Constructive Powers and Regional Security in the Asia-Pacific
ASAN-CIGI-ASPI Seoul Forum Workshop
Seoul, South Korea — October 18-19, 2013

Rule-Making for State Conduct in the Attribution of Cyber-Attacks

Mark Raymond and Aaron Shull

Introduction

This past June, on the 63rd Anniversary of the start of the Korean War, the New York Times reported that offensive cyber operations had been initiated against various websites in both Korean countries.[1] In the North, the affected websites included those of the main Communist Party newspaper and the official Korean Central News Agency. In the South, the website for the presidential office and another government agency were rendered temporarily inaccessible. However, determining who was responsible for these cyber intrusions was complicated by several facts. First, a number of individuals purporting to belong to the "hacktivist" network Anonymous had indicated on Twitter that they would undertake offensive cyber operations against the North on the war's anniversary. Second, Twitter users also "claimed responsibility for the attacks in South Korea, saying they demanded that the South's government stop censoring Internet content."[2] Given the relative ease of compromising authentication for social media accounts and the difficulty of independently establishing the source of origin for data transmitted over the Internet, it is difficult to regard such claims of responsibility as dispositive.[3]

Problems in attributing the cyber-attacks related to the anniversary of the Korean War expose a deeper, and more nuanced problem. The challenge is that attributing cyber intrusions back to a government sponsor is not simply a factual or technical challenge. Rather, it is a multifaceted problem that has technical, legal and political dimensions, all of which must be satisfied for the attribution of a cyber-attack to a particular state to be appropriate. The additional complexity here, as with most things cyber, is that the rules governing conduct in each of these categories are still evolving. However, the complexity of the task should not mask its importance. The increasing importance of the Internet to daily life can be expected over time to amplify the stakes involved in large-scale, persistent disruptions of access. Further, technology already allows the creation of kinetic effects by means of malicious code; such capabilities are highly likely to become more widely available in the foreseeable future.

[1] Choe Sang-Hun, *Cyberattacks Disrupt Leading Korean Sites*, N.Y. Times, available at http://www.nytimes.com/2013/06/26/world/asia/cyberattacks-shut-down-leading-korean-sites.html?_r=0
[2] Ibid.
[3] For more on the difficulties of attributing cyber attacks and on cyber strategy in general, see Joseph S. Nye, Jr., "Nuclear Lessons for Cyber Security," Strategic Studies Quarterly 5.4 (2011): 18-38.

Accordingly, there is real potential for escalation from cyber-attacks to more traditional forms of violent conflict. New rules, norms and institutional structures are needed that can ameliorate the possibility of misattribution and the attendant consequences that would flow from it, and that can act as circuit-breakers inhibiting escalatory spirals even where attribution is properly made.

In order to advance this broad claim, the paper will proceed in the following fashion. The first section will highlight the technical aspects of attributing a cyber-intrusion to a particular actor or set of actors. The second section will discuss the legal rules governing state responsibility and the attribution of acts to state actors. The third section will discuss the international political dimensions involved in determining whether or not to publicly attribute a particular cyber act to a given state. Finally, the paper concludes by arguing clearer rules are needed to avert the possible negative consequences of misattribution and that an impartial international institutional cooperative mechanism is needed to facilitate the development of a meaningful rule based framework for state conduct related to attribution for cyber-attacks.

The Technical Dimension of Attribution

Technical attribution of cyber-attacks is primarily a matter of employing computer forensics to determine the physical point(s) of origin for a particular computer network operation and to locate circumstantial evidence embedded in malicious code that provides information about the likely identity of its author(s), such as languages employed or particular cultural points of reference. For a small number of governments, these means can be augmented with additional intelligence sources and methods.

The expertise necessary to such attribution efforts is, again with the exception of a small number of governments, in the hands of private companies. These large computer network operators, technology companies and computer security firms possess increasingly sophisticated technical means to attribute cyber-attacks.[4] However, the involvement of corporate actors in the attribution of cyber-attacks is likely to prove controversial. Specifically, it is not clear whether states will generally accept the judgment of such companies – particularly when the company is headquartered in the same state making the public attribution of the attack.

Leaving problems of credibility aside, there are also clear limits to technical attribution, particularly when principals employ arms-length agents (potentially located in other states) and/or transnational botnets in the conduct of cyber-attacks. Given the utility of such gambits in frustrating attempts at attribution and the low cost of employing them, these practices are likely to spread. Even when location of origin can be conclusively identified, this is not necessarily a dispositive indication of responsibility either under narrow legal standards or more expansive political ones.

The International Legal Dimension of Attribution

Assuming that cyber-attacks constitute internationally wrongful acts, and assuming that the technical/factual indications provide sufficient evidence that the attack in question has a state connection, then invariably the relevant international legal rules to apply will be the ones governing state responsibility. In short, state responsibility is a fundamental rule governing international relations

[4] For one recent example, see Mandiant, "APT1: Exposing One of China's Cyber Espionage Units." Available at http://intelreport.mandiant.com/.

which holds that if a state commits an internationally wrongful act that is attributable to the state, and the act constitutes a breach of an international obligation, then that state will owe a duty to make reparations for the breach.[5] The relevant question then becomes: what activities are attributable to a state? Unfortunately, the answer is not entirely clear.

Legal attribution is straightforward when the act in question is committed by an organ of the state. In these cases, the law will presume that the act committed by the organ is the act of the state itself. Suppose for example that a cyber-attack was directed by a division of the military or the state intelligence services; this would "according to a well-established rule of international law […] be regarded as an act of that state."[6] Determining responsibility becomes more complex, however, when dealing with entities that are more loosely affiliated with the state. This complexity arises, at least partially, from the fact that there are competing and inconsistent legal rules that apply.

The Draft Articles of the International Law Commission, which potentially reflect customary law,[7] state that the conduct of an entity will be considered an act of a State under international law if that entity "is in fact acting on the instructions of, or under the direction or control of, that State in carrying out the conduct."[8] Thus, under this standard if a state directed a citizen hacker team to carry out a cyber-attack, then those acts would be legally attributable back to that state. However, two international courts have weighed in on what the applicable standard should be when considering the attribution of certain acts to states, and in so doing, they have muddied the analytical waters.

In the *Nicaragua* case, the International Court of Justice was asked to decide whether the acts of the Contras – a rebel group engaged in a conflict with the Nicaraguan government – were attributable to the Government of the United States. In determining whether the relationship between the two was sufficiently close to allow for attribution, the Court held that "to be legally responsible, it would have to be proved that that State had *effective control* of the operations in the course of which the alleged violations were committed."[9] Thus, in order to be found responsible for a particular cyber intrusions, under the *Nicaragua* standard, it must be shown that the state in question had direct control over that very intrusion. In other words, "general overall control would have been insufficient to ground responsibility."[10] Rather, according to the jurisprudence of the International Court of Justice a more exacting standard applies.[11]

However, the International Criminal Tribunal for the Former Yugoslavia sees the status of international law on this issue quite differently. In the *Tadić* case, the Tribunal found that a demonstration of *overall*

---

[5] Responsibility of States for Internationally Wrongful Acts, 2001, at Article 2, http://untreaty.un.org/ilc/texts/instruments/english/draft%20articles/9_6_2001.pdf ; See also *Chorzów Factory*, PCIJ, Series A, No. 17, 1928, p. 29; Malcolm N. Shaw, *International Law*, 6th Ed., at 778 [hereafter Shaw].

[6] *Difference Relating to Immunity from Legal Process of a Special Rapporteur,* ICJ Reports, 1999, pp. 62, 87.

[7] See, e.g., Memorial of Germany (LaGrand Case) (Sept. 16, 1999) at para. 6.15 (referring to the ILC draft articles on responsibility as "the most authoritative statement of customary international law on the matter").

[8] Responsibility of States for Internationally Wrongful Acts, 2001, at Article 8, available at http://untreaty.un.org/ilc/texts/instruments/english/draft%20articles/9_6_2001.pdf

[9] *Case Concerning the Military and Paramilitary Activities in and Against Nicaragua (Nicaragua v. United States Of America)*(Merits) ICJ Reports, 1986, pp. 14, 64-5.

[10] Shaw, *supra* note 5, at 790.

[11] See also *Application of the Convention on the Prevention and Punishment of the Crime of Genocide (Bosnia and Herzegovina v. Serbia and Montenegro) (Judgment of Feb. 26, 2007),* available at *http://www.icj-cij.org/docket/files/91/13685.pdf*

control could be sufficient for a finding of state responsibility. In this way, the Tribunal incorporated a less stringent test under the rules of international law, noting that "the degree of control might vary according to the circumstances and a high threshold might not always be required."[12] As a consequence, the international community is left with a somewhat fragmented set of legal rules to apply to cyber-attacks, when trying to attribute those acts back to state actors. Does one need to show that the cyber attacker was acting on the instructions of the state or that the state had effective control of the impugned operation? Or, is it appropriate to apply some lower threshold for the attribution of acts in cyber-space? Should states, for example, be required to take reasonable steps to prevent attacks that originate from their territory? And if they fail to do so, should that provide a sufficient basis for a finding of responsibility? Or, should there be a more exacting standard in which states are called to account only when there is dispositive proof that an organ of the state committed or directed the attack?
Ultimately, the applicable rules remain somewhat unclear. However, what is clear is that the increasing centrality of cyber operations in the conduct of international relations is going put pressure on the existing set of rules and those rules will – no doubt – evolve in relation to that pressure. Thus, what is required at this point is some forward looking policy making aimed at developing a set of rules that accounts not only for where the subject technology in now, but where it is going to be in the future.

Political Dimensions of Attribution

Technical and legal aspects of attributing cyber-attacks are, respectively, questions of evidentiary sufficiency – whether it is possible to determine the source of the attack to a degree of confidence that can withstand appropriate scrutiny – and appropriate procedure. These dimensions of attribution intersect in a complex fashion with a third, political, dimension. The key characteristic of the political dimension of attribution is publicity; a politically attributed attack, for the purposes of this paper, entails a public accusation.

This political dimension of attribution engages calculations of interest both by state and (increasingly) non-state actors, as well as less understood dynamics of choice driven by emotion and by justice motives.[13] As a result of these considerations, it is possible that attacks that can be reliably attributed from technical and legal perspectives will not be publicly (politically) attributed (i.e. dogs that don't bark). It is also possible that political pressures may lead to the public attribution of cyber-attacks even when relevant technical and legal criteria are not clearly satisfied (i.e. potentially false positives). Both of these outcomes can create problematic impacts on international security.

The primary risk associated with false positives is the creation of escalatory spirals. Accusations, particularly when not accompanied by convincing technical data and when not made according to proper legal procedure, can damage diplomatic relations. The offense-dominant nature of the cyber domain (at least given the current state of technology)[14] suggests strong potential for escalation in the event of a cyber-attack; the perceived necessity of immediate reprisal also creates pressures militating toward hasty attribution. If the attribution is accompanied by self-help measures, the risk of tit-for-tat escalation and the emergence of a crisis atmosphere is increased. In crises, domestic audience costs

[12] Shaw, *supra* note 5, at 790.
[13] On the role of justice in violent conflict see David A. Welch, *Justice and the Genesis of War* (Cambridge: Cambridge University Press, 1993). See also Mark Raymond and David A. Welch, "How Ideas Shape Conflict: Theoretical and Conceptual Considerations," paper presented at the General Conference of the European Consortium for Political Research, 6 September 2013, Bordeaux, France.
[14] On this feature of the cyber domain, see Nye, "Nuclear Lessons," *SSQ*.

have a critical effect on the willingness of parties to back down.[15] Ironically, the Internet's expansion of low-cost communication and its increasing integration with critical infrastructure and other essential aspects of everyday life may act to increase the size of the domestic audience and the stakes of significant cyber-attack disruptions in the eyes of that audience, thus ratcheting up the audience costs associated with conceding in a dispute concerning a publicly attributed cyber-attack and making such crises especially prone to escalatory spirals. Societies with a high degree of Internet penetration may be most prone to escalate in the aftermath of a publicly attributed cyber-attack. As global rates of Internet penetration increase, there is thus also a greater risk of conflict spirals.

These features suggest the dangers inherent both in the employment of cyber-attacks (and conduct that may be mistaken for a cyber-attack, such as some cyber-crime and cyber espionage activities) and in the hasty attribution of cyber-attacks. Less obviously, there is also a significant danger in widespread failure to attribute cyber-attacks when technical and legal criteria have been satisfied. If cyber-attacks are not criticized and no efforts are made to hold bad actors accountable, the establishment of permissive norms (and even permissive customary international law) is likely. Given the escalatory potential of cyber-attacks, the prevalence of permissive norms could pose a significant international security risk. This risk entails, at maximum, the employment of cyber-attacks to create kinetic damage and widespread disruption to critical infrastructure, as well as the potential escalation to more traditional interstate military conflict. While such outcomes are unlikely, they carry high costs. More modestly, permissive norms could lead to higher incidence of increasingly sophisticated cyber-attacks resulting in degradation of Internet services and a loss of trust in the Internet as a means of commerce, communication and delivery of government services. Even this more optimistic scenario involves significant economic opportunity costs. The key point is that, due to its ability to forestall the development of permissive norms, criticism of socially undesirable behavior is important in itself, (at least partially) independent of the ability to effectively enforce international rules.[16]

Conclusion: Rule-Making for State Conduct in the Attribution of Cyber-Attacks

Given that both attribution and non-attribution of cyber-attacks create significant international security risks, what is the appropriate policy response? We argue that thoughtful international efforts at rule-making governing state conduct in the attribution of cyber-attacks have the potential to steer a middle course mitigating the risks associated with both widespread non-attribution and with hasty attribution. Such rules must govern: (1) when and how states publicly attribute cyber-attacks; and (2) the consequences of attribution. While the *ex ante* specification of such rules is impossible due to the unpredictable nature of international negotiations, it is possible to sketch out some general design features of promising rule sets.

With respect to the procedural mechanics of attribution, several such features can be identified. First, attribution should be done on the basis of technical best practices for determining the point of origin for Internet traffic. Second, state and non-state actors seeking to attribute cyber-attacks should bear in mind the ease of concealment either by technical means or by the employment of arms-length proxy agents. Accordingly, attribution should be made only after attempts to engage appropriate authorities in

[15] On international dispute escalation and the role of audience costs, see James D. Fearon, "Domestic Political Audiences and the Escalation of International Disputes," *American Political Science Review* 88.3 (September 1994): 577-592.

[16] On such social pressure tactics, see Alistair Iain Johnston, "Treating International Institutions as Social Environments," *International Studies Quarterly* 45.4 (December 2001): 487-515.

the apparent country of origin. Where such authorities react cooperatively and in good faith, this should be prominently noted when public attribution is made.

There is also potential to develop rules and financing mechanisms for increased international cooperation on the technical and legal aspects of attribution, for example by augmenting national Computer Emergency Response Teams (CERTs) both in industrial and developing states and by augmenting the Forum of Incident Response and Security Teams (FIRST). FIRST, founded in 1990, is the only current global organization coordinating response to computer incidents.[17] It could be given an authoritative role in the investigation and attribution of cyber-attacks in cooperation with national authorities. At minimum, states and other actors could be encouraged to seek its input prior to making public attributions of cyber-attacks. Such steps would increase confidence in the accuracy and independence of attribution. The epidemiological efforts of the World Health Organization may provide a model for the future of cooperative international efforts to track and monitor socially undesirable network activity.[18]

States should also be encouraged to decouple public attribution of cyber-attacks from nationalist and other inflammatory rhetoric. Efforts to attach emotional valences to cyber-attacks may serve domestic political purposes, but their international effects are likely to be undesirable; such rhetoric should therefore be avoided, and should be consistently criticized when undertaken by other actors.

Finally, consistent with the conclusion of the United Nations Group of Governmental Experts (GGE) that the law of armed conflict applies online, response to cyber-attacks should be conducted within the framework of customary international law on state responsibility. The assessment of the GGE represents a start for international rule-making on cyber-attacks rather than a conclusion; however, it suggests that cyber-attacks should be conducted only in self-defence or in accordance with authorization of collective action under the UN Charter. It further suggests that cyber-attacks must distinguish between military and civilian targets and must accord with rules of proportionality. Treating response to cyber-attacks under the legal framework of state responsibility establishes limits on justifiable self-help measures that states may take in response to an attributed cyber-attack. Encouraging adherence to these limits can minimize the escalatory potential of cyber-attacks.

The creation and application of such standards and practices will require responsible global leadership, primarily but perhaps not solely from states. While there will be a need to finance and resource new international facilities for attribution of cyber-attacks, and a need for rule-making both about the procedures and consequences of public attribution, there is also a need for a commitment to the restrained public attribution of cyber-attacks by state actors. The widespread, consistent exercise of such restraint in the face of what are likely to be increasingly common and perhaps increasingly disruptive incidents will take political will and statesmanship.

[17] See www.first.org. Also see the CERT Coordination Center, www.cert.org.

[18] Admittedly, such an epidemiological model would face obstacles related to states' apparent desire to employ the Internet for espionage purposes.