Escalation Dynamics and Conflict Termination in Cyberspace

Herbert Lin

US national security planners have become concerned in recent years that this country might become engaged in various kinds of conflict in cyberspace. Such engagement could entail the United States as the target of hostile cyber operations, the initiator of cyber operations against adversaries, or some combination of the two.

To date, most serious analytical work related to cyber conflict focuses primarily on the initial transition from a preconflict environment to that of conflict. Little work has been done on three key issues: (1) how the initial stages of conflict in cyberspace might evolve or escalate (and what might be done to prevent or deter such escalation), (2) how cyber conflict at any given level might be deescalated or terminated (and what might be done to facilitate deescalation or termination), and (3) how cyber conflict might escalate into kinetic conflict (and what might be done to prevent kinetic escalation). Each of these issues is important to policymakers, both in preparing for and managing a crisis. Before beginning that discussion, it is instructive to consider some relevant terminology and concepts.

Terminology and Basic Concepts

The term *offensive cyber operations* as used here refers collectively to actions taken against an adversary's computer systems or networks that harm the adversary's interests. In general, an offensive cyber operation

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. Of particular note is his role as editor of a 2009 NRC study on cyber attack as an instrument of national policy and a 2010 study on cyber deterrence. He previously served as staff scientist for the House Armed Services Committee (1986–90), where his portfolio included defense policy and arms control issues. He received his doctorate in physics from MIT.

This article is largely based on chapter 9 of *Technology, Policy, Law, and Ethics Regarding U.S. Acquisition and Use of Cyberattack Capabilities* by William Owens, Kenneth Dam, and Herbert Lin (Washington: National Academies Press, 2009). The author is solely responsible for any deviation from the conclusions of that report. This article does not necessarily reflect the views of the research sponsors, the MacArthur Foundation or the Microsoft Corporation.

gains access to an adversary's computer system or network and takes advantage of a vulnerability in that system or network to deliver a payload. In a non-cyber analogy, *access* might be any available path for reaching a file in a file cabinet. A *vulnerability* might be an easy-to-pick lock on the file cabinet—and note that ease of picking the lock is irrelevant to an Earth-bound intruder if the file cabinet is located on the International Space Station where access to the file cabinet would be difficult. The *payload* describes what is to be done once the intruder has picked the lock. For example, the intruder can destroy the papers inside, alter some of the information on those papers, or change the signature on selected documents.

Access is "easy" when a path to the target can be found without much difficulty; a computer connected to the Internet may well be such a target. Access is "difficult" when finding a path to the target is possible only at great effort or may not be possible for any practical purposes. An example of such a target may be the onboard avionics of an enemy fighter plane, which is not likely to be connected to the Internet for the foreseeable future. In general, access to an adversary's important and sensitive computer systems or networks should be expected to be difficult. Furthermore, access paths to a target may be intermittent—a submarine's on-board administrative local area network would necessarily be disconnected from the Internet while underwater at sea but might be connected while in port. If the administrative network is ever connected to the on-board operational network (controlling weapons and propulsion) at sea, an effective access path may be present for an adversary.

A *vulnerability* is a security weakness in the system or network that is introduced by accident (by some party that has a legitimate reason to access the system) or on purpose (by a would-be intruder). An accidentally introduced weakness (a "security bug") may open the door for opportunistic use of the vulnerability by an adversary. Many vulnerabilities are widely publicized after they are discovered and may be used by anyone with moderate technical skills until a patch can be disseminated and installed.[1] Adversaries with the time and resources may also discover unintentional defects that they protect as valuable secrets—also known as *zero-day vulnerability*.[2] A deliberately introduced vulnerability occurs because the intruder takes an action to create one where one did not previously exist. For example, an intruder might deceive a legitimate user of the targeted system or network to disable a security feature (e.g., reveal a password). Both kinds of vulnerability are useful to intruders as long as the weaknesses introduced remain unaddressed.

Payload is the term used to describe the things that can be done once a vulnerability has been exploited. For example, once a software agent (such as a virus) has entered a given computer, it can be programmed to do many things—reproduce and retransmit itself, destroy files on the system, or alter files. Payloads can have multiple capabilities when inserted into an adversary system or network—that is, they can be programmed to do more than one thing. The timing of these actions can also be varied.

Depending on the intent of the intruder, an offensive cyber operation can be classified as cyber attack or cyber exploitation. *Cyber attack* is the use of deliberate information technology (IT)–related actions—perhaps over an extended period of time—to alter, disrupt, deceive, degrade, or destroy adversary computer systems or networks or the data and/or programs resident in or transiting these systems or networks.[3] Such effects on adversary systems and networks may also have indirect effects on entities coupled to or reliant on them. A cyber attack seeks to cause adversary computer systems and networks to be unavailable or untrustworthy and therefore less useful to the adversary. Because so many different kinds of cyber attack are possible, the term *cyber attack* should be understood as a statement about a methodology for action—and that alone—rather than as a statement about the scale of the effect of that action. *Cyber exploitation* is the use of deliberate IT-related actions—perhaps over an extended period of time—to support the goals and missions of the party conducting the exploitation, usually for the purpose of obtaining information resident on or transiting through an adversary's computer system or network. Cyber exploitations do not seek to disturb the normal functioning of a computer system or network from the user's point of view—indeed, the best cyber exploitation is one that goes undetected.

The similarity between these two concepts and the exploitation channel are the most important characteristics of offensive cyber operations. *Cyber attack* and *cyber exploitation* are very similar from a technical point of view. They use the same access paths and take advantage of the same vulnerabilities; the only difference is the payload they carry. These similarities often mean that the targeted party may not be able to distinguish easily between cyber exploitation and cyber attack—a fact that may result in that party's making incorrect or misinformed decisions. The primary technical requirement of cyber exploitation is that delivery and execution of its payload be accomplished quietly and undetectably. Secrecy is often far less important when cyber attack is the mission, because in many cases the effects of the attack will be immediately apparent to the target. All exploitation operations require a channel for reporting the information they collect. If the channel happens to be two-way, payloads can be remotely updated. Thus, the functionality of the operation may be different today than it was yesterday—most significantly, it may be an exploitation payload today and an attack payload tomorrow. In some cases, the initial payload consists of nothing more than a mechanism for scanning the system to determine its technical characteristics and an update mechanism to retrieve the best packages to further the compromise.

Attribution

Attribution is the task of identifying the party that should be held politically responsible for an offensive cyber operation.[4] *Technical attribution* is the ability to associate an attack with a responsible party through technical means based on information made available by the cyber operation itself—that is, technical attribution is based on clues available at the scene (or scenes) of the operation. *All-source attribution* is a process that integrates information from all sources, not just technical sources at the scene of the attack, to arrive at a judgment (rather than a definitive and certain proof) concerning the identity of the intruder.

As a general rule, attribution is a difficult matter. It becomes more difficult as more of the following factors are present:

- The techniques used have never been seen before, so the investigator is unable to link them to other parties that have used similar techniques in the past.
- The intruder leaves no forensic clues and makes no technical mistakes (i.e., tradecraft is error free).
- The intruder maintains perfect operational security, so there are no other sources of intelligence (e.g., SIGINT, HUMINT).
- The motivations for conducting the operation are unknown, or the operation occurs during a time when political circumstances do not suggest conflict or adversarial relations to associate a known party's demands or interests with a possible perpetrator.

- The intrusion requires a rapid response which prevents a thorough investigation, raising the likelihood of a mistaken attribution.

If most or all of these factors are present, then attribution is virtually impossible. On the other hand, it is rare that *all* of these factors are present. One might thus reasonably conclude that although technical attribution is indeed difficult, all-source attribution is sometimes possible. Solving the problem of attribution is not as hopeless as is often portrayed.

The Need for Intelligence Support

Offensive cyber operations against a given system require detailed knowledge about both access paths to and vulnerabilities in the targeted system. The amount of detail should not be underestimated—in principle, it may involve very "small" details such as

- the specific processor model (and even the serial number of the processor) in use on the system;
- the operating system in use, down to the level of specific version, the build number in use, and the history of security patches applied to it;
- IP addresses of Internet-connected computers;
- specific versions of systems administrator tools used;
- the security configuration of the operating system (e.g., whether certain services are turned on or off, or what antivirus programs are running); and
- the physical configuration of the hardware involved (e.g., what peripherals or computers are physically attached).

Note that none of these items of intelligence is easily available from satellite or aerial reconnaissance. As a general rule, a scarcity of intelligence regarding possible targets means that any offensive cyber operation launched against them can only be a "broad-spectrum" and a relatively indiscriminate or blunt attack. Such an attack might be analogous to the Allied strategic bombing attacks of World War II that targeted national infrastructure on the grounds that such infrastructure supported the war effort of the Axis. Substantial amounts of intelligence information about targets and paths to those targets are required if the operation is intended as a very precise one directed at a particular system. Conversely, a lack of such information will result in large uncertainties about the direct and indirect effects of an operation and make it difficult to develop accurate estimates of likely collateral damage.

Active Defense

Defensive measures in cyber security seek to frustrate offensive operations taken against systems or networks**.** Passive defensive measures, such as hardening systems against penetration, facilitating recovery in the event of a successful offensive operation, making security more usable and ubiquitous, and educating users to behave properly in a threat environment, are important elements of a strong defensive posture.[5] Nevertheless, for the defense to be successful, these measures must succeed every time an adversary attacks. The offensive operation need only succeed once, and an adversary who pays no penalty for a failed operation can continue with follow-on operations until it succeeds or chooses to stop. This places a heavy and asymmetric burden on a defensive posture that employs only passive defense.

If passive defense is insufficient to ensure security, what other approaches might help to strengthen one's defensive posture? One possibility is to eliminate or degrade an adversary's ability to successfully conduct offensive cyber operations. In that case, the operation is ultimately less successful than it might otherwise have been because the defender has been able to neutralize the operation in progress or perhaps even before it was launched.

A second possibility is to impose other costs on the adversary, and such a strategy is based on two premises. First, imposition of these costs reduces the adversary's willingness and/or ability to initiate or to continue an offensive operation. Second, knowledge that an operation will prove costly to one adversary deters others from attempting to conduct similar operations—and advance knowledge of such a possibility may deter the original adversary from conducting the offensive operation in the first place. There are many options for imposing costs on an adversary, including economic penalties such as sanctions, diplomatic penalties such as breaking of diplomatic relations, and even kinetic military actions such as cruise missile strikes. In-kind military action—a counteroffensive cyber operation—is also a possibility.

Both of these possible reactions—neutralization of an adversary's offensive operation and imposition of costs to the adversary for the operation—are often captured under the rubric of *active defense*. But note well—the attempt to impose costs on an adversary that conducts offensive cyber operations might well be seen by that adversary as an offensive act itself. This may be especially true in the fog of cyber conflict, where who is actually doing what may be uncertain.

Evolving or Escalating Conflict

The phenomenon of escalation is a change in the level of conflict (where level is defined in terms of scope, intensity, or both) from a lower (perhaps nonexistent) to a higher level. Escalation is a fundamentally interactive concept in which actions by one party trigger other actions by another party to the conflict. Of particular concern is a chain reaction in which these actions feed off one another, thus raising the level of conflict to a level not initially contemplated by any party to the conflict. Escalation can occur through a number of mechanisms which may or may not be operative simultaneously in any instance.[6] It includes four basic types: deliberate, inadvertent, accidental, and catalytic.

Deliberate escalation is carried out with specific purposes in mind. For example, a party may deliberately escalate a conflict from some initial level (which may be zero) to gain advantage, to preempt, to avoid defeat, to signal an adversary about its own intentions and motivations, or to penalize an adversary for some previous action. Offensive cyber operations—specifically, cyber attacks—are one of many possible military options for deliberate escalation.

Inadvertent escalation occurs when one party deliberately takes actions that it does not believe are escalatory but which are interpreted as escalatory by another party to the conflict. Such misinterpretation may occur because of incomplete information, lack of shared reference frames, or one party's thresholds or "lines in the sand" of which other parties are not aware. Communicating to an adversary the nature of any such thresholds regarding activity in cyberspace may be particularly problematic, even under normal peacetime circumstances.

For example, Nation A does X, expecting Nation B to do Y in response. But in fact, Nation B unexpectedly does Z, where Z is a much more escalatory action than Y. Or Nation A may do X, expecting it to be seen as a minor action intended only to show mild displeasure and that Nation B will do Y in response, where Y is also a relatively mild action. However, due to a variety of circumstances, Nation B sees X as a major escalatory action and responds accordingly with Z, an action that is much more significant than Y. Nation A perceives Z as being way out of proportion and, in turn, escalates accordingly.

Accidental escalation occurs when some operational action has direct effects that are unintended by those who ordered them. A weapon may go astray to hit the wrong target; rules of engagement are sometimes unclear; a unit may take unauthorized actions; or a high-level command decision may not be received properly by all relevant units. It is especially relevant here that there is often greater uncertainty of outcome due to a lack of adequate intelligence on various targets when certain kinds of offensive cyber operations are employed.

Catalytic escalation occurs when some third party succeeds in provoking two parties to engage in conflict. For example, Party C takes action against Party A that is not traced to Party C and appears to come from Party B. Party A reacts against Party B, which then believes it is the target of an unprovoked action by Party A. The inherent anonymity of cyber operations may make "false-flag" operations easier to undertake in cyberspace than with kinetic operations.

Through such mechanisms, the escalatory dynamics of conflict show how a conflict, once started, might evolve. Of interest are issues such as what activities or events might set a cyber conflict into motion, what the responses to those activities or events might be, how each side might observe and understand those responses, whether responses would necessarily be "in-kind," or how different kinds of states might respond differently.

Theories of escalation dynamics have been elaborated in the nuclear domain. But the deep and profound differences between the nuclear and cyber domains suggest that any theory of escalation dynamics in the latter would require far more than small perturbations in nuclear escalation dynamics theories, though such theories might be useful points of departure for developing new ones applicable to cyberspace. Some of these differences include the greater uncertainties in attribution of cyber actors, the broad proliferation of significant capabilities for cyber operations to a multitude of states and a variety of nonstate actors as well, and the inherent ambiguities of cyber operations compared to the very distinct threshold of nuclear weapons explosions.

To suggest some of the difficulties involved, consider the following scenarios:

- Nation Blue may believe it has been attacked deliberately by Nation Red, even though Red has not done so. Indeed, because of the ongoing nature of various attack-like activities (e.g., hacking and other intrusions) against the computer systems and networks of most nations, Blue's conclusion that its computer systems are being attacked is certainly true. Attribution of such an attack is a different matter, and because hard evidence for attribution is difficult to obtain, Blue's government may make inferences about the likelihood of Red's involvement by giving more weight to a general understanding of Red's policy and posture toward it than might be warranted by the specific facts and circumstances of the situation. Evidence that appears to confirm Red's involvement will be easy to find, whether or not Red is actually involved. If Red is a technologically sophisticated nation (such as the United States), the lack of "fingerprints" specific to Red can easily be attributed to its technological superiority in conducting such attacks.

- An active defense of its systems and networks undertaken by Nation Red against Nation Blue could have significant political consequences. For example, even if Red had technical evidence that was incontrovertible (and it never is) pointing to Blue's government, Blue could still deny that it had launched such an attack—and in the court of world opinion, its denial could carry some credibility when weighed against Red's past assertions regarding similar issues. That is, Red's cyber attacks (counter–cyber attacks, to be precise) undertaken under the rubric of active defense may not be perceived as innocent acts of self-defense, even if they are. The result could be a flurry of charges and countercharges that would further muddy the waters and escalate the level of political tension and mistrust. The point at which a software agent for cyber attack is introduced or planted on an adversary's computer system or network is, in general, different from the point at which it is activated and begins to do damage. Blue (the nation being attacked) may well regard the hostile action as beginning at the moment Red's agent is planted, whereas Red may believe the hostile action begins only when the agent is activated.
- During periods of crisis or tension when military action may be more likely, it is entirely plausible that Blue would increase the intensity of security scans it conducts on its critical systems and networks. More intense security scans often reveal offensive software agents implanted long before the onset of crisis and that may have been overlooked in

ordinary scans, and yet discovery of these agents may well prompt fears that an attack is pending.

- The direct damage from a cyber attack is often invisible to outsiders. Without CNN images of smoking holes in the ground or troops on the move, an outside observer must weigh competing claims without tangible evidence one way or the other. Under such circumstances, the reputations of the different parties in the eyes of each other are likely to play a much larger political role.
- Nation Red plants software agents in some of Nation Blue's critical networks to collect intelligence information. These agents are designed to be reprogrammable in place—that is, Red can update its agents with new capabilities. During a time of crisis, Blue's authorities discover some of these agents and learn that they have been present for a while, that they are sending back very sensitive information to Red, and that their capabilities can be changed on a moment's notice. Even if no harmful action has yet been taken, it is entirely possible that Blue would see itself as the target of Red's cyber attack.

What follows are some speculations on some of the factors that might influence the evolution of a cyber conflict (see fig. 1).

Crisis Stability

Where kinetic weapons are involved, crisis stability refers to that condition in which neither side has incentives to attack first. Crisis stability is especially important for nuclear weapons, where the existence of an invulnerable submarine-based nuclear missile force controlled by Nation Blue means that Nation Red could not escape retaliation no matter how devastating a first strike it could launch. In terms of cyber weapons, there is no conceivable way for one nation to eliminate or even significantly degrade the cyber attack capability of another.[7] But the question remains whether a second-strike cyber attack capability is the enabling condition for crisis stability in cyberspace.

A related question is that of incentives for preemption. Preemptive attacks by Red against Blue are undertaken to prevent (or at least blunt) an impending attack by Blue on Red. If Blue is planning a cyber attack on Red, a preemptive cyber attack on Blue cannot do much to destroy Blue's attack capability; at best, Red's preemptive attack on Blue might tie up Blue's personnel skilled in cyber operations. On the other hand, it is hard

Crisis Stability

- What is the analog of crisis stability in cyber conflict?
- What are the incentives for preemptive cyber attack?

Escalation Control and Management

- How can intentions be signaled to an adversary in conflict?
- How can cyber conflict between nations be limited to conflict in cyberspace?
- What thresholds of "line-crossing" activity might be created in cyberspace, and how might these be communicated to an adversary?
- How should cyber attack be scoped and targeted so that it does not lead an adversary to escalate a conflict into kinetic conflict?
- How can a modestly scoped cyber attack conducted by a government be differentiated from the background cyber attacks that are going on all of the time?
- How can the scale and scope of a commensurate response be ascertained?
- What confidence-building measures might actually reassure an adversary about a lack of hostile intent?

Complications Introduced by Patriotic Hackers

- How can "freelance" activities on the part of patriotic hackers be handled?

Incentives for Self-Restraint in Escalation

- What are the incentives for self-restraint in escalating cyber conflict?

Termination of Cyber Conflict

- What does it mean to terminate a cyber conflict?

Necessary Capabilities for Escalation Management

- How can national authorities exercise effective command and control of cyber forces in a rapidly evolving conflict environment?
- What is the scope and nature of national capabilities (e.g., technological, command and control, law enforcement/legal capabilities) needed to implement any approach to escalation management and conflict termination in cyberspace?
- How can each side obtain realistic assessments of one's own or an adversary's cyber state and condition (e.g., heavily or lightly damaged)?
- How might other resources/capabilities available to the United States be used to manage escalation of conflict in cyberspace?

Figure 1. Questions about escalatory dynamics of cyber conflict between nation-states

to imagine circumstances in which Red would realize that Blue were planning an attack, as preparations for launching a cyber attack are likely to be invisible for the most part.

A second relevant scenario is one in which Blue is planning a kinetic attack on Red. Intelligence information, such as photographs of troop movements, indicates preparations for such an attack. Under these circumstances, Red might well choose to launch a preemptive cyber attack with the intent of delaying and disrupting Blue's preparations for its own.

Signaling Intentions in Cyber Conflict

Nothing in the set of options above is specific to cyber conflict—such issues have been an important part of crisis management for a long time. But managing such issues may well be more difficult for cyber conflict than for other kinds of conflict. One reason is the constant background of cyber-attack activity. Reports arrive constantly of cyber attacks of one kind or another on US computer systems and networks, and the vast majority of these attacks do not have the significance of a serious cyber attack launched by a party determined to do harm to the United States. Indeed, the intent underlying a given cyber attack may not have a military or a strategic character at all. Organized crime may launch a cyber attack for profit-making purposes. A teenage hacking club may launch a cyber attack out of curiosity or for vandalism purposes.

Thus, if one nation wishes to send a signal to its cyber adversary, how is the latter to recognize that signal? Overtly taking credit for such an attack goes only so far, especially given uncertain communications in times of tension or war and the near certainty of less-than-responsible behavior on the part of one or both sides.

A dearth of historical experience with the use of serious offensive cyber operations further complicates efforts at understanding what an adversary might hope to gain by launching a cyber attack. In the absence of direct contact with those conducting such operations—sometimes even in the presence of such contact—determining intent is likely to be difficult and may rest heavily on inferences made on the basis of whatever attribution is possible. Thus, attempts to send signals to an adversary through limited and constrained military actions—problematic even in kinetic warfare—are likely to be even more problematic when cyber attacks are involved.

Determining the Impact and Magnitude of Cyber Response

If an adversary conducts a cyber attack against the United States, the first questions for US decision makers will relate to impact and magnitude. Such knowledge is necessary to inform an appropriate response. If, for example, the United States wishes to make a commensurate response, it needs to know what parameters of the incoming attack would characterize a commensurate response.

In many kinds of cyber attack, the magnitude of the impact of the first attack will be uncertain at first and may remain so for a considerable period of time. Decision makers may then be caught between two challenges—a policy need to respond quickly and the technical fact that it may be necessary to wait until more information about impact and damage can be obtained. These tensions are especially challenging in the context of active defense and active threat neutralization.

Decision makers often feel intense pressure to "do something" immediately after the onset of a crisis, and sometimes such pressure is warranted by the facts and circumstances of the situation. On the other hand, the lack of immediate information may prompt decision makers to take a worst-case view of the attack and, thus, to assume that the worst that might happen was indeed what actually happened. Such a situation has obvious potential for inappropriate and unintended escalation or kinetic response.

Transparency and Confidence-Building Measures

Where kinetic weapons are concerned, transparency and confidence-building measures such as adherence to mutually agreed "rules of the road" for naval ships at sea, prenotification of large troop movements, and noninterference with national technical means of verification have been used to promote stability and mutual understanding about a potential adversary's intent.

Translating traditional transparency and confidence-building measures into cyberspace presents many problems. For example, generating forces in preparation for offensive cyber operations can be done essentially behind closed doors and with a small footprint, so evidence suggesting impending hostile action will never be evident, except with advance public notice. Thus, there is no reasonable analog for "notification of movement or massing of forces." Because the success of offensive cyber operations is largely dependent on stealth and deception, reassurances of Nation Blue regarding the benign nature of any cyber activity observed, assuming it can be seen and attributed, ring hollow to any parties that have a competitive or politically tense relationship with Blue. Traditional kinetic operations—those military operations on land, sea, and air—are easily distinguishable from most nonmilitary movements. By contrast, it is often difficult to distinguish between military and nonmilitary cyber operations, particularly between cyber attack and cyber exploitation. During a crisis, Blue may consider collecting intelligence on Red as stabilizing and thus lower the likelihood of mistaken escalation. Red may well interpret this as Blue preparing the battlefield as a prelude to attack.

These comments are not meant to suggest that all transparency or confidence-building measures for cyberspace are futile—only that applying traditional measures to cyberspace will be difficult, and new forms of conduct and behavior may be needed to promote transparency and build confidence.

Catalytic Cyber Conflict

Catalytic conflict as mentioned earlier refers to the phenomenon in which a third party instigates or seeks to escalate conflict between two other parties. These could be nation-states or subnational organizations such as terrorist groups. To increase confidence in the success of initiating a catalytic war, the instigator might attack both parties, seeking to fool each into thinking the other is responsible.

Because high-confidence attribution of cyber attacks under all circumstances is highly problematic, an instigator would find it relatively easy to deceive each party about the instigator's identity; thus, a double-sided catalytic attack may be plausible. Also, if a state of tension already exists between the two parties involved, leaders in each nation will be predisposed toward thinking the worst about the other, making them less likely to exercise due diligence in carefully attributing an attack. An instigator might consequently choose just such a time to conduct a catalytic cyber attack.

Complications Introduced by Patriotic Hackers

When traditional kinetic military operations are involved, it is generally presumed that the forces involved engage in armed conflict only at the direction of the cognizant government, only by its authorized military agents, and specifically, not by private groups or individuals. That is, governments maintain their armed forces to participate in armed conflict under the government's direction.

But in the Internet era, it is necessary to consider that nonstate actors may become involved in conflict. During times of conflict (or even tension) with another nation, some citizens may be motivated to support their country's war effort or political stance by taking direct action in cyberspace (see fig. 2). Such individuals—often known as hacktivists or patriotic hackers—are private citizens with some skills in the use of cyber attack weapons, and they may well launch cyber attacks on the adversary nation on their own initiative; that is, without the blessing and not under the direction or control of the government of that nation.

A number of incidents of privately undertaken cyber attacks have been publicized:

- Immediately after the start of the second intifada in Israel in late September 2000, Palestinian and Israeli hackers conducted a variety of cyber attacks on each other's national web presences on the Internet.[8]
- Following the 2001 incident between the United States and China in which a US EP-3 reconnaissance aircraft collided with a Chinese F-8 interceptor, both Chinese and American hackers attacked the web presence of the other nation. In both cases, attacks were mostly aimed at website defacement and denial of service.[9]
- In the wake of the May 1999 bombing by the United States of the Chinese embassy in Belgrade, the US National Infrastructure Protection Center issued an advisory (NIPC Advisory 99-007) noting "multiple reports of recent hacking and cyber activity directed at U.S. government computer networks, in response to the accidental bombing of the Chinese embassy in Belgrade. . . . Reported activity include[d] replacing official web pages with protest material and offensive language, posting similar language in chat rooms and news groups, and denial of service email attacks."[10]
- American hackers have been known to attack jihadist websites. For example, an American was reported by *Wired* magazine to have hijacked www.alneda.com, a widely used website for jihadist recruitment.[11] His motive for doing so was said to be a decision made after the September 11 attacks: "I was going to use every skill I had to screw up the terrorists' communication in any way I could."
- Russian hackers are generally reported to have been responsible for the cyber attacks on Estonia in 2007 and Georgia in 2008.[12]

Allen and Demchek generalize from experiences such as these to predict that future conflicts between nations may involve:

- Spontaneous attack action in cyberspace by "patriots" on each side.
- Rapid escalation of these actions to a broad range of targets on the other side—because hacktivists are interested in making a statement, they will simply attack sites until they find vulnerable ones.
- Involvement of sympathetic individuals from other nations supporting the primary antagonists.

Figure 2. Hacktivism during international conflict and tension. Adapted largely from Patrick D. Allen and Chris C. Demchak, "The Palestinian-Israel: cyberwar" [*sic*], *Military Review*, March–April 2003.

The actions of these patriotic hackers may greatly complicate escalation management. Such actions may be seen by an adversary as being performed under the direction, blessing, tacit concurrence, or tolerance of the state and therefore are likely to be factored into the adversary's assessment of the state's motives and intent. The state's efforts to suppress patriotic hackers may be seen as insincere and are likely to be at least partially unsuccessful as well. In a worst-case scenario, actions of patriotic hackers during times of tension may be seen as an officially sanctioned cyber first strike, even if they have not acted with government approval or under government direction.

Yet another complication involving patriotic hackers is the possibility that they might be directed by, inspired by, or tolerated by their government but in ways in which the government's hand is not easily visible. Under such circumstances, hostile acts with damaging consequences could continue to occur with corresponding benefits to the nation responsible despite official denials. At the very least, the possibility that patriotic hackers may be operating could act as a plausible cover for government-sponsored cyber attacks, even if there were in fact no patriotic hackers doing anything.

Incentives for Self-Restraint in Escalation

One set of incentives is based on concerns about an adversary's response to escalation. Understanding this set of incentives is necessarily based on a sense of what kinds of offensive cyber actions—whether cyber attack or cyber exploitation—might be mistaken for cyber attack and might lead to what kinds of adversary responses, either in cyberspace or in physical space. In this regard, an essential difference between cyber attack and the use of a nuclear, chemical, biological, or space weapon is readily apparent—the initial use of any nuclear, chemical, biological, or space weapon, regardless of how it is used, would constitute an escalation of a conflict under almost any circumstances. By contrast, whether a given cyber attack, or conventional kinetic attack for that matter, would be regarded as an escalation depends on the nature of the operation—the nature of the target(s), their geographical locations, or their strategic significance.

A second set of incentives is based on concerns about blowback—the possibility that a cyber attack launched by the United States against Nation B's computers might somehow affect US computers at a later time. Understanding the likelihood of blowback will require a complex mix of technical insight and intelligence information.

Deescalation and Conflict Termination

Conflict termination presumes the existence of an ongoing conflict to which the participants desire an end. It requires several elements, including:

- a reliable and trustworthy mechanism that can be used by the involved parties to negotiate the terms of an agreement to terminate a conflict,
- a clear understanding on all sides about what the terms of any agreement require each side to do,
- assurance that all parties to an agreement will adhere to the terms of any such agreement, and
- capabilities for each party that can insure all entities taking action on behalf of that party adhere to the terms of any such agreement.

In the cyber environment, these elements may be problematic. National leaders and their representatives will almost certainly be communicating with each other through electronic channels, the reliability of which may be questionable in certain kinds of cyber conflict. A cease-fire agreement in cyberspace presumes each side can know that the other has stopped hostile activity in cyberspace. However, ambiguity and technical limitations create problems. Nation Blue may conduct cyber exploitations seeking to verify that Nation Red is standing down in cyberspace. Red may interpret these operations as prelude to Blue's continuing an attack campaign against it. Patriotic hackers of Blue may press onward against Red even though both Red and Blue have themselves agreed to a cyber cease-fire. During conflict, there is no reason to assume the cessation of continuing cyber operations conducted by others who are not part of the conflict (e.g., criminals). In some cases, ongoing offensive operations by these third parties may be mistakenly attributed to Red or Blue. The two nations may differ in their interpretation of key concepts. What activities constitute an "attack" in cyberspace, or what evidence should be used to determine if an attack is occurring? Differing interpretations and inadequate technical capabilities may impede understanding. For kinetic military forces, a variety of technical means (e.g., photoreconnaissance aircraft and satellites, ocean-scale sonar arrays) are capable of monitoring movements of military personnel and equipment. Most importantly, these means operate from outside territory controlled by an adversary and provide information that is generally regarded as reliable. But because the footprint of cyber forces is so small, movement of adversary forces can take place without signatures that can be externally observed. Based on precedents in kinetic conflict, it is plausible that nations seeking a cease-fire in a cyber conflict would ask for the deactivation of these hostile agents. To comply with such a request (not an unreasonable one in the context of a cease-fire), these nations will need to maintain cyber "demining" capabilities regarding the offensive software and/or hardware agents they implant into adversary systems, networks, and infrastructure. For example, they will need to keep track of where these agents are implanted or be able to communicate with them to disarm them—a capability that may rule out offensive agents that operate in a fully autonomous manner.

Each party will naturally have concerns about its adversary's commitment to adhere to the terms of a cyber cease-fire, especially in the aftermath of a conflict. On what basis would Blue's government believe a claim by Red that it was indeed complying with the terms of a cease-fire? How much would Red tell Blue about system and network penetrations it had made, knowing such information might be used to prosecute an attack or defend more effectively against Red? The availability of effective ways to address the issues described above is almost certainly one aspect of being able to manage conflict termination in cyberspace.

Analysts sometimes raise the issue of how the United States might deter escalation when it has more at stake in cyberspace than its adversaries. The first point to consider is that deterrence of cyber attack does not necessarily entail a threat to respond through cyberspace against an adversary's cyber assets, and when non-cyber threats against an adversary's non-cyber assets are considered, the calculus of deterrence may well be different. For example, kinetic weapons can, in principle, be employed against valuable physical military targets. Although the threshold for such a response may well be higher, an adversary would still have to consider the possibility of a non-cyber response to any attack. Consistent with this point, US policymakers have always noted that the United States reserves the right to respond appropriately in a time, place, and manner of its own choosing. In addition, concerns over blowback may deter an adversary. If an adversary's interests are entangled with those of the United States, it may be deterred from taking actions that might harm US interests because of concerns that one ultimate effect of such actions would be to harm the adversary's interests. For example, a nation that is owed a great deal of money by the United States might well be unlikely to conduct an attack that undermines its financial stability.

Lastly, many analysts note that deterrence is a psychological phenomenon and that threats of retaliation must be focused on assets that an adversary holds dear and values highly. In principle, what an adversary—or more precisely, an adversary decision maker—holds dear can span a wide range, from personal to national (e.g., tools of national power). In the category of personal assets are financial entities (e.g., a leader's bank accounts could be drained), reputation (e.g., a scandal in a policymaker's past might be revealed), and close friends and relatives (e.g., the interests of such individuals could be compromised). Such assets are not typically considered in a traditional military context—but nontraditional approaches to deterrence may well be needed to deal with the nontraditional threats that cyber attacks pose.

The approaches described above may be most useful in deterring hostile cyber operations intended to achieve large-scale effects. They are unlikely to be useful in deterring operations intended to achieve smaller effects, because smaller effects by definition do not cause maximum pain for either side. Put differently, the argument that the United States has more at risk in cyberspace than its adversaries is simply not relevant when the amount of damage that can be done (by definition) is small.

Kinetic Escalation

Issues of escalation and conflict termination in cyberspace are complicated by the fact there may be cross-domain linkages. Although conflict might, in principle, be limited to hostile operations in cyberspace alone, there is no reason this is necessarily so, and policymakers must contemplate the possibility that conflict in cyberspace might spill over into physical space, and might even lead to kinetic actions.

For example, if national command authorities decide to retaliate in response to a cyber attack, an important question is whether retaliation must be based on a "tit-for-tat" response. Assuming the perpetrator of a cyber attack is known to be a hostile nation, there is no reason in principle the retaliation could not be a kinetic attack against the interests of that hostile nation. Allowing a kinetic response to a cyber attack expands the range of options available to the victim. An extreme case is, in the event of a cyber attack of sufficient scale and duration that it threatens the nation's ability to function as a modern society, the attacked nation might choose to respond with kinetic force. On the other hand, the use of kinetic operations during an ostensibly cyber-only conflict is an important threshold. Nations involved in a cyber-only conflict may have an interest in refraining from a kinetic response—for example, they may believe kinetic operations would be too provocative and might result in an undesired escalation of the conflict.

In addition, the logic of offensive cyber operations suggests that such operations are likely to be most successful when the initiator of these operations has the time to gather intelligence on likely targets—such intelligence gathering is obviously time-limited once overt kinetic conflict breaks out.

If understanding the dynamics of cyber-only conflict is difficult, understanding the dynamics of cyber conflict when kinetic operations may be involved is doubly so. To the extent national decision makers have incentives to refrain from conducting offensive operations that might induce a strong kinetic reaction, the obvious approach would be to conduct cyber attacks that are in some sense smaller, modest in result, targeted selectively against less-provocative targets, and perhaps more reversible. The similarity of such an approach to escalation control in other kinds of conflict is not accidental, and it has all of the corresponding complexities and uncertainties.

In keeping a cyber conflict from escalating into physical space, it is important to think about "lines in the sand" beyond which one side warns another not to cross. For example, it is reported that during the first Gulf War, the United States regarded Iraqi use of chemical weapons against US forces as one such threshold of unacceptable activity, one that might well provoke the use of US nuclear weapons against Iraq. When only traditional kinetic forces are involved, lines in the sand might be the use of certain weapons, attacks on or damage to certain targets, movement or placement of armed forces beyond certain geographical lines, and so on. Cyber analogs to these thresholds are hard to construct. Describing a class of cyber weapon whose mere use would be wholly unacceptable is hard to imagine, since there are no real cyber analogs to true weapons of mass destruction where even a single use of a WMD qualitatively changes the landscape of kinetic conflict. And in cyberspace, what is the analog of a geographical border beyond which cyber weapons may not be placed?

Perhaps the most promising analog is the notion of specific targets that might be placed off limits—cyber attacks on such targets could, in principle, be deemed unacceptable. One class of off-limits targets might be cyber assets associated with truly critical infrastructure, such as the bulk power grid or the banking and financial system. But as any bank executive will confirm, some of these targets are under attack quite frequently—so attacks that do not cause large amounts of damage or loss probably should *not* qualify as crossing the threshold of unacceptability. There is also the question of being able to assign *political* responsibility to some perpetrator for the conduct of a successful large-scale attack on some off-limits target—a question whose answer may be in doubt, given the difficulties of rapid attribution of a cyber attack. Finally, one might well ask how a cyber asset would be positively identified as being associated with the bulk power grid or the banking and financial system. Would we provide a computer-readable identification tag on every such computer? Such a tag might make these targets obvious to other parties wishing to do us harm.

Even presuming that the United States could identify specific thresholds, such information would need to be communicated clearly to an adversary. Such communication is difficult even in scenarios of traditional military conflict, and all of these difficulties obtain in the cyber context. But it is worth observing that because cyber conflict is fundamentally based on deception, persuading an adversary to believe any US statement about what is off-limits may be particularly challenging.

The Political Side of Escalation

Despite the focus of the discussion above on escalation dynamics from a primarily military standpoint, escalation dynamics inevitably have a political and psychological component that must not be overlooked. For example, the discussion of active defense above pointed out that US cyber attacks undertaken under the rubric of active defense may not be perceived by others as innocent acts of self-defense, even if they are intended as such. While both sides in most conflicts claim they are acting in self-defense, cyber conflicts are a particularly messy domain in which to air and judge such claims.

Another possible misperception may arise from intelligence-collection activities that might involve cyber-attack techniques. The discussion above noted the problems of misperceiving exploitation as a prelude to continuing cyber operations during a cease-fire. But the problem is broader than that—during conflict or in the tense times that often precede conflict, the needs for current intelligence on the adversary are particularly acute. Knowing what the adversary is doing and the scope and nature of its future intentions are very important to decision makers, and the need to collect such intelligence will almost certainly result in greater pressures to use the entire array of available intelligence-gathering techniques—including techniques of cyber exploitation. If the adversary is unable to distinguish between an offensive operation for exploitation and one for attack—an outcome that seems all too likely—a cyber exploitation may run the risk of being perceived as part of an imminent attack, even if this is not the intent of decision makers.

Finally, it seems likely that escalation issues would play out differently if the other nation(s) involved are or are not near-peer competitors. Escalation to physical conflict is less of a concern to the United States if the nation has weak conventional forces and/or is a nonnuclear state. But a nation with nuclear weapons, or even strong conventional forces in a position to inflict significant damage on US allies, is another matter entirely. Relationships with such states may well need to be explicitly managed, paying special attention to how escalation may be viewed, managed, and controlled, and most importantly, how miscalculation, misperception, or outright error may affect an adversary's response.

Dynamics such as these suggest that factors other than the ones dictated by military or legal necessity play important roles in escalation dynamics, if only because they can strongly affect the perceptions of decision makers on either side.

The Future of Escalation Dynamics

The issues of escalation dynamics, conflict termination, and cross-domain linkages in cyberspace play out against a rapidly changing technological, policy, and geopolitical environment. The substrate of cyberspace—computing and communications technology—is characterized by change on a timescale much shorter than the planning horizon for traditional military acquisitions and planning. Upgrades notwithstanding, major weapons platforms are expected to serve for decades, while the information technology environment changes rapidly in a few years. The growing use of cloud computing is a further—and potentially disruptive—change in possible computing platforms and may require new concepts for assigning responsibility for cyber operations. Mobile computing may present opportunities for determining device location as well as being the enabling technology for many new users of cyberspace. IT will be increasingly embedded, ubiquitous, and connected within all elements of modern society, potentially increasing vulnerabilities to all manner of societal functions. The result is that operational concepts for escalation management must take into account a rapidly evolving set of targets and offensive and defensive capabilities.

In most traditional domains of conflict, US military doctrine has been based on the establishing dominance—that state in which friendly forces have maximum freedom of action and adversary forces have minimal freedom of action. But in the cyber domain, this presumption is not sustainable—and senior US military leaders are beginning to speak publicly about this point.[13] Much of the traditional US approach to escalation control is based on the ability of friendly forces to establish dominance at any level of conflict on the premise that an adversary would not choose to escalate if, at the higher level of conflict, it could not hope to prevail.

Nation-states are increasingly concerned about the risks inherent in involvement in cyberspace. Even apart from the protection of critical national infrastructure and military assets, various nations express deepening worries about traditional criminal activity in cyberspace, protection of intellectual property, and increased connectedness for political movements that may pose a threat to government interests and stability.

Nonstate actors are increasingly important players in cyberspace. Multinational corporations and organized crime syndicates, for example, all have some nontrivial capability to conduct offensive operations in cyberspace to further their interests, and even small groups of individuals can have a large impact by exploiting certain characteristics of cyberspace (e.g., WikiLeaks).

Although existing theories of escalation dynamics and conflict termination may serve as useful points of departure, what is understood very poorly today is how these theories may apply in cyberspace. In the future, finding ways to manage cyber conflict will be even more intellectually challenging than it was for traditional conflict.

Notes

1. The lag time between dissemination of a security fix to the public and its installation on a specific computer system may be considerable, and it is not always due to unawareness on the part of the system administrator. It sometimes happens that the installation of a fix will cause an application running on the system to cease working, and administrators may have to weigh

the potential benefit of installing a security fix against the potential cost of rendering a critical application nonfunctional. Adversaries take advantage of this lag time to exploit vulnerabilities.

2. A zero-day attack is a previously unseen attack on a previously unknown vulnerability. The term refers to the fact that the vulnerability has been known to the defender for zero days. (The adversary has usually known of the attack for a much longer time.) The most dangerous is a zero-day attack on a remotely accessible service that runs by default on all versions of a widely used operating system distribution. This type of remotely accessible zero-day attack on services appears to be occurring less frequently. In response, a shift in focus to the client side has occurred, resulting in many recent zero-day attacks on client-side applications. For data and analysis of zero-day attack trends, see Daniel Geer, "Measuring Security," *Dan@Geer.org*, 278–87, http://geer.tinho.net/measuringsecurity.tutorialv2.pdf.

3. An adversary computer or network may not necessarily be owned and operated by the adversary—it may simply support or be used by the adversary.

4. For purposes of this article, the term *attribution* is used to refer to the identification of the party to which political responsibility should be assigned for the cyber operations that harm the interests of the target. This qualifier is necessary because the entity "responsible" can also be the machine(s) involved in the operation or the specific human beings who took specific actions (at a keyboard) to launch the operation. One of these other meanings may be more relevant, depending on the purposes for which attribution is sought. For more discussion of this point, see David D. Clark and Susan Landau, "Untangling Attribution," *National Security Journal*, 16 March 2011, http://harvardnsj.org/2011/03/untangling-attribution-2/, as well as William Owens, Kenneth Dam, and Herbert Lin, *Technology, Policy, Law, and Ethics Regarding U.S. Acquisition and Use of Cyberattack Capabilities* (Washington: National Academies Press, 2009), chap. 2.

5. The broad topic of how to improve passive cyber defenses and enhance resilience of US computer systems and networks is addressed in a variety of National Research Council (NRC) reports on this topic: *Computers at Risk*, 1991; *Information Technology for Counterterrorism*, 2003; *Cybersecurity Today and Tomorrow: Pay Now or Pay Later*, 2002; *Realizing the Potential of C4I: Fundamental Challenges*, 1998; *Trust in Cyberspace*, 1999; and *Toward a Safer and More Secure Cyberspace*, 2007, all authored by the NRC and published by National Academies Press, Washington, DC. Other important reports include President's Information Technology Advisory Committee, *Cyber Security: A Crisis of Prioritization* (Washington: National Coordination Office for Information Technology Research and Development, February 2005); and Commission on Cyber Security for the 44th Presidency, *Securing Cyberspace for the 44th Presidency* (Washington: Center for Strategic and International Studies, 2008).

6. This taxonomy is based mostly on *Dangerous Thresholds: Managing Escalation in the 21st Century* (Santa Monica, CA: RAND, 2008), though the RAND discussion is silent on escalation in cyberspace per se.

7. Even in the case of a nuclear EMP attack directed against electronic equipment in another nation, there is no reason to assume that all of that nation's cyber-attack capabilities are necessarily resident within its boundaries. Because cyber attacks can originate from anywhere, some cyber attack capabilities may have been deployed in other nations—indeed, some attack agents may already have been clandestinely deployed in US systems.

8. "Cyberwar Also Rages in Mideast," *Associated Press*, 26 October 2000, http://www.wired.com/politics/law/news/2000/10/39766.

9. Michelle Delio, "A Chinese Call to Hack U.S.," *Wired*, 11 April 2001, http://www.wired.com/news/politics/0,1283,42982,00.html.

10. Available at http://www.merit.edu/mail.archives/netsec/1999-05/msg00013.html.

11. Patrick Di Justo, "How Al-Qaida Site Was Hijacked," *Wired*, 10 August 200, http://www.wired.com/culture/lifestyle/news/2002/08/54455.

12. "Expert: Cyber-Attacks on Georgia Websites Tied to Mob, Russian Government," *Los Angeles Times*, 13 August 2008, http://latimesblogs.latimes.com/technology/2008/08/experts debate.html.

13. For example, RADM William Leigher, deputy commander of the US Navy Cyber Command, was recently quoted as saying that "Unlike the physical domain, achieving dominance [in the cyber domain] may be impossible." Amber Corrin, "Dominance in Cyberspace Might not be Possible," *Defense Systems,* 27 January 2011, http://defensesystems.com/articles/2011/01/27 /afcea-west-cyber-warfare-panel.aspx.