Sub Rosa Cyber War

Martin C. Libicki[a,1]
[a] *RAND Corporation*

Abstract. Cyberspace offers the prospect of *sub rosa* warfare, in which neither side acknowledges that they are in conflict with one another or even that one side has been attacked at all. This is possible for two reasons: first, because the battle damage from some types of cyber attack may not be globally visible, and second because attribution can be very difficult. The reason that both sides may keep matters *sub rosa* is to maintain freedom of actions, on the theory that public visibility may complicate negotiations and lead to escalation. Nevertheless, *sub rosa* warfare has it dangers, notably a lack of the kind of scrutiny that may promote actions which cannot bear the light of day, and the overconfident assumption that no third party is aware of what is going on between the hackers of both sides.

Keywords. Cyber space, cyber war, *sub rosa*

Introduction

The last twenty years have seen a burgeoning knowledge base on cyber this and cyber that. We know a good deal more about how to get into other people's systems – and we know a good deal more about how to keep others out. Computer users are far more conscious of security considerations – they have had little choice in the matter. Computer security has risen in the ranks of government – and alliance – issues.

Nevertheless, it is fair to say that this growth has been concentrated at the tactical, which is largely to say technological but also the management end. Shelves are filled with books on how-to, but far fewer tomes explain why to. There is very little open material on how to integrate cyber operations with kinetic operations, which is to say how to use cyber operations to advance the ends for which kinetic operations used to be the exclusive means. Quite possibly, there may not be much behind the green door either, because there is very little intelligent discussion within the literature of professional military integration of *hypothetical* capabilities. As for strategic discussion, there is some, but a great deal is built on the premise that cyber warfare is kinetic warfare (or nihilistic terrorism) by other means. Well, it's not; the two are quite different. One might say they are as checkers and chess – which only look the same because their terrain is the same and some of the pieces have the same name.

This essay expounds on one of the more interesting differences. To wit, cyber war offers the prospect of *sub rosa* warfare, which is a form of combat in which the participation of both sides, or at least one side, is obscured to third parties. *Sub rosa* warfare

[1] Senior Management Scientist, RAND Corporation. Note, the following represents the author's view, not those of RAND or its sponsors or clients.

has some aspects of intelligence operations, and some aspects of special operations – although it is neither. Of note, *sub rosa* warfare is almost impossible to conduct with tanks, much less nuclear weapons.

1. Embracing Ambiguity

Ambiguity, one can argue, is the essence of cyber combat and the exploitation of ambiguity may be central to *some* strategies of cyber warfare. There is a natural human tendency to assume that ambiguity is an epiphenomenon, something that obscures reality, a measurement error, as it were, and thus a temporary irritant to true understanding. Dust the surface, and the essence of the activity shines through and we see things for what they are. Some of this flavor comes from Clausewitz: fog and friction are what differentiate war on paper from war in the field. Those of Platonic bent may see the former as reality and the latter as shadows on the cave's wall. Clausewitz warned that one could not assume fog and friction away; they were embedded. What he did not argue was that fog and friction were central, with all that violence being peripheral – much less that the manipulation of fog and friction played much of a role in operational planning.

With cyber, the opposite can be true. In some cases, ambiguity is central and damage to systems is an artifact. If so, one should embrace ambiguity and not treat it as something of an embarrassment.

1.1. Definitions

Before going further, a few boundary markers may be useful.

First, cyber war will be defined as consisting of computer network (more broadly, systems) attack and defense. An attack succeeds when the target's use of its own systems is hampered – either because such systems fail to work or work very efficiently (disruption) or because systems work but produce errors or artifacts (corruption).

This definition specifically excludes computer network exploitation, which meets neither of these criteria. It is fair to say that CNE accounts for the great preponderance of computer network operations carried out among states and similarly serious non-criminal organizations. Yet it is a different phenomenon. Spying is not an act of war. It never has been, and there's little reason to change that. Furthermore, spying is inherently *sub rosa* and its motives are almost self-explanatory. This is not to say that CNE does not matter – it does – or that it is not interesting – it can be. But, it's not the subject of this essay.

Second, retaliation will be assumed to be an option in wake of a cyber attack, but that such retaliation will be limited to the cyber realm. This is not to say that retaliation must always be in kind, or that both attacker and retaliator can carry on mischief in cyber space without concerning themselves about escalation into the physical realm. Indeed, escalation is a major motive for keeping things *sub rosa.* However, it is difficult to keep physical retaliation *sub rosa*, and considering as much takes us into a different conversation.

Third, the essay limits itself to state-on-state cyber war, mostly because it best frames consideration of attack and retaliation. Non-state actors generally cannot be deterred, or even mildly dissuaded, by putting their systems at risk – because they do not have systems. Furthermore, although *sub rosa* warfare has a legitimate rationale of

sorts, the usual approach to non-state actors generally involves the application of justice, and there are serious problems with *sub rosa* applications of justice that transcend cyber war issues.

1.2. Roshomon

What makes it possible to speak of *sub rosa* attacks is that information systems are generally invisible. The artifacts of a system – such as a personal computer – may be seen, but while some of what systems do is reflected back to the user, a great deal of what they do takes place inside and is not reported. Within an organization, all of the artifacts of computation and even most of the direct results may be hidden from the public, and what they do is visible only to the extent that the owners wish it so. Thus, damage to such a system is often invisible, even if some of the second-order effects may be quite visible. The contrast with physical warfare needs no further elaboration.

Of note, therefore, is the possibility that the target, the attacker, and third parties hold completely different perspectives on the nature of any one cyber attack.

The target includes the system operators, those they may call on for support, and those they report to. All three of *these*, incidentally, may hold views of what happened that diverge from one another's perspectives. In general, the target ought to have some idea that something went wrong, perhaps why, and what the consequences were. As a general rule disruption attacks are easier to see than corruption attacks. However, the target may not know (at least not immediately) what the attack vector did, how it managed to work past defenses, who controlled (or at least launched) the vector, and what the purpose was.

This plausible membership of the set of people who know they have been attacked merits further consideration. One can imagine an attack that only systems administrators notice – one that, for instance, requires them to put in overtime for the purposes of restoring a system's prior functionality and integrity. That being so, how damaging, which is to say, how consequential and thus how strategic can such an attack be? Thus, one must either posit a greater affected population that, nevertheless, keeps silent (such as the intelligence community) or a user base that senses damage but is misinformed about why. Some potential contexts include (1) systems that were planned to go on-stream or start new services but were prevented from doing so (a common enough phenomenon without hackers), (2) systems that went down because of what was believed to be human error, accident, or Mother Nature, or (3) systems that appear to function normally but produce bad information that the public at large is unaware of (e.g., scrambled payments for medical reimbursement checks). For systems operators to keep silent about the last one is quite risky unless they have confidence that they can correct things later without others being the wiser.

The fourth, but partial, possibility is that the problems are blamed on hackers but the hackers are identified as non-state actors (and thus subject to prosecution rather than retaliation). Such an attack is only barely *sub rosa*, in part because many third parties may believe otherwise (it was a state attack) and some of the implications of *sub rosa* attacks, discussed below, do not apply.

The attacker includes the hackers and those they report to (assuming their reports are honest and complete). The attacker will know the attack vector, how it evaded security (or at least the security features they saw), and what the purpose of the attack was (or at least one of the attacker's bosses will know). Depending on what kind of sensors it has emplaced in or near the attacked system, it may know something about the direct

damage, but there may be a great deal it does not know, especially with respect to programs that the target system may have to route around or make up for damage.

Third parties include the public of the target country, the public of the attacker country, third-party states, and third-party publics. If the attack (and perhaps retaliation) were really – which is to say, successfully – *sub rosa* they will not know much about what, if anything is going on.

As we demonstrate, the difference in what each of these parties (broadly defined) knows, coupled with (presumably) their reluctance to share such information, makes the rest of the story possible.

Note that *sub rosa* attacks are not defined simply as those where the attacker's identity is unknown or uncertain – although, if the attack itself is unheard of, the identity of the attacker is moot.

Indeed, it should take little imagination to understand how much of cyber war is subject to ambiguity: not only *did something happen* – but *was it an accident (bad software, human error, Mother Nature), what was the damage, what was left behind, who did it, how they did it (including when they did it, and where was the point of access),* and, most importantly in the long run*, can they do it again?*

2. Basic Principles

Next, we turn to some basic principles of computer network attacks, not necessarily to say anything original, but to emphasize a few things by way of foundation for what comes later.

With one type of exception, the DDoS attack (more on this a little later), attacks are enabled by vulnerabilities on the part of the target.

One can assert, for starters, that there is no forced entry in cyber space. If someone has gotten into a system – or more particularly into the no-go area of a system – from the outside, it is because that someone has persuaded the system to do what its operators did not really want done and what its designers believed they had built the system to prevent. Nevertheless, in any contest between a computer's design and use-model (such as a user's intuition that email is information not instructions) on the one hand and its software code, on the other, the code always wins. Whoever gets into a system gets into a system through paths that the software (to include protocols and firmware) permits. The software may have flaws or may have been misconfigured (for instance, the permissions the administrator established differ from the permissions that the administrator thought had been established). Yet, a system is what it is, not necessarily what it should be. Such a divergence, when it has security implications, is a vulnerability. Whatever the methods, manual or automated, hackers' use, an attempt to take advantage of a vulnerability to gain access to a system or to get it to accept rogue instructions is called an exploit.

A system's integrity dictates how badly a system can be hurt by attacks in cyber space. One might even argue that a system's integrity is a more important determinant of success than the quality of the adversary's exploits—after all, no vulnerabilities, no exploits; no exploits, no cyber attacks.

Thus, in theory, all computer mischief is ultimately the fault of the system's owner – if not because of misuse or misconfiguration, then because of using a system with security bugs in the first place. In practice, all computer systems are susceptible to errors. In that sense all systems are somewhat opaque, unpredictable, and thus, ambi-

guous. The divergence between design and code is a consequence of the complexity of software systems and the potential for human error. The more complex the system – and they do get continually more complex – the more places there are in which errors can hide. Every information system has vulnerabilities—some more serious than others. The software suppliers themselves find a large share of these vulnerabilities and issue periodic patches, which users are then supposed to install – some more expeditiously and correctly than others (notwithstanding those hackers who observe patch releases, reverse engineer them quickly, determine the vulnerabilities the patches were supposed to fix, develop an appropriate exploit, and use it against those slow to patch their systems). Hackers find some vulnerabilities and then spring corresponding exploits on unsuspecting users who have otherwise done everything correctly. Literally thousands of exploits are sitting around. Many of the more devious ones require physical access to the target system. Most of the ones that reach the news do not work on well-patched systems.

In a sense, cyber attacks rely on deception – persuading systems to do what their designers do not want them to do. Fortunately, deception can be its own undoing. An exploit, if discovered, signals to sysadmins that something is not right. If good logs are kept, sysadmins may be able to determine where something unusual took place in the interaction between the hacker and the system. Changes in files (data or instructions), or the presence of unexpected files can also be telling. The process is hardly perfect; it is possible to determine a specific vulnerability and miss the broader design flaw of which the specific vulnerability is just an instance. Nevertheless, any one sysadmin can take advantage of an international community of system defenders with a common interest in minimizing outstanding vulnerabilities.

In contemplating cyber space, it may help to differentiate system peripheries from the system core. Peripheries may be said to contain user equipment; that is, equipment whose function and parameters are established by users. Peripheries, if not air gapped or protected via consistent encryption, tend to be repeatedly vulnerable largely because users are rarely trained in or focused on information security. User systems and privileges can be taken over through password cracking, phishing, social engineering, downloads from bad Web sites, use of corrupted media such as zip drives), etc. Sadly, the security of the periphery as a whole is often no better than the security of the most feckless user. The core, by contrast, is what sysadmins control—monitors, routers, management devices, machinery (such as weapons), and databases. Sysadmins are (or should be) trained and sensitive to security issues; they also set the terms by which users (and their systems) interact with the core. Although it is good personnel practice to sensitize users to security issues, it is good engineering practice to assume that users will not always be sensitive. While it is possible to protect the core from insecure users, it is less clear whether networks can function when enough user systems are compromised badly enough, even though network administration is a function of sysadmins. In general, it is hard to compromise the core in the same precise way twice, but the periphery is always at risk.

DDoS attacks are, as noted, a partial exception to the rule that a system can be attacked only if it has vulnerabilities (the Mafia-Boy attack of February 2000 apparently did take advantage of a certain class of vulnerabilities, since largely cleaned up). However, it is hard to conceive of a *sub rosa* DDoS attack in the sense that the public does not notice. So, we can disregard the exception for our purposes.

sense that the public does not notice. So, we can disregard the exception for our purposes.

2.1. The Attacker's Motive for Going Sub Rosa

An attack can be *sub rosa* only if the effects are limited to entities (such as state entities whose outputs are opaque and who believe in keeping secrets) or if the attacks could conceivably be ascribed to something other than hacking. The target has a good deal to say about whether an attack is *sub rosa;* yet, if attackers want to leave open the possibility of a *sub rosa* attack they have to avoid having such attacks affect the broad public but in ways that cannot be credibly ascribed to accident. They cannot take credit for an attack, which means that it cannot be used for certain forms of coercion.

The overall motive – for both sides – for keeping matters out of the press is that cyber warfare is a negative-sum game. Although this may be said generally true for warfare, it may be doubly true for cyber war (CNE, importantly, aside). Simply put, there is very little to be directly gained, which is to say, seized, in cyber war, unlike kinetic warfare where at least one side can entertain the possibility of a smash and grab (e.g., Kuwait's oil fields). Cyber war cannot even disarm the other side's cyber warfare capabilities, and while it can disarm kinetic warfare capabilities, it can only do so for a limited amount of time. Thus were there to be an extended cyber war, it would inevitably be a contest of attrition, a test of who can, in Wellington's terms, pound longest before someone's spirit gives out.

To go into particulars; an attacker may wish to limit its attacks to those that offer the target the opportunity to keep quiet in part to forestall retaliation. The attacker believes that while the state's elites may be able to handle things rationally – for instance, understand when they have been back-footed and thus retreat from some position – the same cannot be said for the target's publics. Thus, informing such publics will put pressure on the state to retaliate publicly when state elites may think other courses are less costly to the state. Worse is the possibility of escalation; elites may have a consensus among themselves to keep things in the cyber realm, but the public may not favor such limits. More generally, getting one or both publics involved introduces the possibility that events may spin out beyond the elites' ability to keep things under some sort of control. Many observers of war – for instance, of Gelb's book, *The Irony of Vietnam, the System Worked* – have concluded that state decision makers often prefer to risk losing a war than to risk losing *control* over a war. Finally, if the war is controlled, it is possible for elites of both sides to engineer a de-escalation of hostilities. All this manages the risk that the attacker faces in a cyber confrontation – for both sides.

One should also note the possibility that the effects of the cyber attack can be fit into the attacker's narrative *but only if* the results of the attack can be blamed on something other than the attack. Of course, if no one notices the effects of the attack, there's nothing to narrate about. A *sub rosa* attack whose effects are felt but not explained tends to shed focus not on the attacker but on the incompetence of the target – one unable, for instance, to protect sensitive health records from being scrambled.

As noted, a high form of *sub rosa* warfare is to make the attack look like an accident. One should not count too highly on anything more than momentary success; investigations tend to be pretty good at getting at root phenomena.

Incidentally, for some purposes the attacker may want its identity known to its opposite number. A few tricks such as mailing a letter before the attack that is received afterwards, leaving a "Kilroy was here" in the target machine, or revealing knowledge that only a penetration could provide should suffice.

Here are a few scenarios for a *sub rosa* attack:

One, they can be used to put others on notice that their systems are not so reliable that they can afford to engage in such a fight. Consider this. In step one, an attacking state creates anomalous behavior in a key system, be it government or a government-linked entity. The act (rather than the attacker, which is kept as ambiguous as possible) gets the attention of the leaders of the target state, which perceives its infrastructure at risk.

Subsequently, the system owners and their engineers claim that it was an accident and vow that such an act will never happen again. They get large sums of money to work hard on the problem. After this team starts to claim success, the attacker again creates anomalous behavior, preferably to the first victim, but perhaps to another comparably important system. This signals that problems persist (admittedly, step two is hard, precisely because the target state is working diligently against the possibility – certainly on the attacked system and quite likely on similar others). This not only reduces the credibility of the target's information system security, it also, and more importantly, reduces the credibility of those who promised to achieve that security.

Yet the attacker does not reveal itself or what it has done. This is unnecessary and even gets in the way. Doing so would make getting back at the attacker a more visible centerpiece of the target's strategy than simply misleadingly reassuring those who know they rely on the attacked system. Indeed, the attack itself is not so much the issue as it is to foster a general sense that the other side's information systems are fragile and unreliable. The attacker's message then becomes not "Cower before us!"—which requires identifying "us"—but the more impersonal, "You live in glass houses; are you sure you want to invest so much in stones?"

Perhaps the whole point of the attack is to make the target extra wary of expanding or opening up its networks, especially to outsiders, such as allied militaries, other government agencies, or support contractors. Further wariness may result from making the attack appear to come from a trusted source. Such a strategy presumes a skewed response from the target: not that networking should not be done naively but that networking is bad. It is easy to see why such a strategy can backfire and thus why cyber strategists, thinking over an extended period, must keep second and nth-order effects in mind.

Scenario two, cripple, test, or exercise someone else's military. Cyber attacks on the target's military may be used to impede the target's ability to respond to crises. A large, successful attack may retard the target's ability to wage war; if the target's military deployment can be delayed long enough (e.g., after everything has been decided and after the aggressor's forces have dug in for defense), the target's military intervention in a crisis started by the cyber attacker may be deemed pointless.

Such an attack can be a prelude to aggressive military action, or it can be in response to fears, however ill-founded, that the target is about to start something. In the former case, if the attack disarmed the target's military enough to allow successful kinetic combat, the *sub rosa* nature of the cyber attack may be temporary, and basically irrelevant if the cyber attack is quickly followed by violence of an obvious sort ("quickly" because the effects of *any* cyber attack are temporary and measured in hours or days). However, if the cyber attack fails to dent the target's military capability the attacker may call off its dogs and has no reason to publicize what it has done. In the latter case, cyber attack as pre-emption, the result of a successful cyber attack may be exactly nothing – in contrast to the violence that might have happened if the target's systems were intact. Since the effects of the cyber attack are temporary, war may take place anyway later – or not, if the cyber attacker (who is the presumed impending vic-

tim of the target's military) has used the time gained to rush to the front, so to speak, and discourage the war's outbreak.

Complicating this logic are attacks that look like they are meant to cripple another's military but are not. For instance, what if the cyber attacks were meant to persuade the target military that war was imminent, draw it to the ramparts for no reason, and repeat the cycle often enough to exhaust or spoof the target (as Egypt did when it carried out exercises in early-1973 but not attack until October of that year)? In contrast to physical feints, however, cyber feints may be poor strategy. By hardening the target's systems, every attack makes a subsequent attack more difficult. The choice of targets, if not masked by noise, may also suggest what the attacker finds important to disrupt and thus hints at how the cyber attacker would fight if war turned physical.

Attacks may be launched on military systems to see how well their operators react, in preparation for some later, larger attack. Can enemy sysadmins determine what happened and why? What workarounds do they use? Will corruption be detected? If the target knows it has been so tested, should it retaliate? Conversely, attacks may well reveal a great deal about the attacker and what it knows about the target's vulnerabilities.

There are solid grounds for believing that attacks on military can retain their *sub rosa* character. The attacker has the usual motives for keeping quiet, with the possible exception that it may wish to whisper about the attack to the target's allies so as to reduce their faith in the target's military. For the target, on its part, to reveal that it was attacked – and successfully so – is apt to reduce rather than increase confidence in its military capabilities. The latter may not have much of a choice if the damage is so widespread that a universe of witnesses defeats all thoughts of keeping them silent. The target may also broadcast the attack for purposes of supporting a "hate the enemy" campaign, regardless.

Cyber attacks that cripple intelligence assets do not have to lead to war. They may be justified if they blind the target's systems long enough for the attacker to carry out operations (e.g., moving missile parts) safe from prying eyes. Perhaps needless to add, intelligence assets are extremely hard targets for cyber war.

Coercion – especially against democratic states – normally requires the damage to be publicly visible and clearly associated with the coercer and its cause. Adversary actions need not affect the public, though, if there are other ways to compel governments to accede to demands. Indeed, the opposite may be true: the less the public knows, the easier it may be to garner concessions, especially invisible ones.

The case for *sub rosa* cyber war for the purposes of coercion rests on the belief that publicly visible attacks could lead to more popular pressure on the state to stand firm than to concede. The attacker counts on the possibility that the target's leaders are less afraid to make concessions whose true rationale can be hidden than to be blamed when, say, the economy hits an air pocket. As long as the new policy (which contains concessions) does not appear unwise *per se* or does not contradict earlier policies too much, the target's leadership need merely hide the fact that their policy choices were driven by fear. Keeping mum has other advantages for the target. Reducing the public itch for revenge (or their desire to demonstrate resolve) may facilitate negotiations or mutual de-escalation. Obscuring the fact or at least the damage from the attack may also mask the state's vulnerabilities from the eyes of third parties (presumably, the attacker will have a better sense of which vulnerabilities it had, in fact, exploited).

One ought not forget in all this that the *sub rosa* strategy has a serious Achilles heel from the cyber attacker's point of view. It assumes or, more to the point, *requires*

that the target reacts as expected and maintains its silence. This requires that the cyber attacker have sufficient insight into the target to operate below the threshold past which it decides to mobilize against the cyber attacker – an act that generally requires the target being open about the attack and its consequences. The larger the cyber attacker's gain vis-à-vis the target, the less likely the target is to restrict its own activities. In effect, the attacker's strategy is hostage to the target's behavior, the basis for which we now turn.

2.2. Should the Target Reveal the Cyber Attack?

The likelihood that any attack is visible is the likelihood that the effects of an attack are visible multiplied by the likelihood that these effects will be publicly ascribed to a cyber attack (rather than to error, accident, or bad design). Both parts of the equation are anything but given. CNE is rarely apparent until an investigation reveals it. Corruption may go unnoticed until it reveals itself as a discrepancy between what a system is doing and what it should be doing. Sometimes even disruption may go unnoticed; for example, if a sensor is silent, is it silent because it has nothing to report, or has someone tampered with its reporting channel? If it is not people but machines or other processes which consume certain services, their loss may be noticed only when the processes they feed behave incorrectly.

Normally, full disclosure is the best policy. It is too easy for governments to believe they can control information much better than they actually succeed in doing – witness Chernobyl. Post hoc revelation eats at government credibility—not to mention competence, if playing catch-up with events makes the government look bad. Screaming helps mobilize the citizenry to support the government and (less cynically) pay attention to information security. It raises the seriousness level of the whole cyber war contest and thus gives the government more scope for implementing domestic security measures that the citizenry would otherwise object to. If the fact of the damage is evident, but not the cause, revealing the cause may enhance the credibility of infrastructure owners by switching attention from their own fecklessness as sysadmins to factors (portrayed as) outside their control. Revelation is necessary if the target state is going to respond visibly, either with retaliation or without (using legal, diplomatic, or economic measures, for example). Going public provides an opportunity to be clear about the aims of the response; it also subjects them to the test of knowing whether it can bear scrutiny. Incidentally, revelation may also be necessary for *sub rosa* retaliation: just because the retaliator did not want to make a fuss about how it hit back does not mean that the attacker (as target of retaliation) will do likewise.

Yet silence may still be golden. Revelation may expose the fecklessness of the target's system security, reducing the public confidence in it and making it a target for repeat attacks (a case for discretion comes from the public's tendency to overestimate the risks of cyber insecurity; there is considerable agreement that the public is wildly inconsistent in how it reacts to low-probability, high-impact risks). Evidence to support the attack claim may reveal sensitive information about system security.

2.3. Should Cyber Retaliation Be Obvious?

In cyber space, the target can hit back against the attacker, and no one (aside from the security establishments on either side) need be the wiser. This sort of *sub rosa* retaliation tends to make more sense if the attack is not public or if public attribution is not

viable. In the latter case, the evidence behind attribution may be of the sort that is not easily released or not easily argued if released. *Sub rosa* retaliation avoids having to make the choice of what to reveal. This is no small matter. Reveal one's forensics and one has given all attackers a clue about what to avoid leaving behind the next time. Information about sources and methods is among the most closely-guarded secrets of the intelligence community. Furthermore, the attacker, as the victim of retaliation, could be under subsequent public pressure to counter-retaliate. If the effects of retaliation were not obvious, the attacker could therefore conclude that letting things drop after the retaliation is wiser than carrying on.

States that would employ *sub rosa* retaliation have to manage the expectations of those in the know who are looking for revenge. Retaliation could still convey the target's displeasure over the attacker's leadership and could change the latter's calculus to discourage further attacks.

Sub rosa retaliation, however, may be too seductive, particularly if the retaliator feels no need to convince the attacker of its guilt—after all, the attacker knows that it struck first, right? One danger is that, if the intelligence or law enforcement agency does not need to worry about defending its attribution to others, its case to national command authorities (that is, those who control the retaliation capability) may go unchallenged. The agency may thus claim its attribution is correct when the evidence suggests a higher degree of skepticism is warranted. Furthermore, a decision to retaliate *sub rosa* – like the decision to attack *sub rosa* – takes certain targets off the list (e.g., power plants) or at least demands they be hit in ways that do not look like a hit (which then fails to communicate displeasure reliably). The remaining targets may be those thought to be important to the other side's intelligence and law enforcement communities but do not directly affect the public at large. Finally, the entire strategy rests on the attacker's willingness not to make a fuss. Again, but in reverse this time: the wisdom of the strategy is hostage to the discretion of the state that (supposedly) engineered the attack in the first place.

2.4. Sub Rosa Retaliation against a Sub Rosa Attack Has One Big Advantage

To wit, the requirements for attribution are not nearly so great. One does not even need that much confidence in the quality of attribution. So, in hitting back, one may consider two possibilities. One; it was the attacker that suffered retaliation. Two; it was an innocent third party.

Take the first case. The attacker, knowing that it started things, will have a fairly good idea of why it was hit and take the message (subject to all the other caveats). If retaliation is to be reliably read as retaliation by the attacker, the "accident" would have to occur rather quickly after the original attack. Thus, the capacity to retaliate has to be maintained at a fairly high degree of readiness (that is, one must ascertain that the vulnerabilities still exist and that the victim's reaction will be roughly as predicted). Furthermore, the normal deliberation that might take place after an attack to increase the odds that the retaliation was well-directed would have to be short-circuited. All the previous caveats about the difference between what you think others do not know and what they actually do not know also apply.

In the second case, the innocent third party, unaware of what may have motivated an unprovoked attack (which the retaliation may look like to the victim) can only trot out its usual suspects and look for forensic evidence. As noted, this requires the original attack be unknown to any but the attacker and the target.

Managing the consequences of any venture that assumes ignorance among others is always contingent on third parties not spilling the beans. For instance, if retaliation against a third-party state is discovered by the attacking state, the attacker now has a very valuable piece of information – who attacked the third-party state. If the attacking state can figure out how to profit from implicating the retaliating state, it may well do so. Telling the third-party state that it started things may not be the smartest move, but it may be able to downplay its own role to suggest the retaliator over-reacted and was stupid about things to boot. It may maintain its innocence but circulate hints that make it easier for the innocent victim to identify the attacker (finding something is a lot easier when you know exactly what you are looking for). Or, the attacking state may blackmail the retaliator lest its actions be revealed to the innocent victim. The assumption that no one in the third-party state knows about the original attack may be in error; it is not unknown for two states with little in common but their dislike of the United States to swap intelligence (Iraq and Serbia, for instance, traded information on how to defeat U.S. aircraft and avoid anti-radiation missiles). More generally, the original attack may not be so secret prior to the attack or its existence may be revealed after the fact. Such revelation may be deliberate (perhaps someone here in the know is bothered by the retaliation or the possibility that it was misdirected), or simply reflect the universal difficulty of hiding secrets. Finally, the retaliator may have overstated its ability to keep itself anonymous. The third party does not have to know who did it, but it may have serious enough suspicions to affects its relationship with the retaliator – and if it did not know why the retaliator acted as it did, it may be angrier than if it understood that retaliator's motivation.

Again, perhaps *sub rosa* cyber war may be too clever by half – and one does not gain points for upholding the rule of law in cyber space by being sneaky.

What about being even cleverer and making retaliation look like an accident? The last technique is a variant of the first. Not only is the retaliation anonymous but it appears to be an accident. It is two steps rather than one step removed from something that the innocent third-party victim may find actionable. Again, the true attacker will presumably suspect that the accident was too closely timed to the original attack to be an accident, while the innocent victim of misguided regulation will have even less indication of what happened much less why. Indeed it would be most cool if the reprisal could be made to look like something caused by the original attack going haywire – all the dissuasive impact, and none of the risk.

It is unclear how to make an attack look like an accident in the first place. True, many attacks are initially hard to distinguish from accidents – which argues against hasty reactions all around. But there are techniques that can distinguish the two. If the problem is faulty software (such as the DSC bug that crippled phone service in the 1990s) then the fault can often be replicated by simulating the conditions at the time of failure. Human error can often be detected in various process logs. The greater the pain, the greater are the resources likely to be devoted to its elucidation. Thus, safeguards against the victim's (whether the original attacker or the unfortunate third party) detecting that it has been attacked may be temporary.

Finally, a state that wishes to establish principles – such as, do not hack – and then enforces them surreptitiously communicates either that it lacks sufficient faith in such principles or the strength to maintain and defend them openly.

Subtlety, nay sneakiness, in retaliating against a cyber attack absent strong attribution is normally difficult, but the exigencies of cyber space – the high level of ambiguity everywhere in the medium – only make things harder. Thus, while there are some

notional ways to ways to work around the attribution problem they require a great deal of certainty about matters (the effect of cyber attacks, or the perception of attackers and third parties) that stand in stark contrast to the *uncertainty* about who did it. This leaves us with approaches that our British friends might call frightfully clever, with the emphasis on "frightful."

2.5. Ending Sub Rosa Warfare

How does one end a war that one does not admit one is fighting? In general, wars can end in one of four ways: through the destruction of one or both parties, through a formal peace agreement, through an informal tacit peace agreement, or as a series of bilateral decisions not to attack.

Cyber war generally lacks the power to destroy one or more parties to a conflict, and all the more so when the warfare is *sub rosa* – which not only takes certain types of attacks off the table (notably those that put pressure on populations), but also lies below the level where either side has reason to escalate into at least explicit warfare.

A formal peace agreement that pledges each side to halt an activity appears inconceivable if neither party admits to being the victim, much less the perpetrator of the acts in question. Yet, the transition from *sub rosa* to explicit cyber war is easy to make. Third parties may discover as much and make their findings public. Each side may also discover reasons for changing its mind and announcing as much.

An informal peace agreement requires that each side of a fight that is not public is nevertheless willing to discuss such secret maneuverings with its counterpart. At a minimum this requires some confidence on each party's part that it is not kidding the other.

Both formal and informal peace agreements in cyber space, however, can be problematic to enforce, or even state the terms of. Monitoring peace pacts in cyber space poses challenges not found in physical space. If either side still believes it can, if unpunished, reap unilateral advantages from new attacks, then attribution and damage assessment will likely remain as difficult afterward as they were beforehand (if attacks are extended to include CNE, the odds that one or another side finds attacks useful only go up). Each could cheat by shifting from visible disruption attacks to more-subtle corruption attacks.

Unfortunately, tacit de-escalation presents many of the same validation problems as negotiations – only made worse by the fact that there would only be a rough consensus rather than an explicit statement of what actions were and were not considered a violation. How could one tell that the other side is even cooperating, without clarity on what constituted cooperation?

In physical wars, peace pacts are often followed by unilateral disarmament (after World War I, for instance, Germany's army was limited to 100,000) or multilateral disarmament (for example, the Washington Naval Treaty). But disarmament in cyber space is virtually meaningless because cyber war is less about arms (exploits) than about vulnerabilities. So, disarmament cannot bulwark a peace agreement that applies to cyber space.

Mutual transparency may help keep the peace (in much the same way that formerly warring sides exchanged hostages), but no state (not even a friendly one) exposes the secrets of its security architecture to another. Besides, if the war is still *sub rosa* both sides have amply demonstrated the virtue of transparency. If it did, the transparency would have to be bilateral rather than public, lest mischievous third parties profit from

the new-found knowledge. Even then, each side could attack the other from third parties outside the transparency agreement.

Thus, the least problematic outcome is for neither side to find any especial reason to commit serious resources to breaking the systems of the other. This may ensue because the broader ends that led at least one of them into cyber war in the first place have been met or because further cyber war will get no party closer to meeting them than the last spate of cyber war did.

The part of the equation in which one side decides that the effort no longer pays is not strategically problematic because it does not require the other side to recognize that anything has changed. But it is hard to believe that the party that quit making the effort would not hope to see some rewards for its restraint. As long as the one side had not made either explicit (that is, negotiated) or implicit commitments to restraint, the other side would not be able to hold up some future system malfunction as evidence that it had been lied to or cheated. Furthermore, if the other side still found advantage in computer attacks – or if it was engaged in other forms of hostilities – it may have no motive to acknowledge such restraint. But if the other side also finds that the advantages of hacking have waned or that they are trumped by the rewards of friendly engagement, it too might work itself into a *modus vivendi*.

3. Conclusions

Cyber space is a medium in which the absence (or, more specifically, unimportance) of physical artifacts permits a form of warfare that is generally unavailable in other media. The closest analog to *sub rosa* warfare would be a campaign of espionage, but even there, the potential exposure of the saboteurs (whilst hackers can be sheltered by the attacking country or in the anonymity of the Internet) makes it hard to keep matters quiet for terribly long. That such *sub rosa* warfare is possible, however, makes it neither probable nor particularly wise. Paradoxically, maintaining *sub rosa* warfare requires the tacit assent of the other side, and is therefore quite fragile. More practically, the very shadowy nature of the whole enterprise (coupled with the difficulty of getting policymakers to understand the requisite ins and outs of cyber war in general) creates enormous temptation to take risks without adequate political consideration of their cost.