2013 5th International Conference on Cyber Conflict
K. Podins, J. Stinissen, M. Maybaum (Eds.)
2013 © NATO CCD COE Publications, Tallinn

# Exploring the Prudent Limits of Automated Cyber Attack

**Jeffrey L. Caton**
President
Kepler Strategies LLC
Carlisle, Pennsylvania, U.S.A.
Jeff.Caton@keplerstrategies.com

**Abstract:** The notion of cyber conflict occurring at network rates that surpass the speed of decision-making by national leaders has bolstered the possibility of introducing automated cyber attacks as part of their spectrum of response. This paper's objective is to identify some prudent limits to govern the incorporation of automated cyber attack as an instrument of policy in national and collective defense. For this paper, the concept of automated cyber attack focuses on nations' in-kind responses to strategic-level attacks by actors that use cyber means. The main aspects of the paper explore the theoretical roles of critical thinking in the development and operation of such systems. Topics include the context, points of view, and cognitive biases of the cyber actors; the assumptions and inferences inherit in their decision making; and the implications of decisions related to automated cyber attack.

The structure of research utilizes the Gerras critical thinking model to identify the factors to evaluate. It outlines how techniques such as the analysis using *Tallinn Manual* criteria may be used to identify assumptions and inferences for categorizing national response actions as cyber attack. It examines several historical incidents involving decisions related to strategic attack for implications to automated cyber attacks. It also investigates the implications of adopting a policy of cyber resilience, focusing on how it could be integrated with automated cyber responses measures. Finally, it studies the implications of automated cyber attack connected to the philosophy and ethics of evolving Just Cyber Warfare theory, such as that proposed by Taddeo.

**Keywords:** *critical thinking, escalation, resilience, automated response, attack assessment*

The views expressed are those of the author and do not necessarily reflect official policy or position of any Department or Agency within the U.S. Government. All information and sources for this paper were drawn from unclassified material.

# 1. INTRODUCTION

When contemplating the topic of cyber warfare, there is general consensus supporting the primacy of offensive over defensive actions [1]. In more common parlance, it is often said that “the best defense is a good offense.” But should such a tenet be implemented in service of a nation’s security in cyberspace? And how should this tenet be characterized in an environment where thrusts and parries may occur at network speeds? This paper’s objective is to identify some prudent limits to govern the incorporation of automated cyber attack as an instrument of policy in national and collective defense. A key aspect of the paper is to explore the role of critical thinking in the development and operation of such systems.

# 2. CRITICAL THINKING

The framework for analysis in this paper utilizes the Gerras [2] model (derived from the work of Paul and Elder) which defines critical thinking as “deliberate, conscious, and appropriate application of reflective scepticism.” Gerras applies the context-dependent school of thought and focuses on factors important to the decision making of strategic leaders. The model is broken into six main elements: clarify concern, point of view, assumptions, inferences, evaluation of information, and implications. These elements are considered to be interactive and are not necessarily linear or sequential in application of assessing the deliberate use of critical thinking.

The element of *clarify concern* concentrates on the desire to separate the root causes of problems from their symptoms; this should be done in such a way as to not preclude or limit potential responses. When evaluating the actions of nations, a significant aspect of the *point of view* element is egocentrism, which Gerras calls the “tendency to regard oneself or one’s own opinions or interests most important.” He offers four specific applications of this principle—egocentric memory (forgetting information that does not support one’s thoughts); egocentric myopia (narrowing point of view in assessment to support one’s thoughts); egocentric righteousness (considering one’s thoughts to be superior); and egocentric blindness (disregarding information that does not support one’s thoughts). Making *assumptions* is an inherent trait that humans use to provide boundaries for decision making; clearly stating and understanding such assumptions aids the critical thinking process. *Inferences* are logical perceptions of how available facts and evidence fit together in the environment being considered. In ideal applications, the *evaluation of information* is an objective process. However, decision makers often employ cognitive strategies such as heuristics (“rules of thumb”) to simplify the process; but these useful tools may also introduce unknown and undesired biases.

Considering *implications* of any decision should include potential effects (desired and undesired) beyond or collateral to the anticipated outcomes.

# 3. RESPONSE AND ESCALATION

For this paper, the concept of automated cyber attack focuses on nations' in-kind responses to strategic-level attacks by actors that use cyber means. Such automated responses would go beyond merely defending or mitigating the effects on an ongoing attack, but would instead be an offensive or proactive counter-strike to thwart any future attacks. The intent to have a cyber attack response capability is made clear by such statements as General Keith Alexander's recent testimony [3] to the U.S. Senate as Commander, U.S. Cyber Command: "We feel confident that foreign leaders believe a devastating attack on the critical infrastructure and population of the United States by cyber means would be correctly traced back to its source and elicit a prompt and proportionate response." This paper is a theoretical study that assumes that the desire and technical capability to automate such cyber attacks is feasible in the near future.

## A. *ASSESSING POTENTIAL ATTACKS*

It is critical to ensure a cyber attack has occurred before considering a cyber attack as an in-kind response. How does one differentiate a coincidental incident in cyberspace with negative consequences from an actual attack? One of the best tools supporting this complex task is the framework of the Schmitt [4] criteria which considers the intensity of damage in each of seven areas to provide a composite assessment of the effects of a potential cyber attack. These are considered in the perspective of *jus ad bellum* and compared to international norms and agreements such as those established by charters of the United Nations and the North Atlantic Treaty Organization (NATO) as well as humanitarian law [5]. These criteria have been further refined and expanded to eight areas in their recent adoption as an integral part of the *Tallinn Manual on the International Law Applicable to Cyber Warfare* [6]. Figure 1 [7] depicts the *Tallinn Manual* criteria and related elements as a framework to assess incidents in cyberspace which may put them into categories of hostile events ranging from use of force to armed conflict. If the determination is made for cyber attack, then any response should apply *jus in bello* tenets, such as those codified in the Law of Armed Conflict.

Even learned scholars may disagree on the practical application of this framework in complex and dynamic geopolitical environments. The implications of cyber attack characterization are potentially dangerous, as Ziolkowski [8] notes, "the threshold of endangering the (physical) security of a State is a high one and should not be diluted." It may become a mostly academic issue if a nation opts to implement an automated cyber attack responses based on pre-determined indicators and criteria.

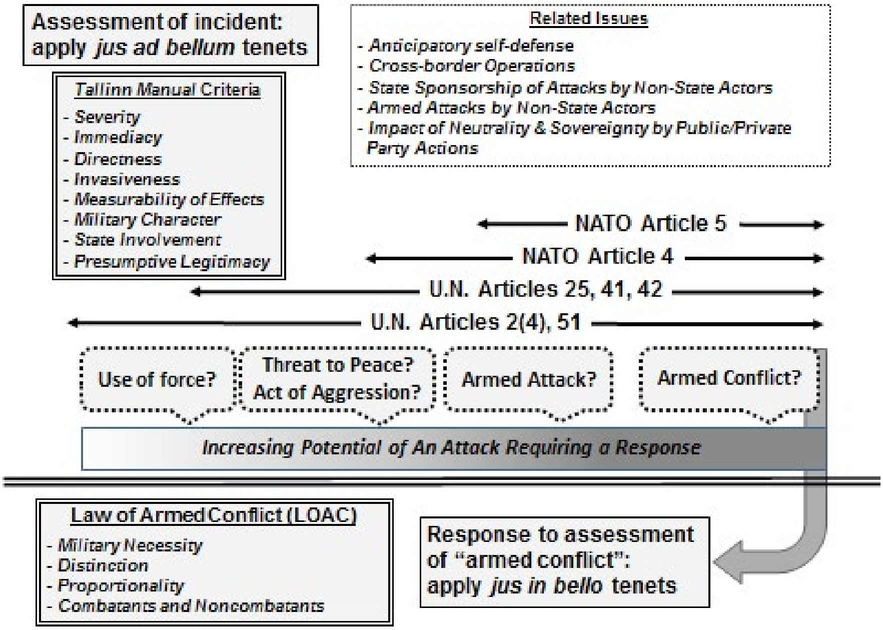
Assessment of incident: apply *jus ad bellum* tenets
Tallinn Manual Criteria
- Severity
- Immediacy
- Directness
- Invasiveness
- Measurability of Effects
- Military Character
- State Involvement
- Presumptive Legitimacy
Related Issues
- Anticipatory self-defense
- Cross-border Operations
- State Sponsorship of Attacks by Non-State Actors
- Armed Attacks by Non-State Actors
- Impact of Neutrality & Sovereignty by Public/Private Party Actions
NATO Article 5
NATO Article 4
U.N. Articles 25, 41, 42
U.N. Articles 2(4), 51
Use of force?
Threat to Peace? Act of Aggression?
Armed Attack?
Armed Conflict?
Increasing Potential of An Attack Requiring a Response
Law of Armed Conflict (LOAC)
- Military Necessity
- Distinction
- Proportionality
- Combatants and Noncombatants
Response to assessment of "armed conflict": apply *jus in bello* tenets

Figure 1. Cyber incident assessment and escalation

## B. CONTEXT AND ACTORS IN THE ATTACK RESPONSE PROCESS

Once an incident is assessed to be an attack, the analysis shifts to consider appropriate response. This is accomplished at two levels: the immediate and local effects, and the long-term and global impacts. The Law of Armed Conflict (LOAC) establishes the international norms that define how the use of force in responses should be planned and implemented. Fanelli and Conti [9] explore cyber operations effects in terms of their severity and persistence. Examining longer term and global impacts may require the methodical exploration of the dynamic context of cyber attack to assess policy options for using either continuous or discrete automation. This evaluation should consider possible consequences that build upon previous outcomes and thus intensify global tensions. Such a framework is the Kahn [10] escalation ladder which codifies in 44 metaphorical rungs the range of nuclear-related conflict between nations from subcrisis maneuvering up through various manifestations of military and civilian central nuclear war.

Any response must consider the actor nations that will be targeted. Was the initial attack conducted by actors that were rational or irrational, or could it have been an accidental initiation? Does automated decision making take all these possibilities into consideration? Any actor in the process is capable of rational or irrational decisions and as Gerras [11] notes, "logically fallacious arguments can be psychologically compelling." Such critical thinking flaws may influence the design or operation of automated systems by propagating biases into the beliefs-desires-intentions (BDI) reflective properties of automated agents [12].

If dealing with rational actors, automated response may enhance cyber deterrence by punishment [13] or perhaps even enable cyber coercion [14]. However, even with rational state actors, there is a range of state responsibility for the cyber activity that occurs within their sovereign borders [15]. But is there really a legitimate concern that nations may not apply critical thinking to decision making for the use of strategic weapons? Before exploring the implications for cyber attack situations, let us first look at how automated defenses may have affected recent historical events not directly related to cyberspace.

# 4. LESSONS FROM RECENT HISTORY

The benefit of hindsight allows us to examine how errors and shortfalls in critical thinking almost led to catastrophic effects in three cases studies that occurred over the last three decades.

## A. ABLE ARCHER (1983)

In November 1983, NATO initiated a command-post exercise to test the procedures and communications necessary for theater nuclear war in Europe. Many historians assert that this exercise culminated a series of events that accidently led the world to the brink of nuclear exchange akin to the Cuban Missile Crisis of 1962 [16]. As facts surrounding this case continue to come forward, it is still not clear how this eventually resolved itself as a fortuitous "non-event." Perhaps its origin and closure are best thought of as "normal accidents" [17]—that is, there was no single clear cause or effect.

In this case, the key concern to clarify was for the U.S.S.R. to determine if NATO was going to launch a pre-emptive nuclear attack using the *Able Archer* exercise as a cover for the preparation and initiation. The Soviet point of view included an aging leadership that was biased to view U.S. actions as part of a conspiracy to eliminate their country. U.S. President Ronald Reagan adopted a tough stance that included stationing intermediate range nuclear missiles in Europe coupled with the new AirLand battle doctrine, perhaps due in part to Soviet deployment of SS-20 nuclear missiles. Both sides assumed the worst of the other's actions, setting in motion a vicious cycle of escalating mistrust and misinterpretation of events. The U.S. added Soviet political and military command structure to its nuclear targeting, inferring that it would induce caution in Soviet leadership. The U.S.S.R. inferred that they could prognosticate U.S. nuclear intentions based on the model of their Operation RYAN, which used extensive and diverse information gathering and indication-based decision making. Unfortunately, the model's design had inherent egocentric myopia and blindness which encouraged the reporting of potential crises [18]. Reagan later came to recognize his own misunderstanding of Soviet intentions that were also fueled by ethnocentrism. Fortunately, based on advice from his advisors, he agreed not to have himself or other principals in Washington participate in the exercise [19]. Hampered by biases, both sides appeared to be able to discern the others' *capabilities* but not *intentions*. Some historians contend that the role of a KGB agent turned by British intelligence provided the critical insight that prevented *Able Archer* from escalating to catastrophe [20]. Regardless, it appears that fortuitous circumstances rather than critical thinking prevailed.

## B. NORWEGIAN RESEARCH ROCKET (1995)

Almost twelve years after *Able Archer* came another nuclear close-call between NATO and Russia. The routine launch of a research rocket on 25 January 1995 was mischaracterized as a possible prelude to nuclear attack on Russia [21]. The situation occurred during a time of increased tension between Russia and Norway (and perhaps the world in general) that caused failures in the critical thinking of tactical and strategic intelligence as well as communication systems.

The concern to clarify for Russian leadership was simple—was the Norwegian rocket the first step in a NATO nuclear attack? After the collapse of the Soviet Union, a Russian government was formed in 1991 with much of its military structure—the Strategic Rockets Forces, specifically—mostly intact, but declining in capability [22]. From their point of view, the fledgling Moscow leadership was struggling with governing crises, lingering Warsaw Pact issues, and a war brewing in Chechnya, while U.S. and allied efforts in Desert Storm were being hailed as successful examples of next-generation warfare. Russia assumed that the world was hostile to their new place on the global stage and that the well-publicized eastward expansion of NATO might be an existential threat. Also, Norway was pressing an old claim dispute for over 150,000 square miles of territorial waters that were rich in resources, further fueling speculation that it was becoming the preferred springboard for rapid deployment of Western forces into Russia.

Norway informed the Russian embassy in Oslo of their scientific rocket launch plans on 21 December 1994 and again on 16 January 1995; based on past experience, they inferred that this was sufficient to reduce risk between the countries. Unfortunately, the Norwegians also inferred that the launch would be monitored and assessed by the Russians in the same way as previous such launches (over 600 since 1962). But the new *Black Brant XII* was almost twice the size of any previous rocket, with specifications similar to a Pershing II nuclear missile; they did not consider how its longer range and higher trajectory might be viewed by Russian early warning assets. The immediate evaluation of the launch data was that the flight profile fit that of an electromagnetic pulse attack—the anticipated prelude to knock out Russian command and control systems before a nuclear strike. Unfortunately, the Norwegian launch notification did not get passed internally by the Russians to the proper military or civilian authorities and so the Russian nuclear launch briefcases were activated by President Yeltsin and General Kolesnikov as a precaution. While the exact details are still coming forth, it appears that these leaders waited for almost seven tense minutes until it was clear that the rocket was not headed toward Russia [23]. Fortunately for all, the Russian release of nuclear weapons still required deliberate initiation by its civilian leader.

## C. CHINESE ANTI-SATELLITE WEAPON TEST (2007)

Going forward twelve years after the *Black Brant XII* launch there was another rocket flight with significant international implications. On 11 January 2007, China conducted its first kinetic-kill anti-satellite (ASAT) test, destroying its own Fengyun-1c weather satellite and causing extensive collateral damage of spacecraft debris that poses collision hazards for operational satellites. China miscalculated both the magnitude of the damage they would cause as well as the negative international ramifications [24].

The concern to clarify is to determine the purpose for China to conduct such a destructive test with long-term negative effects on the commons of space. From the Chinese point of view, this test was simply part of a larger ASAT program that included electronic jamming and laser dazzling of satellite systems. They may have assumed that it was an acceptable operation since the U.S. and U.S.S.R. both conducted similar destructive tests in the 1970s and 1980s with little residual effects [25]. China inferred incorrectly that the test would not cause long-term damage, despite the fact that it occurred at an orbital altitude significantly higher than other ASAT tests. It is unclear if the evaluation of the operation went beyond military leadership; China gave no advance warning of the test and did not issue a public statement until twelve days later [26]. The implications of this test are still significant six years later as other nations' satellites must contend with a more hazardous space environment; although China has less than 4 percent of the world's active payloads on orbit, it accounts for almost 28 percent of the on-orbit debris, the majority of which was generated by this one event [27].

## D. IMPLICATIONS FOR AUTOMATED CYBER ATTACK

Table I summarizes the key elements of the historical cases. In any of these vignettes, one must consider how the outcomes may have changed if the leaders' responses had been automated (by either side). These cases were selected to illustrate where lapses and fallacies in critical thinking leading up to the crises were actually contributing factors to the development of the actual crises. To examine how this might apply to situations where automated cyber attack may be considered, let us look at the critical thinking factors from three possible perspectives summarized in Table II. These theoretical analyses are illustrative, not comprehensive.

Table I. Summary of Historical Cases With Strategic Attack Issues

| **Critical Thinking Factors** | **Examples of Strategic Attack Concerns from Recent History** | | |
|---|---|---|---|
| | **Able Archer (1983)** | **Norway Research Rocket (1995)** | **Chinese Anti-Satellite (2007)** |
| **Clarify Concern** | NATO nuclear attack on USSR? | NATO nuclear attack on Russia? | Purpose of Chinese destruction of satellite? |
| **Point of View** | - USSR deploy SS-20s.<br>- US tougher stance with nuclear weapons.<br>- AirLand doctrine. | - NATO expansion.<br>- Chechnya war.<br>- Tensions between Norway & Russia. | - PRC: logical progression of military space development. |
| **Assumptions** | - US & USSR doctrines more aggressive.<br>- Aging USSR leadership more offensive-minded. | - Hostile world opinion toward Moscow.<br>- NATO making Norway a springboard for attack on Russia. | - PRC: no long-term damage expected?<br>- ASAT development similar to that of US and USSR programs. |
| **Inferences** | - US nuclear targeting of USSR leadership.<br>- USSR Operation RYAN use of indicators. | - Routine research rocket notification and launch.<br>- Issues with new rocket size and trajectory. | - Failure at technical level (to predict collateral effects).<br>- Failure at decision level to consider implications. |
| **Evaluation of Information** | Still debated. Reagan made right call not to have principals play. Possible intervention by Soviet spy. | Launch assessed as possible pre-emptive strike on Russian communication.<br>Yeltsin made right call not to respond. | Wrong call by PRC to destroy satellite. Unclear if military leadership had permission of civilian leaders. |
| **Implications** | **Fortuitous Non-Event** as part of a vicious cycle of mistrust; escalated near to point of nuclear exchange. | **Benign Event** misinterpreted by military—almost to point of nuclear exchange. | **Serious Event** that polluted space environment and increased risks for all space-faring countries. |

The first perspective is the U.S. internal view to clarify whether automated cyber attack is necessary for its existential defense. This could be framed by a point of view of cyberspace as a domain where attacks may occur at network speeds and may cause devastating surprise attacks (e.g., "cyber Pearl Harbor"). Assumptions may include current defenses being too slow and dispersed, and that their automation and centralization will increase their effectiveness. The inference is that the use of pre-determined indicators and automated cyberspace agents that can attack threat systems is sufficient and appropriate. If such a system is deployed, it may be difficult to determine when decision makers will know that an attack and response have occurred as well as what their role will be during the hostilities. The implications are that the value of the automated attack system must be viewed not only regarding their effects on tactical threats, but also on how they shape the strategic defense and deterrence posture.

The second perspective is that of U.S. allies view to clarify if automated cyber attack responses are suitable for collective or cooperative defense. A logical point of view is one where cyber attacks on one partner nation may affect all nations and that pooled resources for cyber defense will enhance the security of all. Allies may assume that automated responses may limit the extent of effects from adversarial attacks. They may also assume that design criteria and implementation methods can be shared to help ensure unity of effort. The inference is that the use of pre-determined indicators and automated cyberspace agents requires significant cooperation and coordination among allies. Evaluation of this inference raises issues regarding how the roles and responsibilities are assigned for the development, maintenance, and application of the automated capability. The implication is that, if properly implemented, the use of automated attack responses can improve collective cyber defense and deterrence.

Table II. Critical Analyses of Possible Automated Cyber Attack

| **Critical Thinking Factors** | **Possible Perspectives of Automated Cyber Attack** | | |
|---|---|---|---|
| | **U.S. Internal** | **U.S. Allies** | **Other Countries** |
| **Clarify Concern** | Necessary for existential defense of US? | Suitable for collective/ cooperative defense? | Level of threat posed by primary and collateral effects? |
| **Point of View** | - Attacks may occur at network speeds.<br>- Devastating surprise attacks possible. | - Attacks on one partner may affect all.<br>- Pooled resources will enhance security. | - US and allied attacks primarily for their own interests. |
| **Assumptions** | - Current defenses too slow and dispersed.<br>- Centralized and automated defenses better. | - Automated responses can limit extent of attacks.<br>- Design criteria and implementation can be shared. | - Automated responses have no direct control.<br>- No warning provided in advance of their use. |

| | | | |
|---|---|---|---|
| **Inferences** | Use of pre-determined indicators and cyberspace agents is sufficient and appropriate. | Use of pre-determined indicators and cyberspace agents requires coordination among allies. | Do any of the US responses inadvertently violate national sovereignty? |
| **Evaluation of Information** | When do decision makers know an attack and response have occurred? | Who is responsible for the coordinated development and maintenance of automated response systems? | Can countries receiving collateral damage respond? |
| **Implications** | Cyber national deterrence and defense enhanced? | Collective cyber deterrence and defense enhanced? | Potential escalation by automated means? |

The third perspective is that of other countries that may be concerned about the level of threat posed by primary (intentional) and collateral (unintentional) effects caused by the automated attack systems. They may have the point of view that the systems are designed to support interests other than their own, and assume that the automated responses have no deliberate control and thus will issue no advance warning of their use. The inference is that the automated attack response of others may inadvertently violate their own national sovereignty, thus giving cause to evaluation if they can respond in kind to any collateral damage absorbed. The implication is that such responses to automated responses may lead to a cycle of escalation largely driven by mechanisms detached from deliberate decision making.

## E. RECENT ACTIVITY REGARDING MILITARY CYBER RESPONSE

General Keith Alexander's March 2013 testimony to the U.S. Senate [28] outlined recent activity of U.S. Cyber Command worthy of critical analysis. The concern to clarify is how the U.S. military will respond to activities perceived as cyber attack. Alexander stated, "the Department of Defense and U.S. Cyber Command are being integrated in the machinery for National Event responses so that a cyber incident of national significance can elicit a fast and effective response to include pre-designated authorities and self-defense actions where necessary and appropriate." The point of view with regard to "fast and effective responses" in unclear, but Alexander mentioned that the inter-agency and international exercise CYBER FLAG "introduced new capabilities to enable dynamic and interactive force-on-force maneuvers at net-speed." From this perspective, can "pre-designated authorities and self-defense actions" include automated responses? If so, who determines if they are "necessary and appropriate," and what criteria do they utilize?

Two implicit assumptions in the testimony are that traditional organizational structures can handle the challenges in cyberspace and that negative events in cyberspace are threat-based. The inferences lead to traditional military approaches such as establishing three main levels of forces: a Cyber National Mission Force, a Cyber Combat Mission Force (supporting Combatant Commands), and a Cyber Protection Force (for DoD systems). These forces are pursuing normalized cyber operations for "a more reliable and predictable capability to be employed." Following such ethnocentric approaches may open vectors for manipulation by other actors. The evaluation of information includes the drive for increased operational awareness by such means as "a weekly Cyber Operating Directive (CyOD) across the DoD cyber enterprise…so that all 'friendlies' can understand what is happening in cyberspace." However, such useful measures may unknowingly foster ethnocentrism akin to the Operation RYAN activities surrounding *Able Archer*.

The implications are that U.S. cyber forces may be leaning toward a threat-based viewpoint of cyberspace that encourages the rapid identification and response to perceived aggressive action with little account for the broader dynamics of the information environment. But is this a realistic concern? The U.S. Department of State Legal Advisor, Harold Koh [29], stated the U.S. may legally respond to cyberspace activities "that amount to an armed attack or imminent threat thereof." Regarding capability and intent, he notes that the "United States has impressive cyber-capabilities" and "that adherence to established principles of law does not prevent us from using those capabilities to achieve important ends." Koh's views on the international legal aspects regarding the use of such capabilities is largely congruent with *Tallinn Manual* principles [30], and he stresses that the preferred use of such capabilities considers multilateral and regional issues.

# 5. POLICY RECOMMENDATIONS

Many other questions and implications can be examined using the critical-thinking framework. This section provides recommendations for cyberspace-related policy summarized from the historical and hypothetical cases as well as current trends examined above.

## *A. ROLE OF RESILIENCE*

Although automated cyber response measures may provide added security and deterrence, they also risk interacting with other mechanisms and indicators that may create reactive and escalatory vicious cycles such as those in case studies. Perhaps, instead the focus should be on fostering resilience, such as that proposed in the U.S Department of Homeland Security's healthy cyber ecosystem model, specifically calling for cyberspace resilience in critical infrastructure as well as business, social, and civic process [31]. The current NATO Policy for Cyber Defence [32] also lists resilience as an overarching principle (with prevention and non-duplication). Having sufficient resilience measures in place can provide strategic leaders with adequate time for critical thinking in their decision making. This can include evaluating information and options with the goal of keeping responses from becoming escalatory. Balancing the combination of resilience and automated responses should be evaluated in the context of a dynamic cyberspace environment where the success of a nation's strategy depends on the strategy of other nations, and their interaction and behavior will change the environment [33].

## B. ETHICAL IMPLICATIONS

Most of the debate among nations regarding cyber attack in general—let alone when such attack is automated—focuses on protecting their fundamental national purpose and interests. Thus, expanding the decision making to include international repercussions may only be done through the lens of *realpolitik* pragmatism. However, to nurture a more open and cooperative cyberspace environment, nations should also contemplate an ethical-based framework, possibly adopting first principles for Just Cyber Warfare proposed by Taddeo [34]. These principles state that Just Cyber Warfare should only be waged "against entities that endanger or disrupt the well-being of the Infosphere" and that it seeks to preserve, but not necessarily promote, the well-being of the Infosphere.

Leder and others [35] examine the struggle between what is technically feasible for the application of automated responses and the concern that they "may interfere with law or current ethical beliefs depending on their invasiveness and impact on third-parties." They examine ethics issues related to automated and proactive botnets that target control servers, traffic, or infected systems. Other researchers cite legal opinions that conclude that applying automated methods, such as "white worms" which enter systems to disinfect them from malicious software may be illegal if they operate without express consent of the owners [36].

## C. SIGNALLING BETWEEN NATIONS

If automated response options are being considered or are in place, this fact can be communicated to other countries as a sign of commitment as well as deterrence against escalation. Such clear signalling of intent among nations can help mitigate tension; as noted in the *Able Archer* case, knowledge of intent is more difficult to discern than knowledge of capability. Also, too much secrecy may work against clear deterrence and signalling whereas simple declaratory statements may enhance effectiveness [37]. For example, the announcement of the establishment of U.S.

Cyber Command by Secretary of Defense Robert Gates in June 2009 caught much of the world by surprise; it may have been more effective if it was coordinated with the State Department's diplomatic connections.

For like-minded states, communication could be enhanced by establishing terms of reference, such those explored by Prescott [38], regarding participation in cyber hostilities. Such communication should include factors regarding the nature of the diffusion and interdependence of cyber attacks across global regions [39]. It may also be useful to develop hypothetical escalation models such as a cyber form of the Kahn nuclear ladder; signalling may include publically stating where a nation has its automated responses enabled based on the details of such a construct. Such standards of nation signalling may help to form the basis to facilitate agreements that could eventually lead to formal cyber weapon treaties [40].

# 6. SUMMARY

A very dangerous event would be an accidental incident in cyberspace that was interpreted as an attack during a period of heightened tensions between two world powers--adding automated attack systems could make a bad situation into a catastrophic one.  Since such incidents have occurred in the physical domains of warfare during the last three decades, it is reasonable to assume they will happen in cyberspace. The application of critical thinking can help mitigate risk, but only if time is available for leaders to reflect. Adopting a policy that emphasizes cyber resilience may help provide time for decision makers to thoughtfully consider the situation and weigh alternatives. If automated attack responses are deemed necessary they should be implemented in a graduated manner that is signalled to potentially hostile nations. Adopting ethical principles of just cyber war may provide overarching guidance for the development and deployment of automated cyber attack responses that strive to preserve the overall well-being of cyberspace while protecting nation purposes and interests.

## REFERENCES

[1] T.Rid and P. McBurney. "Cyber-Weapons." RUSI Journal, vol. 157, no. 1, pp. 6-13, February/March 2012.

[2] S. Gerras. "Thinking Critically about Critical Thinking: A Fundamental Guide for Strategic Leaders." Carlisle, Pennsylvania: U.S. Army War College, August 2008.

[3] K. Alexander. Statement before the Senate Committee on Armed Services, Washington, D.C., 12 March 2013, pp. 3.

[4] J. Michel et al. “Measured Responses to Cyber Attacks Using Schmitt Analysis: A Case Study of Attack Scenarios for a Software-Intensive System.” *Proc. of Twenty-seventh Annual International Software and Applications Conference, IEEE*, 2003.

[5] M. Schmitt. “Attack as a Term of Art in International Law: The Cyber Operations Context.” *Proc. 4th International Conference on Cyber Conflict*, 2012, pp. 283-293.

[6] *Tallinn Manual on the International Law Applicable to Cyber Warfare*. General editor M. Schmitt. New York: Cambridge University Press, 2013.

[7] J. Caton. “Cyberspace and Cyberspace Operations” in *Information Operations Primer: Fundamentals of Information Operations*, AY12 Edition. Carlisle, Pennsylvania: U.S. Army War College, November 2011, pp. 19-32.

[8] K. Ziolkowski. “*Ius ad bellum* in Cyberspace-Some Thoughts on the “Schmitt-Criteria” for Use of Force.” *Proc. 4th International Conference on Cyber Conflict*, 2012, pp. 295-309.

[9] R. Fanelli and G. Conti. “A Methodology for Cyber Operations Targeting and Control of Collateral Damage in the Context of Lawful Armed Conflict.” *Proc. 4th International Conference on Cyber Conflict*, 2012, pp. 319-331.

[10] H. Kahn. *On Escalation: Metaphors and Scenarios*. New York: Praeger, 1965.

[11] S. Gerras. “Thinking Critically about Critical Thinking: A Fundamental Guide for Strategic Leaders.” Carlisle, Pennsylvania: U.S. Army War College, August 2008, pp. 9.

[12] E. Tyugu. “Command and Control of Cyber Weapons.” *Proc. 4th International Conference on Cyber Conflict*, 2012, pp. 333-343.

[13] K. Geers. Strategic Cyber Security. Tallinn, Estonia: NATO Cooperative Cyber Defence Centre of Excellence, 2012, pp. 119.

[14] F. Hare. “The Significance of Attribution to Cyberspace Coercion: A Political Perspective.” *Proc. 4th International Conference on Cyber Conflict*, 2012, pp. 125-139.

[15] J. Healey. *Beyond Attribution: Seeking National Responsibility for Cyber Attacks*. Washington D.C: The Atlantic Council, 2011.

[16] A. Manchanda. “When Truth is Stranger Than Fiction: The *Able Archer* Incident.” *Cold War History*, vol. 9, no. 1, pp. 111-133, February 2009.

[17] T. Czerwinski. *Coping with the Bounds: Speculation on Nonlinearity in Military Affairs*. Washington, D.C.: National Defense University, 1998, pp. 98-99.

[18] A. Manchanda. “When Truth is Stranger Than Fiction: The *Able Archer* Incident.” Cold War History, vol. 9, no. 1, pp. 117-118, February 2009.

[19] D. Hoffman. *The Dead Hand: The Untold Story of the Cold War Arms Race and Its Dangerous Legacy.* New York: Doubleday, 2009, pp. 94-96.

[20] B. Fisher. “Anglo-American Intelligence and the Soviet War Scare: The Untold Story.” *Intelligence and National Security*, vol. 27, no. 1, pp. 75-92, February 2012.

[21] P. Pry. *War Scare: Russia and America on the Nuclear Brink*. Westport, Connecticut: Praeger, 1999, pp. 214-241.

[22] D. Hoffman. The Dead Hand: The Untold Story of the Cold War Arms Race and Its Dangerous Legacy. New York: Doubleday, 2009, pp. 399-400.

[23] P. Pry. *War Scare: Russia and America on the Nuclear Brink*. Westport, Connecticut: Praeger, 1999, pp. 214-241.

[24] S. Kan. *China's Anti-Satellite Weapon Test*. CRS Report for Congress RS22652, Washington D.C.: Library of Congress, 23 April 2007.

[25] D. Ball. *Assessing China's ASAT Program*. Austral Special Report 07-14S. Melbourne: RMIT University, 14 June 2007.

[26] S. Kan. China's Anti-Satellite Weapon Test. CRS Report for Congress RS22652, Washington D.C.: Library of Congress, 23 April 2007.

[27] "An update of the FY-1C, Iridium 33 and Cosmos 2251 Fragments." *Orbital Debris Quarterly News*, vol. 17, no. 1, pp. 4-7, January 2013.

[28] K. Alexander. Statement before the Senate Committee on Armed Services, Washington, D.C., 12 March 2013.

[29] H. Koh. "Remarks as Prepared for Delivery By Harold Hongju Koh to the USCYBERCOM Inter-Agency Legal Conference, Ft. Meade, MD, Sept 18, 2012." *Harvard International Law Journal (Online)*, vol. 54, pp. 1-12, December 2012.

[30] M. Schmitt. "International Law in Cyberspace: The Koh Speech and Tallinn Manual Juxtaposed." *Harvard International Law Journal (Online)*, vol. 54, pp. 13-37, December 2012.

[31] "Enabling Distributed Security in Cyberspace: Building a Healthy and Resilient Cyber Ecosystem with Automated Collective Action." Washington, DC: Depart. of Homeland Security, Mar. 23, 2011, pp. 8, 26.

[32] "Defending the Networks: The NATO Policy on Cyber Defence." Brussels, Belgium: NATO Public Diplomacy Division, Jun. 2011.

[33] R. Jervis. "From Complex Systems: The Role of Interactions." in *Coping with the Bounds: Speculations on Nonlinearity in Military Affairs*, Washington, D.C.: National Defense University, 1998, pp. 259-277.

[34] M. Taddeo. "An Analysis For A Just Cyber Warfare." *Proc. 4th International Conference on Cyber Conflict*, 2012, pp. 209-218.

[35] F. Leder et.al. "Proactive Botnet Countermeasures: An Offensive Approach." *Proc. of the Conference on Cyber Warfare*, 2009, pp. 12-14.

[36] L. Vihul et al. *Legal Implications of Countering Botnets*, Joint report from the NATO Cooperative Cyber Defence Centre of Excellence and the European Network and Information Security Agency (ENISA). Tallinn: CCDCOE, 2012, pp. 45-46.

[37] D. Alperovitch. "Toward Establishment of Cyberspace Deterrence Strategy." *Proc. 3rd International Conference on Cyber Conflict*, 2011, pp. 87-94.

[38] J. Prescott. “Direct Participation in Cyber Hostilities: Terms of Reference for Like-Minded States?” *Proc. 4th International Conference on Cyber Conflict*, 2012, pp. 251-266.

[39] Kim et al. “Comparative Study of Cyberattacks.” *Communications of the ACM*, vol. 55, no. 3, pp. 66-73, March 2012.

[40] L. Arimatsu. “A Treaty for Governing Cyber-Weapons: Potential Benefits and Practical Limitations.” *Proc. 4th International Conference on Cyber Conflict*, 2012, pp. 91-109.