarly, there are relevant legal obligations other than the prohibition on the use of force, with the general due diligence obligation being especially relevant. A comprehensive examination of these is beyond the scope of this essay, but all have relevance in the context of the legal framework surrounding cyberattacks.

In the specific context of cyber armed attacks, however, conventional principles could be adapted by developing a model that varies attribution requirements based upon the victim state's contemplated response. Such a model would build on existing legal foundations, minimize the risk of misattribution and conflict escalation, and incentivize cooperation between states. It would also go some way towards addressing some of the existing challenges with regards to attribution that this essay has identified.

[28] Vincent-Joël Proulx, *Babysitting Terrorists: Should States Be Strictly Liable for Failing to Prevent Transborder Attacks?*, 23 Berkeley J. Int'l L. 615, 643–59 (2005).

[29] ARSIWA, *supra* note 16, at ch. 2.

[30] Jonathan Berr, *"WannaCry" Ransomware Attack Losses Could Reach $4 billion*, CBS News (May 16, 2017).

[31] Tallinn Manual, *supra* note 5, at 104–11.