I/S: A JOURNAL OF LAW AND POLICY FOR THE INFORMATION SOCIETY

Thoughts on Threat Assessment in Cyberspace[†]

HERBERT LIN[*]

Abstract. Hard information about the cyber-threat posed by adversaries is scarce and difficult to obtain. Thus, threat assessment in cyberspace is an inherently more uncertain endeavor than for more traditional domains of potential conflict. Under circumstances of information scarcity and faced with potential threats, there are many influences on analysts to make worst-case assessments. Greater information scarcity about possible threats (as is generally true for threats in cyberspace) would increase the likelihood that worst-case assessments would be forthcoming and also increase the uncertainty inherent in those assessments.

[†] This paper is based in part on material drawn from NAT'L RESEARCH COUNCIL OF THE NAT'L ACADS., TECH., POLICY, LAW, AND ETHICS REGARDING U.S. ACQUISITION AND USE OF CYBERATTACK CAPABILITIES, (William A. Owens, Kenneth W. Dam & Herbert S. Lin eds., 2009) [hereinafter NRC CYBERATTACK REPORT]. To the extent that it diverges from the conclusions of the NRC CYBERATTACK REPORT , the views expressed herein reflect those of the author alone. Although the NRC CYBERATTACK REPORT was supported by the MacArthur Foundation and the Microsoft Corporation, the views reflected in this article and in the referenced publication do not necessarily reflect the views of either of these sponsors.

[*] Dr. Herbert Lin is chief scientist at the Computer Science and Telecommunications Board, National Research Council of the National Academies, where he has been study director of major projects on public policy and information technology. These studies include a 2007 study on cybersecurity research (Toward a Safer and More Secure Cyberspace), a 2009 study on offensive information warfare (Technology, Policy, Law, and Ethics Regarding U.S. Acquisition and Use of Cyberattack Capabilities), and a 2010 study on cyber deterrence (Proceedings of a Workshop on Deterring Cyberattacks: Informing Strategies and Developing Options for U.S. Policy). Prior to his NRC service, he was a professional staff member and staff scientist for the House Armed Services Committee (1986–1990), where his portfolio included defense policy and arms control issues. He received his doctorate in physics from MIT.

> *From the standpoint of policy formulation and decisionmaking regarding possible adversary operations in cyberspace (as opposed to traditional domains of conflict), policymaking consumers of threat analyses would be well advised to exercise even greater caution in interpreting the assessments they receive.*

I. THE THREAT ASSESSMENT PROCESS

Threat assessment is a process through which analysts assemble and interpret information in order to assess the threats faced by a potential target. A target may be a nation with diplomatic, military, or economic interests; a company with valuable intellectual property that it wishes to keep confidential; a power station with computer-controlled generators; a factory producing enriched uranium using computer-controlled centrifuges; and so on. Threats to these targets emanate from adversaries, which range from nations to individuals (e.g., criminals or terrorists).

Entities that may be targeted have an interest in knowing the scope and nature of the threats they face so that they can allocate resources prior to a hostile action and plan what actions they should take if they are in fact targeted. To draw an analogy from everyday life, a family living in the inner city may equip its doors with high-security dead-bolt locks and alarm systems, whereas a family living in a rural area may not lock its doors at all. The difference between these two scenarios reflects a difference in each family's understanding of the threat it faces—that is, in its threat assessment.

In principle, a threat assessment incorporates the following:

- the effects that would ensue if an adversary were able to take advantage of vulnerabilities to valued assets;
- the feasibility of specific adversaries being able to exploit (i.e., take advantage of) those vulnerabilities; and
- the likelihood that specific adversaries will in fact exploit those vulnerabilities.

The information used to undertake threat assessments comes from many sources. Forensic analysis of actual events provides one kind of information, yielding, when successful, information about attacker methodology, identity, resulting damage, and so on. Other kinds of information may include analysis of adversary documents (both open-source and secret), communications and other signals intercepts, interviews with those knowledgeable about adversary doctrine or operations, photo reconnaissance, reports from intelligence agents, public writing and speeches by adversary leaders, and expertise available to the adversary.

In general, threat assessment focuses on the negative consequences of adversary action. Threat assessments are also inherently uncertain depending on the nature of the available information. Judgments about adversary intent are generally more uncertain and tentative than judgments about adversary capability, especially in those cases where the buildup of capability requires long lead times. Judgments about intent also influence assessments of likelihood.

Threat assessments are intended to help policymakers understand the scope and nature of a threat and assist them in formulating appropriate responses to the threat. For example, a threat assessment may suggest that more resources should be deployed to combat threat X as compared to threat Y, or that strategy A would be more effective than strategy B in responding to a given threat.

To the extent that policymakers rely on a threat assessment that does not accurately characterize a threat, they may allocate resources suboptimally. Policymakers tend to regard as useless threat assessments that acknowledge large degrees of uncertainty and thus analysts have incentives to minimize the uncertainty expressed in an assessment.

Threat assessments are most prone to “inflation” (that is, exaggerated or worst-case depictions of a threat compared to the threat that actually exists) when the information supporting them is thin and analysts assume the worst about adversary capabilities for hostile action and the capabilities of a nation for defending itself.

For example, in 1967, President Lyndon B. Johnson noted:

> We've spent between thirty-five and forty billion dollars on space . . . but if nothing else had come from that program except the knowledge that we get from our satellite photography, it would be worth ten times to us what the whole program has cost. Because tonight I know how many missiles the enemy has and . . . our guesses were way off. And we were doing things that we didn't need to do. We were building things that we

> didn't need to build. We were harboring fears that we didn't need to have.[1]

In other words, in the absence of good knowledge about the number of enemy missiles, the United States was building up its military forces unnecessarily (i.e., to a level that provided military capability in excess of what was warranted by the actual threat). This tendency for analysts to prepare worst-case estimates is well-known in the intelligence community.[2] Many analysts believe they have an obligation to present the "worst-case" scenario so that policymakers will know the outer limits of the harm that the United States could face under any set of circumstances. But at the same time, the worst-case scenario is rarely the same as the "most likely" scenario, which by definition paints a picture that is less dire than (or, at most, only equally as dire as) the worst-case scenario. Policymakers are thus likely to be led into overreaction by reliance on the worst-case scenario, just as President Johnson indicated.

One key element driving worst-case analysis is the sole focus on adversary capabilities and intent. When they focus only on capabilities, analysts omit the adversary's operational skill (also known as tradecraft) from the scope of their analysis and necessarily assume that the adversary will not make mistakes. Because the adversary could perform its missions without error and exploit its capabilities to their fullest, the analyst, under a worst-case analysis, must assume that it will do so.

An adversary's intent (what it would like to do) can change more rapidly than its capabilities (what it is capable of doing), and thus it is sometimes alleged that intent is more ephemeral. Nevertheless, analysts do pay significant attention to it. For example, Robert Jervis argues that, during the Cold War, analysts often exhibited perceptual vigilance—a high sensitivity to information that an undesired outcome was likely (in other words, worst-case assessment of Soviet intent).[3]

[1] Smithsonian, *Satellite Reconnaissance: Secret Eyes in Space*, NAT'L AIR & SPACE MUSEUM, http://www.nasm.si.edu/exhibitions/gal114/SpaceRace/sec400/sec400.htm (last visited Dec. 7, 2011).

[2] *See, e.g.*, Wayne G. Jackson, *Scientific Estimating*, 9 STUD. INTELLIGENCE (1965), *available at* https://www.cia.gov/library/center-for-the-study-of-intelligence/kent-csi/vol9no3/html/v09i3a02p_0001.htm (approved for release by the Central Intelligence Agency Historical Review Program on Sept. 22, 1993).

[3] ROBERT JERVIS, PERCEPTION AND MISPERCEPTION IN INTERNATIONAL POLITICS 372–78 (1976).

Robert Mandel notes that "[w]orst-case planning has especially been evident within the United States government since 9/11 due to the prevailing view that the American intelligence community underestimated Al Qaeda's capabilities to carry out the terrorist attack."[4] In short, Mandel argues that intelligence analysts and national security policymakers have "a natural tendency to . . . hedge their bets by thinking in worst-case terms" regarding adversary intent.[5] Mandel cites Barry Buzan as noting that the very existence of the nation-state as a political entity depends on the presence of security threats that call for collective, rather than individual, response.[6]

II. INFORMATION ON ADVERSARY PREPARATIONS FOR HOSTILE ACTION IN CYBERSPACE

Information regarding adversary preparations for action in cyberspace is very difficult to obtain.

A. *Information on Adversary Cyber-Weapons*

One basic element of threat assessment is information on the capabilities of adversary weapons. For traditional kinetic weapons, some plausible "ball-park" estimates of performance can be generated based on our own experience with comparable weapons. For example, we know that any plausible adversary tank is not likely to travel on the ground through rough terrain at 100-mph speeds and that an adversary fighter jet will not travel at Mach 5.

But such confidence cannot extend to capability assessment of cyber-weapons. Cyber-weapons—a term which is used in this paper to refer to an information technology artifact that is designed to cause harm to an information technology system or the information resident in the system—are closer to thought than to material object.

[4] Robert Mandel, *On Estimating Post-Cold War Enemy Intentions*, 24 INTELLIGENCE & NAT'L SEC. 194, 197 (2009).

[5] *Id.*

[6] *Id.* at 197, n. 16 (citing BARRY BUZAN, PEOPLE, STATES AND FEAR: AN AGENDA FOR INTERNATIONAL SECURITY STUDIES IN THE POST-COLD WAR ERA 140–41 (2d ed. 1991)).

For example, Frederick P. Brooks, Jr., architect and manager of the legendary OS/360 operating system for the IBM System/360 mainframe computer, has written that:

> The programmer . . . works only slightly removed from pure thought-stuff. He builds his castles in the air, . . . creating by exertion of the imagination. . . . Yet the program construct . . . is real in the sense that it moves and works, producing visible outputs separate from the construct itself. . . . The magic of myth and legend has come true in our time. One types the correct incantation on a keyboard, and a display screen comes to life, showing things that never were nor could be.[7]

Thus, understanding what a cyber-weapon can actually do requires a direct inspection of the artifact that reveals its innermost workings. It is as though one would not know anything—literally anything—about a tank's capabilities without having the tank in front of you and taking it apart piece by piece and reverse engineering what every single component does and what the entire assembly does when every component is in place.

To illustrate, consider the Sapphire/Slammer worm of January 2003:

> [T]he Sapphire worm was the fastest computer worm in history (infecting more than 90 percent of vulnerable hosts within 10 minutes)[—]a defective random number generator [(RNG)] significantly reduced its rate of spread. (The worm targeted IP addresses chosen at random, and the [RNG] produced [addresses] that were improperly restricted in range.)[8]

The reach and impact of this worm would have been vastly increased if its RNG were working properly—and the only way that one would know of its defective RNG is by inspecting the worm's code.

[7] FREDERICK P. BROOKS, JR., THE MYTHICAL MAN-MONTH: ESSAYS ON SOFTWARE ENGINEERING 7–8 (1975).

[8] NRC CYBERATTACK REPORT, *supra* note 1, at 122 (footnote omitted) (citing David Moore et al., *The Spread of the Sapphire/Slammer Worm*, *available at* http://www.caida.org/publications/papers/2003/sapphire/sapphire.html).

Furthermore, improvements to the code of a cyber-weapon can, in principle, be implemented quickly—much more rapidly than improvements to traditional kinetic weapons, which are produced on an assembly line. Changes in the manufacturing process are necessarily infrequent—a tank that rolls off the assembly line is very similar to every other tank. Improvements are sometimes made, of course, but such improvements tend to be incremental in nature.

A second source of uncertainty regarding adversary weapons is the relationship between the characteristics of the target and the characteristics of the weapon. Weapons seek to cause effects to targets and thus the characteristics of a target affect the performance of a weapon.

In the kinetic world, operational planners must match the characteristics of a weapon (e.g., explosive yield, fusing, and likely miss distances) against target characteristics (e.g., target hardness, size, and shape) and characteristics of surrounding environment (e.g., terrain and weather).[9] Weapons effects in the kinetic world can in principle be calculated on the basis of computational models that are based on physics-based algorithms. Because the fundamental physics of explosives technology and of most targets is well known and kinetic effects on a given target can be calculated with a high degree of confidence, these calculations can be empirically validated (e.g., at test ranges where weapons can be directed against high-fidelity replicas of targets).

> But there is no comparable formalism for understanding the effects of cyberweapons. The smallest change in the configuration and interconnections of [the target] IT system can result in completely different system behavior, and the direct effects of a cyberattack on a given system may be driven by the behavior and actions of the human system operator and the specific nature of that system as well as the intrinsic characteristics of the cyberweapon involved.[10]

Thus, a threat assessment is necessarily dependent on assumptions about the cyber-hardness of the target. Because such knowledge *is*

[9] *Id.* at 122.

[10] *Id.*

available (the target belongs to the same nation that employs the analyst), many analysts are likely to assume that the defense will exhibit its average (that is, its most-likely) performance.[11]

Thus, threat assessments for cyberspace threats are likely to depict a match between an adversary armed with every weapon it could possibly have, each with the maximum possible individual capability, operated perfectly in an error-free environment against a defender with known capabilities that operates in an average manner. Under such conditions, an analysis of an adversary and a defender who are evenly matched will always show the adversary to be superior.

B. *Information About an Adversary's Order of Battle*

As used by some components of the U.S. Department of Defense, the term "order of battle" refers to information about an adversary's combat capability.[12] Order of battle traditionally includes information regarding the adversary's strength, composition, tactics, and training.[13]

In assessing strength for kinetic conflict, intelligence analysts consider the number and capabilities of adversary weapons. Because kinetic weapons are tangible objects, they can, in principle, be counted as they come off the assembly line and are more or less identical to each other. Even the number of bullets or bombs manufactured can be counted. Although research and development is important for these weapons, the production line is what counts for combat strength because a larger number of weapons results in greater strength.

But cyber-weapons—information technology artifacts, especially software artifacts—are of a very different nature. Cyber-weapons instantiated as software can be reproduced at zero incremental cost and time and "bean counts" of such weapons would more logically count different types of weapons for different purposes (on the assumption that such counts would be meaningful at all). In other cases, a cyber-weapon may be designed to strike at very specific targets and to ignore any others. Such a weapon would not have any generally usable capability and so counting the number of such

[11] *See* Richmond M. Lloyd, *Force Planning for the 1990s*, *in* FUNDAMENTALS OF FORCE PLANNING, VOL. 1: CONCEPTS 105 (Naval War Coll. Force Planning Faculty ed., 1990).

[12] E.g., U.S. DEP'T OF THE ARMY, FIELD MANUAL 34-3: INTELLIGENCE ANALYSIS 3-1 (1990).

[13] *Id.*

weapons of different types would not provide any indication of the adversary's strength.

With both capability and quantity of cyber-weapons being difficult to ascertain or even to define, an intelligence analyst might examine personnel trained to conduct cyber-operations. For example, one might try to determine the number of individuals graduating with degrees in computer science who subsequently serve in an adversary's armed forces. Alas, this approach is also flawed because the potency of a cyber-attack is almost certainly far more a function of the skill and expertise level of the best adversary operators than of their sheer number.[14] And obtaining good information on such intangible factors is difficult indeed.

In some cases, even knowing who counts as "someone serving in an adversary's armed forces" is problematic. A recent article in *Foreign Policy*, for example, suggests the existence of a large number of "patriotic hackers" in China who may conduct cyber-attacks to further Chinese interests, but who operate without close coordination with the Chinese government.[15] Assuming this report to be true, who, if any, of these patriotic hackers should be counted as part of Chinese military forces, even if they could be identified?

As for adversary tactics, tactics for pursuing kinetic conflict can, in principle, be observed and thus information gained. Military units deploy and exercise, for example, but exercises involving cyber-operations are difficult if not impossible to observe if they occur on systems and networks controlled by the adversary. Some analysts suggest that at least some the hostile cyber-operations experienced by the United States may in fact be adversaries conducting cyber-exercises and thus that the United States can provide useful intelligence information on tactics that might be used during a real cyber-conflict. On the other hand, to the extent that such exercises were successful in penetrating U.S. cyber-defenses and that such successes were known to the United States, it is likely that the specific weaknesses in U.S. defenses would be remediated. This suggests that, in a real conflict, an adversary is likely to use cyber-tactics that have not been seen before. (The same argument holds for cyber-weapons for the same reasons and, in a real conflict, it is likely that an

[14] A suggestive data point is the fact that the productivity difference between the best and worst programmers exceeds a factor of ten. *See, e.g.*, BARRY W. BOEHM ET AL., SOFTWARE COST ESTIMATION WITH COCOMO II (2000).

[15] Mara Hvistendahl, *China's Hacker Army*, FOREIGN POLICY, Mar. 3, 2010, *available at* http://www.foreignpolicy.com/articles/2010/03/03/china_s_hacker_army.

adversary will use cyber-weapons with capabilities that have not been seen before.)

C. *Information About Adversary Intent*

Identifying predictive analysis with forecasts of adversary intent and courses of action, Joint Publication 2-0 notes that predictive analysis "is not an exact science and is vulnerable to incomplete information, adversary deception, and the paradox of warning."[16] The publication further notes that predictive analysis is more difficult and risky than assessments of adversary capabilities because the former "deals more extensively with the unknown" and thus "the chances of analytic failure are greater."[17] Drawing on interviews with uniformed and civilian intelligence officers, Gary Schaub, Jr. echoes this point, noting that these individuals believe that "producing such analyses [of adversary intent] is considered more of an art than a science."[18]

Absent specific information about adversary intent, analysts must draw more heavily on general principles that are plausibly relevant to conditions in the environment of today. For example, Robert Mandel argues this point in the context of the post-Cold War environment:

> [I]nternational coercion has gone well beyond formal threats of direct military attack from states and often has taken on the guise of far more subtle and varied unorthodox modes of disruption by non-state groups. Emerging threats have been typically covert, dispersed, decentralized, adaptable, and fluid, with threat sources

[16] JOINT CHIEFS OF STAFF, JOINT PUBLICATION 2-0: JOINT INTELLIGENCE II-10 (2007), *available at* http://www.dtic.mil/doctrine/new_pubs/jp2_0.pdf.

[17] *Id.*

[18] Gary Schaub, Jr., *When Is Deterrence Necessary?: Gauging Adversary Intent*, 3 STRATEGIC STUD. Q. 49, 67–68 (2009).

> relatively difficult to identify, monitor, target, contain, and destroy.[19]

That is, the utility of war between major powers has declined due to norms against interstate aggression and adversaries are thus more likely to contemplate unorthodox aggressive actions short of direct military invasion.

Offensive cyber-operations undertaken by adversaries certainly fall into this category. Although much effort has been expended on trying to understand what might define an "armed attack" or a "use of force" in cyberspace (terms used in the U.N. Charter[20]), experience suggests that few, if any, hostile or unfriendly actions in cyberspace have risen to the threshold that would lead many nations to assert that uses of force or armed attacks have occurred. Mandel's argument above would suggest that hostile actions in cyberspace below U.N. Charter thresholds of what is forbidden are well suited as instruments of adversarial competition between states and the canons of worst-case thinking imply that adversaries will exploit all possible instruments to gain advantage whenever possible.

Given the high and increasingly growing U.S. dependence on information technology for both military and civilian purposes, it is logical to conclude that adversaries would target U.S. information technology whenever possible. Furthermore, given overwhelming U.S. military advantages in traditional spaces of military competition, adversaries would be highly motivated to conduct asymmetric warfare against the United States in a conflict—warfare that takes advantage of specific U.S. vulnerabilities, such as those in cyberspace. Furthermore, they would be likely to draw from adversary writings for information to support (or confirm) their analysis and dismiss writings contradicting their analysis as merely political propaganda intended

[19] Mandel, *supra* note 5, at 195.

[20] U.N. Charter, art. 2, para. 4, prohibits nations from using "the threat or use of force against the territorial integrity or political independence of any state, or in any other manner inconsistent with the Purposes of the United Nations." The Charter also contains two exceptions to this prohibition, permitting the Security Council to authorize uses of force in response to "any threat to the peace, breach of the peace, or act of aggression," U.N. Charter art. 39, in order "to maintain or restore international peace and security," U.N. Charter art. 42; and U.N. Charter art. 51 provides as follows: "Nothing in the present Charter shall impair the inherent right of individual or collective self-defen[s]e if an armed attack occurs against a Member of the United Nations, until the Security Council has taken measures necessary to maintain international peace and security." The self-defense contemplated by Article 51 does not require Security Council authorization.

to lull the U.S. into complacency. A common approach is to interpret an adversary's military doctrine as indicating that adversary's intent.

Further complicating an assessment of adversary intent in cyberspace are two other realities: First, even knowing that the United States as a nation is the victim of hostile cyber-operations originating with a specific adversary is problematic.[21] Various entities within the United States, belonging both to the private and public sectors, are subject to a variety of hacking activities, virus propagation, distributed denial-of-service attacks, and other activities conducted for many possible purposes, including illicit monetary gain, sport, or pure maliciousness. Of course, covert intelligence gathering or "preparing the battlefield" for possible future attacks may be going on as well. But knowing which hostile operations are associated with which purpose is highly problematic because analysis of the hostile cyber-operations themselves in the absence of contextual factors (such as the party responsible for them, as discussed below) is unlikely to indicate how any given operation fits into a bigger picture that might indicate a serious national security threat.

Second is the well-known problem of attribution, which is the effort to identify the party responsible for a cyber-operation. As discussed in Chapter 2 of the NRC Cyberattack Report, technical attribution of a cyber-attack is very difficult to perform effectively.[22] For example, an attempt to identify the original perpetrators of a hostile operation against the United States might find that the proximate source of the operation was computers located in another nation, which will be called Zendia in the discussion below. However, there may well be no technical way to differentiate among a number of different scenarios consistent with this discovery. These scenarios include the following:

- The attack against the United States was launched by agents of the Zendian government with the approval of the Zendian national command authority;

[21] *See* NRC CYBERATTACK REPORT, *supra* note 1, at 79–160.

[22] *See id.* at 138–39 ("Technical attribution is the ability to associate an [operation] with a responsible party through technical means based on information made available by the fact of the [operation] itself—that is, technical attribution is based on the clues available at the scene (or scenes) of the attack.").

> - The attack against the United States was launched by low-level agents of the Zendian government without the approval or even the knowledge of the Zendian national command authority;
>
> - The attack was launched through the efforts of computer-savvy citizens of Zendia who believe that the United States oppresses Zendia in some way. Although the efforts of these citizens are not initiated by the Zendian government, the Zendian government takes no action to stop them;
>
> - The Zendian computers used to conduct the attack against the United States have been compromised by parties outside Zendia (perhaps even from the United States . . .), and Zendia is merely an innocent bystander on the international stage;
>
> - The attack was launched at the behest of the Zendian government, but not carried out by agents of the Zendian government. For example, it may have been carried out by the Zendian section of an international criminal organization.[23]

By taking into account information from other sources and not just technical sources at the scene of the operation, one might obtain enough information to make plausible judgments concerning the identity of the responsible party. For example, other useful information might be available from intelligence sources inside possible adversary governments, other technical information (e.g., similarities between a given operation and previous operations that had been attributed or mistakes made by the perpetrator), and temporal proximity to other coercive or aggressive actions that can be attributed to an actor. But the operative term in attribution based on

[23] *Id.* at 139–40 (footnote omitted).

all-source analysis is “judgment,” as contrasted to definitive and certain proof.

To summarize, the environment for assessing adversary intent in cyberspace is one in which there are incentives for analysts to draw worst-case conclusions. Various entities within the nation are subject continually to hostile cyber-operations of unknown origin and intent. Technical attribution to specific adversaries is exceedingly difficult and all-source attribution is based on human judgments about an adversary. Overlaid on all of these factors is a high degree of uncertainty in the information used to generate the assessment and thus a high degree of uncertainty in the assessment itself.

III. RHETORIC AND THE CONSUMERS OF THREAT ASSESSMENTS

Contributing to an atmosphere of hyperbole regarding threats in cyberspace is imprecise use of terminology. In particular, hostile cyber-operations conducted by adversaries are often lumped together under the generic label of “cyber-attack.” In fact, a cyber-operation could be a cyber-attack or a cyber-exploitation.

- Cyber-attack: the use of deliberate actions against adversary computer systems or networks to alter, disrupt, deceive, degrade, or destroy these systems or networks or the information and/or programs resident in or transiting these systems or networks. A cyber-attack seeks to cause adversary computer systems and networks to be unavailable or untrustworthy and therefore less useful to the adversary;

- Cyber-exploitation: the use of deliberate actions against adversary computer systems or networks to obtain putatively confidential information resident on or transiting through these systems or networks. Cyber-exploitations do not seek to disturb the normal functioning of a computer system or network from the user’s point of view. Cyber-exploitation is, in essence, a form of espionage.

As noted in Chapter 2 of the NRC Cyberattack Report, cyber-attack and cyber-exploitation are often conflated in public discourse and, in particular, cyber-exploitations are reported and discussed using the term "cyber-attack." Such conflation can be seen in Congressional debate,[24] press articles,[25] and even in notices from national laboratories.[26] Conflation of these terms in the public

[24] For example, Representative Frank Wolf (R-VA) stated on the House floor in June 2008:

> [I]n August 2006, four of the computers in my personal office were compromised by an outside source. . . . On these computers was information about all the case work I've done on behalf of political dissidents and human rights activists around the world. . . . [T]he FBI . . . revealed that the outside sources responsible for this attack came from within the People's Republic of China.

154 CONG. REC. H5281 (daily ed. June 11, 2008) (statement of Rep. Wolf), *available at* http://www.gpo.gov/fdsys/pkg/CREC-2008-06-11/pdf/CREC-2008-06-11-pt1-PgH5280.pdf (emphasis added).

[25] For example, a 2005 Time magazine article stated:

> Carpenter had never seen hackers work so quickly, with such a sense of purpose. They would commandeer a hidden section of a hard drive, zip up as many files as possible and immediately transmit the data to way stations in South Korea, Hong Kong or Taiwan before sending them to mainland China. They always made a silent escape wiping their electronic fingerprints clean and leaving behind an almost undetectable beacon allowing them to re-enter the machine at will. An entire attack took 10 to 30 minutes.

Nathan Thornburgh, *The Invasion of the Chinese Cyberspies*, TIME, Aug. 29, 2005, *available at* http://www.time.com/time/magazine/article/0,9171,1098961,00.html (emphasis added).

[26] In December 2007, the Oak Ridge National Laboratory posted a notice labeled "Potential Identity Theft" stating:

> Oak Ridge National Laboratory (ORNL) recently experienced a sophisticated cyber attack that appears to be part of a coordinated attempt to gain access to computer networks at numerous laboratories and other institutions across the country. A hacker illegally gained access to ORNL computers by sending staff e-mails that appeared to be official legitimate communications. When the employees opened the attachment or accessed an embedded link, the hacker planted a program on the employees' computers that enabled the hacker to copy and retrieve information. The original e-mail and first potential

discourse tends to overstate the actual threat, thus inflaming public passions and beating the drums of war unnecessarily. It is certainly true that cyber-exploitations are not friendly acts, but they are not armed attacks either. Most nations engage in espionage against each other and such actions do not lead to war or even armed conflict. Indeed, espionage does not even constitute a violation of international law.

From the perspective of making policy, such impolitic language is part of shaping the environment in which threat assessments from professional intelligence analysts are received. Although professional analyses are likely to be more sober and precise in their use of language than news reports, the policy-making consumers of threat assessments are, in general, non-specialists regarding the subject of these analyses and it would be surprising if they themselves were not influenced by the larger public discourse.

IV. Discussion and Conclusion

Section I pointed to influences on analysts to draw worst-case conclusions. The discussion of Sections II.A and II.B suggests that threats in cyberspace are much more tenuous, ephemeral, and uncertain than those found in traditional domains of military conflict. Thus, the influences pushing towards worst-case analysis originating from information scarcity apply even more strongly when cyberspace is the domain in question. Indeed, given the shadowy nature of cyber-operations, it is not unreasonable to regard them as actions that compromise knowledge, information, and certainty.

When inexperienced human beings with little hard information are placed into unfamiliar situations in a general environment of tension, they will often make worst-case assessments. In the words of a former senior Justice Department official involved with critical infrastructure protection, "I have seen too many situations where government officials claimed a high degree of confidence as to the source, intent, and scope of an attack, and it turned out they

> corruption occurred on October 29, 2007. We have reason to believe that data was stolen from a database used for visitors to the Laboratory.

OAK RIDGE NAT'L LAB., U.S. DEP'T OF ENERGY, *Potential Identity Theft*, http://www.ornl.gov/identifytheft (last visited Dec. 7, 2011) (emphasis added).

were wrong on every aspect of it. That is, they were often wrong, but never in doubt."[27]

The above paragraph refers to situations in which policymakers are responding to reports of hostile operations against U.S. interests, such as protection of U.S. critical infrastructure, often in an atmosphere of crisis. But the reader might consider the possibility that such thinking is also influential in conducting threat assessments under non-crisis conditions. That is, in the absence of hard information, there is little that can be done analytically to argue against worst-case threat assessments of intent and, without countervailing pressures, threat assessments about adversary intent in cyberspace could be expected to drift towards more and more dire predictions.

None of these comments should be interpreted to mean that the threat to U.S. interests emanating from cyberspace is not serious. Today, the debate over the defensive cybersecurity posture is between those who think the present situation is dire and those who think it is very dire. No analyst argues that the gap between the defensive cybersecurity posture of the United States and the threats it faces is shrinking; the only serious debate is over how fast that gap is growing.

Against this backdrop, what should the policy-making consumer of threat assessments in cyberspace make of the assessments they receive? Although a capabilities assessment that pits adversary weapons that perform perfectly against friendly defenses that demonstrate only average performance may be misleading as an indicator of likely outcomes, worst-case assessments of adversary intent may in fact be reflected in reality. That is, a "God's eye view" of the intent of U.S. adversaries may indeed be as bleak as a worst-case assessment would portray.

What should policymakers do in the face of such uncertainty? Despite very high degrees of uncertainty about the scope and nature of the cyber-threat, policymakers must still act and nothing in this analysis suggests that uncertainty should paralyze the decisionmaking process. Threat assessments are undertaken for many purposes, but one of the most important is to inform the preservation or maintenance of some important functionality. The presence of an adversary threat by definition endangers some U.S. functionality and a worst-case threat assessment reflects a subjective estimate that the importance of the functionality at risk is high (i.e., the likelihood of a

[27] *See* NRC CYBERATTACK REPORT, *supra* note 1, at 142.

bad outcome is high and/or the magnitude of the loss implied by that bad outcome is high).

As a matter of logic, there are two (not necessarily mutually exclusive) responses to a threat: (1) to defend that functionality by keeping the adversary at bay, or Category 1; and (2) to develop a capability for working around the loss of that functionality, or Category 2.

- Category 1 are responses such as passive defense measures (e.g., intrusion and anomaly detection and more robust software that is less likely to contain security vulnerabilities) and anticipatory measures (e.g., obtaining early warning of an adversary's intent to attack and anticipating the attack through an immediate change in defensive posture or a preemptive attack that degrades the adversary's ability to carry out an attack);

- Category 2 are responses such as measures to enhance rapid recovery (e.g., capabilities for rapid rebooting of affected computer systems and rapid restoration of data after a compromise has been detected) and resilience (e.g., capabilities for dropping lower priority functionalities while under attack); to deploy backup or alternative capabilities; and to train organizations that might be affected by the loss of cyber-functionality to work without it. While one cannot expect that such organizations would be able to continue to function at peak effectiveness, it is surely not unreasonable to expect that they should be able to perform at least some of their critical functions by carrying out emergency procedures manually.

How much should be invested in Category 1 vis-à-vis Category 2, and how much should be invested in their aggregate? In the absence of metrics that tie investment to capability (a difficult problem that has bedeviled the cybersecurity community for forty years and remains unsolved today), the answers to these questions cannot be found through the sort of quantitative analysis used in dealing with kinetic threats.

In the absence of a more traditional quantitatively analytical basis for making investment decisions, policymakers ask questions such as "how much seems reasonable to spend on X vs. Y given all of the budget constraints?" Such "level of effort" decisions are necessarily based on assessments of likelihood and consequences.

If a policymaker believes that worst-case outcomes are likely, then the balance of investments in Categories 1 and 2 should probably be more tilted in favor of investments in Category 2. That is, if the worst-case outcomes are more likely to occur, the nation should be better prepared to do without those capabilities and/or to reconstitute them quickly.

At a high level of abstraction, the thoughts offered in this note are not new—many researchers and commentators have noted the propensity of many analysts to offer worst-case analyses that paint an excessively pessimistic picture regarding adversaries. Others, notably Robert Jervis, have pointed to psychological factors that may account, at least in part, for such tendencies. This note draws on such work to underscore the importance of understanding these factors in a problem domain—that of cyberspace—which is even more subject to the uncertainties and incomplete information that characterize traditional subjects of analysis and thus that policymakers interpreting threat assessments would be well advised to consider the influences that often push analysts towards worst-case scenarios.