8

Trust but Verify: Diverse Verifiers Are a Prerequisite to Cyber Peace

Rob Knake and Adam Shostack[*]

"Trust but verify." Students of history and readers of a certain age will recall those words being spoken by former US President Ronald Reagan. His argument was that peace required verification mechanisms so that each side could be confident in the actions of the other side. There are important lessons for cyber peace. While Reagan was speaking in the context of strategic nuclear arms control, many papers have been devoted to the difficulties of arms control in the cyber domain (Maybaum and Tölle, 2016). Cyber weapons do not require the large physical infrastructure of nuclear programs and can far too easily be kept secret to allow for meaningful validation of adherence to arms control commitments. Moreover, many "cyber weapons" are dual use in nature, being deployed for the administration of computers and networks, or for security testing. Yet, as we discuss in this chapter, arms control is only one area in which verification is an important tool for maintaining international peace in cyberspace and other domains.

This chapter starts with a discussion of the role played by verifiers in peace. We discuss some of the many types of verifiers, and how those whose roles are outside the formal political process can help to construct peace. Many of these have scientific or investigatory roles whose work informs the state of the world. There are interesting models in aviation, including not only the National Transportation Safety Board (NTSB) but also a variety of others including institutions dedicated to telemetry analysis and near miss analysis. We examine each and suggest how a cyber equivalent could contribute to our understanding of the state of the world and in doing so, support peace.

1 THE NEED FOR VERIFIERS IN CYBERSPACE

A state of peace is a social construction. Peace exists because all parties act as if it exists, but it can be broken or threatened by the actions of any party. As long as peace and a belief in peace exists, it acts as an inhibitor to the initiation of violence,

[*] The authors would like to thank Scott Shackelford, Steve Luczynski, and the participants in the colloquia for providing helpful comments and feedback on the draft of this chapter and Ben Lefkowitz for his research and editorial assistance.

because peace is worth preserving. Parties inhibit their activities to maintain peace (or they act to break it). We take these ideas as axiomatic to allow us to investigate the idea of diverse verifiers and investigate several categories of verifiers whose existence would support the construction of peace. Both the construction and effects of peace have many aspects that are explored elsewhere in this volume, particularly in Chapters 1–3.

War and peace are frequently paired with terms of probability, duration, and time: An impending war, an uneasy truce, a stable peace. If people are uncertain about the existence of peace, if they are uneasy about it, then their willingness to make threats, to plan to carry out threats, and to impose their will on others will be higher. If societies are worried about a lack of peace they will invest in security. Building walls and forts takes substantial resources and takes those resources away from other possible investments. They will invest in arming, training, and maintaining military forces. In times of peace, those investments are reduced. The frames assigned to such things (the "peace dividend," "to maintain peace, prepare for war," and the like) are usually normative and closely relate to the speaker's belief in the stability and longevity of peace. A more widely shared belief that the world is at peace and that peace is stable will free resources for nondefense spending. To the extent that such a belief is accurate and well founded, those other areas of spending will reflect the desires (rather than the fears) of the public. Wide sharing of a belief in peace will be enhanced if many groups with different perceived motives are reporting similar things. Contrariwise, if some sources are reporting signs of war and others are not, there will be disagreement over spending.

A cyber peace dividend might consist of several components, including reduced corporate investment, reduced national investment, and reduced cost to the general public. Today, a widely cited rule of thumb is that commercial entities spend about 8 percent of IT budgets on security (Nash, 2019). Not all of that could be reclaimed by a cyber peace dividend. National investments by governments include both attack and defense. The former is easier to unilaterally reduce. We note, in passing, that the offense budgets are often "black budgets" and hard for outsiders to understand. The cost to the public is a mix of anxiety and the inhibition of productive work because security is hard.

In 2020, there is extreme distrust both across and between societies. The Trump administration announced that the United States would withdraw from the World Health Organization (WHO), a move that the Biden administration reversed. The United Kingdom has withdrawn from the European Union. Many people are refusing to wear masks, refusing to believe in climate change, the list goes on. Attacks on the credibility of news organizations ("it's fake news") augment and bolster other attacks on credibility. In order to overcome this distrust, the world would be better with a series of neutral, trustworthy, and trusted institutions that are less subject to political or market forces and must adhere to strict protocols for verifying the claims of actors in cyberspace. "Governments and diplomats," as Roger Hurwitz

(2012) notes, "… have been less clear in recognizing how foundational public trust is for cyberspace." Similarly, Elinor Ostrom has commented that "trust is the most important resource" (ESCOTET Foundation, 2010). In that spirit, diverse verifiers are the soil in which trust grows.

We look for inspiration to aviation. Among the reasons to look to aviation is that while aviation is inherently risky, deeply technical, and still relatively new, it has evolved into a set of trusted and trustworthy institutions. In addition, other research projects we have done over the last few years have familiarized us with the institutions there, and on consideration they seem to be perhaps both interesting and inspirational.

2 BUILDING OFF OF THE AVIATION MODEL

In other contexts, international mechanisms exist to investigate claims of activity that violate international agreements or norms of behavior. Interpol and the International Criminal Court both investigate allegations of war crimes and human rights violations. The International Atomic Energy Agency (IAEA) investigates violations of the nonproliferation treaty. Given the limitations we note above on applying the arms control model to cyberspace, a better analogy than nuclear site inspections may be international civil aviation. In the domestic context, the authors have separately and collectively promoted the development of cyber incident investigations, modeled on the National Transportation Safety Board's process for investigating aviation incidents and the processes for sharing "near misses" within the aviation community. In concert with the development of national mechanisms for investigating cyber incidents, the international community is also in need of international mechanisms to coordinate and referee international cyber incidents involving multiple states.

For international aviation incidents, the Convention on International Civil Aviation (1994) dictates that the jurisdiction of the crash site will have primary responsibility but allows that jurisdiction to cede authority to a different authority. Such arrangements are managed through the International Civil Aviation Organization (ICAO), the organization established by the convention. In the case of the Malaysian Airlines Flight 17, which was shot down by Russian-backed rebels over Ukraine on July 17, 2014, Ukraine delegated the Netherlands to conduct the investigation given that the flight originated in Amsterdam and had a large number of Dutch citizens onboard (Parker and Olearchyk, 2014). The decision may also have created the perception of improved capability and objectivity by bringing in a third country that was not embroiled in the ongoing conflict to conduct the investigation. In the case of Malaysian Airlines Flight 370, which disappeared over the Indian Ocean on March 8, 2014, Malaysia assembled a Joint Investigative Team of experts from Malaysia, China, the United Kingdom, and the United States, led by an independent investigator under ICAO standards.

In contrast, when international cyber incidents occur, investigations are conducted in an ad hoc manner, usually under the authority of the victim state or by private firms. The findings of such investigations are often the subject of political machinations by the victim company or organization who may wish to avoid negative market reactions for failing to prevent the incident; by the victim's government, which may either seek to downplay or promote the narrative depending on the geopolitical concerns of the moment; and, of course, by the attacker or the attacker's country. In the vast majority of cases, however, no investigative report is ever published. Incident response will be carried out for the purposes of containing an ongoing incident, recovering systems, and preventing future incidents at the victim company. Incident handlers are not, however, in the business of fact finding and reporting so that lessons can be learned and, thus, similar incidents being prevented at other companies.

Some incident handlers generate or contribute to a product labelled "threat intelligence." These "feeds" are often commercial and include the attacker's given names like "Dynamite Panda" (MITRE ATT&CK, 2020). Many times, these products include attribution information, such as "this group uses these tactics," or "the Panda set of attackers are Chinese Government affiliated." The quality of these products have not fared well under scrutiny (Bouwman et al., 2020).

On attributing an attack to a specific state, attribution is also typically carried out in ad hoc manner, as was discussed more fully in Chapter 7. Cybersecurity firms may choose to attribute the incidents they discover, or prevent the actions of specific states, if they see it in their commercial interest, or believe that they have a patriotic duty to do so. More often than not, however, cybersecurity firms will choose to avoid attributing activity to a specific nation state so as not to hurt their commercial prospects in that state, or to avoid becoming a target themselves of that state. When national governments make a claim attributing malicious cyber activity to an adversary state, those claims are typically rebuffed by the accused state and largely ignored by the international community.

3 BACKGROUND: HISTORICAL INCIDENT INVESTIGATIONS

In the United States, investigations of cyber intrusions are typically conducted by private, for-profit cybersecurity firms. In rare cases, when a significant incident occurs, the federal government will investigate and report out on the incident. When the incident involves a federal computing system, such as the incident at the Office of Personnel Management (OPM) in 2015, Congress may investigate. In other cases, Congress asks the Government Accountability Office (GAO) to investigate. These reports are often slow to be produced and can be highly political in nature. While they may provide lessons learned to the cybersecurity community, that is not their primary purpose. Instead, their goal is to assign blame, sometimes in a highly partisan fashion. In the case of the OPM data breach in 2014, the House Oversight and Government Report Committee issued a 241-page report on the incident titled "The OPM Data Breach: How the Government Jeopardized Our National Security for

More than a Generation." While the report provides a comprehensive review of the incident that is valuable from a historical context, its partisan tone undermines its legitimacy as an even-handed fact-finding effort. Its timing, two years after the incident and a month before a hotly contested presidential election, also led to questions about its motivation and purpose.

On the international front, as with the downing of Malaysian Airlines Flight 17 in the air domain, Ukraine has proven to be the focus of significant international conflict within the cyber domain due to the protracted conflict between Russian-backed separatists and the western Ukraine government. Offensive cyber operations that were conducted against electric sector targets caused widespread power outages on two occasions. Ukraine was also the target of the NotPetya malware attack. Given the global spread of NotPetya and international concern over attacks on critical infrastructure, this analysis will focus on the attacks on the power grid. In the first of those incidents (in December of 2015) offensive cyber operators took thirty substations and two power distribution centers offline. The Ukrainian government sought international assistance to investigate the matter. According to reporting by Wired Magazine (Greenberg, 2017), the investigation into the incident was conducted by Ukrainian officials with the assistance of the US Federal Bureau of Investigations and the US Department of Homeland Security. At least two private sector experts were brought in to assist the investigation. They were Robert Lee, a former National Security Agency technical operator and CEO of the industrial control systems security firm Dragos, and Michael Assante, the former chief information security officer (CISO) for the North America Electric Reliability Corporation. Both Lee and Assante were also instructors at the private SANS Institute.

Following the investigation, Lee and Assante published a publicly available report, "Analysis of the Cyber Attack on the Ukrainian Power Grid" (2016), under the auspices of the SANS Institute and the Electricity Information Sharing and Analysis Center (E-ISAC). That report addressed one of the two main purposes for conducting such an investigation, relating to other security professionals what happened so that lessons could be learned to prevent other, similar incidents in the future. It did not, however, address attribution of the attack. The Ukraine government asserted that the attack was carried out by Russia, but no international body validated that claim and the Ukrainian government offered no proof to substantiate the claim. For its part, the US government has never publicly attributed the attack to Russia, but leaks to the media have substantiated the claim (Park et al., 2017).

While the 2015 attack could have been the launching point of an effort to investigate incidents at critical infrastructure and disseminate lessons learned, no such virtuous cycle of process development and ongoing improvement began. When the Ukrainian power grid was attacked a second time, in December of 2016, the incident garnered far less attention. A standout example of dissemination of findings following a cyber incident was the March 2019 breach of Norsk Hydro, a Norwegian aluminum maker. Norsk Hydro made the unprecedented decision to be fully transparent

about the incident, hosting web conferences to disseminate findings to the security community. In this incident, Microsoft's Detection and Response Team led the response and authored the main report on it (Briggs, 2019).[1]

4 INVESTIGATING DOMESTIC INCIDENTS: THE NEED FOR A NATIONAL CYBERSECURITY BOARD

When a major security incident happens, victims are strangely incented to lavish praise on the attackers. After all, there is little shame in being hacked by the pros – "how were we supposed to fight the Russians?" So, some attacks that were performed by criminals or even teenage hackers will be blamed on professionals. If the Acme Company blames the KGB, who is to contradict them? From where do we get our facts? This misattribution is not harmless. The act of blaming the Russians (the Israelis, the Chinese, and the North Koreans) undercuts our assurance of a state of peace.

An investigatory board could help provide those facts. Reports from the NTSB, for example, are seen as authoritative and trustworthy. An investigatory board that invested in gaining and maintaining a reputation for competence could be a substantial counterbalance to organizations spreading self-serving claims. For example, a cyber board could conduct an investigation and release a report that assessed the sophistication displayed by an attacker on a scale from "not sophisticated" to "exceptionally sophisticated." It could assess the idea that an attack was carried out by a nation state or the reliability of a claim that it was a particular nation state.

As this is being drafted, the United States, United Kingdom, and Canada released a joint statement claiming that Russian Intelligence is trying to steal vaccine information (NCSC et al., 2020), but such statements are unusual. The process for releasing intelligence information is opaque. Is the absence of such an announcement the result of peace or a geopolitical decision by intelligence agencies to withhold information?[2] By credibly communicating facts, a cyber board could be a stabilizing force for peace.

4.1 *Why Do We Not Already Have a Cyber NTSB?*

This subsection starts with a brief summary of what the NTSB does, examines some of the objections to a cyber analog, continues with some of the ways those objections might be addressed, and ends with some practical, achievable steps to create a cyber version NTSB. The NTSB is best known for investigating *accidents* in *aviation*.

[1] We do not mean to cast aspersions on Microsoft, but having the creator of the operating system that was attacked may introduce bias.

[2] An intelligence agency might withhold information to protect sources and methods, or to continue an operation to meet additional objectives.

https://doi.org/10.1017/9781108954341.008 Published online by Cambridge University Press

Aviation is a regulated sector. For an airplane to exist (in the United States) requires permission from the FAA; taking off requires a qualified pilot at the controls before leaving an airfield. Each of these is a term of both law and art and, while exceptions exist, these many constraints also act as constraints on the NTSB. An accident is something that leads to the death or injury of someone on a plane, or meaningful damage to one, and these are usually prerequisites to, and provide scope for, an investigation.

The first call we know of for a cyber investigations board was in the 1991 National Research Council report, *Computers at Risk*. Yet no such board exists thirty years later, and the reason, we think, is primarily industry opposition.[3] The core of that opposition is concern. No one wants to have their actions judged with 20/20 hindsight. No one wants to have their innovation judged by those who've never operated a business or been responsible for a profit and loss account. And while such judgements may or may not be real, the perceived threat inhibits the creation of such a board. In contrast, the NTSB was created when accidents in aviation were frequent, and those accidents inhibited the growth of the sector. The aviation industry came together in support of an investigatory body. In contrast, the technology sector seems to be generally opposed. It may be that there is also support, for example, from the insurance industry, but such support has not caused a cyber version of the NTSB to come into existence.

The fear of being judged can be a real problem. An interesting quote from *Roving Mars* (Squyres, 2005 discusses the choice to launch the Mars Exploration Rovers (Spirit and Opportunity)). Before we reach this scene, there was one prelaunch review board after another, examining the engineering choices that had been made:

> Chris Scolese, Ed's deputy, was still in the room, and he explained what had happened. Chris is an engineer, and he has managed space flight projects. What Chris knew is that practically every spacecraft that's ever flown has had some kind of weird problem that popped up once or twice during testing, never to be seen again. You have to take some risks in this business, and the risk we were taking with the transponder was lower in Chris's judgement than the risks we'd already decided we were willing to take on launch day and landing day. Chris had told Ed that he thought we should fly, and Ed had accepted Chris's advice. But it had been a tough call by both of them.

With 20/20 hindsight, Scolese's decision was right, but imagine if the rocket had blown up. Was "you have to take some risks" and "the risks were lower with the transponder" really justifiable? The prospect of such questioning inhibits experimentation and risk-taking. Sometimes that inhibition is appropriate. We would all agree that it is important to have test systems that mirror the production system as

[3] There have been many analogies made to such a system, under a variety of acronyms. For this chapter, we generally will refer to such things as a board, an investigations board, or even a cyber investigations board, using the terms interchangeably with specifics to improve readability.

https://doi.org/10.1017/9781108954341.008 Published online by Cambridge University Press

closely as possible, and to test with those systems, right? Take a moment to think and see if you agree. Sometimes that inhibition is appropriate. That being said, progress requires innovation and experimentation, and blame and second-guessing inhibit such experimentation.

As it turns out, the real world is a strange and complex place. It turns out that companies like Facebook and Netflix have moved to a practice of rolling out changes slowly across subsets of their production systems. This practice is often derisively called "testing in production," which was a shocking strategy when these companies first admitted to it (Mappic, 2011). If those trying it had been worried about an external review board, they might have been prevented from experimenting. Testing in production is now accepted practice; it is considered by some to be a leading approach.

Industry concerns about having their practices judged are strong and real, as is the regular reinvention of the idea. It may be that there are ways to square this circle.

4.2 *Getting to a Cyber NTSB*

To stand up to a cyber incidents investigation board, we must balance the real and perceived concerns with an understanding of the myriad benefits, which include the ability to learn from the misfortune of others and to support the construction of peace. A board does not have to investigate everything to be useful to the cause of peace. The NTSB's role is strictly constrained to accidents involving transportation; thus, a cyber version could be created in a way that aids in peace while addressing corporate concerns.

For example, such a board could initially limit its investigations to breaches involving US Government computers and limit its investigation of more complex incidents to the government computer subset of those cross-entity incidents.[4] As the capability of the organization grows, and as processes mature, the scope could be expanded to other critical infrastructures or other organizations could be created for this purpose. Today, these might be investigated by the FBI, and the attackers might be the subject of surveillance or other operations by intelligence agencies. Each of these agencies has limited resources, and different goals. Managing the overlap of such investigations may carry some complexity. However, this is a reality of complex incidents. For example, the Air Force already imposes such complexity on itself. Accidents are investigated by both a Safety Investigation Board and an Accident Investigation Board, each with different goals (Air Combat Command, 2013).

Another key question area would be the ability of a board to compel participation by either or both an organization and specific staff. Obviously, the participation of the victim organization is important, but to what extent is it expected and

[4] One of our reviewers commented that limiting to "just" US government computers seems quite narrow. We agree, and it would be much broader than what we have today.

https://doi.org/10.1017/9781108954341.008 Published online by Cambridge University Press

reasonable? What about their staff? To what extent should an investigations board be able to compel participation from suppliers to that victim? Would Microsoft, Google, and Amazon need staff dedicated to answering the board when their products are involved in a breach? Would investigators be limited to "what's in the manual" or can they delve into product design decisions?[5] Even with regard to the manual, it is not always obvious what section of a complex product's technical documentation is relevant. The two volumes of the latest edition of "Windows Internals" (Yosifovich et al., 2017) comprise 1,568 pages, and those are books. The more voluminous technical documentation is now largely online and updated frequently. What is a reasonable expectation of an operator of such systems? These questions are not insurmountable, but some versions of them need to be addressed to move proposals forward.

What about the participation of staff? Can that be compelled? What about the right against self-incrimination? As we write this, Uber's former Chief Security Officer has just been charged with obstruction of justice. What are the expectations for staff of a breached organization in terms of participation in an investigation? Is it "answer three questions by email" or "be deposed for a day or more?" How are software development staff to be trained, and whose staff would receive training? For example, the Air Force delivers annual training to pilots on the various investigations that will happen after an accident.

4.3 *What Could a Cyber NTSB Do for Peace?*

Calls for a cyber investigations board have traditionally focused on learning and disseminating lessons from incidents. This is inherently useful in the creation and preservation of cyber peace because it makes future attacks more difficult. And there are many other ways in which a board could support the cause of peace, including the following:

- Publishing lessons learned reports (as opposed to sharing them under NDAs)
- Bring different goals to incident investigation
- Investigating more/different cases than police or intelligence agencies
- Provide attribution with different biases
- Report on the state of the world
- Provide international assistance
- Support a construction of peace

The primary reason for previous calls for a cyber investigations board has been to find and distribute lessons. The incredible safety record of aviation is commonly attributed to these and other learning systems. An investigations board could

[5] Even suggesting this discomforts the author, Shostack. Having each of the product tradeoffs judged raises issues discussed elsewhere.

https://doi.org/10.1017/9781108954341.008 Published online by Cambridge University Press

establish consistency and credibility, and stand in complement to the information released by police and prosecutors. That information is focused on literally "making the case" for prosecution and conviction, rather than learning lessons or informing. Analysis that is designed to be objective could better support peace by informing debate about the state of the world. It could potentially do so in a larger set of cases if the investigators are not required to testify, be subjected to cross-examination, and perform other tasks in the judicial system. The cases that a board investigates might be quite different than the ones that the police investigate. (There would need to be a deconfliction/equites process to ensure that investigations did not accidentally cross paths with other investigations. That process, like all the others, requires training for the involved participants.)

A board could provide attribution information about cases with a different authority than either private or prosecutorial analysis. Such analysis might be read with less skepticism or read with different skepticism by different parties, providing information that either supports or undercuts the construction of peace through a better understanding of the state of the world.

In addition to information about specific attacks, additional high-quality information about the frequency and intensity of international attacks would illustrate the state of the world at a given time and add information about the actors who are violating the peace, increasing the likelihood that they would be either caught[6] or meaningfully made to take the blame for their actions.

The NTSB provides help and assistance to air crash investigations around the world. It would not be unreasonable to expect that once a board had established itself and its competence, it could, when asked, help investigate "important incidents" outside of the federal government, including state and local governments, as well as, perhaps, private enterprises. This assistance to entities within national borders could raise the cost of attacks via exposure. International assistance could be an act of goodwill, bolstering peace.

Additionally, a stream of analytic reports that establish norms and expectations would inform industry's position on the impact of investigations. While it is reasonable to think that more data would aid in the understanding of the state of the world as was described in Chapter 3, it is similarly reasonable to think that most industry benefits from peace and trade.

5 A SYSTEM FOR REPORTING NEAR MISSES

The NTSB is the best known of a polycentric constellation of aviation safety programs which complement and overlap to make hurtling through the air at hundreds of miles per hour incredibly safe. There are others including the Aviation

[6] Methodological analysis of incidents might cause attacks that had been attributed to criminals to be correctly attributed to state actors, or vice versa.

https://doi.org/10.1017/9781108954341.008 Published online by Cambridge University Press

Safety Reporting System (ASRS) and the Aviation Safety Information Analysis and Sharing System (ASIAS). One of the authors (Shostack) has argued at length for a Cyber Security Reporting System (CSRS),[7] and we believe that such a system could also enhance and preserve peace (Bair et al., 2017). Before discussing near misses at some length, we will first briefly explain the ASIAS system, and some of the limits an ASIAS analog would face. This helps illustrate the value of an ASRS-like system.

5.1 *ASIAS: Telemetry Analysis*

The ASIAS program collects telemetry from aircraft in operations, analyzes it, and reports back to the operators. For example, if flights operated by one airline have substantially different wing flutter than those operated by other airlines from that same airfield, then that might be interesting for each airline to know. Our ability to compare telemetry is built on a scaffolding of similarities. Aircraft and their components are made by a small number of manufacturers. The operational systems are defined by flights of a limited number of types (general, cargo, and military) from one field to another. This leads to similarity between the telemetry each emits. Computer systems run a far more varied set of workloads. A mail server might run on Windows, Linux (Ubuntu, Debian, RedHat, etc.), FreeBSD, OpenBSD, or others (McKusick et al., 1996). The mail software might be sendmail, postfix, qmail, or Exchange, or even Gmail or Hotmail, which are (reputedly) unique software. Each of these operating systems and mail packages logs differently. Similarly, there is diversity in each "stack" of software, and that software delivers diverse values.

Despite this diversity, aggregated analysis of attacks could produce useful information. For example, if logs of rejected emails were collected, then we could learn about spam campaigns. There is a difference between mail from northeastem.com going to northeastern.edu and it going to shostack.org. On first blush, the former is much more likely to be a targeted campaign, and the latter to indicate a broad spamming campaign. But if we gathered rejection data from many recipients about email domains, we could tell recipients about the unusual campaigns they receive. Unusual might be determined algorithmically based on those whose sending domains are unusual, and there are standard computer science techniques that would help determine what counts as unusual relative to each recipient.[8] The data sent back to participants could motivate their participation, and the agency performing the analysis could provide information about the state of conflict in the world and possibly between states and semi-state and nonstate actors.

[7] Since there are fewer calls for such a thing, we will use the CSRS acronym.

[8] There are standard techniques that could be applied, for instance, term frequency/inverse document frequency, or "small edit distance."

https://doi.org/10.1017/9781108954341.008 Published online by Cambridge University Press

5.2 ASRS: Near Miss Reporting

We believe we can develop broader, and perhaps less expected lessons, from a cyber version of ASRS. In aviation, if there is an incident, then anyone involved can submit a short, two-page form to the ASRS, operated by NASA.[9] An incident is anything short of an accident, which, again, is the death or injury of a person or damage to an aircraft. The reports go to NASA to isolate them from accidental disclosure to the regulators. (There are important additional protections in both law and agreements between NASA and the FAA.) NASA ingests the reports, analyzes them, and publishes data that are carefully anonymized.[10] NASA also sends back a receipt. The reporter can use that receipt to demonstrate "evidence of constructive engagement" in a disciplinary proceeding. This evidence is one of the factors that the FAA takes into account in its administrative law proceedings. This incentive, which might seem small, adds to each participant's desire for a safe aviation system and is enough to motivate roughly 100,000 reports each year to the ASRS (ASRS, 2019).

5.3 Cyber Near Misses and What We Might Learn

Near miss reporting, both within and between organizations, is an important building block in safety programs in a great many industries. Similarly, many of these programs use blamelessness as a tool to demonstrate their prioritization of learning over retribution.

The nature of near misses in cybersecurity makes them easier to report and discuss, and that eases open doorways to understanding the state of the world. The sorts of things we might understand include (but are not limited to) attacks that progress too close to a meaningful target or attacks that gain the interest of investigators for their distinctiveness. In doing so, near-miss reporting makes more measurable what is commonplace and effective, such as phishing and the techniques in use. These are nominally reported on, but what's almost working can be lost in the noise.

We can learn useful things about what works to protect, detect, and respond to problems by tracking which tools are reliably reported for each. Such analysis can be broad and helps us to better preserve peace by prioritizing effective defenses. For example, while the NIST CSF contains over 900 controls (Reciprocity Labs, 2019),[11] the Australian Signals Directorate recommended a "top 4," now transformed into an "essential eight" (Coyne, 2017).[12] Even if we believe that the controls in each set

[9] The form can be found at https://asrs.arc.nasa.gov/report/electronic.html

[10] The anonymization has both a technical component and a review component.

[11] The NIST CSF is the National Institute for Standards and Technology's Cyber Security Framework, one of the primary ways the United States specifies the cybersecurity defenses (controls) that organizations are expected to deploy and maintain.

[12] If the Australians double their list every three years, it will still take till roughly 2042 before they're closing in on 900 controls.

https://doi.org/10.1017/9781108954341.008 Published online by Cambridge University Press

are at different levels of abstraction, and thus each of the eight represents a dozen in the NIST set, there remains a massive difference in the control recommendations. Either one of these standards is missing crucial controls, or the other standard includes investments that do not do very much good.[13] Knowing what does not work can be an important step forward. Stopping ineffective investments makes room for new ones. So, both positive and negative reports can be useful. A mix allows for interesting science: Why does measure A work for some organizations but not others?

5.4 *The Contribution of a CSRS to Cyber Peace*

The first contribution of a CSRS to peace would be the ability to improve defenses or to reduce costs without reducing the quality of defenses. The former makes attacks harder, and the latter allows us to invest in other things. Today in cyber warfare, the attacker has tremendous advantages. Improving the effectiveness of defenses would shift the balance somewhat. Making attacks more difficult, more likely to be detected, or more attributable would shift the logic against launching attacks and thus contribute to peace.

The second contribution could be an assessment of attacker activity. If a CSRS-adjacent body had access to confidential descriptions of "tactics, techniques, and procedures," then it could analyze near miss information to report on rates of attacks or attack intensity.[14] This would be a very different function than aviation's ASRS, but streams of near miss information in cybersecurity could be leveraged for this. Such variation may cause problems for multinational companies reporting to local authorities.

6 AN INTERNATIONAL MECHANISM TO INVESTIGATE AND ATTRIBUTE CYBER INCIDENTS

Building off of the ICAO model, what is needed in the international context is a mechanism for requesting international support for investigating significant cyber incidents. These investigations would be carried out for a dual purpose. First, they would provide a standard process and rapid timeline for disseminating findings useful to cyber defenders. Second, they would provide a means for determining attribution and releasing such findings to the public, allowing other international bodies

[13] There is another possibility, which is that they are aiming at different levels of security, but since we have no measure of what that means, we exclude it.

[14] TTPs and "indicators of compromise" are things such as domains used by attackers, email subjects, IP addresses, and malware identifiers. They are useful for detecting and grouping attacker behavior. They are often kept close to the vest to prevent attackers from becoming aware that defenders are using them. Collective reporting of an analysis might be easier to report on than specific comments like "the Acme corp managed an attack by the Drunken Bear APT group."

to censure or penalize the offending state. These findings could also serve as the basis for organizing coalitions of governments to sanction or otherwise condemn the actions of the offending state should international institutions fail to act.

At this stage, rather than funding a standalone organization to investigate international cyber incidents, a more modest approach would be to establish a concept of operations for how such investigations should take place and who should take part in them. As in the successful example of the 2015 Ukraine investigation, such investigations will need to rely heavily on private sector expertise. Particularly in the area of industrial control systems, expertise on the security and forensic methods for such systems is exceedingly rare. Thus, keeping experts with the knowledge to carry out these investigations on the sidelines while waiting for the phone to ring would not be practical. Instead, ad hoc teams should be formed at the behest of the victim state. These teams would be invited to investigate and issue initial findings in a rapid fashion, followed by a comprehensive final report issued by the international body sponsoring the effort.

Some of these functions might be picked up by a "Cyber Peace Corps," as discussed elsewhere in this volume, including in the essays section. But such a group, with room for everyone, carries a different function and requires a different culture from an organization with strong leadership focus on producing investigative reports. A Peace Corp could be a feeder to such an investigative body, helping to respond to problems, preserving evidence, and bringing forward interesting cases.

On determining attribution, significant conclusions can typically be achieved by comparing the tradecraft of the attacker to other known historic incidents. This process has led ESET (2016) and Dragos (2017), among others, to conclude that the team behind the Ukraine attacks was the same team behind the attacks on the Democratic National Convention and other political targets in the lead up to the 2016 US presidential election. Thus, without the benefit of national intelligence capabilities, investigators should be able to make preliminary conclusions on attribution. Intelligence agencies could then provide their own findings to the team, agreeing to release some, all, or none of the evidence uncovered through intelligence collection to the public. This process would allow for sources and methods to largely be protected, while providing an independent verification mechanism of the claims.

CONCLUSION

In this chapter, we have argued that trusted verifiers are essential for cyber peace. By creating trusted national mechanisms for investigating cyber incidents, lessons learned can be shared with the wider community and confidence that problems that caused one incident can be corrected elsewhere before more such incidents occur. By creating trusted verifiers for near misses, all members of the cybersecurity community can provide telemetry to determine the current level of

https://doi.org/10.1017/9781108954341.008 Published online by Cambridge University Press

hostility in cyberspace. With a strong international mechanism for investigating significant cross border cybercrime, determining lessons learned, and attributing malicious activity, more consequences can be created for states that engage in such activity. As norms of conduct in cyberspace are developed, it is essential that verifiers are enabled at multiple levels to ensure that they are being upheld, and when they are not to verify that claims of malfeasance are proved true and taken seriously. Trust but verify is, now more than ever, essential to the preservation of peace.[15]

BIBLIOGRAPHY

Air Combat Command. (2013, January 11). *Air Force Safety* and *Accident Board Investigations*. www.acc.af.mil/About-Us/Fact-Sheets/Display/Article/199117/air-force-safety-and-accident-board-investigations/

Aviation Safety Reporting System. (2019, July). *ASRS Program* Briefing. National Aeronautics and Space Administration. https://asrs.arc.nasa.gov/docs/ASRS_ProgramBriefing.pdf

Bair, J., Bellovin, S. M., Manley, A., Reid, B., & Shostack, A. (2017). That was close: Reward reporting of cybersecurity near misses. *Colo. Tech. LJ*, *16*, 327.

Bouwman, X., Griffioen, H., Egbers, J., Doerr, C., Klievink, B., & van Eeten, M. (2020). A different cup of {TI}? The added value of commercial threat intelligence. In 29th {USENIX} Security Symposium ({USENIX} Security 20) (pp. 433–450).

Briggs, B. (2019, December 16). *Hackers hit Norsk Hydro with ransomware. The company responded with transparency.* Microsoft. https://news.microsoft.com/transform/hackers-hit-norsk-hydro-ransomware-company-responded-transparency/

Cherepanov, A., & Lipovsky, R. (2017, June 12). *Industroyer: Biggest threat to industrial control systems since Stuxnet*. WeLiveSecurity. www.welivesecurity.com/2017/06/12/industroyer-biggest-threat-industrial-control-systems-since-stuxnet/

Committee on Oversight and Government Reform. (2016, September 7). *The OPM Data Breach: How the Government Jeopardized Our National Security for More than a Generation*. U.S. House of Representatives. https://republicans-oversight.house.gov/wp-content/uploads/2016/09/The-OPM-Data-Breach-How-the-Government-Jeopardized-Our-National-Security-for-More-than-a-Generation.pdf

Convention on International Civil Aviation. (1994, December 7). International Civil Aviation Organization (ICAO), U.N. Document 7300. www.icao.int/publications/pages/doc7300.aspx

Coyne, A. (2017, February 6). *Overhaul of ASD's Top 4 cyber threat strategies*. itnews. www.itnews.com.au/news/drastic-overhaul-of-asds-top-4-cyber-threat-stategies-449787

Dragos. (2017). *CRASHOVERRIDE: Analysis of the Threat to Electric Grid Operations*. www.dragos.com/wp-content/uploads/CrashOverride-01.pdf

E-ISAC. (2016). Analysis of the cyber attack on the Ukrainian power grid. https://ics.sans.org/media/E-ISAC_SANS_Ukraine_DUC_5.pdf

ESCOTET Foundation. (Fall of 2010, reproduced fall of 2020). *Interview with Nobel Laureate Elinor Ostrom*. https://escotet.org/2010/11/interview-with-nobel-laureate-elinor-ostrom/

[15] Since this chapter was written, the authors have released a technically focused report on the subject of learning systems: Robert Knake Adam Shostack Tarah Wheeler, Learning from Cyber Incidents: Adapting Aviation Safety Models to Cybersecurity, Belfer Center for Science and International Affairs, Harvard Kennedy School, November 12, 2021, www.belfercenter.org/learning-cyber-incidents

https://doi.org/10.1017/9781108954341.008 Published online by Cambridge University Press

Greenberg, A. (2017, June 20). How an entire nation became Russia's test lab for cyberwar. *Wired*. www.wired.com/story/russian-hackers-attack-ukraine/

Hurwitz, R. (2012). Depleted trust in the cyber commons. *Strategic Studies Quarterly*, 6(3), 20–45. www.airuniversity.af.edu/Portals/10/SSQ/documents/Volume-06_Issue-3/Fall12.pdf

Lemos, R. (2020, August 14). Research casts doubt on value of threat intel feeds. *Dark Reading*.www.darkreading.com/threat-intelligence/research-casts-doubt-on-value-of-threat-intelfeeds/d/d-id/1338676

Mappic, S. (2011, December 6). Why testing in production isn't as stupid as it sounds. Appdynamics blog, www.appdynamics.com/blog/product/why-testing-in-production-isnt-as-stupid-as-it-sounds/

Maybaum, M., & Tölle, J. (2016, May). Arms control in cyberspace-architecture for a trust-based implementation framework based on conventional arms control methods. In 2016 8th International Conference on Cyber Conflict (CyCon) (pp. 159–173). IEEE.

McKusick, M. K., Bostic, K., Karels, M. J., & Quarterman, J. S. (1996). *The design and implementation of the 4.4 BSD operating system (Vol. 2)*. Addison-Wesley.

MITRE ATT&CK. (2020, March 30). *APT18*. https://attack.mitre.org/groups/G0026/

Nash, K. (2019, December 30). Tech chiefs plan to boost cybersecurity spending. *The Wall Street Journal*. Retrieved from www.wsj.com/articles/tech-chiefs-plan-to-boost-cybersecurity-spending-11577701802

National Cyber Security Centre (U.K.), Communications Security Establishment (Canada), & National Security Agency (U.S.A.). (2020). Advisory: APT29 targets COVID-19 vaccine development. United States Department of Defense. https://media.defense.gov/2020/Jul/16/2002457639/-1/-1/0/NCSC_APT29_ADVISORY-QUAD-OFFICIAL-20200709-1810.PDF

National Research Council. (1991). *Computers at risk: Safe computing in the information age*. National Academy Press.

Park, D., Summers, J., & Walstrom, M. (2017, October 11). *Cyberattack on Critical Infrastructure: Russia and the Ukrainian Power Grid Attacks*. The Henry M. Jackson School of International Studies: University of Washington. https://jsis.washington.edu/news/cyberattack-critical-infrastructure-russia-ukrainian-power-grid-attacks/

Parker, A., & Olearchyk, R. (2014, July 21). Netherlands to lead MH17 investigation. *The Financial Times*. Retrieved from www.ft.com/content/19c29f34-10e1-11e4-b116-00144feabdc0

Reciprocity Labs. (2019, December 10). *What Are NIST Controls and How Many Are There?* https://reciprocitylabs.com/resources/what-are-nist-controls-and-how-many-are-there/

Squyres, S. (2005). *Roving mars: Spirit, opportunity, and the exploration of the red planet*. Hachette Books.

Treaty on the Non-Proliferation of Nuclear Weapons, 1970.

Yosifovich, P., Solomon, D. A., & Ionescu, A. (2017). *Windows internals, part 1: System architecture, processes, threads, memory management, and more*. Microsoft Press.