2018 10th International Conference on Cyber Conflict
CyCon X: Maximising Effects
T. Minárik, R. Jakschis, L. Lindström (Eds.)
2018 © NATO CCD COE Publications, Tallinn

Drawing Inferences from Cyber Espionage

Martin C. Libicki
Center for Cyber Security Studies
U.S. Naval Academy
Annapolis, MD, United States
libicki@usna.edu; libmazo@gmail.com

Abstract: To survive a confrontation, it helps to understand other side's capabilities and intensions. Estimates of opposing capabilities rest on an empirical basis but understanding the other side's intentions is inferred from words and deeds.

Therein lies a dilemma common across all military domains: acts to alter the balance of a confrontation can also shape the inferences that the other side draws about one's intentions. The dilemma also operates in cyberspace, but in unique ways.

First, efforts by one side to acquire information on the other can be read by the other side as preparations for a cyber attack prefatory to a military attack.
Second, others may draw inferences from the fact of cyber espionage alone, even though the basis for believing in a cyber security dilemma is weak.
Third, there are ways of carrying out cyber espionage that can mitigate inferences that others draw about the imminence of cyber attack by, for example, limiting which components within a network are targeted for eavesdropping or by using penetration methods that do not leave arbitrary code behind.
Fourth, defenders themselves can also modulate their reactions in ways that limit drawing unnecessary inferences.
Fifth, expectations of how well modulating cyber espionage can convey peaceful intentions should be very modest.

All these are complicated by difficulties in the target's ascertaining a penetration's date, characterization, and authorship. We conclude with a call for those who would penetrate military-related systems to think about the inferences that the other side may draw if such penetrations are discovered.

Keywords: *cyber espionage, cyber attack, signaling*

1. INTRODUCTION

To survive a confrontation, it helps to be mindful of the other side's capabilities and intentions. Estimates of opposing capabilities often take painstaking work, but at least rest on an empirical basis. But understanding the other side's intentions is something that needs to be inferred from words and deeds.[1]

Therein lies a dilemma present across conflict domains. Acting can alter the terms of a confrontation to the actor's advantage, but it can also shape the inferences that the other side draws about one's intentions. Some inferences can both help *and* harm. One side may want to signal its resolve to attain and defend some objective. It does so by demonstrating capability, readiness, and a willingness to put people and assets in harm's way. It hopes that the other side backs off. But the other side may infer not only that its potential foe is prepared and willing, but also that it is facing a now higher level of aggression. Perhaps the objective has grown or the willingness to take risks to achieve it has risen. So, the other side sees a growing threat – one that forces it to do something to recover its former level of security. Therefore, it decides to bolster its own capability, readiness, and willingness to fight.[2] The advantages that one side reaps from its actions can be thereby nullified by the indirect disadvantages because the other side is drawing unhelpful inferences about its adversary's intensions.

We contend that the dilemma operates in cyberspace, but in a unique way – efforts by one side to acquire information on the other can be read by that other side as preparations for a cyber attack prefatory to a military attack.[3] It hardly helps stability when the high degree of ambiguity present in cyberspace combines with the thin experience base of cyber attacks and its non-physical (hence non-intuitive) nature. Perhaps needless to add, what happens in cyberspace matters to conventional military affairs more than it used to.

This essay walks through the problems and issues that may arise when inferences are drawn from activity in cyberspace, particularly those that take place during a crisis or confrontation. One might imagine, for reference purposes, that China and the United States are at odds over the South China Sea; neither is certain what the other side wants or how far it is willing to go, even if each has a good idea of what physical assets are to hand. So, what considerations should go into each side's rules of engagement in cyberspace?

[1] The classic treatment being Robert Jervis, *Perception and Misperception in International Politics*, Princeton NJ (Princeton University Press), 1978.

[2] Elsa Kania, "Cyber Deterrence in Times of Cyber Anarchy: Evaluating the Divergences in U.S. and Chinese Strategic Thinking," November 11, 2016; unpublished paper.

[3] The logic that links a cyber attack to a kinetic attack is that because many of the effects of a cyber attack are temporary and reversible, carrying one out is pointless unless the intent is to exploit a temporary interruption or degradation of the other side's information services by using kinetic forces to make permanent changes in the military balance or outcomes.

In addressing this question, this paper distinguishes cyber espionage, which is unauthorized access to systems in order to acquire information, from cyber attack, which entails accessing systems in order to disrupt their operations or corrupt their information. To put this in the language of the CIA triad: cyber espionage affects only confidentiality while cyber attacks affect integrity and availability. Unfortunately, popular use generally applies "cyber attack" to a broad array of mischief in cyberspace, including the manipulation of social media. Cyber attack, in this paper, is also distinguished from "attack," which is used to mean kinetic attack using physical force.

2. INFERRING CYBER ATTACKS FROM CYBER ESPIONAGE

Cyber espionage can create knowledge *and* help set up cyber attacks; yet, if discovered, it may alter the target's assessment of the intruder's capabilities and intentions. The first is generally helpful. The second is generally harmful, in that the target may conclude that the intruder is preparing to fight and to do so soon.

Although caution is therefore advised in penetrating systems whose disturbance may enflame the other side, in a crisis a country may want to carry out *more* cyber espionage in order to determine the status, readiness, and intentions of the other side's armed forces. Indeed, as with spy satellites in the 1960s whose imagery persuaded U.S. leaders not to panic over the size of Soviet ICBM arsenals, or as former British intelligence officials would argue,[4] better intelligence tends to foster stability. It substitutes fact for doubt in situations in which leaders believe they must assume the worst, and hence gird for conflict. Some risk is inescapable. Even if traditional espionage uses tools clearly different from those used in warfighting, the heightened effort to collect intelligence prefatory to bolstering defense is nearly indistinguishable from efforts to collect intelligence prefatory to offense. Thus, any discovery of heightened intelligence efforts may lead the target to react badly.

Moreover, because a malware implant designed for cyber espionage is often identical to one designed for cyber attack, discovering and attributing[5] one in a critical system could easily be viewed as a *direct* precursor to attack. This normally would lead the target to raise its alert levels, which, in and of itself may exacerbate tensions.[6] In a crisis, not only are alert levels high to begin with, but so is suspicion of the other side's motives.

[4] Based on remarks by Nigel Inkster (personal communications) and Sir David Osmand (http://carnegieendowment.org/2017/03/20/concurrent-session-i-cyber-weapons-and-strategic-stability-pub-67884).

[5] Although attribution can be uncertain, the paper focusses on two countries in a confrontation at the time of discovery. Thus, the target is probably more apt to blame the intrusion on the other side (because it is easier to impute a motive) than if there were no confrontation.

[6] Paul Bracken described how ominous signs could make the other side raise its alert level in his "Strategic War Termination," in Ashton B. Carter, John D. Steinbruner, and Charles A. Zraket, eds., Managing Nuclear Operations (Washington, D.C.: The Brookings Institution, 1987), pp. 197–214.

One important facet in drawing inferences from an implant is that its implantation would reflect conditions true at the time of its implantation rather than at the time of its discovery. Good forensic teams working on well-monitored networks can often figure out when an intrusion took place, and hence shed light on why.[7] If the penetration predated the crisis, it may be deemed not to be part of a dynamic of escalating alert levels. Nothing, of course, prevents one country from implanting malware against the day it might be needed for attack, but discovery alone cannot support the supposition that any such attack will take place imminently.

However, because many countries lack access to good forensics or fail to monitor their networks assiduously, the age of the intrusion may not be obvious to *them*. And until the other side figures out *when* the first penetration that resulted in a system's compromise took place, it may, in fear, conclude that the penetration was recent enough to have been motivated by the crisis itself.

The target need not be not forced into one conclusion. Perhaps what looks like cyber espionage was just fact-finding. Yet even cyber espionage unrelated to any possible cyber attack is not necessarily innocent. If the compromised system tracks military units in real time, an implant into it is still cyber espionage, but can also be used for later adversary targeting. Discovering that such a system was compromised regardless of how long ago, *should* raise concerns, just not ones that require going onto a war footing.

Now, what if the target infers that the intrusion was meant to be seen?[8] Granted it is difficult to distinguish between: (1) the desire to be seen; (2) an indifference to being seen which leads to a relaxation of operational security, thereby raising the likelihood of being seen; and (3) simple bad luck on the intruder's part. The target, in drawing inferences from what it has discovered, may also forget that the characteristics of discovered intrusions are not necessarily characteristics of undiscovered ones.[9]

[7] The fact, for instance, that intrusions against the DNC started in the summer of 2015 strongly suggests that their motivation was more anti-Clinton than pro-Trump, whose nomination was hardly assured at that point.

[8] The DNC had been penetrated for roughly a year before discovery (Dmitri Alperovitch, "Bears in the Midst: Intrusion into the Democratic National Committee", June 15, 2016: At the DNC, COZY BEAR intrusion had been identified going back to summer of 2015, while FANCY BEAR separately breached the network in April 2016). Yet the FBI still argued, "The most startling exchange at this week's hearing involved questions about why Russian hackers were so indiscreet when they stole e-mails from the Democratic National Committee and from the head of the Clinton campaign. That 'loudness' looks deliberate, Mr Comey replied." (source: "The FBI says it is investigating the president's campaign," March 23, 2017; http://www.economist.com/news/united-states/21719491-slice-country-hears-president-victim-government). See also Julian Borger, "Trump-Russia collusion is being investigated by FBI, Comey confirms", March 20, 2017; https://www.theguardian.com/us-news/2017/mar/20/fbi-director-comey-confirms-investigation-trump-russia: "The Russian intervention in the election was 'unusually loud', as if Moscow did not care about being caught."

[9] Presumably, intrusions that are discovered are those that are easiest to discover. Their discoverability may not characterize the discoverability of the average intrusion (unless all of them are eventually discovered).

Still, the target's perception that the intruder was brandishing its capabilities by allowing its implants to be discovered – when spies normally go to great lengths to hide *theirs* – may persuade it to see coercion taking place. It could then ask: for what purpose? And why now? This could have been a periodic reminder and hence not indicative of an imminent threat. Logically, it should not indicate an imminent attack, since the attacker should be at pains to mask its intentions until they are suddenly revealed. But it could be a warning to back down, by containing the implicit message that failure to do so would be dangerous.

Another complicating factor with cyberspace operations arises from the question: how can countries underscore the credibility of deterrence instruments (such as retaliatory cyber attacks) without revealing the particulars of such capabilities and thereby inducing countermeasures?[10] Because countermeasures do not emerge immediately when systems prove broadly vulnerable, the target may infer that the other side is signaling its urgency by revealing what it can do *and* that it will not be needing such capabilities for long. If the target concludes from the intruder's presumed willingness to burn exploits that the intruder needed to make a quick impression, the target may then ask what the occasion is or will be.

The target may also conclude that the intrusion was undertaken to test the efficacy of and reaction to a cyber attack to be launched at some later date. This conclusion would be reinforced if it was a cyber attack, albeit a small one, that had taken place. Evidence for that may include the location of the intrusion, the identity of the affected systems, or the presence of attack code within the implant. Its placement or characteristics may persuade the target that the attacker has little confidence of being able to access the implant once the system goes to war.[11] But even such a discovery would not be particularly good evidence of an imminent attack, especially if the characteristics of the implant suggested the attacker's confidence that it could persist indefinitely without discovery.

Conversely, if the target concludes that a nominal cyber attack was carried out primarily as a final test prior to deployment, it may expect that use to be imminent. Its fears may rise if the implant's placement, characteristics and, especially, its implantation date suggest that the attacker was risking a high likelihood of discovery to validate or characterize a particular type of cyber attack. It is but a short step for the target to infer that discovery is evidence of discoverability, and thereby conclude that detonation is coming sooner rather than later. Further evidence of imminent use may be an implant's fragility, in that it is not robust against the run of changes that systems undergo. Other indications are recent rises in the frequency or scale of communications between the

[10] See, for instance, Austin Long, Brendan Green, "Clandestine Capabilities and Deterrence in World Politics", unpublished.

[11] This raises the question of how to activate the cyber attack if the implant is unreachable, but the answer may be that activation – a one-bit decision – can be triggered on the malware's assessment of network events in cases where malware cannot build attack code on the fly.

implant and its controller, or tests of the ability of the implant to support a certain payload. The latter can sometimes be inferred from reading logs.

Finally, any particular intrusion may serve several purposes. Concluding that one purpose may have been relatively benign hardly proves that more malign purposes are absent.

3. INFERENCES FROM THE FACT OF CYBER ESPIONAGE ALONE

A country's reaction to having simply been spied on may reflect its take on the security dilemma. Countries that believe that someone else's gain is automatically their loss are apt to interpret intrusions more darkly than those that believe that both sides can simultaneously be more secure. Those inclined to believe that the other is implacably hostile will read events as proof of dark design; those inclined to impute a mix of motives to the other side will hold many differing interpretations and delay imputing malevolence to system intruders pending further evidence. Some will see Munich in 1938; others, Sarajevo in 1914. The usual caveats apply: countries with different political cultures may draw inferences differently, the various bureaucracies within a single country may disagree with one another, and members of the public, elite opinion, and private organizations may each have their own opinion.

Furthermore, what seems innocent after the crisis has passed may seem otherwise during the crisis. The human tendency to impute intent to random circumstance may lead to conclusions that *because* the discovery of implants happened to produce fear, they were meant to induce fear and their discovery was part of that plan.

That noted, the technical basis for imagining a security dilemma *in cyberspace* is weak, particularly compared to contests such as nuclear missiles versus nuclear missiles or WWI-era land forces versus similar land forces. There are several reasons why. *First*, the contest in cyberspace is asymmetric: the best measures against cyber attack are cyber defenses, not an opposing cyber attack capability used for counterforce purposes.[12] Most measures that increase defense do not allow one's own attackers to enjoy greater success.[13] *Second*, because the element of surprise is intrinsic to the

[12] In other words, the cost-effectiveness of carrying out cyber attacks on the attackers themselves would be low, in large part because the primary assets used in cyber attacks, computer code and intelligence, are essentially indestructible, and the hardware used is easily replaced. This consideration has nothing to do with the relative cost-effectiveness of offense versus defense, or with deterrence in cyberspace.

[13] Ben Buchanan (in *The Cybersecurity Dilemma: Hacking, Trust and Fear Between Nations*, Oxford 2017) has argued that NSA intrusions have provided information on adversary intrusion (and hence attack) capabilities that have permitted stronger defenses. Thus, stronger defenses *by potential attackers* against penetration would have yielded weaker defenses on the part of defenders allied with the penetrators. But even if true, information is available only on some actors not all, such information is only part of what it takes for defense, and networks that benefit from NSA-acquired information are only a fraction of the total networks in the United States (albeit perhaps disproportionately important ones).

success of a cyber attack, it would take great confidence in such defenses before one side is sufficiently emboldened by the prospect of impunity to launch its own cyber attacks. *Third*, even if all system defenses were perfect, the logic that in cyberspace impunity emboldens aggression must also presume that the other side will not escalate into physical combat. This presumption is valid only if the stakes involved are too small to merit violence. *Fourth*, the strong commercial consensus on the need for better cyber security in general means that actions that improve cyber security for one (e.g., the discovery of a vulnerability that leads to a patch, an improved understanding of cost-effective practices) often improve cyber security for all.

Cyber espionage, like espionage in general, also permits information to be transmitted in particularly credible ways. If one side in a confrontation were to aver that it lacked active planning for aggression, the other side may well dismiss its avowals as motivated. But if one were to *steal* corroborating information from potential foes, one would have to be very suspicious indeed to conclude that such information was deliberately planted there, particularly if finding it was hard.

Such deception *could* happen,[14] but carrying on ostensibly confidential communications under the assumption they were wiretapped and would therefore be transmitted to the other side's leadership requires either giving up all confidential channels or knowing in advance which channels would stay confidential and which would be penetrated. The same holds with even more weight if the deception involved physical evidence, such as the disposition of military forces. Thus, however irritated one side may be at being penetrated, a salve on this irritation is the presumption that one's peaceful intentions have been more credibly communicated than mere narrative would allow.

4. HOW TO KEEP ON WITH CYBER ESPIONAGE WITHOUT SO MUCH RISK

How might cyberspace spies suppress unhelpful inference-making? One way is to loosen the correlation between being spied on and being attacked. Presumably, countries will not credibly promise never to attack in cyberspace; doing so forgoes a potentially significant military advantage and anyway would not be believed. Nevertheless, the correlation between espionage and attack *can* be weakened by copious acts of cyber espionage *not* correlated with a cyber attack. But this may backfire if the other side thinks that this is being done deliberately – that is, to inhibit the target from raising its guard after discovering intrusions that really were prefatory to cyber attack. Besides, being caught spying a lot tends to make one look unfriendly to begin with.

[14] A great deal depends on how widely system owners start using deception. One case is France's then-candidate Emmanuel Macron suspecting that Russia would penetrate his campaign's networks and lacing false documents in his networks. See Adam Nossiter, David Sanger, and Nicole Perlroth, "Hackers Came but the French were Prepared," May 9, 2017; https://www.nytimes.com/2017/05/09/world/europe/hackers-came-but-the-french-were-prepared.html.

Another possible way to reduce the risk is to ensure that one's cyber espionage implants lack the characteristics that would permit leveraging them for cyber attack. The implant may be placed, say, in a router for the purpose of capturing messages from an internal office system; a cyber attack launched against a router would, at worst, be an inconvenience that lasted no longer than it takes to round up and install a replacement. So, no reasonable inference about a future cyber attack could be made. In practice, making such fine distinctions requires: (1) that the target has systems worth eavesdropping on that can be distinguished from those worth attacking; (2) that the intruder knows which are which; (3) that the target (the network owner) also knows which are which and believes the intruder may want to make that distinction; and (4) that such differences can and will be communicated correctly to the target's leadership. The first condition is clearly not up to the penetrators. The second is an assumption that requires a great deal of prefatory cyber espionage in the first place, reintroducing the very risks of discovery that the strategy was attempting to modulate. The third may require insight into the intruder, since the point is to understand whether the intruder meant simply to spy or to also set up a cyber attack. As for the fourth, one can only guess.

The target's technical experts may point out that a penetration in, say, a well-guarded albeit Internet-linked network is no indication of how well the more critical and hence often air-gapped (i.e., electronically isolated) military systems can survive attack. This is particularly true for a cyber attack whose effectiveness depends on good timing, hence on an ability to exercise real-time command and control over the implants. But might such leaders also remember the same technical experts arguing that these dearly-acquired guards would protect their conversations? And while technical experts may remind leaders of the many caveats that follow all assessments of cyber security, lay-folk often disregarded them or view them as attempts to evade responsibility for being wrong. Leaders may therefore be skeptical of arguments that a penetration here does not mean an attack there. Again, the essential role played by surprise in cyber operations erodes assurances of all sorts.

Lastly, is it in one country's interest to improve another country's *confidence* in the resilience of its armed forces in the face of cyber attack? Success at calming the other side would reduce the risks of overreaction that might follow penetrations into the networks of its military. Confidence makes it easier to dismiss the implications of having found the implants, because the target will conclude that they cannot affect a military force resilient to cyber attack. But feeding such confidence also obviates the value of brandishing one's weapons in cyberspace and vitiates the corresponding deterrence value of one's cyberspace capabilities. Furthermore, unless the argument is generic – we are resilient to such attacks, so you probably are also resilient – demonstrating the resilience of another side's military systems with any credibility would have to show a level of insight into the details of their systems which would be anything but reassuring.

So, increases in cyber espionage unavoidably create risks if getting caught raises fears.

5. THE DEFENDER'S OPTIONS

Although the target of a discovered intrusion may well infer an imminent attack and raise its alert levels in ways that lead to mutual escalation which culminates in war, nothing *compels* defenders to act that way. Wars are costly and risky and actions such as raising alert levels are not risk-free. The questionable value of running these risks because intrusions *might* be precursors to attack and pre-emption *might* improve the odds of surviving an attack suggests a place for alternative reactions.

A great deal depends on whether such intrusions are an *indicator* of future aggression (specifically, evidence that the odds of physical aggression need to be revised upward) or just an enabler. If an indicator, then countries need to attend to what happens on the ground, so to speak. If an *enabler*, then policies to stop intrusions merit consideration, as they always should.

Warning against further intrusions may bolster deterrence; it signals discovery, displeasure, and, most importantly, that the target takes these intrusions as indicators of potential attack. Although the standard cyber deterrence challenges apply, such as what constitutes an infraction that merits a response and what the response should be, the issue of grandfathering also merits note. Contrast cyber attacks with cyber espionage; if you warn the other side to stop immediately, then later attacks can be assumed to reflect acts of volition that took place *after* the warning; attacks tend to announce themselves at the time. Intrusions, however, do not announce themselves. An intrusion discovered tomorrow may have been carried out yesterday. Thus, being able to time-stamp the last *hostile* volitional activity (not simply the first intrusion) is important in a coherent deterrence posture.

Unfortunately, correct characterization of the intruder's post-warning activity is not trivial, and the problem is worse if the intrusion leaves behind an autonomous implant, one that takes some actions on its own. The intruder can try to erase or deactivate the implant, but then imagine a target's ire in discovering the intruder's post-warning footprints. Even if discovery does not activate reprisals, it could provide a clue as to how the intruder penetrated otherwise inaccessible systems. After all, if the intruder was confident that, even in wartime, it could command and control the intrusion in real-time, then the implanted code would not need autonomous capabilities. Thus, the existence of such capabilities suggests that the system is hard to access. *Telling* the target about the intrusion so that the target can de-activate it runs into similar problems *and* connotes an obeisance that one rival may not wish to convey to another.

So, unless the target *wants* to build a narrative that would justify fighting the intruder, it needs to exercise forbearance or even forgiveness when it catches what look like violations following a warning.

6. DELIBERATING SIGNALING

Similar issues bedevil using cyber espionage to signal broader intent, in contrast to using it to brandish capabilities. A 2016 study[15] suggests that, if given what they think is the opportunity, policy-makers will try to signal their intentions through cyberspace. In the words of then-CIA-director John Deutsch, they may believe that the "electron is the ultimate precision-guided munition",[16] allowing precision signaling. Or, they may conclude that signaling in cyberspace is far cheaper than moving, say, warships. In one war game examined by the study:

> Strict rules of engagement—to include no network exploitation of strategic command and control and limited military command and control—were placed on computer network exploitation with the assumption that these activities would be detected and would be interpreted as signals of the United States' [lack of] desire to escalate the crisis.

There are two reasons for being skeptical that such signaling would have the desired effects.

One is general to all signaling: there is no guarantee that they will correctly infer what you imply.[17] Some inferences are contrary to fact; for example, that you have forces hidden when in fact you do not. Other inferences are contrary to what you were signaling: you brandish cyber attack capabilities to show how prepared you are, but they think you emphasized non-lethal capabilities because you are afraid to use lethal capabilities. A litany of fairly prosaic reasons can be adduced to explain inaccurate inference, but the simplest is that people make mistakes: they do not see all the evidence or they do not know how to evaluate everything they see. Being busy, as decision-makers typically are, they fail to pay the requisite attention to what they

[15] Jacquelyn Schneider, U.S. Naval War College, *Cyber and Crisis Escalation: Insights from Wargaming*, unpublished paper, January 2017.

[16] U.S. Senate Committee on Government Affairs on the subject of "Foreign Information Warfare Programs and Capabilities." June 25, 1996.

[17] See, for instance, Max Fisher, "Do U.S. Strikes Send a 'Message' to Rivals? There's No Evidence", April 21, 2017; www.nytimes.com/2017/04/21/world/do-us-strikes-send-a-message-to-rivals-theres-no-evidence.html.

do see. Being people, they have confirmation bias: they see what they want to see and when evidence comes along they emphasize their prior perceptions and discard what contradicts it. They themselves may be good evaluators but work for organizations that, collectively, exercise confirmation bias. People also tend to mirror-image: if they see you doing something that they could have done, they may well infer that you are doing it for the same reasons they would have. Leaders with a high regard for their own personal perspicacity (which is reinforced by sycophantic assistants) may rely on their intuition over the painstakingly-generated insights of their intelligence community. Finally, the signal's receivers may be aware of things that signalers are not – and they, in turn, may be aware of things that they think the receivers should have been aware of but were never exposed to. What you see as a signal of yours, they interpret as arising from internal machinations at their end.

Unfortunately for clarity, signalers may have too little idea of what things look like from the perspective of receivers (who, themselves, often take pains to keep others in the dark). Signalers have too little idea of why recipients would think the signal should be read in a certain way. In the end, the signaler may be wrong, but error is beside the point. The reactions of those receiving the signal are entirely determined by facts and circumstances as *they* see them. Neither reality nor what the signaler intended to signal count, if the point is to influence their thinking.

The other set of reasons is specific to cyberspace. Even though cyber espionage may be misinterpreted as preparations for cyber attack, the failure to discover cyber espionage may not necessarily be correctly interpreted as a lack of desire to carry out a cyber attack. Such an interpretation would require that the other side *expects* to find evidence of cyber espionage and then concludes that an absence of a discovery means the absence of activity. It also assumes that they do not find cyber espionage from third parties and erroneously conclude that it came from their potential foes, the most likely guess under the circumstances. They may easily conclude that penetrations carried out *because of the* crisis would not be discovered, because advanced persistent threats even from countries as casual about operational security as China has been can linger undiscovered for several months. Those from more careful penetrators such as Russia or the United States may linger undetected far longer. Even if the penetrators made themselves easy to find in the more benign parts of the other side's network and scarce in the more sensitive areas, the more likely conclusion may be that they took greater pains to be stealthy in the latter case.

Hostile signals – look at us in your system – *should* have a greater fidelity than non-signals. At least there is something to work with. And penetrators should want to take more pains going in than going out, lest they be blocked prematurely. But, to reverse all the cautions noted above, unless the penetration was found where it would clearly be prefatory to a cyber attack, the other side could interpret their finding as evidence of mere cyber espionage, which may imply nothing out of the ordinary.

Perhaps the difficulty of drawing the correct inferences from discoveries of penetrations in general, or implants in particular, may be eased as cyberwar examples accrete. But would they? While cyberspace is a very dynamic place, few cyber attacks have taken place at nation-state scale, as distinct from cyber espionage and cybercrimes.[18] Thus, by the time enough incidents have accumulated to support conclusions, years may have passed and, more importantly, the world that such incidents describe may have changed so much that earlier evidence is immaterial. The problem is not that the technological basis of computation and communication is so fluid – with the possible exception of what artificial intelligence *might* bring, there is a fair degree of year-to-year stability – but that the interaction between people and markets and between attackers and defenders is constantly evolving. Consider the many ways of creating flooding attacks: volunteers on their own computers, large botnets (involuntarily recruited zombie computers), medium-sized botnets amplified by packet reflection, web servers (e.g., those that support WordPress), cloud servers, and networked devices (e.g., video cameras) – with no guarantee that novel techniques may not be added to the list. The technology behind ransomware was largely available twenty years ago, but did not take off[19] until someone showed that it could work; then many others jumped into the business aided, in part, by the emergence of digital currencies such as Bitcoin. Because measures beget countermeasures which beget counter-countermeasures, techniques may morph rapidly in the hothouse environment that is cyberspace. Meanwhile, other tricks die off. Spam is no longer the problem for consumers that it once was,[20] and changes in Microsoft Windows over the last ten years have complicated any strategy that relies on USB sticks as an infection vector. Correctly interpreting any one penetration against such a dynamic background is difficult.

Speculatively, future years may see a shift from first-order attack methods (the insertion of arbitrary executable code into target systems) to second-order (shaping inputs to yield unexpected outputs in the target system). This could arise because preserving the integrity of a system's code base is a workable problem (e.g., by burning instructions into hardware, if nothing else) while ever-increasing system complexity leads to an exponential increase in the interaction space. Furthermore, the NSA at least (according to the former head of its Tailored Access Office, Rob Joyce[21]) tends to rely on hijacking credentials as much as or more than inserting malware into

[18] Notably, system intrusions for the ultimate purpose of getting money, the best example of which was the theft of $81 million from the Bank of Bangladesh, putatively by North Korea (which has also been associated with bitcoin-related theft).

[19] For instance, Dan Bilefsky and Yonette Joseph, "Cyberattack in U.K. Hits 16 Health Institutions," May 12, 2017; https://www.nytimes.com/2017/05/12/world/europe/uk-national-health-service-cyberattack.html.

[20] "Spam email levels at 12-year low," July 17, 2015; http://www.bbc.com/news/technology-33564016.

[21] See his address to the USENEX Enigma 2016 conference: https://www.youtube.com/watch?v=bDJb8WOJYdA.

systems, and hijacked credentials are less useful for cyber attack because the damage you can do with them is limited to the damage that the credential's true owner can carry out. So, credentials may be good enough for tapping the flow of information but not for altering it. If so, the methods used for cyber espionage and cyber attack may diverge, making the world free for cyber espionage.

7. CONCLUSIONS

In a crisis, countries will be looking at indicators of all sorts, not just from within their network. But, as with all things cyberspace, intrusions into networks are likely to garner greater importance over time. As long as the methods of cyber espionage – notably implants – look like the methods of cyber attack, the discovery of one will raise fears about the imminence of the other. Unfortunately for stability, the link between the two is unpredictable. Discovery may or may not happen, but it is more likely to happen in a crisis when systems are being scrubbed more diligently. Figuring out *when* the intrusion took place (the earlier, the more benign) is a forensic art not possessed by all, and without such information the target may assume the worst. The target's reaction, in turn, may be colored by its understanding of the security dilemma in cyberspace. If so, the course of wisdom may be to counter with one's own signals, perhaps deterrent signals. Conversely, signaling through the manipulation of cyber espionage traces likely offers less fidelity than other signaling methods, which themselves have often been misread.

The lesson is to consider what message you want your cyber espionage to carry if and when it is discovered. If you do not want to inflame tensions, double down on operational security, but do not assume success. Thus, also avoid adding military targets to spy on when in crisis, or at least approach them with techniques that look very different from those used to set up cyber attacks. If you are brandishing capabilities or signaling intent, generate a narrative that anticipates discovery. But think this through *beforehand*.