A HOOVER INSTITUTION ESSAY

Attribution of Malicious Cyber Incidents

FROM SOUP TO NUTS

Attribution of malicious cyber activities is a deep issue about which confusion and disquiet can be found in abundance. Attribution has many aspects—technical, political, legal, policy, and so on. A number of well-researched and well-executed papers cover one or more of these aspects, but integration of these aspects is usually left as an exercise for the analyst. This paper distinguishes between attribution of malicious cyber activity to a machine, to a specific human being pressing the keys that initiate that activity, and to a party that is deemed ultimately responsible for that activity. Which type of attribution is relevant depends on the goals of the relevant decision-maker. Further, attribution is a multidimensional issue that draws on all sources of information available, including technical forensics, human intelligence, signals intelligence, history, and geopolitics, among others. From the perspective of the victim, some degree of factual uncertainty attaches to any of these types of attribution, although the last type—attribution to an ultimately responsible party—also implicates to a very large degree legal, policy, and political questions. But from the perspective of the adversary, the ability to conceal its identity from the victim with high confidence is also uncertain. It is the very existence of such risk that underpins the possibility of deterring hostile actions in cyberspace.

HERBERT LIN

Aegis Paper Series No. 1607

Attribution of malicious cyber activities is a deep issue, about which confusion and disquiet can be found in abundance. Attribution has many aspects, and a variety of well-researched and well-executed papers cover one or more of these aspects; these papers are referenced in the body of the paper and are called out again in the Acknowledgments section. This paper tries to synthesize the best aspects of these works with some original thoughts of the author's own into a coherent picture of how attribution works, why it is both important and difficult, and how the entire process relates to policymaking.

The primary takeaway messages of this paper are that (1) attribution has a different meaning depending on what a relevant decision-maker wants to do (i.e., attribution of malicious cyber activity can be to a machine, to a specific human being pressing the keys that initiate that activity, and to a party that is deemed ultimately responsible for that activity); (2) attribution is a multidimensional issue that draws on all sources of information available, including technical forensics, human intelligence, signals intelligence, history, and geopolitics, among others; (3) all attribution judgments are necessarily accompanied by some measure of uncertainty; and (4) an adversary cannot be fully confident of its ability to conceal its identity from the victim.

What Is Attribution About?

Every parent who has ever broken up a fight between two children and tried to figure out what happened has asked, “Who started this?” The question expresses our very basic concerns about responsibility for actions that lead to conflict or harm.

Concerns about responsibility for actions or for events are embedded in domestic law. A person is found on the street with a bullet through his head, and we want to know who fired that shot. Much of our criminal justice system is devoted to “fair” processes that we believe can determine the identity of that person with sufficient certainty to mete out an appropriate punishment. International law is concerned with questions of responsibility as well, especially as it relates to matters involving conflict. With a number of important (and controversial) exceptions, states are usually regarded as accountable for actions that emanate from within their borders.

Similar concerns about responsibility are also present in cyberspace, but just how they play out is often quite different, for reasons both technical and historical. Usually captured under the rubric of attribution, concerns about responsibility generally arise when a malicious cyber activity or incident is known to have happened.[1] “Who (or what) is responsible?” is then often the question of interest.

If this question cannot be answered, it may be hard for victims to mitigate ongoing harm: to do so would require the victim to be able to quickly and correctly identify the instrument or mechanism causing the harm and find a way to stop its malicious activities. Further, it would be impossible to punish the parties responsible for causing the incident. And, if punishment is impossible, deterrence of malicious activity in the future is also difficult to achieve.[2]

We begin with a working definition of a cyber incident. We recognize a cyber incident when something “bad” happens to an information technology-based system. In this context, badness involves errant behavior of the victim’s computer (or a system involving a computer)—that is, the computer or system behaves in a way that it should not behave. Examples abound: the computer freezes; commands given to the computer do not have the expected result; the printer spews out paper with gibberish.[3] More serious examples of badness include the following: a drive-by-wire car does not slow down when the driver presses the brake pedal; the computer-controlled missile misses a target when it should have hit it; or the ATM machine at the corner bank dispenses hundreds of $20 bills onto the street.

Investigations are usually (but, alas, not always) triggered by errant computer (or system) behavior. But apart from routine inspections, investigations will not occur if the errant behavior occurs and we have no clues that it has occurred. Similarly, clues may be noticed only long after the precipitating actions or events have occurred, making investigations much more difficult.[4]

The first part of the investigation is determining that something "errant" has happened at all. In all of the examples above, it is pretty clear that an undesirable outcome has occurred, and the undesirability demonstrates (or at least suggests) a breakdown in the program's functionality. But consider the case in which a computer system (and anything that is controlled or affected by that system) produces an undesirable result or outcome that is what would be expected given the inputs. (Most people who have tried to balance a checkbook by hand, or even with a calculator, can speak to such an experience.) In such cases, it is far more likely that the result—though undesirable—is correct and inevitable because the user has provided bad inputs than it is that the program used to calculate that result is in error.

Similarly, if the missile misses its target or the car does not slow down when the driver presses the brake pedal, it is possible that a human operator aimed the missile at a shadow or the driver pressed the accelerator when he thought he pressed the brake. In such cases, it is hard to associate "errant" behavior with the computer or system per se, since the system was given the wrong input.[5] It is also possible that the errant behavior is the result of a flaw in the program, introduced by accident rather than intentionally.

Errant behavior resulting from factors other than foul play does not usually play a part in traditional attribution concerns. Attribution usually arises as a concern when an incident is determined to have resulted from foul play (i.e., intentional harm). When the determination is made that foul play was involved, what was previously a cyber incident involving errant system behavior becomes a malicious cyber incident (or, equivalently, an intrusion)—and attribution is the process by which it is determined who or what is responsible for the intrusion.

Attribution sometimes goes hand in hand with determining if a cyber incident is malicious, a descriptor that usually implies bad intention on the part of some actor. That is, an investigation regarding the cause of errant system behavior may (or may not) reveal it to be the deliberate and intentional action of an actor. But identification of the specific actor is not necessarily required to infer bad intention—in many cases, a particular behavior of the system is so likely to be the result of an intentional bad action that investigators presume maliciousness.

Suppose that Bill is the legitimate user of a computer in the human resources department of a large defense contracting firm. He has been putting together a spreadsheet with all of the names, addresses, e-mail addresses, and salaries of the other employees of this firm. One day, he opens his computer to discover that the spreadsheet has been deleted from his hard drive. He reports this to IT support, which then begins an investigation. What happened? How did the file get deleted?

The IT support staff may begin by examining who had access to the file. Susan, Bill's direct supervisor, also had access to the file. Susan, however, claims that she did nothing to the file. Network records demonstrate that Susan's computer did access and delete the file the evening before Bill reported it missing. Is Susan forgetful or lying? Or was she somehow tricked into deleting the file? Or did someone else access Bill's file, pretending to be Susan?

Perhaps the investigators determine—or make an educated guess—that Susan is indeed telling the truth, and that she inadvertently deleted the file without knowing it. Who set this action in motion? In this case, misdirection is involved: on the surface, Susan appears responsible, but she did not wish for the file to be deleted and does not actually bear any meaningful responsibility for ill intent.[6]

But the IT support staff may determine that an intruder engineered this attack through Susan's computer. Attribution has two goals: to distinguish between errant behavior that is malicious and deliberate and errant behavior that is accidental and, if the former, to distinguish between intentional, real, and meaningful responsibility on one hand and apparent responsibility on the other. The latter goal focuses on the question of who set this event in motion. However, determining "real" responsibility is much more difficult than it may initially seem. This paper explores different ways of understanding attribution and, subsequently, responsibility for a malicious cyber incident.

What Does Attribution Mean?

Ascertaining responsibility for malicious cyber activity can be understood in a variety of different ways because the term "responsibility" has a number of possible meanings, any or all of which may (or may not) be relevant in any given situation.

Working through a concrete scenario helps to unpack the meaning of "responsibility." The following scenario, as known from a God's-eye perspective,[7] involves Tony, the systems administrator for a Department of Defense (DOD) computer system in San Francisco. This computer system is attacked (and, in this instance, has been the subject of a remote-access attack in which an unauthorized party—George—took direct control of it as if he were sitting at the keyboard in San Francisco). The attack traffic came from a computer based in Arkansas, owned by Karen, an 84-year-old woman. The computer in Arkansas, however, was compromised through a computer in Greece. George sat at the keyboard in Greece and pressed the keys that set into motion the attack against the DOD computer in San Francisco. George is a citizen of China. However, he is also a member of a Russian organized crime group. The head of that crime group, Sergey, is a close personal friend of a senior operative named Ivan in the Federal Security Service (FSB) in Russia. Ivan and Sergey had dinner two weeks ago, and while Ivan and Sergey did not talk about computers or hacking, Ivan did tell his close friend that he was having problems with some activity happening at a DOD facility in San Francisco.

Who is "responsible" for the attack on the US computer in San Francisco?

Three meanings of attribution

In principle, the question of "who is responsible?" can be answered in three ways, which are not mutually exclusive. The possible types of answers are a machine, a specific human being pressing the keys or otherwise setting the intrusion into motion, and an ultimately responsible party.[8] To distinguish between the human being and the ultimately responsible party, the reader should understand the term "intruder" (or, equivalently and interchangeably, "perpetrator") to mean the former and the term "adversary" to mean the latter. Some degree of uncertainty attaches to any specific answers. Which possible type of answer should be sought depends on the goal of the relevant decision-maker.

Attributing malicious cyber activity to a machine (or machines) In the above example, attributing the intrusion to a machine would require identifying the computers used to perpetrate it on the DOD computer in San Francisco. The easiest machine to identify is Karen's computer, since that computer is proximate (in cyberspace) to the DOD computer. Any other computers through which the intrusion was routed are also of interest because each computer in the path points to one or more additional links. The trail will eventually stop somewhere, either at George's computer because the evidence collected along the way suggests that George's computer is in fact the originating point of the attack (a good outcome) or somewhere else because the trail peters out (a bad outcome). Following Clark and Landau,[9] an intrusion in which multiple computers are used in a chain to reach the intended target is called a multi-stage intrusion.[10]

Ascertaining the machines associated with a malicious cyber incident usually involves technical forensics—the art and science of looking for technical clues left behind in an intrusion.[11] In tracing the origin of the activity, it may be necessary to gain access to Karen's computer to obtain any relevant information it might have. Technical forensics could also be performed at the network level without needing direct access to Karen's computer, e.g., by examining various logfiles that document what has been happening on the servers in the network. (In general, technical forensics at the network level must examine large volumes of mostly irrelevant information to find the few—if any—relevant entries.)

For example, technical forensics applied to the DOD computer may reveal the IP address of Karen's computer, which was the one most immediately and proximately connected to the DOD computer in San Francisco. By consulting a service that provides geocoded IP addresses, investigators may learn that this computer is in Arkansas. Internet address assignment authorities will show the name of the Internet service provider associated with that specific IP address—call the ISP in question Castcom. Using a subpoena, investigators may then ask Castcom to reveal the name of the subscriber using that IP address at that time. Castcom may or may not be able to provide that information. For example, the logs containing a dynamic assignment of their customers to IP addresses may only be retained by them for a brief time, or they may be using a technology called Carrier Grade NAT (Network Address Translation) that shares a single IPv4 address among a multiplicity of customers. Should Castcom reveal that the name of the subscriber is Karen, and that her address is 132 Main Street in Little Rock, Karen may receive a visit from investigators armed with a search warrant who demand access to her computer to gather further information.

On the other hand, if Karen's computer is found in another country rather than in the United States, it is likely that a different set of procedures would obtain. Under some circumstances, investigators may ask law enforcement authorities in that country for assistance. Under other circumstances (such as the refusal of that country's authorities to cooperate), they may simply find a technical way to gain access (e.g., they hack into it by sending an authorized user of the computer an e-mail that grants them access when the victim clicks on a link in the e-mail).

In either case, the proximate computer may well hold additional clues that help to identify the next link in the chain. For example, the investigators may find malware on Karen's computer that periodically contacts a particular IP address in Greece.

Technical forensics can be challenging,[12] especially in an environment in which multi-stage cyber intrusions are conducted. Complicating the technical forensics job even more,

BOX 1: DARPA'S EFFORTS ON ENHANCED ATTRIBUTION

In April 2016, DARPA announced a solicitation for proposals related to enhanced attribution. The announced program aims to make currently opaque malicious cyber adversary actions and individual cyber operator attribution transparent by providing high-fidelity visibility into all aspects of malicious cyber operator actions and to increase the government's ability to publicly reveal the actions of individual malicious cyber operators without damaging sources and methods.

The program will develop techniques and tools for generating operationally and tactically relevant information about multiple concurrent independent malicious cyber campaigns, each involving several operators, and the means to share such information with any of a number of interested parties (e.g., as part of a response option). The program seeks to develop:

- technologies to extract behavioral and physical biometrics from a range of devices and vantage points to consistently identify virtual personas and individual malicious cyber operators over time and across different endpoint devices and C2 infrastructures
- techniques to decompose the software tools and actions of malicious cyber operators into semantically rich and compressed knowledge representations
- scalable techniques to fuse, manage, and project such ground-truth information over time, toward developing a full historical and current picture of malicious activity
- algorithms for developing predictive behavioral profiles within the context of cyber campaigns
- technologies for validating and perhaps enriching this knowledge base with other sources of data, including public and commercial sources of information

Source:

Broad Agency Announcement on Enhanced Attribution, DARPA-BAA-16-34, April 22, 2016, https://www.fbo.gov/utils/view?id=138959e641d75afda40b9bedb5ec8d2b

anonymity-enhancing tools can be used; such tools obscure technical information that might be used for forensics. Impeding technical forensics may serve a socially desirable goal when it protects people who engage in politically controversial dialogue, but anonymity-enhancing tools can also be problematic when they help malicious cyber actors to evade responsibility for their actions and get in the way of identifying the actual machines involved in perpetrating an intrusion.

TOR is a good example. TOR is a system that enables users to communicate more anonymously across the Internet with ease.[13] TOR traffic is automatically encrypted and routed through many different nodes around the world rather than being routed directly. A list of anonymity-enhancing tools is maintained by the Electronic Privacy Information Center;[14] the proper use of such tools increases the difficulty of performing technical forensics.

At the same time, anonymity-enhancing tools are only one side of the coin. Efforts to improve technical forensics are also underway. A contemporary example is described in Box 1 entitled "DARPA's Efforts on Enhanced Attribution."

A second source of information that can contribute to an attribution judgment is honeypots. A honeypot is, in essence, a decoy configured to look attractive to an intruder but instrumented so that the intruder's behavior can be clandestinely observed and monitored. If and when the same intruder returns to the targeted installation, his behavior can be recognized more easily.

A third source of information useful for machine attribution consists of pre-positioned instrumentation. In some cases, pre-positioning of instrumentation occurs in systems and networks that an adversary might use to launch an intrusion. Thus, if that adversary initiates an intrusion, the pre-positioned instrumentation can record data streams that, when properly interpreted, indicate the nature and source of malicious activity underway. Such instrumentation was reportedly part of the attribution to North Korea of the attack against Sony Pictures Entertainment in 2014.[15] Use of pre-positioned instrumentation obviously presumes a prior policy decision that a particular adversary may launch future intrusions and that an investment in anticipatory emplacement of such instrumentation is therefore justified.[16]

In other cases, instrumentation is pre-positioned as a matter of good security practice on the part of others or even good luck. In the first instance, consider the possibility that an intruder is able to successfully launch an intrusion that appears to be coming from Institution A. If Institution A has installed instrumentation that monitors traffic in and out of its networks (a good security practice of A), Institution A may be able to show that it was not in fact the source of the intrusion. That fact may in turn provide information on the techniques used by the intruder. Good luck may contribute if the intruder unwittingly reveals actions that may be preparatory to the intrusion. In both cases, information potentially relevant to attribution is uncovered; if shared among the relevant parties, that information may actually be relevant.

Two observations about this process are noteworthy. First, attributing to a machine or an IP address is not the same as identifying the human being who perpetrated the attack. Technical information can point to a computer located at IP address 62.217.69.62 and note that this particular IP address is associated with someone calling himself George.[17] While that piece of information is suggestive, it does not imply that George was necessarily the individual who pressed the keys initiating the attack.

Second, as Clark and Landau point out, the use of one or more intermediaries (in this case, Karen's computer) through which to route an intrusion greatly complicates the technical forensics task. Investigators start with information found on the DOD computer, and this information points to Karen's computer. They need information from Karen's computer, but their access rights to that privately owned computer in Arkansas are more limited than if they had full control over it (which they would have if it were a DOD computer). In addition to their technical tasks, they now also face tasks based on law and policy about how and to what extent, if any, they may access Karen's computer. If the law and policy are clear in any given instance, those tasks may be relatively easy to complete. But if they are not (e.g., what if Karen's computer is in Brazil rather than Arkansas and the investigators need Brazilian permission to access Karen's computer due to a bilateral agreement between the two nations?),[18] carrying out the full range of technical forensics needed may be much more difficult.

Attributing malicious cyber activity to a human intruder Attributing malicious cyber activity to a human intruder means ascertaining the identity of the person or persons directly involved in perpetrating it. In the example above, attributing the activity to its human intruder means identifying George as the person who pressed the keys on the keyboard located in Greece needed to launch it.

Since anyone could be sitting at that keyboard in Greece, technical forensics alone cannot definitively determine the identity of that person because technical forensics usually look only at information that may have been left behind on the various computers in the wake of an intrusion.[19] However, someone else may have stolen George's login credentials to pretend that she is George, and the identity of the credentials thief may not be discoverable using only technical forensics. (A non-cyber analogy: the fact that John Doe's car may have been the car that killed a pedestrian does not mean that John Doe was the one driving the car.)

In the example above, investigators might consult historical records and find that this particular Greek IP address has been identified many times in the past as an originating point for a variety of Chinese and Romanian hackers. But the particular malware found on Karen's computer has been used primarily by Chinese hackers in the past, thus suggesting that Chinese rather than Romanian involvement in this attack is more likely.

Yet another clue might be found in a Chinese online discussion forum that is ostensibly private but that has been secretly infiltrated by a US intelligence agency for a number of years. In this forum is a question from George asking for the most recent information about security measures taken at the DOD computer facility in San Francisco—and the date on which this question was posted is eight days before the attack on the San Francisco computer.

If enough such clues can be accumulated, the investigators may have sufficient confidence to point to George as the most likely perpetrator of the intrusion on the DOD computer in San Francisco. Of course, how many and what kinds of clues are "enough" is an important question and is the focus of a later section of this paper entitled "How Attribution Judgments Are Made." Another important question is the strength of these clues, since no one clue is likely to be definitive (i.e., investigators of such incidents rarely, if ever, find a "smoking gun"). For purposes of attribution, investigators may require a large number of clues that point only weakly to a given person or a fewer number of clues that point strongly to that person.

There are many instances in which technology can help facilitate attribution to a human intruder. Authentication is the process through which specific individuals can be better tied to technical online activities and actions. Most people are familiar with the ritual of entering a login name followed by a secret password. If the login process is successful (and the user's login credentials have not been compromised), the user is granted access to a variety of privileges on the relevant computer system, and many of that user's actions on the system can be associated with him or her personally.

If the user goes beyond the local computer system onto the Internet, an Internet service provider (ISP) will have provided Internet access. That ISP will often have information on file about the individual to provide access (e.g., where the individual is) and to receive payment (e.g., through the individual's credit card). Thus, the ISP may have some insight into the Internet activities of its subscriber as individuals. (The ISP may not have complete insight into activities carried out on its networks. For example, if the individual sends e-mails with attachments encrypted locally, the ISP will know about their recipients, but not know about their contents. But such information might not be necessary for an attribution judgment, depending on the particular pattern of facts and circumstances that obtain at the time.) Using the ISP's records on its subscribers, an investigator would be in a better position to attribute some activity carried on its network to a particular individual.

And technical means do sometimes point directly to specific individuals. For example, the way an individual types on a keyboard may be sufficient to specify that individual uniquely—that is, no other person in the world would type a particular passage of text with the same timing of keystrokes.[20] If the human intruder is using a remote-access tool to explore the victim's computer system or network, a keystroke monitor may be able to capture such data. (Indeed, the DARPA program on enhanced attribution described above uses keyboard dynamics as one aspect of identifying virtual personas of intruders.) Similarly, hacking into the computer in Greece to turn on its camera and capture a picture of the person at the keyboard would also yield useful information.

Such means can indeed provide useful information about an individual's identity, just as a DNA signature (a specific genomic sequence of As, Ts, Cs, and Gs belonging to an individual) or fingerprints can point to specific individuals. But none of these signatures—keyboard, pictures, DNA, or fingerprints—is of any value in *identifying* the individual unless there is some database against which the given signature can be compared and an identity uncovered. That database is the essential link between specifying an individual and identifying that individual, and technical forensics applied to any one incident, cyber or otherwise, cannot populate that database. In the absence of such a database, the most that can be said is that the same individual perpetrated two or more intrusions, but this individual will not be identifiable.

Compromising this link is also the intent of stealing credentials. Someone may have used George's credentials to gain access to the computer in Greece, but how do we know if that someone was actually George? Two-factor authentication is a stronger form of authentication than a username-and-password combination that calls for the user to present something he or she knows (e.g., a password) with something he or she has (e.g., a token or a smartphone). The use of two-factor authentication reduces the likelihood that an attempt to impersonate George will succeed. But two-factor authentication is not foolproof, as a gun held to George's head will also probably serve the same purpose for someone determined to use George's credentials.

More generally, even if George can be identified as the human perpetrator of the intrusion, it is often important to know why George did it and who asked him to do it. That is, for many purposes, the identity of the party responsible for setting the intrusion into motion is quite important. Who is the party that is ultimately responsible for the intrusion?

Attributing malicious cyber activity to the ultimate responsible party On whose behalf was George acting? George may be acting on his own—that is, he alone chose to carry out the intrusion and acted accordingly. But in the most general case, George acts on behalf of another party—usually an organization, such as his employer, his gang, or his government. Attributing malicious cyber activity to a specific adversary as the ultimate responsible party answers the question "who is to blame?" rather than "who did it?" (which is the focus of attributing an intrusion to its human perpetrator).[21]

Considered in this light, it is clear that the party on whose behalf George is acting cannot be determined by technical forensics alone. Indeed, in some cases it is possible that technical forensics play only a minimal role in making this determination.

A non-cyber example is a good place to start. If a missile fired from an Elbonian navy ship caused damage to a US Navy ship during peacetime in the Atlantic Ocean, the United States would hold Elbonia responsible. If Elbonia asserted that the ship's captain was a rogue actor and not acting on orders from the Elbonian government, it would be up to the Elbonian government to demonstrate that this claim were true. For example, in no particular order, the Elbonian government could prosecute and punish the captain; allow the United States to interview the captain and members of the crew; pay reparations; formally apologize; show the United States the orders under which the captain was operating; or share the message traffic to and from the ship to Elbonian authorities before and after the incident or recordings made on the bridge of the Elbonian ship during the incident.

Some combination of these (and/or other steps) might suffice to persuade the United States that the missile firing was the act of a rogue captain and that the Elbonian government should not be held responsible for what would otherwise be an illegal use of force. But the reason that the Elbonian government would be required to demonstrate its lack of culpability in such an incident is the international convention that states that, in general, states are responsible for the acts of their armed forces.[22] Units of these forces are clearly marked with national insignias, partly for this reason. The rationale for this presumption is that, historically, only states have had the wherewithal to build and use weapons that are capable of threatening national security.

But it is unclear how to apply present conventions for state responsibility to cyber incidents and the extent to which, if any, cyber-specific rules would be needed for such application. Is Greece the responsible party because George launched the attack from Greece? Is China the responsible party because George is a Chinese citizen? Is the Russian organized crime group the responsible party because of George's involvement with the group? Is Russia the responsible party because of ties between the FSB and the organized crime group of which George is a part? In principle, a plausible case could be made for any of these possibilities. But in the absence of a broad political agreement or convention that argues for one over the other, the determination of "the responsible party" is necessarily based on policy and political judgments that take into account the relevant facts known from all sources.

The relationship among the three types of attribution As noted above, the question of "who is responsible?" can be answered by pointing to a specific machine (or machines), a specific human being pressing the keys, and a specific adversary as the ultimately responsible party. But the discussion above should make clear that the last kind of attribution is different from the first two in that the notion of a party that is "ultimately responsible" implicates legal, policy, or political issues to a much greater degree. Sections below on nation states and subnational entities as the ultimately responsible entity will build on this point.

There is not necessarily a direct connection between these different types of attribution. Knowing the machine responsible (i.e., the machine causing the damage being suffered by the victim) does not necessarily provide the identity of the human perpetrator, and knowing the identity of the human perpetrator does not necessarily reveal the party that is ultimately responsible, i.e., the adversary.

Nevertheless, although these three types of attribution are conceptually distinct, they are often related in practice. Knowing the machine from which the intrusion initially emanated may provide some clues that can help uncover the identity of the human perpetrator, and knowing the human perpetrator may provide some clues that can help identify the party ultimately responsible for setting the entire intrusion into motion.

For example, if the machine originating an intrusion is definitively located in Nation A, it suggests that the human perpetrator has access to machines in Nation A. If Nation A is a country in which only a small segment of the population has easy access to computers, the search for the perpetrator's identity may entail examining fewer possible suspects than if Nation A made it easy for large segments of the population to access computers. A common clue picked up by technical forensics is the language setting for the keyboard of a particular computer. Despite the fact that many people in the world are multilingual, such a clue is nevertheless suggestive and raises the likelihood that the human perpetrator is from a nation in which that language is used.

It may also be the case that responsibility cannot be allocated cleanly to a specific party. For example, in decentralized organizations, it is common for the leader to express his or her intent and then leave it to subordinates to execute in accordance with that intent. A subordinate operator may well do something that he believes is consistent with that intent but in fact may be "too much" from the perspective of the leader. In such a situation, responsibility is diffused among the individuals involved in an unclear manner.

A worked example of attribution In 2013, Mandiant released a report called "APT1: Exposing One of China's Cyber Espionage Units,"[23] identifying a group it called "APT1" as a single organization of operators that conducted a cyber espionage campaign against a broad range of victims between 2006 and 2013. Mandiant concluded that APT1 was most likely sponsored by the Chinese government. Mandiant was also able to develop profiles ("personas") on several individuals within APT1, though it was not able to determine with any certainty their real names or identities.

The attribution process in which Mandiant engaged touched on all three meanings of attribution: specific machines, specific human beings (perpetrators) pressing the keys, and an ultimately responsible party.

For example, the Mandiant report notes:[24]

> [C]yber intruders leave behind various digital “fingerprints.” They may send spear-phishing emails [in this case, emails to specific individuals within the targeted company containing malicious links or files] from a specific IP address or email address. Their emails may contain certain patterns of subject lines. Their files have specific names, MD5 hashes, timestamps, custom functions, and encryption algorithms. Their backdoors may have command and control IP addresses or domain names embedded.

All of these indicators were used by Mandiant in its identification of the specific machines used by APT1 in its intrusions.

Mandiant used a variety of other information to associate these machines with Chinese actors. For example, it noted large volumes of intrusion traffic associated with blocks of IP addresses known to be assigned to Chinese Internet service providers operating in Shanghai. APT1 hackers also used a remote desktop client from Microsoft to manage its remote access to targeted systems; in the majority of such cases, the keyboard language setting was “simplified Chinese.”

Public domain registration information (e.g., who is the registered owner of the domain example.com) also helps to identify specific individuals; such information includes names, addresses, phone numbers, and e-mail addresses. Of course, an intruder may provide false registration information when asked, but systematic errors (e.g., misspellings) can provide valuable clues as well.

To identify individuals, Mandiant searched the web for various e-mail addresses uncovered through domain registration and other sources. In many cases, these e-mail addresses were also found on other sites providing additional information about the individual, often apparently supplied by the individual. Mandiant was confident in its identification of personas, but far less certain about the actual names associated with those personas.

As for an ultimately responsible party, Mandiant pointed to a specific unit of the People’s Liberation Army (PLA). Mandiant first identified a group of operators who perpetrated a large number of intrusions, resulting in the exfiltration of large volumes of information. It found that the industries targeted matched industries that China has identified as strategic to its growth. Mandiant then identified a unit of the PLA (Unit 61398) that was similar to this group in its mission, capabilities, and resources, as well as being located in the same geographical area from which many APT1 activities appeared to have originated. Mandiant identified individuals with a connection to Unit 61398, which appears to be actively soliciting and training English-speaking personnel specializing in a wide variety of cyber topics, such as covert communications, operating system internals, digital signal processing, and network security. Unit 61398 also recruits new talent from the science and engineering departments of Chinese universities and associates various “profession codes” describing positions within Unit 61398 with competence in highly technical computer skills. Lastly, Mandiant found a memo describing a special fiber optic communication infrastructure provided by the state-owned enterprise China Telecom in the name of national defense.

In sum, Mandiant asserted high confidence that APT1 should be associated with Unit 61398 of the PLA. But it also acknowledged the possibility that “a secret, resourced organization full of mainland Chinese speakers with direct access to Shanghai-based telecommunications infrastructure [had] engaged in a multi-year, enterprise scale computer espionage campaign right outside of Unit 61398’s gates, performing tasks similar to Unit 61398’s known mission.”

Attribution for different types of intrusion For simplicity of discussion, this section uses a particular scenario involving an intrusion on a DOD computer in San Francisco to illustrate some aspects of the attribution process. But although the scenario is based on a multi-stage intrusion in which intermediate computers are used to mask the computer from which a remote-access intrusion originated, other types of intrusion are possible and, indeed, as common—or even more common—than that depicted. In practice, the attribution process unfolds differently with different types of intrusion.

For example, intrusions result from the sending of an e-mail to a user, who then clicks on a malicious link or attachment and inadvertently launches malware that takes destructive action in his computer. That e-mail may be sent from a Gmail address, and it is well known that a Gmail address can be created from anywhere (e.g., a WiFi-equipped coffee shop) with near total anonymity. In this case, there are no intermediate stepping stones that will lead back to an originating computer. Technical forensics may thus of necessity focus more on characteristics of the malware being used.

Intrusions can occur when a user merely surfs the web on ostensibly safe sites. Because many sites display advertisements, the content that a user sees on his or her screen is not entirely under the control of the operator of the website to which the user navigated. Ad content can be poisoned, so that when the image from the ad is displayed, malware is downloaded to the user’s computer. Investigation in this instance may require approaching the party who obtained ad display rights on that website.

Other types of intrusion may not involve the Internet at all. An adversary may be able to compromise the hardware supply chain, leading to the delivery to the intended victim of a clandestinely modified computer that is never connected to the Internet. The modification might cause the computer to destroy itself on a specific date. In this case, technical forensics would focus on the characteristics of the compromised hardware that was delivered to the user, which is not the focus in investigations involving malware. Another scenario involves inducing a user to insert into his computer a USB key that is contaminated with malware and runs upon insertion. In such cases, technical forensics directed at Internet activity may not reveal useful information, depending on what the malware did (e.g., if it destroyed files without accessing Internet services), but the manufacturer of the USB key may be able to provide insight. In such scenarios, technical forensics coupled to other investigations might yield useful information about the human perpetrator immediately responsible for the intrusion.

Legal authorities for gathering information related to attribution

For the United States and its various law enforcement and intelligence agencies, gathering information that might be used to make an attribution judgment does not take place in a vacuum—that is, US law and policy govern the information-gathering activities of US law enforcement and intelligence agencies, and in particular recognize several key distinctions. These include information-gathering undertaken domestically versus that undertaken on foreign soil; information-gathering undertaken to investigate domestic criminal activity versus that undertaken for national security and foreign intelligence purposes; and information-gathering involving US citizens versus foreigners.

In today’s security environment, the activities of adversaries often blur these lines. The September 11, 2001, terrorist attacks on the World Trade Center and the Pentagon were clearly matters of national security, but they were criminal acts as well. Terrorists may seek to fund their operations by engaging in criminal activity such as human or drug trafficking. Foreign terrorists may operate from US territory, thereby gaining some of the default protections afforded to US citizens on US soil. And US citizens may undertake criminal activity on behalf of foreign governments or terrorist movements. Operating in cyberspace further complicates these distinctions, as communications traffic (and intrusions) freely transit national borders, while jurisdiction and legal authorities to gather information do not.

Describing existing law relevant to gathering information useful for attribution, John Carlin, assistant attorney general for national security in the Obama administration, writes in a recent article that “online” investigations are in fact conducted mostly offline and thus use investigative tools for obtaining information related to attribution such as physical examination of servers, conversations with network users, and requests for—or compelled production of—copies of records from service providers.[25] (He also notes, somewhat cryptically and no doubt constrained by classification, "the important (and sensitive) tools that the IC [Intelligence Community], beyond just the FBI, brings to the effort to attribute . . . cyber activity.") Carlin points to several legal instruments governing the domestic use of these tools by the law enforcement community, including:

- The Stored Communications Act (SCA), which sets out the procedures for law enforcement agencies to obtain voluntary or compelled disclosure of stored communications from domestic communications-service providers, e.g., whether a search warrant or a subpoena is necessary in a given instance to compel disclosure of the information sought.

- The Foreign Intelligence Surveillance Act of 1978, which allows electronic surveillance conducted in the United States for national security or foreign intelligence investigations. In such instances, the target of surveillance must be a foreign power or an agent of a foreign power; the facilities or places at which the electronic surveillance is directed must be used, or must be about to be used, by a foreign power or an agent of a foreign power; and a significant purpose of the surveillance must be to obtain foreign intelligence information.

- Search warrants (or, in the case of national security and foreign intelligence investigations, FISA orders) for the search and seizure of physical devices—e.g., phones, computers, or servers.

Outside the United States, activities of the intelligence community (as opposed to the law enforcement community[26]) are governed by Executive Order 12333,[27] which is intended to "provide for the effective conduct of United States intelligence activities and the protection of constitutional rights." Because constitutional rights do not attach at all to foreigners unless they are within the United States, intelligence collection activities directed against foreigners are largely unconstrained by US law and policy except to the extent that US persons[28] may be involved.[29] (When US persons are involved, the executive order and other laws—notably the Foreign Intelligence Surveillance Act—do place some constraints on US intelligence agencies.) International law has traditionally placed no constraints on intelligence collection activities (aka espionage),[30] though such activities against foreigners abroad may violate the domestic laws of other nations.

US law enforcement agencies also operate outside the United States in cooperation with their counterparts abroad by "exchanging information, investigating attacks or crimes, preventing or stopping harmful conduct, providing evidence, and even arranging for the rendition of individuals from a foreign state to the United States."[31] Sometimes such cooperation is governed by treaty, e.g., extradition treaties or mutual legal assistance treaties (MLATs) that generally apply to a list of agreed crimes. MLATs also require "state parties to assist one another by providing information, evidence, and other forms of cooperation when requested to do so in such situations."[32]

The Budapest Convention, also known as the Council of Europe Convention on Cybercrime, is an international agreement that seeks to harmonize national laws explicating offenses that constitute cybercrimes, to improve national capabilities for investigating such crimes, and to increase international cooperation among the signatories on investigations.[33] The Convention's provisions on cooperation are a rough substitute for pairs of signatory nations that do not have an MLAT in place, but existing MLATs between other pairs of nations supersede the Convention's provisions. Increased international cooperation on investigations may well increase the amount and quality of useful information available for attribution judgments.

Nation-states as the ultimately responsible party?

As noted earlier, the consensus that exists for the presumed responsibility of states for the acts of their armed forces does not necessarily apply when a state is associated in some way with—or somehow connected to—malicious cyber activity.

Identifying a particular nation-state as the party ultimately responsible for a cyber intrusion hinges on what it means to be "responsible." A variety of different forms of state responsibility can be imagined. The following hierarchy of national involvement as it corresponds to responsibility closely follows Jason Healey's taxonomy in "Beyond Attribution: Seeking National Responsibility for Cyber Attacks."[34]

- A state could *prohibit* hacking activities (defined here as conducting cyber intrusions of various kinds), but have no ability to enforce this prohibition against third-party actors.

- A state could *tolerate* hacking activities. States could decide not to outlaw these actions, or not to prosecute those who launch attacks.

- A state could *encourage* hacking activities. In this scenario, a state may provide under-the-table support (intelligence, operational guidance, or "suggestions"), or simply promote a culture whereby these actions are lauded.

- A state could *direct* hacking activities. For example, a state could ask organizations within its jurisdictional reach or contract with non-state organizations to conduct specific hacking activities.

- A state could *conduct* hacking activities. A state uses its military or intelligence assets to conduct offensive cyber operations, perhaps integrated with third-party hackers.

A refinement on the above list is that these different types of responsibility might vary by the specific kind of hacking activity involved. For example, a state might conduct cyber-enabled espionage but prohibit destructive cyberattacks.[35]

A second related dimension along which to characterize state responsibility is the actor conducting any of the hacking activities described above. Responsibility could in principle also attach to hacking activities initiated by parties within the state's geographic borders and/or by parties who owe some form of allegiance or loyalty to the state (e.g., citizens of that state).

With respect to the case of responsibility attaching to activities initiated by parties within the state's geographic borders, a body of international law related to terrorism may be relevant.[36] Prior to the September 11, 2001, attacks on the United States, a nation-state was responsible for the acts of private groups inside its territory over which it exercised "effective control."[37] In the aftermath of those attacks, the United States took the position that the mere harboring of these actors, even in the absence of control over them, suffices to make the state where the terrorists are located responsible for their actions.[38] Many parts of the international community, including the UN Security Council, concurred with this position.[39] How and to what extent, if any, such a law applies to subnational or transnational groups perpetrating acts of cyber intrusion is uncertain, but the law as it relates to its original context of terrorism is at least suggestive.[40]

To the best of this author's knowledge, there is no body of international law that holds a nation accountable for the actions of its citizens per se. On the other hand, various nations can and do assert jurisdiction over their own citizens in many instances even when these citizens are abroad; in such cases, a citizen of Nation A is subject to the domestic law of Nation A even if he or she is located in Nation B. Moreover, various Nation Bs have from time to time sought, using diplomatic and other means, to influence or persuade Nation A to exert more control or influence over A's citizens when A's citizens are responsible for harm to B.

This paper does not seek to resolve the "proper" definition for state responsibility, but three observations are pertinent.

- Technology has very little to say about the proper definition for state responsibility. No amount of technical forensic information will point to the proper definition.

- For all practical purposes, the definition that a nation-state will adopt in any given instance will almost certainly depend on the facts and circumstances of that instance. It may be that, over time, an international consensus or norm may develop for the level in the Healey hierarchy that corresponds to the minimum level of involvement needed to declare that a state is "responsible." But we are not there yet.

- Multiple parties could be responsible depending on how norms for assigning responsibility evolve. For example, if citizenship and the geographic location from which an intrusion was initiated both become important norms in determining a responsible state party, then perhaps China and Greece would both bear some responsibility for the intrusion.

Subnational entities as the ultimately responsible party?

As a general rule, nations are the subject of international law. However, from time to time, the UN Security Council has identified particular subnational entities engaged in international terrorism as threats to the maintenance of international peace and security. For example, UN Security Council resolution 1267 called out Osama bin Laden and others associated with him as terrorists who were being protected by the Taliban, and called upon member nations to deny permission for Taliban-operated aircraft to take off from or land in their territory and to freeze Taliban funds and other financial resources.[41]

Such actions suggest that under international law, subnational entities could at some point be recognized as the ultimately responsible party for serious cyber intrusions in a way that certain subnational entities are held responsible for terrorism. But there is no history that is directly on point regarding this matter.

Arguments have also been made that individuals could even be responsible under international law for cyber "war crimes." For example, Fidler has argued that the videos showing the killing of human beings by the Islamic State are themselves violations of international humanitarian law (IHL) and constitute war crimes.[42] Under the Rome Statute (which establishes the International Criminal Court and gives it jurisdiction over individuals charged with war crimes),[43] Fidler argues that "those making and posting the Islamic State's videos are criminally accountable" under IHL.

How Attribution Judgments Are Made

In a 2014 paper on attribution,[44] Rid and Buchanan argue that thinking about attribution is currently based on three assumptions, two of which are relevant to the discussion of this section: first, that attribution is a largely intractable problem because of the technical characteristics and the geography of the Internet (as described in Box 2 entitled "The Design of the Internet and the Difficulty of Attribution"), and second, that attribution is either possible or not possible in any given case of interest. The third assumption—that the main challenge in attribution is finding the evidence itself and not in interpreting or using it—is relevant to the section below on "The Relationship between Attribution and Action."

In short, the conventional wisdom holds that one cannot attribute a malicious cyber activity to its perpetrator with high confidence.[45] A saying in the technology community is that "electrons don't wear uniforms"—there's no inherent binding of any given IT

BOX 2: THE DESIGN OF THE INTERNET AND THE DIFFICULTY OF ATTRIBUTION

The difficulty of attribution is often held to be the result of the design of the Internet. For example, Clark and Landau note that "there have been calls for a stronger form of personal identification that can be observed in the network. A non-technical version of this view was put forward as: 'Why don't packets have license plates?' which they describe as 'the attribution problem.' "[46] Hunker et al. assert, "The Internet's architecture and its evolving administrative and governance systems make the attribution of cyber attacks extremely challenging. . . . The Internet has no standard provisions for tracking or tracing. A sophisticated user can modify information in IP packets and, in particular, forge the source addresses of packets (which is very simple for one-way communication). Attackers often employ a series of stepping stones where compromised intermediate hosts are used to launder malicious packets. Packets can also be changed at hops between hosts; thus, attempting a traceback by correlating similar packets is ineffective when sophisticated attackers are involved."[47]

These assertions about the Internet's design are entirely true. But to the extent that they are even relevant to the threat environment of today, they relate primarily to the technical forensics dimension of attribution. Also, it should be noted that many kinds of cyberattack were propagated even before the Internet existed; pre-Internet vectors for cyberattack included human beings exchanging floppy disks and computers using modems to connect to dial-up bulletin boards; both floppy disks and bulletin boards could be (and were) contaminated with malware of various kinds from time to time. Analysts trying to find the origin of a given instance of malware still faced the problem that malware did not generally carry the signatures of individuals. Intrusions can also originate in a supply chain compromise, in which a security vulnerability can be introduced into a product or service at any point from initial design and manufacture to delivery or use at the customer's door.

Clark and Landau make a second, related point as well: the attribution discussion of this box refers to packet-level (or, equivalently, network-level) attribution—that is, association of sender identity with the content carried on the network in packet form. It is silent on application-level attribution (e.g., between a bank and its customers), which is discussed above ("Attributing malicious cyber activity to a human intruder";) and can be carried out regardless of whether packet-level authentication is in place.

activity to specific actors. Anyone could be at the computer in Greece that launched the attack against the DOD computer in San Francisco, evidence could have been planted to mislead investigators, and the perpetrator could even have been a computer program, set by someone to run autonomously.

The conventional wisdom has a grain of truth to it—technical forensics alone cannot lead to high-confidence attribution.[48] Caloyannides goes so far as to assert that “forensics’ presumed usefulness against anyone with computer savvy is minimal because such persons can readily defeat forensics techniques. Because computer forensics can’t show who put the data where forensics found it, it can be evidence of nothing.”[49]

At the same time, that grain of truth does not come close to being the full story of how attribution judgments can be—and are—made. One important point to consider is that while an intruder may have many counter-forensics measures at his or her disposal, he or she may not take all of the necessary measures; we return to this point below. Most importantly, only when the goal is attribution—to a machine—are technical forensics the primary source of evidence.

In trying to attribute an intrusion to a human perpetrator or an ultimately responsible party, technical forensics by themselves are generally inconclusive and the information they provide must often be combined with other sources to be genuinely useful.

For example, a given intrusion may be similar or even identical to a previous intrusion—the same code could be executed, the same IP addresses used, the same technical signatures found. Such similarity would suggest that the same party could be behind the intrusion at hand.[50] If that party had been previously identified, that identification might be carried over to the present case—or perhaps allies or associates of that other party might be implicated. Is such similarity conclusive or dispositive? Absolutely not. But neither should the clue it provides be thrown away.

Behavioral information can also contribute to attribution judgments. For example, Carlin notes that useful clues may be found in the kinds of malware that intruders use and in the way they communicate with their victims.[51] Behavioral patterns have been used in criminal investigations for a wide variety of offenses, and many of the analytical techniques used to understand these patterns have proven useful in attribution.

In the case of the 2014 Sony hack, the perpetrators left a “splash screen” on infected Sony computers with the name “Guardians of Peace” and various logos. Carlin points out that the perpetrators behaved in ways that were similar to the behavior of criminals like serial killers who “stage” the crime scene, arranging it to send a message or conceal involvement. Such staging goes beyond what is necessary to commit the crime, and they thus provide extra information that can be helpful in attribution.

An intruder can also make errors of tradecraft. For example, text stings can sometimes be extracted from the binaries used in an intrusion. When an investigator examines the binary used in the intrusion on the DOD computer in San Francisco, she finds the text string “Linsong9862.” An Internet search reveals that this string is also the user name associated with a dating profile of a Chinese computer scientist who says he lives in Greece. Another indicator may be the time of day that certain malicious cyber incidents occur—a time, possibly, that correlates with working hours in Greece. In neither case is such evidence conclusive, but that evidence constitutes additional data points that may point to the human intruder.

Sometimes intruders make mistakes of operational security. For example, an intruder may discuss his or her plans on insecure channels that are monitored. A hacker may look to others for advice, or seek recognition for his or her bravado and skill in perpetrating a successful intrusion, or upload or download files to or from known, previously used locations. Because intelligence agencies collect information from a variety of different sources in different parts of the world, sometimes such information is available; if so, such information could prove useful in identifying the human intruder.

The style and methodology of an intrusion may be helpful. For example, a cyberattack aimed at destroying or disrupting cyber physical systems that are part of a nation’s physical infrastructure is likely to require significantly more expertise than one directed at deleting files on computer systems; while both require expertise in penetration techniques, only the former requires expertise regarding the specific cyber physical systems involved. One reason the Stuxnet attack on Iran’s nuclear program was attributed to state actors was the sophistication of the attack in precisely targeting particular configurations of Siemens controllers (and leaving others alone), in concealing from centrifuge operators what was happening to the targeted centrifuges, and in the profligate use of zero-day vulnerabilities, which are usually regarded as a resource to be conserved and used sparingly.[52]

Other intelligence and information-gathering activities may also provide information useful for attribution. According to the CIA,[53] human intelligence (HUMINT)—information that can be gathered from human sources—is collected through “clandestine acquisition of photography, documents, and other material, overt collection by people overseas, debriefing of foreign nationals and U.S. citizens who travel abroad, and official contacts with foreign governments.” For example, a spy in the office of a senior political leader in another nation could provide information that the intrusion was ordered by that nation's leadership—such information could well be conclusive when coupled with technical forensics. Information about adversary plans and capabilities for cyber operations may be found in a dumpster and used later to investigate an intrusion.

HUMINT is not necessarily clandestine. As suggested in the section above on legal authorities, informal conversations or formal interviews with operators, service providers, and other users can also generate useful information. Debriefing a US citizen who had conversations with foreign network operators on a recent trip abroad can provide useful tips. Interviews with victims of cyber intrusions can provide valuable context for an intrusion, as investigators might learn more about why the intruders wanted to do what they did when they did it. For example, investigators might learn of demands that the intruder made of the victim in connection with the intrusion. Sharing information about similar intrusions might be useful as well; one victim might have one part of the information necessary to attribute an intrusion and a second victim might have another part.

Pre-positioned implants for cyber-enabled intelligence collection may provide useful information regarding the connection between the intrusion and agencies of the nation's government—for example, these implants may have revealed communications regarding an intrusion between decision-makers in that government's military department. Such implants were mentioned in the above section on attributing activity to a machine.

Geopolitical circumstances could provide clues as to who would want to launch a particular intrusion. What nation would most benefit from gaining access to the DOD computer in San Francisco? Are there particular tensions between a company and a state, or between the United States and another international actor? Is another international actor making demands of the United States, demands that are serious enough to warrant the use of force or cyber force? Who would benefit most from this intrusion? This information could provide a helpful lens for determining who would be most motivated to launch a certain attack.

Finally, historical relationships help to frame the attribution process. It is less likely that a non-adversarial nation would conduct, support, or tolerate malicious cyber activity against the United States as compared to an adversarial nation.

None of these methods or sources of evidence alone can be used to determine the responsible party. However, together, these pieces of data could pull together into a compelling analysis. A useful analogy is that of big data analytics, in which no individual datum is by itself significant, but instead large volumes of data are analyzed to draw conclusions.

BOX 3: ALL-SOURCE ANALYSIS AND THE SINKING OF THE *CHEONAN*

Outside of cyberspace, consider a radar-based surveillance and reconnaissance system involving several different independent radars, each of which detects a target in the same location and at the same time but with low confidence. Even though each individual sighting has a low probability that a target is actually present, the likelihood that *all* of them are incorrect—if the radars truly operate independently of each other—is very low.

In the language of this paper, each sighting is merely a suggestive clue. Aggregating these clues provides higher confidence that a collective sighting is correct. Although radar target sightings are usually brought to the attention of system operators with a probability of detection associated with them, the same principle applies to any process of evaluating independent threads of evidence.

As a real-world example of combining technical forensics with other information in a non-cyber domain, consider the investigation of the sinking of the *Cheonan,* a South Korean corvette, on March 26, 2010. Drawing on experts from South Korea, the United States, Australia, the United Kingdom, and Sweden, one international report on this incident noted the collection of “propulsion parts [from a torpedo], including propulsion motor with propellers and a steering section from the site of the sinking” and noted that “the evidence matched in size and shape with the specifications on the drawing presented in introductory materials provided to foreign countries by North Korea for export purposes. The marking in Hangul, which reads ‘1번 (or No. 1 in English),’ found inside the end of the propulsion section, is consistent with the marking of a previously obtained North Korean torpedo. The above evidence allowed the JIG [the investigators] to confirm that the recovered parts were made in North Korea.”[54] On this basis, the investigators concluded that the *Cheonan* was sunk by a torpedo made in North Korea.

The report further noted that the North Korean military had a variety of submarines and torpedoes capable of causing the same level of damage suffered by the *Cheonan,* and that a few small submarines and a mother ship supporting them left a North Korean naval base in the West Sea two to three days prior to the attack and returned to port two to three days after the attack. Finally, it noted that all submarines from neighboring countries were either in or near their respective home bases at the time of the incident. They thus concluded that the torpedo was fired by a North Korean submarine.

In short, attribution is an all-source issue—no one method or source of information can be used to point fingers, but multiple sources taken as a whole may paint a convincing picture. Box 3 entitled “All-Source Analysis and the Sinking of the Cheonan” illustrates how the all-source intelligence process can be applied to attributing putatively anonymous non-cyber incidents.

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. That is, more investigation may reveal additional useful information, which may (or may not) reinforce attribution judgments made earlier.

One important reason for the improvement in capabilities for attribution over the past several years is that as the importance of cybersecurity has grown, more people are paying attention. Given the likelihood of malicious cyber activity in the future, they are more willing to make investments in intelligence and to build investigative capacity that will pay off in the future. Put differently, capabilities for attribution are partly a function of the investment a nation (or, indeed, third parties, such as private cybersecurity companies) is willing to make in those capabilities, both in infrastructure and in the effort that any given case demands.[55]

Lastly, it is important to understand that the all-source intelligence process described in this section has a different focus than the discussion of the ultimate responsibility of states and non-state actors in the sections above on nation-states and subnational entities as ultimately responsible parties. The all-source intelligence process seeks to approximate the God's-eye understanding of an intrusion, whereas the discussions of those sections are legal and policy discussions. In short, understanding who did what (the focus of the intelligence process) is different, though relevant to, who is to blame.

Evolving US Government Views on Attribution

US government views of attribution have evolved over the past half-dozen years.

In 2010, then-deputy secretary of defense William Lynn emphasized the difficulties of attribution in cyberspace.[56] He said that "whereas a missile comes with a return address, a computer virus generally does not. The forensic work necessary to identify an attacker may take months, if identification is possible at all."

In 2012, then-secretary of defense Leon Panetta said that the DOD "has made significant advances in solving a problem that makes deterring cyber adversaries more complex: the difficulty of identifying the origins of an attack. Over the last two years, DOD has made significant investments in forensics to address this problem of attribution and we're seeing the returns on that investment. Potential aggressors should be aware that the United States has the capacity to locate them and to hold them accountable for their actions that may try to harm America."[57]

In 2015, the DOD Cyber Strategy stated, "Attribution is a fundamental part of an effective cyber deterrence strategy as anonymity enables malicious cyber activity by state and non-state groups. On matters of intelligence, attribution, and warning, DOD and the intelligence community have invested significantly in all source collection, analysis, and dissemination capabilities, all of which reduce the anonymity of state and non-state actor activity in cyberspace. Intelligence and attribution capabilities help to unmask an actor's cyber persona, identify the attack's point of origin, and determine tactics, techniques, and procedures. Attribution enables the Defense Department or other agencies to conduct response and denial operations against an incoming cyberattack." The 2015 articulation is thus more measured and moderate in tone than the Panetta comments of 2012.

Also in 2015, Director of National Intelligence James Clapper testified, "Although cyber operators can infiltrate or disrupt targeted ICT [information and communications technology] networks, most can no longer assume that their activities will remain undetected. Nor can they assume that if detected, they will be able to conceal their identities. Governmental and private-sector security professionals have made significant advances in detecting and attributing cyber intrusions."[58] He testified in 2016 that "Information security professionals will continue to make progress in attributing cyber operations and tying events to previously identified infrastructure or tools that might enable rapid attribution in some cases. However, improving offensive tradecraft, the use of proxies, and the creation of cover organizations will hinder timely, high-confidence attribution of responsibility for state-sponsored cyber operations."[59]

One significant development in the attribution landscape in the past several years is the increasing involvement by private-sector firms in rendering attribution judgments. Regarding the value of private-sector attribution, the DOD cyber strategy of 2015 notes that private-sector parties (e.g., security firms) reporting on attribution "can play a significant role in dissuading cyber actors from conducting attacks in the first place" and states that "The Defense Department will continue to collaborate closely with the private sector and other agencies of the U.S. government to strengthen attribution. This work will be especially important for deterrence as activist groups, criminal organizations, and other actors acquire advanced cyber capabilities over time."[60]

In addition to the Mandiant APT1 report described above, examples of private-sector involvement in attribution include:[61]

- FireEye's report, "APT28: A Window Into Russia's Cyber Espionage Operations,"[62] indicating Russian involvement in a variety of espionage activities against private-sector and government actors.

- Novetta's report, "Operation SNM: Axiom Threat Actor Group Report,"[63] indicating Chinese government involvement in cyber espionage against a variety of private companies, governments, journalists, and pro-democracy groups.

- CrowdStrike's report, "CrowdStrike Intelligence Report: Putter Panda,"[64] identifying Unit 61486 in the Chinese PLA as being responsible for the cyber-enabled theft of corporate trade secrets primarily relating to the satellite, aerospace, and communication industries.

Private-sector involvement in attribution has advantages and disadvantages.[65] Among the advantages are:

- The unclassified nature of such reports. Because such are unclassified in their entirety, they can be used by government officials in responding to questions about the attribution of any given cyber incident. They also make available to independent analysts substantial information that would not otherwise be available and thus contribute to a more informed public debate about such matters.

- The potential increase in analytical and collection resources that can be brought to bear on tracing the origin of hostile cyber operations. Additional resources will be necessary as the volume of hostile cyber operations conducted by parties with advanced cyber capabilities increases.

- Continuing concealment of sensitive sources and methods of government intelligence, which are not revealed in private-sector attribution reports.

- The attenuation of government responsibility for an attribution judgment. When the actual judgment is associated with a private party, government officials can distance themselves from it, even if they point unofficially to that analysis when questioned about a given incident. The resulting ambiguity may have diplomatic benefits.

Some of the disadvantages include the following:

- The marketing aspect of private-sector attribution reports. Such reports often gain considerable media attention, especially if government officials have not been particularly forthcoming about cyber incidents. These reports are thus valuable marketing tools that elevate the authoring firms in the public eye, and the incentives motivating these firms to produce such reports quickly and ahead of their competitors may degrade the quality of their research and analysis.

- Lack of independent quality control and independent oversight. Authoritative government reports are usually subject to an interagency process that challenges evidence and conclusions. The private-sector security market is robust enough to provide some independent scrutiny, and since each firm has its professional reputation to uphold, all firms have incentives to produce high-quality work. Whether market forces are sufficient to uphold quality in such reports remains to be seen.[66]

- The possible lack of true independence of private-sector reports. Given the semi-permeable membrane between private-sector security firms and government authorities, it would not be surprising if, from time to time, government officials talking to their colleagues in the private sector suggest that looking for X rather than Y in their investigative efforts could prove more fruitful. That is, such reports may be produced with some measure of government input, even if such input is not apparent.

Finally, nations other than the United States often do not appreciate fully the separation between the public and private sectors that operates in the United States. In particular, more authoritarian regimes that exert a high degree of control and influence over civil society may well regard private-sector entities as being willing to speak or act in accordance with US government wishes under many or most circumstances.

How Attribution Relates to Policy

The discussion up to this point has presumed that the attribution task is to determine as best as possible the machine, human intruder, and/or ultimately responsible parties that are behind a given malicious cyber incident. In this context, the word "determine" is relative to a God's-eye perspective—to determine the machine, intruder, and/or party that was/were actually involved in and responsible for undertaking the intrusion. As noted earlier, attribution to a machine or a perpetrator turns on factual issues, whereas attribution to an ultimately responsible party strongly depends on the legal, policy, and political definition of "ultimately responsible."

Determining factual reality—important as it is—is only the beginning of the attribution process from a policy perspective. Three key points need to be made.

- A "determination" is rarely definitive. God may know who "really" did it, but our determinations of who did it will be associated with some degree of uncertainty or confidence about it—and it is very hard to be 100 percent confident about a determination. The use of the word "judgment" underscores this point.

- The necessary degree of confidence in an attribution judgment depends on the nature of the malicious activity being attributed and the action that is contemplated in its aftermath.

- The audience that an attribution judgment seeks to persuade has a significant impact on how subsequent aspects of the attribution process unfold.

These points are fundamentally policy points rather than technical ones, and are at the heart of the political challenges of attribution.

Confidence in attribution

An attribution judgment is a statement with an inherent degree of uncertainty. To describe that uncertainty, different professions use different sets of words to convey such uncertainty.[67] For example, in the US legal community, the following words are used regarding the persuasiveness of evidence that a given person is in fact responsible for an event (e.g., "There is evidence that *John Doe robbed the bank yesterday,*" where the italicized words refer to the event in question.)

- Reasonable suspicion: There is reasonable suspicion that . . .
- Probable cause: The police officer had probable cause to believe that . . .
- Substantial evidence: There is substantial evidence that . . .
- Preponderance of the evidence: The preponderance of the evidence indicates that . . .
- Clear and convincing evidence: There is clear and convincing evidence that . . .
- Beyond reasonable doubt: The evidence indicates beyond a reasonable doubt that . . .

The audience in question for these statements is an impartial and unbiased judge or jury, and advocates for each side try to persuade this audience to draw some conclusion about the responsibility of the alleged perpetrator of some event that happened in the past. The relevant standard of evidence that the judge or jury applies depends on the nature of the case. If the event in question is a criminal matter, the judge or jury must be convinced beyond a reasonable doubt about the party responsible, whereas in a civil matter the judge or jury need only be convinced by a preponderance of the evidence.

The legal process of ascertaining responsibility is also intended to be fair. Due process requirements seek to ensure that state action occurs only in accordance with law and that justice is administered fairly, i.e., that prejudicial or unequal treatment does not occur.[68] Due process also protects the rights of an accused party, e.g., by excluding improperly gathered evidence from a trial.

In short, if a malicious cyber incident is regarded as a matter for domestic law enforcement authorities to address, then legal requirements for process, standards of evidence, and degrees of certainty about attribution obtain. But outside this context, there is much less clarity.

Consider, for example, the attribution issue from the standpoint of international law. International law operates in an environment of sovereign nations. Nations sometimes have interests in using international bodies such as the International Court of Justice or the United Nations to adjudicate their political and diplomatic positions with respect to other nations, and thus they grant these bodies jurisdiction in certain contexts. But few if any of these nations are willing to subordinate important national interests to the judgments of such bodies. Moreover, unlike domestic courts that are backed by police forces, these bodies generally lack the enforcement authorities associated with the use of force. (It is true in principle that the UN Security Council may authorize the use of force to enforce a judgment, but it is exceedingly rare in practice and any one of the Permanent Five can veto a resolution containing such authorization.)

An important legal lacuna in the ability of an international tribunal to make attribution judgments is underscored by Tsagourias,[69] who argues that the nations that may be involved may for security reasons be unwilling to make relevant information available or may make it available only in truncated or abstracted form. For example, in the 1986 Nicaragua case, the International Court of Justice noted:[70]

> One of the Court's chief difficulties in the present case has been the determination of the facts relevant to the dispute. First of all, there is marked disagreement between the Parties not only on the interpretation of the facts, but even on the existence or nature of at least some of them. . . . Thirdly, *there is the secrecy in which some of the conduct attributed to one or other of the Parties has been carried on. This makes it more difficult for the Court not only to decide on the imputability of the facts, but also to establish what are the facts* (emphasis added). Sometimes there is no question . . . that an act was done, but there are conflicting reports, or a lack of evidence, as to who did it. *The problem is then . . . the prior process of tracing material proof of the identity of the perpetrator.*

Tsagourias also argues that "International law does not lay down any specific standards of evidence with regard to issues involving the use of force or self-defence," citing the separate opinion of a judge in the Oil Platforms case of the ICJ.[71] He suggests (but does not defend) a generic threshold that "claims against a State involving charges of exceptional gravity must be proved by evidence that is fully conclusive. The same standard applies to the proof of attribution for such acts." He notes that this standard is less strict than "beyond a reasonable doubt" but is higher than the "balance of evidence."

Tsagourias's overall conclusion: "standards concerning the availability and probity of evidence in cases involving armed attacks, uses of force or interventions are rather lax." Nevertheless, he argues, "even if the standard of proof is not the same as the one required for the criminal prosecution of individuals and even if 'a more political approach to attribution . . . might accept less exacting standards,' it should be stressed that a State should not resort to self-defence on the basis of casual evidence or wild political inferences."

No national policymakers would agree that any action of theirs, let alone actions related to self-defense, can or should ever be justified "on the basis of casual evidence or wild political inferences." Nevertheless, if the malicious cyber incident in question is regarded as a national security matter, determining the necessary degree of certainty is more complex. When national security is at stake, policymakers may have to make decisions that have a wide range of potentially significant and nation-transforming consequences. But unlike the unbiased judge or jury that is the linchpin of decision-making in the legal community, national security policymakers are highly biased in the sense that they are predisposed to make decisions that they believe best protect and advance national interests. Nor does national security decision-making recognize good analogs to "rights of the accused" or "due process." To take one obvious example, information is not excluded from consideration if it has been gathered "improperly."

To support national security decision-making, the intelligence community provides information, often in the form of assessments. For example, the National Intelligence Estimate for Iran's nuclear intentions and capabilities stated:[72]

> We judge **with high confidence** that in fall 2003, Tehran halted its nuclear weapons program; we also assess **with moderate-to-high confidence** that Tehran at a minimum is keeping open the option to develop nuclear weapons. . . . We assess **with moderate confidence** Tehran had not restarted its nuclear weapons program as of mid-2007, but we do not know whether it currently intends to develop nuclear weapons. . . . We continue to assess **with low confidence** that Iran probably has imported at least some weapons-usable fissile material, but still judge **with moderate-to-high confidence** it has not obtained enough for a nuclear weapon. **We cannot rule out** that Iran has acquired from abroad—or will acquire in the future—a nuclear weapon or enough fissile material for a weapon (emphasis added).

The words in bold above are words of estimative probability that are intended to convey the degree of uncertainty (or, conversely, the degree of confidence) in various assessments and judgments made by analysts.[73] Assessment guidelines call for ascribing high, moderate, or low levels of confidence to assessment as follows:[74]

- “High confidence generally indicates that our judgments are based on high-quality information, and/or that the nature of the issue makes it possible to render a solid judgment. A ‘high confidence’ judgment is not a fact or a certainty, however, and such judgments still carry a risk of being wrong.

- “Moderate confidence generally means that the information is credibly sourced and plausible but not of sufficient quality or corroborated sufficiently to warrant a higher level of confidence.

- “Low confidence generally means that the information’s credibility and/or plausibility is questionable, or that the information is too fragmented or poorly corroborated to make solid analytic inferences, or that we have significant concerns or problems with the sources.”

This background on how the intelligence community operates is important because it frames how the policymaker approaches attribution judgments in a national security context. Given that national security decisions are a matter of sovereignty (i.e., there is no world government body that serves the role of impartial judge or jury, and there are no due process requirements on national decision-making imposed by international law), the standard that governs national security decision-making is not controlled by legal terms such as “beyond a reasonable doubt” or “preponderance of the evidence” but is rather one of reasonableness—taking everything that is known into account, is the decision a reasonable one?

Policymakers are also quite often in the position of having to take a responsive action, even when only low or moderate confidence assessments are available. And a further complicating factor is that the degree of confidence required to take any given action depends on the nature of the action—and the putative actor—involved. This point is discussed further in the section below on “The Relationship between Attribution and Action.”

The persuasiveness of attribution judgments

Based on intelligence information and shaped by their own biases and judgments about what is best for the national interest, policymakers need to satisfy themselves about attribution. But it is a different—and often more difficult—task to persuade others who may be skeptical about official US positions.

One major reason for such difficulty is that much of the public believes that legal standards of evidence are applicable for national security decision-making. These individuals thus conclude that because publicly offered evidence (which in practice cannot include *all* sources of information) would not “stand up in a court of law,” the US government does not have a legitimate basis for acting. For example, in the wake of the Sony hack in December 2014, public critics of the US government, which had attributed the hack to North Korea,[75] asserted that the evidence presented in favor of the attribution to North Korea was weak and that the available evidence pointed instead to a disgruntled insider at Sony.[76] In a telling commentary, one security expert said:[77]

> [C]alling out a foreign nation over a cybercrime of this magnitude should never have been undertaken on such weak evidence. The evidence used to attribute a nation state in such a case should be solid enough that it would be both admissible and effective in a court of law. As it stands, I do not believe we are anywhere close to meeting that standard.

This stance is somewhat ironic, given that even international courts have ruled that the standards for evidence in disputes between nations may not be as stringent as disputes aired in domestic courts. For example, in the 1949 *Corfu Channel* case, even an international court—the International Court of Justice—recognized the difficulties in providing evidence if that evidence had to be obtained from territory under the control of another state that was unwilling to cooperate. The court wrote:[78]

> By reason of this exclusive control, the other State, the victim of a breach of international law, is often unable to furnish direct proof of facts giving rise to responsibility. Such a State should be allowed a more liberal recourse to inferences of fact and circumstantial evidence. This indirect evidence . . . must be regarded as of special weight when it is based on a series of facts linked together and leading logically to a single conclusion.

Other nations are also a potential audience for an attribution judgment. In the wake of a malicious cyber incident, a state may want to persuade allies and unaligned nations that it has been wronged. To do so, the victimized state will not follow legally prescribed procedures, but instead will use tools of diplomacy and persuasion to convince necessary actors that a particular event occurred. Individual states may require different levels of evidence before siding with the supposed victim state.

In this context, it is worth recalling that during the Cuban missile crisis in 1963, President Kennedy asked former secretary of state Dean Acheson to seek French support for the US position. He traveled to Paris and offered to show French President Charles de Gaulle the CIA's surveillance photos of the Cuban missiles. According to Theodore Sorenson, then counselor to Kennedy, de Gaulle declined to view the photographs, saying, "The word of the president of the United States is good enough for me."[79] Today, in the wake of the Edward Snowden disclosures and a history of public failure such as US government claims of "yellow rain" in Southeast Asia and weapons of mass destruction in Iraq, a similar scenario of trust, either between the US government and other nations—even friendly nations—or even between the US government and its citizens, seems unlikely in the future under most circumstances. Yet, diplomatic dealings often necessitate a different interpretation of trust and evidence.

Against this backdrop, it is fair to say that whether or not the public does mistakenly apply domestic legal standards to the national security decision-making process (it does, but it should not), skepticism about attribution judgments increases pressures on policymakers to make public more evidence for attribution judgments than they might otherwise prefer. Jack Goldsmith said it well on *Lawfare:*[80]

> [E]ven if the attribution problem is solved in the basement of Ft. Meade and in other dark places in the government, that does not mean the attribution problem is solved as far as public justification—and defense of legality—is concerned.

Policymakers are not legally constrained in their freedom of action by such considerations, but politically they may very well be—and in the long run, they will almost certainly have to reveal some amount of hitherto secret information relating sources and methods for gathering evidence used in attribution judgments. Goldsmith notes further that we will almost certainly see in the future an increase "in the demand for publicly verifiable attribution before countermeasures (or other responses) are deemed legitimate. In this small but significant sense, the United States has lost a battle in the early days of cyber conflict."[81] Similarly, Paul Rosenzweig argued, "In the post-Watergate post-Snowden world, the USG can no longer simply say 'trust us.' Not with the U.S. public and not with other countries. Though the skepticism may not be warranted, it is real."[82]

In this context, it is not without irony that private-sector entities such as Google and Facebook are also sensitive to the need to protect sources and methods of information used to attribute compromises of user accounts to nation-states.[83] These entities warn users if they believe a nation-state compromise has occurred, but also do not provide the evidence underlying such a judgment. For example, Google tells compromised users, "You might ask how we know this activity is state-sponsored [but] we can't go into the details without giving away information that would be helpful to these bad actors."[84] Facebook tells compromised users, "To protect the integrity of our methods and processes, we often won't be able to explain how we attribute certain attacks to suspected attackers. That said, we plan to use this warning only in situations where the evidence strongly supports our conclusion."[85]

Lastly, it is highly unlikely that any amount of evidence made public would persuade a nation to publicly acknowledge its own responsibility for an untoward event, cyber or otherwise, if such an acknowledgment would not be in its interests. Demands for such public acknowledgment are common,[86] but are unrealistic and are not a matter of "sufficient evidence" in any case. These demands are again rooted in an expectation derived from a legal system in which an impartial court standing in judgment of an individual can require such acknowledgment from a party found responsible for some misdeed. (Box 4 entitled "A Possible Attribution "Court" describes proposals for such a court.)

Note that even if an adversary has openly claimed responsibility for an incident, decision-makers would still have to ascertain the scope and nature of that claim—and intelligence analysts would go through exactly the same process of gathering and sifting evidence to arrive at a judgment with low, medium, or high confidence.[87] This point is addressed further below.

The Relationship between Attribution and Action

Attribution is a key element of taking responsive action, but attribution and responsive action are not independent variables. Indeed, and as noted earlier, even the type of attribution at issue in any given instance—that is, whether attribution should be to a specific machine, to a specific human perpetrator, or to a specific adversary—depends on the goal of the relevant decision-maker.

The section titled "What Does Attribution Mean" began with a specific scenario. If the goal of the decision-maker faced with that scenario is action to stop or mitigate the pain being caused by the intrusion as soon as possible, then what is most relevant is **machine** attribution—to find the machine causing the pain as quickly as possible and to take action against it. If Tony—the operator of the targeted computer—discovers that files are being deleted from his computer mid-attack, his immediate concern may be to simply stop this from happening further. In this moment, he may not care that Karen—the owner of the attacking computer in Arkansas—is not truly responsible for initiating the attack. Instead,

BOX 4: A POSSIBLE ATTRIBUTION “COURT”

At least two noteworthy proposals for an attribution court have surfaced in the past few years. In 2012, the Atlantic Council proposed the establishment of a Multilateral Cyber Adjudication and Attribution Council (MCAAC) that would “provide an international mechanism for arriving at a consensus attribution of illegal cyber campaigns by states and a formal process for adjudicating associated interstate disputes.” In June 2016, Microsoft advanced a similar proposal for an international non-governmental body that could weigh in credibly on attribution judgments for cyberattacks that exceeded a certain threshold of consequence.

Both proposals emphasize the importance of strong technical competence and multilateral participation. (Microsoft suggests that all of the permanent members of the United Nations Security Council should be represented, while the Atlantic Council argues for states with “higher cyber attribution and forensics capacities,” and then identifies all of the permanent members of the United Nations Security Council as examples of such states.) Both proposals also cite the International Atomic Energy Agency (IAEA) as precedent for an international non-governmental body that addresses disputes of a highly technical nature and note the value that the IAEA has had in verifying compliance with the Nuclear Non-Proliferation Treaty.

Whether nation-states themselves would be willing participants in such a body remains to be seen. Microsoft notes that governments may be reluctant to empower an independent body to make findings that may be both politically important and politically charged. (An even more sensitive issue would be granting such a body any enforcement powers.) The Atlantic Council raises state (and private-sector) concerns about protecting intelligence sources and methods or indicators that could be used in making attribution judgments, and notes that without the capability to force the sharing of relevant attribution information, investigators may not be able to follow the chain of evidence in its totality. And, of course, there is a problem with scale—on what basis would this body accept cases for review, given the plethora of cyberattacks seen every day?

Microsoft believes that, nevertheless, if such a body were to achieve for attribution the kind of legitimacy that the IAEA has with respect to nuclear proliferation matters, it could address in part many of the difficulties in “the attribution problem” that today stem from the lack of a widely recognized internationally authoritative court to handle such matters. For its part, the council argues that one of the most valuable services that the MCAAC could provide is to rule on the extent and nature of state responsibility for actions undertaken by non-state actors operating from national territories.

Sources:

Jason Healey, John C. Mallery, Klara Tothova Jordan, and Nathaniel V. Youd, “Confidence-Building Measures in Cyberspace: A Multistakeholder Approach for Stability and Security,” Atlantic Council, Brent Scowcroft Center on International Security, Washington DC, 2012, http://www.atlanticcouncil.org/images/publications/Confidence-Building_Measures_in_Cyberspace.pdf

Scott Charney, Erin English, Aaron Kleiner, Nemanja Malisevic, Angela McKay, Jan Neutze, and Paul Nicholas, “From Articulation to Implementation: Enabling progress on cybersecurity norms,” Microsoft Corporation, June 2016, https://mscorpmedia.azureedge.net/mscorpmedia/2016/06/Microsoft-Cybersecurity-Norms_vFinal.pdf.

Tony simply is concerned that a computer in Arkansas is deleting files from his computer, and intends to disrupt further infiltration by said computer. The human perpetrator or the specific adversary ultimately responsible is not important.

If the goal of the decision-maker is action to prosecute someone for an attack that has occurred, then he will care about **human** attribution—to ascertain the identity of the human perpetrator as the first step in taking the person into custody. In this case, identifying George as the perpetrator is crucial: as the actor who set the attack in motion, he is the person who can be charged with committing an actual crime. Of course, the ability to prosecute someone depends on the relevant legal regime that governs his or her actions—and the ultimately responsible party may have some influence over the specifics of that legal regime. Note also that Tony is most likely not the one who will decide that prosecution is the appropriate path to take. Someone else, higher in the chain of command, will almost certainly make that decision.

If the goal of the decision-maker is action to deter malicious cyber activity in the future from being perpetrated against him, then he cares most about the **party** that is ultimately responsible for motivating and initiating the activity. Identification of the responsible party is a prerequisite for administering the punishment that is required to dissuade it from conducting similar actions in the future. Identification of the responsible party is also a prerequisite in convincing an adversary that not undertaking the action to be deterred will result in an outcome acceptable to him.[88] The human perpetrator is not the most relevant party in deterring future malicious activity, because anyone with sufficient technical skill can be hired, persuaded, or amused enough to press the right keys—that is, the individual person is likely to simply be one cog in the machine. Because the ultimately responsible party could easily act through other humans or machines in the future, only the ultimately responsible party can actually be meaningfully deterred from initiating and conducting further malicious activity. Moreover, a decision to pursue deterrence rather than prosecution will be made at an even higher level up the chain of command—very much removed from Tony, the person operating the computer that suffered the attack.

Regardless of the type of attribution involved, the confidence required of an attribution judgment depends on the nature and target of that action. For example, policymakers would usually require a higher degree of confidence if the action contemplated were a kinetically destructive action than if the action were a diplomatic démarche—in general and all else being equal, the more “severe” or “serious” the action, the higher the confidence in an attribution judgment would have to be. Under some circumstances, the response action may simply be a public announcement pointing the finger at an ultimately responsible party—public “naming and shaming” may be effective in deterring future action, especially if the ultimately responsible party conducted its actions believing it could do so anonymously.

Similarly—and, again, all else being equal—policymakers would usually require a higher degree of confidence if the putative actor involved were a powerful nation or one with whom the United States had a relationship with multiple important threads than if it were a relatively weak or relatively isolated nation.

The connection between attribution and action also has a temporal dimension. As noted above, attribution judgments are made on the basis of multiple sources of information, and integrating multiple sources of information takes time. Filtering through technical forensic details, comparing a given incident to previous incidents, extracting information obtained from human and signals intelligence sources, and so on are not easy tasks. Attributing a cyber incident may take weeks or months under some circumstances even when the analytical skills are available. Put differently, what is hard is *prompt* high-confidence attribution.

What is the significance of the difference between prompt and delayed attribution? For what purposes and under what circumstances is prompt attribution necessary (and by implication delayed attribution inadequate)? The answer depends on the nature of the response at issue for policymakers.

Consider first the tactical response to a malicious cyber incident. As noted above, machine attribution will be needed to mitigate the immediate harm being caused by the intrusion; the malicious operation of the machines involved in the intrusion must be blocked or disrupted. (Mitigation may well only be temporary if other machines are available to the adversary.) Choosing which courses of action would be most appropriate or wise is another matter.[89]

If a response is to arrest the perpetrator(s) or hold them criminally responsible for the incident, the conventions and rules of law enforcement hold sway. Because we hold individuals responsible for criminal acts, attribution to specific individual human beings is needed. Under these circumstances, rapid response may be desirable, but law enforcement authorities may work for years to identify, pursue, and take into custody individuals believed to be responsible for criminal acts.

If the response is to impose costs on a nation-state ultimately responsible for an intrusion, the conventions and rules of national decision-making are relevant, especially those of making such decisions in a security context. In the aftermath of a cyberattack, national security decision-makers may respond to punish or to retaliate for an adversary's attack. There are limits on such responses—retaliation or punishment for a hostile act once the act has stopped is prohibited under UN Charter Section 2(4) if it rises to the level of a use of force. Nevertheless, forceful actions are allowable under Article 51 of the UN Charter if they can be regarded as acts of self-defense in the face of an armed attack. Such actions are often justified as acts of self-defense that deter future attacks—and it is a matter of stated US policy that a sufficiently severe cyberattack would indeed qualify as an armed attack under the UN Charter.[90]

Note also that responses even to an armed attack may not entail the use of military force. As noted in the "International Strategy for Cyberspace,"[91] the United States reserves the right to use "all necessary means—diplomatic, informational, military, and economic—as appropriate and consistent with applicable international law, in order to defend our Nation, our allies, our partners, and our interests" in response to hostile acts in cyberspace.

Appropriate responses are a central element of deterrence, but what makes a response appropriate? US Strategic Command identifies three important factors for achieving deterrent effects; one is the US Strategic Command's "Deterrence Operations: Joint Operating Concept," and the other two factors are credibility of a threat to impose costs on a would-be adversary and costs that the adversary regards as too painful to incur.[92] (Credibility is equivalent to certainty—a more credible response is one that an adversary regards as more certain, and painful costs are equivalent to severity of response.) These two factors are also identified in the traditional deterrence literature in international relations.[93] If these conditions are met, an adversary faced with a credible threat to impose too-painful costs should the adversary act in a certain way will choose not to act in that way, i.e., will be deterred from that action.

By definition, an action that has already happened cannot be deterred. But future actions can be deterred, and an appropriate response to an action that has already happened serves to reinforce the credibility of a deterrent threat in the future. Thus, when faced with a decision about how to respond to a given hostile action, decision-makers must identify the party against which to respond (i.e., they must attribute the hostile action correctly) and then respond in a sufficiently painful way so that the adversary will be deterred from similar actions in the future.

Curiously, the temporal element is missing from this calculus. Traditional theories of deterrence in international relations as well as the US Strategic Command's construct for deterrence are silent on the impact on deterrence, if any, of the elapsed time between the hostile action and the response. It is intuitively plausible that long delays between hostile action and response will change the deterrent effect of a response, but whether this intuition is in fact true is not at all clear.[94] For example, consider that an attribution effort that requires many months may cover the transition from one political administration to another, and a second administration may well have different policy preferences, some of which might drive different responses with different costs. A "tougher" administration might choose to impose costs that are even more painful than a "softer" one, or vice versa.

Delays in attribution may implicate international law as well.[95] An extended period of time passing after an intrusion likely weakens the case for forceful responsive actions being regarded as legitimately acting in self-defense, since actions taken in self-defense are supposed to be only the minimum necessary to restore the status quo. A similar argument holds true for countermeasures, which are acts that would be forbidden under international law except for the fact that they are taken in response to a prior illegal act by another nation and are intended to induce the cessation of that illegal act. For a sufficiently extended period of time (imagine in the limit a decade or two), a forceful "response" would likely be regarded as a new (and illegal) use of force in its own right.

Perhaps of greatest significance are the political dimensions. In some cases, the speed of a response—such as publicly calling out an adversary—is important for geopolitical reasons, since other events in the world will continue to play out and silence regarding an important intrusion will have negative consequences. Under such circumstances, policymakers are likely to accept a higher degree of uncertainty in an attribution judgment than they would prefer, especially if history suggests that a suspected adversary would benefit from silence. An overt signal to that adversary (or perhaps to an influential ally) sent promptly could help to forestall those negative consequences.[96]

In other cases, policy makers may have an attribution of a malicious cyber incident in hand (indeed, perhaps a high-confidence attribution) and choose not to make it public. One obvious reason for not "going public" is the reality that a public attribution will generate demands for public evidence, a point discussed earlier. But another reason for not going public is that the relationships between many nations that act against each other in cyberspace are complex and multidimensional. "Going public" may result in demands to take retaliatory action that, in the view of senior policy makers, may be unwise given the range of interests at stake. The consequences of any retaliatory action—i.e., the significance of a possible adversary response—must be taken into account, and before policy makers decide to retaliate, they must be willing to face the consequences of any such action.

BOX 5: RISK COMMUNICATIONS WITH THE PUBLIC

Communicating to the public about technology-driven problems or issues is often done poorly. One reason is often that the knowledge of individual public-facing policymakers about the underlying technology is inadequate. For example, they may not know enough to answer questions posed by reporters or they may use inappropriate analogies that undermine public confidence in their capabilities to make good decisions. On the other hand, technical experts often have poor intuitions about and/or understanding of their audiences' knowledge and needs and don't know how to communicate effectively with the public.

Scientific approaches to such communications have been developed over the past forty years.[97] In general, these approaches call for developing and vetting a strategic approach to communication, a defensible risk/benefit analysis in advance of any controversy, and communication activities that are both audience-driven and interactive.

This process calls for:

- Identifying the information regarding context and scientific background that is most critical to members of the audience.
- Conducting empirical research to identify audience members' current beliefs, including the terms they use and their organizing mental models so as to craft messages that will reach the desired audiences.
- Designing messages that close the critical gaps between what people know and what they need to know, taking advantage of existing knowledge and the research base for communicating particular kinds of information (e.g., uncertainty).
- Evaluating those messages until the audience reaches acceptable levels of understanding.
- Developing in advance multiple channels of communication to the relevant audiences, including channels based on media contacts, opinion leaders, and Internet-based and more traditional social networks, and avoiding undue dependence on traditional media and public authorities for such communication.
- Ensuring that messages reach the intended audiences in a prompt and timely fashion. Controversies can emerge and grow on the time scale of a day, requiring responses on similar time scales.
- Persisting in such public engagements over long periods of time.

Source:
The contents of this box are loosely adapted from Jean-Lou Chameau, William F. Ballhaus, and Herbert S. Lin, eds., *Emerging and Readily Available Technologies and National Security: A Framework for Addressing Ethical, Legal, and Societal Issues,* National Academies Press, Washington, DC, 2014, pp 158–159.

Lastly, the effects of the intrusion may manifest themselves quickly and force political leaders to face public pressures to "do something" even in the face of incomplete information. If one accepts that active cyber defense is likely to be technically ineffective, pressures for rapid response are in the end political in nature. Under these circumstances, the consequence of this conclusion is unpleasant for political leaders—they must be prepared to resist public pressures until the necessary judgments are in hand and to communicate to the public their rationales for waiting (Box 5: Risk Communications with the Public).

Attribution from the Standpoint of the Adversary

Up to this point, this paper has focused on the victim's perspective in attribution. But it is also necessary to consider the adversary's perspective on attribution. For example, most discussions of attribution (including this one) assume that the adversary wishes to conceal its involvement in an intrusion. This assumption may not always be valid—an adversary (Nation A) may conduct an intrusion and *deliberately* engage in sloppy tradecraft to signal the victim (Nation B) that it has the ability to conduct such an intrusion. A may send such a signal to B in the hope that knowledge of A's capabilities would deter B from taking some action that would be undesirable to A.[98]

Assuming the adversary wishes to conceal its involvement in an intrusion, it is important to consider any given intrusion in a larger context. Specifically, any given intrusion may be only one in a set of intrusions,[99] and an adversary may well change its approach to later intrusions depending on the defending victim's actions in attempting to attribute and/or thwart earlier intrusions. That is, the adversary's techniques, tactics, and procedures may be adaptive to the defense's actions.

Thus, if the adversary's personnel make mistakes of tradecraft that give the victim enough information to attribute the intrusion publicly, they will try not to make those mistakes again if they can figure out what those mistakes are. They may use different tools to conduct future intrusions to frustrate historical comparisons. Such actions may make the attribution judgment more difficult for the victim.

On the other hand, the adversary may not know what mistakes he made that revealed useful information to the victim. New tools may be unfamiliar to the adversary's human perpetrators, thus increasing the likelihood of making a mistake in using them. Such actions may increase the likelihood that an attribution judgment will be successful.

In short, while the victim faces a number of uncertainties in reaching an attribution judgment, the adversary faces a number of uncertainties in seeking to mask its responsibility. It is true that the victim cannot always be highly confident in the success of its attribution process, but although the cyber terrain favors the adversary under many circumstances, the adversary still cannot always be confident that it will remain anonymous. Put differently, even if the victim cannot always have high confidence in its ability to attribute an intrusion to a specific adversary, the adversary always runs some risk that the victim will be able to attribute hostile intrusions successfully. It is the very existence of such risk that underpins the possibility of deterring hostile actions in cyberspace.

If an adversary affirmatively wants its use of cyber weapons to be attributed to it for some reason, a somewhat different set of considerations applies. In this scenario, Nation A uses its cyber weapons against Nation B, but also wants B to know that A is responsible. In this context, one would usually speak of A's taking credit for the cyberattack.

A could persuasively take credit simply by informing B that it was responsible for the cyberattack on target X belonging to B on a particular time and date, and providing B with details that only A would know about that particular attack. In this case, B would almost certainly want to verify A's claims—and B would have to go through the all-source intelligence process described above to confirm A's involvement. However, seeking to confirm A's involvement is an easier task than determining A's involvement, because in the former case, A has provided information that would not be available in the latter case.

In principle, it is also possible for A to use "loud" cyber weapons that self-attribute, much like nationality markings on aircraft assert that an airplane using the US nationality marking is in fact a US military airplane and national uniforms worn by soldiers assert that a soldier wearing a US military uniform is in fact a member of the US armed forces. But even if such cyber weapons are used (and US Cyber Command has expressed an interest in obtaining such weapons), B might still have to go through the process of determining A was indeed responsible, even if the weapon was eminently traceable to A.[100] (The technical challenge for self-attributing cyber weapons is two-fold. First, the self-attributing characteristic must not enable an adversary's defenses to identify the weapon as hostile before it acts. Second, the self-attributing characteristic must not be usable by another Nation C.)

Conclusion

This paper began with the observation that attribution is a deep issue. In 2009, the National Research Council wrote, "The bottom line [on attribution] is that it is too strong a statement to say that plausible attribution of an adversary's cyberattack is impossible, but it is also too strong to say that definitive and certain attribution of an adversary's cyberattack will always be possible."[101] Fast forwarding to 2016, Clapper's observation above is consistent with that view—in some ways attribution is becoming easier, and in other ways it is becoming harder.

On one hand, attribution capabilities are increasing because more attention and resources are being devoted to the topic. Indeed, attribution capabilities are better than they were a decade ago in large part because nations are more attentive to the possibility of malicious cyber activity. They are thus more likely than before to collect data that might be useful in the investigation of a present—or a future—intrusion, and collection efforts have resulted in a decade's worth of data, providing a historical corpus against which to compare future cyber intrusions.. The tools for attribution are better, and analysts are more experienced. Put differently—given the likelihood of malicious cyber activity in the future, many nations are more willing to make investments in intelligence and to build investigative capacity that will pay off in the future, and capabilities for attribution are in large part a function of the investment a nation is willing to make in those capabilities, both in infrastructure and in the effort that any given case demands.[102]

On the other hand, adversaries are more aware than ever that they are being tracked, and given the ease with which false clues can be planted and false-flag operations conducted, they may well be more likely to carry out countermeasures to throw investigators off the attribution trail, especially as the stakes grow larger. And the number of skilled adversaries is growing. Adversaries that are identified can also exploit the uncertainty inherent in an attribution judgment. An adversary can deny its activities outright, secure in the knowledge that even if the information underlying the judgment is publicly revealed, that information is highly unlikely to contain any "smoking guns" pointing to its involvement.[103] It can discredit each individual inference and piece of circumstantial evidence by pointing to alternative story lines. Such an approach to discrediting an attribution judgment may be especially valuable in the court of public opinion, in which individuals have little expertise on which to base their own judgments.

Policymakers are accustomed to making decisions about what to do or not to do under conditions of uncertainty—this is the reality of their daily lives. But the reality of some degree of irreducible uncertainty about attribution judgments has important political ramifications. If policymakers are forced to "go public" with an attribution judgment, skeptics and adversaries alike will pounce on any expressed uncertainty to dispute it and to set forth alternative theories and conclusions. Thus, they may be forced to assume a public posture that appears to be more certain than the actual evidence warrants.

The center of gravity of informed judgment seems to indicate greater confidence in attribution overall today than was true a decade ago, but the future remains cloudy as intruders and attributers advance their respective capabilities. Nevertheless, and regardless of how these competing factors compare in the future, a number of fundamental propositions will remain. To be successful, attribution will always entail an all-source proposition, and technical forensics will be only one part of an attribution judgment. Attribution judgments will always have some degree of uncertainty associated with them, and the significance of such uncertainty is a political and policy matter rather than a technical one. Victims will have to live with the possibility that they will not be able to arrive at accurate attribution judgments with high confidence, and adversaries will have to live with the possibility that their victims will be able to attribute their malicious cyber activities to them.

Acknowledgments

I am grateful to Taylor Grossman's services as research associate for this paper. Steven Bellovin, Eileen Donahoe, Kristen Eichensehr, David Elliott, John Gerth, Jack Goldsmith, Chris Jacobi, Alex Keller, Susan Landau, Hal Murray, Joseph Nye, Mark Seiden, Ashwin Sreenivas, Eli Sugarman, Wesley Tiu, and Benjamin Wittes provided valuable comments on an early version of this paper that helped to improve it.

Furthermore, this paper drew heavily on and built on work by W. Earl Boebert,[104] John Carlin,[105] David Clark and Susan Landau,[106] Clement Guitton and Elaine Korzak,[107] Jason Healey,[108] Thomas Rid and Ben Buchanan,[109] Jon Lindsay,[110] Nicholas Tsagourias,[111] and David Wheeler and Gregory Larsen.[112]

NOTES

1 Malicious cyber activities or incidents are also sometimes known as "intrusions"; these terms are meant to include both what are called cyberattacks and cyber exploitations in much of the literature. Attacks are intended to destroy, degrade, damage, disrupt, manipulate, usurp, or reduce the availability of information and/or the computer and communications systems handling such information. Exploitations are intended to surreptitiously exfiltrate information that is meant to be kept confidential by the owners or operators of the system or network storing or transmitting such information. For more discussion of the difference between these two, see William Owens, Kenneth Dam, and Herbert Lin, eds., *Technology, Policy, Law, and Ethics Regarding U.S. Acquisition and Use of Cyberattack Capabilities,* Washington DC: National Academies Press, 2009, Chapter 1.

2 Many conceptualizations of deterrence include deterrence by denial, a strategy that seeks to deny an adversary the benefits it may realize by conducting malicious or hostile activities. According to the logic of deterrence by denial, an adversary will refrain from malicious actions if he knows he will not gain the benefits of those actions. In cyberspace, this approach is essentially equivalent to having cyber defenses that are sufficient to make it not worth the adversary's while to act maliciously. The problem today is that we don't know how to design, build, or operate cyber defenses that are sufficiently effective to deter.

3 Sometimes, the misbehavior or badness is not apparent. A computer can be compromised in a way that allows it to be misused in ways that cause no change in the computer's behavior that is apparent to the user—that is, a machine can be compromised and still be fully and properly functional from the user's standpoint. Such a compromise can nevertheless cause the machine to behave in a way that the user would not like if he or she knew about it. For example, say a machine is compromised to serve as a clandestine sender of spam or a proxy in an attack on another machine; the user would not experience direct harm, but his or her machine would be being used for nefarious purposes without his or her knowledge.

4 For example, when a long time elapses between intrusion and the manifestation of a clue that something is wrong, many more system log entries may need to be examined to find the two or three useful entries that relate

to the initial intrusion. Or multiple system updates performed during this time may have destroyed information that could have been useful.

5 A complementary point of view is that computers or computer-based systems that allow the user to do the wrong thing are in fact defective in some sense themselves, even if the computers per se worked properly. Further, as such systems become more sophisticated, knowing whether a "bad outcome" is the result of human error or computer error becomes harder. And if the problem is "computer error," we won't know what the cause of the error is—and in particular whether it's due to a malicious actor or some unanticipated quirk from a big data analysis or something similar. This point, for which the author has considerable sympathy, will not be further addressed in this paper because it is not usually regarded as falling within the ambit of attribution as a security concern.

6 Whether she bears responsibility for being careless in her security precautions is a different question—and if she does, it would be fair to call her carelessness an indirect cause of or a contributing factor to the incident. (On the other hand, a system that makes it easy to inadvertently delete a file and not know it is poorly designed, and thus a deletion of a file could arguably reflect a system design problem rather than foul play.)

7 A God's-eye perspective describes what actually happened. The attribution process is intended to reveal to investigators as much of that perspective as possible.

8 This particular way of formulating answers to this question owes much to a discussion found in David Clark and Susan Landau, "Untangling Attribution," *Harvard National Security Journal* 2(1):323–352, 2011, http://harvardnsj.org/2011/03/untangling-attribution-2/.

9 Clark and Landau, "Untangling Attribution."

10 The term "stepping stones" is also used in the literature. See Yin Zhang and Vern Paxson, "Detecting Stepping Stones," *Proceedings of the 9th USENIX Security Symposium,* pp. 171–184, August 2000, https://www.cs.utexas.edu/~yzhang/papers/stepping-sec00.pdf.

11 A good, if dated, treatment of technical means that can yield information useful for attribution can be found in David Wheeler and Gregory Larsen, "Techniques for Cyber Attack Attribution," Institute for Defense Analyses, October 2003, http://www.dtic.mil/cgi-bin/GetTRDoc?AD=ADA468859&Location=U2&doc=GetTRDoc.pdf. This report presages a number of the conclusions drawn in the present paper.

12 See, for example, W. Earl Boebert, "A Survey of Challenges in Attribution," in *Proceedings of a Workshop on Deterring Cyberattacks: Informing Strategies and Developing Options for U.S. Policy,* Washington DC: National Academies Press, pp. 41–52, 2010, http://www.nap.edu/catalog/12997/proceedings-of-a-workshop-on-deterring-cyberattacks-informing-strategies-and.

13 See TOR website, www.torproject.org/.

14 See EPIC guide at www.epic.org/privacy/tools.html.

15 David E. Sanger and Martin Fackler, "N.S.A. Breached North Korean Networks Before Sony Attack," *New York Times,* January 18, 2015, http://www.nytimes.com/2015/01/19/world/asia/nsa-tapped-into-north-korean-networks-before-sony-attack-officials-say.html.

16 Given the capability to pre-position instrumentation to surveil traffic in a potential adversary's network, an interesting question is why one could not also pre-position other tools to shut down an intrusion by that adversary as it is being launched. A full answer to this question is beyond the scope of this paper; for now, two observations must suffice. First, it may not be possible to immediately recognize traffic associated with the start of an intrusion as such, especially if that information is collected and analyzed without knowledge of what is about to happen. Second, even if it were possible to do so, the scope and nature of the intrusion's negative effects may not warrant exposing the intelligence capability in place. Weighing those equities (preventing the presumed negative effects of the intrusion versus maintaining the secrecy of the intelligence capability in place) is not something that policymakers would do quickly or leave to an automated system to decide.

17 And at worst, an adversary may be able to hijack an IP address so that intrusion traffic appears to originate from that address, making the IP address much less useful as evidence for attribution. A similar outcome may result under circumstances in which IP addresses are assigned dynamically.

18 Similar issues even arise in a purely domestic context that crosses state lines. For example, a 2013 decision of the Fifth Circuit Court of Appeals found that federal district judges may not authorize wiretaps of cell phones outside of their jurisdiction (www.ca5.uscourts.gov/opinions/pub/11/11-60763-CR0.wpd.pdf). This ruling conflicted with a 1997 decision of the Seventh Circuit Court of Appeals stating that district judges did have some authority to do so under certain circumstances (https://law.resource.org/pub/us/case/reporter/F3/112/112.F3d.849.96-2340.96-2276.96-2257.96-2237.html). For a newspaper account of this story, see Joe Palazzolo, "Court Curbs Authority to Issue Wiretap Warrants," *Wall Street Journal,* August 27, 2013, http://blogs.wsj.com/law/2013/08/27/court-restricts-judicial-authortity-to-issue-wiretap-warrants/. More recently, controversy has arisen over a proposed change to Rule 41 that some analysts believe grants judges anywhere, regardless of jurisdiction, the authority to "issue a search warrant to remotely access, seize, or copy data relevant to a crime when a computer was using privacy-protective tools to safeguard one's location." See Rainey Reitman, "With Rule 41, Little-Known Committee Proposes to Grant New Hacking Powers to the Government," Electronic Frontier Foundation, April 30, 2016, www.eff.org/deeplinks/2016/04/rule-41-little-known-committee-proposes-grant-new-hacking-powers-government.

19 Whether actions such as turning on a web camera to capture a picture of the person sitting at the keyboard should count as technical forensics is an interesting edge case.

20 It has been observed that the same "neuro-physiological factors that make written signatures unique, are also exhibited in a user's typing pattern," and thus, "When a person types, the latencies between successive keystrokes, keystroke durations, finger placement and applied pressure on the keys can be used to construct a unique signature (i.e., profile) for that individual. For well-known, regularly typed strings, such signatures can be quite consistent." See Fabian Monrose and Aviel D. Rubin, "Keystroke dynamics as a biometric for authentication," *Future Generation Computer Systems* 16(4):351–359, February 2000, www.cs.columbia.edu/4180/hw/keystroke.pdf.

21 This formulation ("who is to blame"versus "who did it") is due to Jason Healey, *Beyond Attribution: Seeking National Responsibility in Cyberspace,* Atlantic Council issue brief, February 22, 2012, www.atlanticcouncil.org/publications/issue-briefs/beyond-attribution-seeking-national-responsibility-in-cyberspace.

22 Article 4 of The International Law Commission's Draft Articles on State Responsibility states that "The conduct of any State organ shall be considered an act of that State under international law, whether the organ exercises legislative, executive, judicial or any other functions, whatever position it holds in the organization of the State, and whatever its character as an organ of the central Government or of a territorial unit of the State." The Draft Articles are a UN-sponsored attempt to codify international law in this area, but although a UN General Assembly Resolution in December 2001 (https://documents-dds-ny.un.org/doc/UNDOC/GEN/N01/477/97/PDF/N0147797.pdf) took note of the Draft Articles and commended them to the attention of governments without prejudice to the question of their future adoption or other appropriate action, no further action has been taken on these articles.

23 Mandiant, "APT1: Exposing One of China's Cyber Espionage Units," www.fireeye.com/content/dam/fireeye-www/services/pdfs/mandiant-apt1-report.pdf.

24 Mandiant, "APT1," 62.

25 John P. Carlin, "Detect, Disrupt, Deter: A Whole-of-Government Approach to National Security Cyber Threats," *Harvard National Security Journal* 7(2): 391–436, 2016, http://harvardnsj.org/wp-content/uploads/2016/06/Carlin-FINAL.pdf.

26 The Federal Bureau of Investigation and the Drug Enforcement Agency are both federal law enforcement agencies and members of the intelligence community. See Office of the Director of National Intelligence, www.dni.gov/index.php/intelligence-community/members-of-the-ic.

27 Federal Register, Executive Order 12333—United States intelligence activities, www.archives.gov/federal-register/codification/executive-order/12333.html.

28 EO 12333 defines US persons as US citizens, US permanent resident aliens, an unincorporated association substantially composed of US citizens or permanent resident aliens, or a corporation incorporated in the United States, except for a corporation directed and controlled by a foreign government or governments. See https://fas.org/irp/offdocs/eo/eo-12333-2008.pdf.

29 The only known public and explicit constraint on US intelligence activities regarding foreigners is contained in PPD-28, which states, "To the maximum extent feasible consistent with the national security, these policies and procedures [in this PPD] are to be applied equally to the personal information of all persons, regardless of nationality." In other words, PPD-28 states that foreigners *do* have some legitimate privacy interests against US intelligence agencies, and that these agencies will treat that data (in the absence of national security concerns) as it treats data about US citizens. For more on this point, see Benjamin Wittes, "The President's Speech and PPD-28: A Guide for the Perplexed," *Lawfare* blog, www.lawfareblog.com/presidents-speech-and-ppd-28-guide-perplexed.

On the other hand, both US law and policy do forbid other activities (e.g., EO 12333 forbids assassinations). To the extent that intelligence collection activities might run afoul of US law, the US Constitution, or executive order, they may not be undertaken. Also, other international law not specifically related to intelligence collection could prohibit certain collection activities—for example, torture is prohibited as a matter of international law and US intelligence agencies are prohibited from torturing individuals to collect intelligence information. This paper does not address the distinction between torture and enhanced interrogation techniques, but for more information on this point, see Anne Daugherty, Congressional Research Service, "Perspectives on Enhanced Interrogation Techniques," Library of Congress, Washington DC, January 8, 2016, www.fas.org/sgp/crs/intel/R43906.pdf.

30 A May 2015 blog post on *Lawfare* by Ashley Deeks, "The Increasing State Practice and Opinio Juris on Spying," notes that in the wake of the Snowden revelations, many states have expressed views on the relationship between surveillance and international law and that these expressions are "an important development in the process of understanding how intelligence activities are and should be regulated by international law." See www.lawfareblog.com/increasing-state-practice-and-opinio-juris-spying. In August 2015, Deeks also argues that "adopting a number of procedural norms to regulate foreign surveillance would help states and their citizens begin to balance the competing equities of privacy and security in concrete and observable ways." See Ashley Deeks, "An International Law Framework for Surveillance," *Virginia Journal of International Law* 55(2):291–368, August 2015, http://www.vjil.org/articles/an-international-legal-framework-for-surveillance.

31 Abraham Sofaer, "Cyber Security and International Agreements," in *Proceedings of a Workshop on Deterring Cyberattacks: Informing Strategies and Developing Options for U.S. Policy,* National Academies Press, Washington, DC, pp. 179–207, 2010, http://www.nap.edu/catalog/12997/proceedings-of-a-workshop-on-deterring-cyberattacks-informing-strategies-and.

32 Abraham Sofaer, "Cyber Security and International Agreements," in *Proceedings of a Workshop on Deterring Cyberattacks: Informing Strategies and Developing Options for U.S. Policy,* National Academies Press, Washington, DC, pp. 179–206, 2010, www.nap.edu/catalog/12997/proceedings-of-a-workshop-on-deterring-cyberattacks-informing-strategies-and.

33 Michael Vatis, "The Council of Europe Convention on Cybercrime," in *Proceedings of a Workshop on Deterring Cyberattacks: Informing Strategies and Developing Options for U.S. Policy,* National Academies Press, Washington, DC, pp. 207–223, 2010, www.nap.edu/catalog/12997/proceedings-of-a-workshop-on-deterring-cyberattacks-informing-strategies-and.

34 Jason Healey, "Beyond Attribution: Seeking National Responsibility for Cyber Attacks," Atlantic Council issue brief, February 22, 2012, www.atlanticcouncil.org/publications/issue-briefs/beyond-attribution-seeking-national-responsibility-in-cyberspace.

35 I am indebted to Chris Jacoby for this point.

36 The discussion of this paragraph is taken from William Owens, Kenneth Dam, and Herbert Lin, eds., *Technology, Policy, Law, and Ethics Regarding U.S. Acquisition and Use of Cyberattack Capabilities,* National Academies Press, Washington DC, 2009, p. 186, footnote 30.

37 See, for example, Article 8 of the ILC (International Law Commission) State Responsibility Articles, available at http://untreaty.un.org/ilc/texts/instruments/ english/commentaries/9_6_2001.pdf, pp. 47 ff; and the ICJ (International Court of Justice) Nicaragua decision (arguing for "effective control") and the ICTY (International Criminal Tribunal for Yugoslavia) Tadic decision (arguing for "overall control").

38 UN Security Council, "Letter Dated 7 October 2001 From the Permanent Representative of the United States of America to the United Nations Addressed to the President of the Security Council," UN Doc. No. S/2001/946 (2001).

39 Derek Jinks, "State Responsibility for the Acts of Private Armed Groups," *Chicago Journal of International Law* 4(1):83–96, Spring 2003, http://papers.ssrn.com/sol3/papers.cfm?abstract_id=391641.

40 Clarke and Knake propose an explicit treaty assigning responsibility to nations for cyber activities emanating from their territories. See Richard Clarke and Robert Knake, *Cyber War: the Next Threat to National Security and What to Do About It,* New York: Harper Collins, 2010.

41 UN Security Council, Resolution 1267, www.un.org/ga/search/view_doc.asp?symbol=S/RES/1267(1999).

42 David Fidler, "Cyber War Crimes: Islamic State Atrocity Videos Violate the Laws of War," blog post on *Net Politics,* April 8, 2015, http://blogs.cfr.org/cyber/2015/04/08/cyber-war-crimes-islamic-state-atrocity-videos-violate-the-laws-of-war/.

43 International Criminal Court, "How the Court Works," www.icc-cpi.int/about/how-the-court-works/Pages/default.aspx#legalProcess.

44 Thomas Rid and Ben Buchanan, "Attributing Cyber Attacks," *Journal of Strategic Studies* 38(1–2):4–37, December 23, 2014, www.tandfonline.com/doi/abs/10.1080/01402390.2014.977382.

45 For a canonical expression of this perspective, see Jeffrey Carr, "Responsible Attribution: A Prerequisite For Accountability," Tallinn Paper No. 6, NATO Cooperative Cyber Defence Centre of Excellence, Tallinn, Estonia, 2014, https://ccdcoe.org/sites/default/files/multimedia/pdf/Tallinn%20Paper%20No%20%206%20Carr.pdf.

46 David Clark and Susan Landau, "Untangling Attribution," in *Proceedings of a Workshop on Deterring Cyberattacks: Informing Strategies and Developing Options for U.S. Policy,* Washington DC: National Academies Press, pp. 25–40, 2010, http://www.nap.edu/catalog/12997/proceedings-of-a-workshop-on-deterring-cyberattacks-informing-strategies-and.

47 Jeffrey Hunker, Robert Hutchinson, and Jonathan Marguiles, "Attribution of Cyber Attacks on Process Control Systems," in *Critical Information Protection II,* International Federation for Information Processing, Volume 290, Mauricio Papa and Sujeet Shenoi (eds.), Springer, Boston, MA, March 2008, pp. 87–99.

48 A view often heard in the technical community and presented here in oversimplified form holds that definitive attribution is essentially impossible. Many in the technical community believe that only technical evidence speaks for itself, and that it is somehow "purer" and "less tainted" than information gained from some source whose motives were suspect and who could lie. They further assert that "a mountain of weak or poor-quality evidence" is inherently unpersuasive and non-authoritative. As someone who once held this view, I (the author of this paper) would assess each piece of weak evidence on its own ("weak" in this context meant "not bullet-proof") and, because it was weak, I would throw it away. At the end of the process, because I insisted on throwing away every piece of weak evidence, and only weak evidence was available, I was left with no evidence at all. And, of course, with no evidence, attribution is impossible. This is not to say that conclusions that emerge from analyzing weak evidence are necessarily reliable. An important caveat is that pieces of weak evidence collectively point to a stronger conclusion only when they are independent. For example, an intruder determined to mislead forensic

investigators will plant a variety of false clues. Thus, at the moment of collection, the investigator cannot presume the independence of any given clue, and he or she must take into account the probability that a newly gathered clue is not in fact independent. On the other hand, that probability is not unity, and it would have to be probability 1.0 to discard the new clue entirely. In general, the higher the probability of non-independence, the greater the necessity of obtaining other corroborating sources that are not technical in nature.

49 Michael Caloyannides. “Forensics is so ‘yesterday,’ ” *IEEE Security & Privacy* 7(2):18–25, March/April 2009, https://www.computer.org/csdl/mags/sp/2009/02/msp2009020018-abs.html. Some empirical work undertaken by Nunes et al. found that in an exercise where ground truth about identities was known, the majority of misidentifications of an intruder resulted from deceptive activities. See Eric Nunes, Nimish Kulkarni, Paulo Shakarian, Andrew Ruef, and Jay Little, “Cyber-Deception and Attribution in Capture-the-Flag Exercises”, *2015 IEEE/ACM International Conference on Advances in Social Networks Analysis and Mining,* ASONAM ’15, August 25–28, 2015, Paris, France, http://dl.acm.org/citation.cfm?doid=2808797.2809362.

50 Similarities between the malware used in the 2014 hack on Sony Pictures Entertainment and malware used in other cyber intrusions previously attributed to North Korea were in part responsible for the FBI’s attribution of the Sony hack to North Korea. See James B. Comey, director, Federal Bureau of Investigation, Remarks at the International Conference on Cyber Security, Fordham University, January 7, 2015, www.fbi.gov/ news/speeches/addressing-the-cyber-security-threat.

51 John P. Carlin, “Detect, Disrupt, Deter.”

52 Guitton and Korzak elaborate on this point, arguing that the correlation between “sophistication” and likelihood of a nation-state actor being involved is not perfect, at least in part because “the lack of clarity and inconsistency around the term ‘sophistication’ ” means that sophistication is context-dependent and is therefore an unreliable guide to associating a nation-state with any given intrusion. See Clement Guitton and Elaine Korzak, “The Sophistication Criterion for Attribution: Identifying the Perpetrators of Cyber-Attack,” *The RUSI Journal* 158(4):62–68, 2013, www.tandfonline.com/doi/abs/10.1080/03071847.2013.826509.

53 Central Intelligence Agency, “Human Intelligence,” www.cia.gov/news-information/featured-story-archive/2010-featured-story-archive/intelligence-human-intelligence.html.

54 The Joint Civilian-Military Investigation Group, “Investigation Result on the Sinking of ROKS ‘Cheonan,’ ” May 20, 2010, http://news.bbc.co.uk/nol/shared/bsp/hi/pdfs/20_05_10jigreport.pdf (a more readable form can be found at http://www.globalsecurity.org/military/library/report/2010/100520_jcmig-roks-cheonan/100520_jcmig-roks-cheonan.htm).

55 For more on this point, see Jon R. Lindsay, “Tipping the scales: the attribution problem and the feasibility of deterrence against cyberattack,” *Journal of Cybersecurity* 1(1): 1–15, 2015, http://cybersecurity.oxfordjournals.org/content/1/1/53.

56 William Lynn III, “Defending a New Domain: The Pentagon’s Cyberstrategy,” *Foreign Affairs* 89(5): 97–108, September/October 2010, www.foreignaffairs.com/articles/united-states/2010-09-01/defending-new-domain.

57 Leon Panetta, “Defending the Nation from Cyber Attack,” remarks on cybersecurity to the Business Executives for National Security, New York City, October 11, 2012, http://archive.defense.gov/speeches/speech.aspx?speechid=1728.

58 James Clapper, “Worldwide Threat Assessment of the US Intelligence Community,” testimony to the Senate Armed Services Committee, February 26, 2015, https://www.dni.gov/files/documents/Unclassified_2015_ATA_SFR_-_SASC_FINAL.pdf.

59 James Clapper, “Worldwide Threat Assessment of the US Intelligence Community,” testimony to the Senate Armed Services Committee, February 9, 2016, www.dni.gov/files/documents/SASC_Unclassified_2016_ATA_SFR_FINAL.pdf.

60 US Department of Defense, “The DoD Cyber Strategy,” Washington, DC, April 2015, http://www.defense.gov/Portals/1/features/2015/0415_cyber-strategy/Final_2015_DoD_CYBER_STRATEGY_for_web.pdf.

61 The examples described here are taken from Kristen Eichensehr's blog post, "The Private Frontline in Cybersecurity Offense and Defense," October 30, 2014, www.justsecurity.org/16907/private-frontline-cybersecurity-offense-defense/.

62 www2.fireeye.com/apt28.html.

63 www.novetta.com/wp-content/uploads/2014/11/Executive_Summary-Final_1.pdf.

64 https://cdn0.vox-cdn.com/assets/4589853/crowdstrike-intelligence-report-putter-panda.original.pdf.

65 The discussion below of private-sector attribution is derived from Herbert Lin, "Reflections on the New DOD Cyber Strategy: What It Says, What It Doesn't Say," *Georgetown Journal of International Relations,* forthcoming 2016.

66 This point relates only to process, and should not be read to imply that private-sector analyses are necessarily less accurate or rigorous than government analyses.

67 In the annals of intelligence, these words are called "words of estimative probability." See Central Intelligence Agency, "Words of Estimative Probability," https://www.cia.gov/library/center-for-the-study-of-intelligence/csi-publications/books-and-monographs/sherman-kent-and-the-board-of-national-estimates-collected-essays/6words.html.

68 See www.law.cornell.edu/wex/due_process and http://dictionary.law.com/Default.aspx?selected=595#ixzz4A442UwI7.

69 Nicholas Tsagourias , "Cyber attacks, self-defence and the problem of attribution," *Journal of Conflict and Security Law* 17 (2): 229–244, 2012.

70 International Court of Justice, CASE CONCERNING MILITARY AND PARAMILITARY ACTIVITIES IN AND AGAINST NICARAGUA (NICARAGUA V. UNITED STATES OF AMERICA), MERITS, JUDGMENT OF 27 JUNE 1986, http://www.icj-cij.org/docket/files/70/6503.pdf, page 28 (emphasis added); cited in Tsagourias, note 66.

71 Separate Opinion of Judge Higgins in Case Concerning Oil Platforms (Islamic Republic of Iran v USA) (Merits) [2003] ICJ Rep 161, paragraph 30; cited in Tsagourias, note 24.

72 https://www.dni.gov/files/documents/Newsroom/Reports%20and%20Pubs/20071203_release.pdf.

73 The term "words of estimative probability" comes from Sherman Kent's classic 1964 piece "Words of Estimative Probability," https://www.cia.gov/library/center-for-the-study-of-intelligence/csi-publications/books-and-monographs/sherman-kent-and-the-board-of-national-estimates-collected-essays/6words.html.

74 See https://www.dni.gov/files/documents/Newsroom/Reports%20and%20Pubs/20071203_release.pdf.

75 See https://www.washingtonpost.com/world/national-security/why-the-sony-hack-drew-an-unprecedented-us-response-against-north-korea/2015/01/14/679185d4-9a63-11e4-96cc-e858eba91ced_story.html.

76 Even assuming that a disgruntled insider at Sony was involved, there is no reason in principle that government operatives from North Korea might not have compromised such an individual. Indeed, when asked whether other individuals may have assisted North Korea or were involved in the assault on Sony, but not ultimately responsible for the damage that was done, an FBI spokesperson said, "We're not making the distinction that you're making about the responsible party and others being involved." See www.thedailybeast.com/articles/2014/12/30/fbi-won-t-reject-a-sony-insider-hack.html and http://dailycaller.com/2014/12/29/cybersecurity-firm-identifies-six-in-sony-hack-one-a-former-company-insider/.

77 Marc Rogers, "No, North Korea Didn't Hack Sony," *The Daily Beast,* December 24, 2014, www.thedailybeast.com/articles/2014/12/24/no-north-korea-didn-t-hack-sony.html,

78 Corfu Channel case, Judgment of April 8, 1949: I.C. J. Reports 1949, p. 18, http://www.icj-cij.org/docket/files/1/1645.pdf.

79 Theodore Sorenson, *Counselor: A Life at the Edge of History,* New York: Harper Collins, 2008, p. 291.

80 Jack Goldsmith, “The Sony Hack: Attribution Problems, and the Connection to Domestic Surveillance”, December 19, 2014, https://www.lawfareblog.com/sony-hack-attribution-problems-and-connection-domestic-surveillance.

81 Jack Goldsmith, “The Consequences of Credible Doubt About the USG Attribution in the Sony Hack,” December 30, 2014, https://www.lawfareblog.com/consequences-credible-doubt-about-usg-attribution-sony-hack.

82 Paul Rosenzweig, *Was it North Korea?* December 24, 2014, https://www.lawfareblog.com/was-it-north-korea.

83 I am grateful to a blog post by Kristen Eichensehr on this point; see Kristen Eichensehr, “ ’Your Account May Have Been Targeted by State-Sponsored Actors’: Attribution and Evidence of State-Sponsored Cyberattacks,” *Just Security* blog, January 11, 2016, https://www.justsecurity.org/28731/your-account-targeted-state-sponsored-actors-attribution-evidence-state-sponsored-cyberattacks/. The examples in footnotes 84 and 85 are from her blog post as well.

84 https://security.googleblog.com/2012/06/security-warnings-for-suspected-state.html.

85 https://www.facebook.com/notes/facebook-security/notifications-for-targeted-attacks/10153092994615766/.

86 See, for example, http://freebeacon.com/national-security/china-says-opm-hack-was-not-state-sponsored/.

87 It is not coincidental that the same process occurs when various groups claim credit for an act of terrorism.

88 This is one of three factors in US Strategic Command’s formulation of the requirements for deterrence. See U.S. Strategic Command, “Deterrence Operations: Joint Operating Concept,” version 2.0, December 2006, http://www.dtic.mil/doctrine/concepts/joint_concepts/joc_deterrence.pdf.

89 Certain types of active cyber defense call for just such action, and a number of analyses have asserted the value of such action. This particular author is skeptical about the actual value of such action, but this point will not be addressed in this paper.

90 Harold Hongju Koh, legal advisor, US Department of State, “Remarks on International Law in Cyberspace,” delivered at the USCYBERCOM Inter-Agency Legal Conference of September 18, 2012, Fort Meade, MD, http://www.state.gov/s/l/releases/remarks/197924.htm.

91 White House, “International Strategy for Cyberspace,” 2011, https://www.whitehouse.gov/sites/default/files/rss_viewer/international_strategy_for_cyberspace.pdf.

92 US Strategic Command, “Deterrence Operations.”

93 See, for example, Thomas Schelling, *Arms and Influence,* New Haven, CT: Yale University Press, 2008 (originally printed 1966).

94 The criminal deterrence literature does address the impact of celerity (or swiftness of punishment) on the deterrence of crime, but here too the outcome is mixed. In a review of the criminal deterrence literature, Paternoster (Raymond Paternoster, “How Much Do We Really Know about Criminal Deterrence,” *The Journal of Criminal Law and Criminology* 100(3):765–824, 2010, available at http://www.jstor.org/stable/25766109) cites early theories of criminal deterrence (where “early” refers to theories of 1764!) arguing that punishment must be swift in order to be effective and more recent experimental work (1987) suggesting that “given the choice, people would like to get their punishment over as quickly as possible and that punishment delayed is seen as more costly than if given immediately”—that is, dread induced by delay increases rather than decreases the perceived cost of punishment. Paternoster concludes that the criminal deterrence literature has no real knowledge base about the celerity of punishment. Moreover, it is unclear how and to what extent, if any, the psychological mechanisms of would-be criminals driving their cost estimates are applicable to how nations account for potential costs. As an example of a criminal investigation running a very long time, consider that in the case of Benjamin Arellano-Felix,

the leader of the Tijuana Drug Cartel, fifteen years elapsed between his indictment (See “Under New Law, Mexico Extradites Suspect to U.S.,” *New York Times,* May 5, 2001, http://www.nytimes.com/2001/05/05/world/under-new-law-mexico-extradites-suspect-to-us.html) in 1997 and his incarceration in 2012 (Richard Marosi, “Former Drug Kingpin Arellano Felix Gets 25-Year Prison Term,” *Los Angeles Times,* April 3, 2012, http://articles.latimes.com/2012/apr/03/local/la-me-arellano-felix-20120403). (Both of these citations are from Carlin.)

95 I am grateful to Kristen Eichensehr for this point.

96 Joseph S. Nye Jr., “Deterrence and Dissuasion in Cyberspace,” *International Security* 41(3), Winter 2016/17.

97 Some notable sources include National Research Council, “ Improving Risk Communication” (1989), and Baruch Fischhoff and Dietram Scheufele, “The Science of Science Communication,” 2012, http://onlinedigeditions.com/publication/?i=174803.

98 Nye, “Deterrence and Dissuasion in Cyberspace.”

99 It is often the case that one intrusion is conducted to establish continuing access, thus facilitating later intrusions.

100 Chris Bing, “US Cyber Command Director: We Want ‘Loud,’ Offensive Cyber Tools,” August 16, 2016, http://www.fedscoop.com/us-cyber-command-offensive-cybersecurity-nsa-august-2016.

101 William Owens, Kenneth Dam, and Herbert Lin, eds., *Technology, Policy, Law, and Ethics Regarding U.S. Acquisition and Use of Cyberattack Capabilities,* Washington DC: National Academies Press, 2009.

102 For more on this point, see Jon R. Lindsay, “Tipping the scales: the attribution problem and the feasibility of deterrence against cyberattack,” *Journal of Cybersecurity* 1(1): 1–15, 2015, http://cybersecurity.oxfordjournals.org/content/1/1/53.

103 For example, North Korea explicitly denied that it was responsible for the Sony hack. See http://www.reuters.com/article/us-sony-cybersecurity-nkorea-idUSKCN0JI1NZ20141204.

104 W. Earl Boebert, “A Survey of Challenges in Attribution,” in *Proceedings of a Workshop on Deterring Cyberattacks: Informing Strategies and Developing Options for U.S. Policy,* Washington DC: National Academies Press, pp. 41–52, 2010, http://www.nap.edu/catalog/12997/proceedings-of-a-workshop-on-deterring-cyberattacks-informing-strategies-and.

105 Carlin, “Detect, Disrupt, Deter.”

106 Clark and Landau, “Untangling Attribution.”

107 Clement Guitton and Elaine Korzak, “The Sophistication Criterion for Attribution: Identifying the Perpetrators of Cyber-Attack,” *The RUSI Journal* 158(4):62–68, 2013, http://www.tandfonline.com/doi/abs/10.1080/03071847.2013.826509.

108 Jason Healey, *Beyond Attribution: Seeking National Responsibility for Cyber Attacks,* Atlantic Council issue brief, February 22, 2012, http://www.atlanticcouncil.org/publications/issue-briefs/beyond-attribution-seeking-national-responsibility-in-cyberspace.

109 Thomas Rid and Ben Buchanan, “Attributing Cyber Attacks,” *Journal of Strategic Studies* 38(1–2):4–37, December 23,2014, http://www.tandfonline.com/doi/abs/10.1080/01402390.2014.977382.

110 Jon R. Lindsay, “Tipping the scales: the attribution problem and the feasibility of deterrence against cyberattack,” *Journal of Cybersecurity* 1(1): 1–15, 2015, http://cybersecurity.oxfordjournals.org/content/1/1/53.

111 Nicholas Tsagourias , “Cyber attacks, self-defence and the problem of attribution,” *Journal of Conflict and Security Law* 17 (2): 229–244, 2012.

112 David Wheeler and Gregory Larsen, *Techniques for Cyber Attack Attribution,* Institute for Defense Analyses, October 2003, http://www.dtic.mil/cgi-bin/GetTRDoc?AD=ADA468859&Location=U2&doc=GetTRDoc.pdf.

The publisher has made this work available under a Creative Commons Attribution-NoDerivs license 3.0. To view a copy of this license, visit http://creativecommons.org/licenses/by-nd/3.0.

Hoover Institution Press assumes no responsibility for the persistence or accuracy of URLs for external or third-party Internet websites referred to in this publication, and does not guarantee that any content on such websites is, or will remain, accurate or appropriate.

Copyright © 2016 by the Board of Trustees of the Leland Stanford Junior University

The preferred citation for this publication is:
Herbert Lin, "Attribution of Malicious Cyber Incidents," Hoover Working Group on National Security, Technology, and Law, Aegis Series Paper No. 1607 (September 26, 2016), available at https://lawfareblog.com/attribution-malicious-cyber-incidents.

About the Author

HERBERT LIN

Herb Lin is Research Fellow at the Hoover Institution and Senior Research Scholar at the Center for International Security and Cooperation at Stanford University. He also served at the Computer Science and Telecommunications Board of the National Academies directing major projects on public policy and information technology. Previously he was a staff scientist for the House Armed Services Committee. He received his doctorate in physics from MIT.

Jean Perkins Foundation Working Group on National Security, Technology, and Law

The Working Group on National Security, Technology, and Law brings together national and international specialists with broad interdisciplinary expertise to analyze how technology affects national security and national security law and how governments can use that technology to defend themselves, consistent with constitutional values and the rule of law.

The group focuses on a broad range of interests, from surveillance to counterterrorism to the dramatic impact that rapid technological change—digitalization, computerization, miniaturization, and automaticity—are having on national security and national security law. Topics include cybersecurity, the rise of drones and autonomous weapons systems, and the need for—and dangers of—state surveillance. The working group's output, which includes the Aegis Paper Series, is also published on the *Lawfare* blog channel, "Aegis: Security Policy in Depth," in partnership with the Hoover Institution.

Jack Goldsmith and Benjamin Wittes are the cochairs of the National Security, Technology, and Law Working Group.

For more information about this Hoover Institution Working Group, visit us online at http://www.hoover.org/research-teams/national-security-technology-law-working-group.

Hoover Institution, Stanford University
434 Galvez Mall
Stanford, CA 94305-6003
650-723-1754

Hoover Institution in Washington
The Johnson Center
1399 New York Avenue NW, Suite 500
Washington, DC 20005
202-760-3200

HOOVER INSTITUTION