2011 3rd International Conference on Cyber Conflict
C. Czosseck, E. Tyugu, T. Wingfield (Eds.)
Tallinn, Estonia, 2011 © CCD COE Publications

Towards Establishment of Cyberspace Deterrence Strategy

Dmitri Alperovitch
McAfee Inc.,
dmitri_alperovitch@mcafee.com

Abstract— **The question of whether strategic deterrence in cyberspace is achievable given the challenges of detection, attribution and credible retaliation is a topic of contention among military and civilian defense strategists. This paper examines the traditional strategic deterrence theory and its application to deterrence in cyberspace (the newly defined 5th battlespace domain, following land, air, sea and space domains), which is being used increasingly by nation-states and their proxies to achieve information dominance and to gain tactical and strategic economic and military advantage. It presents a taxonomy of cyberattacks that identifies which types of threats in the confidentiality, integrity, availability cybersecurity model triad present the greatest risk to nation-state economic and military security, including their political and social facets. The argument is presented that attacks on confidentiality cannot be subject to deterrence in the current international legal framework and that the focus of strategy needs to be applied to integrity and availability attacks. A potential cyberdeterrence strategy is put forth that can enhance national security against devastating cyberattacks through a credible declaratory retaliation capability that establishes red lines which may trigger a counter-strike against all identifiable responsible parties. The author believes such strategy can credibly influence nation-state threat actors who themselves exhibit serious vulnerabilities to cyber attacks from launching a devastating cyber first strike.**

Keywords: cyberdeterrence, cyberspace, strategy, first-strike, counter-strike, confidentiality, integrity, availability

I. Introduction

Deterrence is a psychological 'game of chicken' that attempts to influence the cognitive state of the potential adversary actor and prevent them from embarking on course of action that they may wish to take. US Department of Defense defines

deterrence as the 'prevention from action by fear of the consequences. Deterrence is a state of mind brought about by the existence of a credible threat of unacceptable counteraction[1].' This is accomplished through either a directed or latent coercion that convinces the opponent that the costs of action are extremely prohibitive. Much of the strategic deterrence theory, which had been developed in the aftermath of World War II and enhanced over the duration of the Cold War, had focused on nuclear deterrence, achieved through the overt or opaque threat of nuclear force retaliation. This paper examines the traditional strategic deterrence theory and it's application to deterrence in cyberspace, which is the newly defined 5th battlespace domain, following land, air, sea and space domains and is being used increasingly by nation-states and their proxies to achieve information dominance and gain tactical and strategic economic and military advantage[2]. The paper does not attempt to address tactical cyberdeterrence for attacks that do not cause strategic damage to national or economic security, although it is possible to foresee a 'death by a thousand cuts' scenario, where numerous smaller cyberattacks can amount to a strategic threat.

II. DETERRENCE THEORY

There are two essential components to any viable deterrence strategy. In order for deterrence to be effective and credible, one must convince the potential adversary that you possess both the **capability** and the **will** to either retaliate or initiate a first preemptive strike to thwart an eminent attack[3].

The retaliation capability comprises of the ability to timely detect a threat (before the counter-strike assets or the C2 (command and control) to launch them are destroyed in the attack), rapid C2 decision-making and execution to launch a retaliatory or preemptive strike and ability to inflict prohibitively costly damage on the aggressor through such a strike. This capability can be demonstrable and overt, as with the unconcealed nuclear triad forces of the United States and Soviet Union during the Cold War, or unannounced and opaque, as with the widely assumed but unacknowledged nuclear arsenal of the state of Israel.

The will to retaliate, on the other hand, is a more nebulous psychological concept that is comprised of reputation (such as past willingness to use nuclear arms for the United States or historic proclivity to not put significant value on human life for the Soviet Union) and declaratory first-strike or second-strike policy. In the case of Israel's nuclear deterrence strategy, while there is no declaratory policy of nuclear retaliation due to its policy of ambiguity, there is a declaratory 'Shoah-proof' or 'Never Again' security doctrine which is designed to influence an adversary's thinking on the willingness of the country's leaders to use its undeclared nuclear arsenal as the last resort scenario[4].

III. TAXONOMY OF CYBERATTACKS

The rules of deterrence do not change once we move from a nuclear domain to a cyberspace one. The goal remains to influence the opponent's evaluation of one's capability and will to retaliate in response to an imminent cyber threat in order to dissuade them from ever launching the attack.

In order to evaluate what capability must exist for effective cyberdeterrence, we must first examine the types of cyberattacks that are in the realm of possible and analyze which ones have potential to be deterred.

For over two decades, the CIA (Confidentiality, Integrity, Availability) triad has been used in industry and academia to model the fundamental principles of information security[5]. It states that the goals of an information security system are to provide Confidentiality, Integrity and Availability of information. Conversely, cyberattacks can be classified using the same model as they exploit one or more of these attributes as they attempt to either steal information, or attack the Confidentiality part of the triad, modify information, or attack its Integrity, or prevent access to information, attacking its Availability.

IV. CONFIDENTIALITY ATTACKS

Confidentiality attacks (also referred by the US Department of Defense to as CNE, or Computer Network Exploitation) are nothing more than traditional espionage achieved through high tech means. Most of the sophisticated cyber attacks that are seen launched by either nation states or criminal groups fall into this category. History has shown us that espionage, known as the second oldest profession, has been around for nearly as long as there has been a human civilization and is an act that, while considered to be a sign of unfriendly relations, has become an internationally accepted norm that typically does not trigger more than a diplomatic retaliatory response[6]. Even during the darkest days of the Cold War and while faced with pervasive Soviet espionage activity in its most sensitive national security area – the 'Manhattan Project' initiated to design, create and test a nuclear weapon, the United States had not considered any type of retaliation beyond criminal prosecution of the spies and occasional expulsion of diplomats. It is widely acknowledged that even friendly nations spy on each other and when such activity is detected, it rarely has any effect, other than perhaps a fleeting chill placed on diplomatic relations. The international norms of just war theory dictate that retaliation must be proportional to the harm suffered from the attack[7]. Thus, it is unimaginable to envision a non-pariah state on the international scene responding with an overwhelming destructive attack to a case of cyber-espionage activity, no matter how damaging the loss of information had been to vital national security interests. And without the threat of massive response, the key pillar of the deterrence strategy is removed, preventing effective deterrence of confidentiality attacks in cyberspace.

V. INTEGRITY ATTACKS

Attacks on integrity are much more insidious as they are designed to achieve a tactical or strategic advantage over an adversary by sabotaging the operation of their critical civilian or military information systems. The sabotage can involve manipulation of data inside information systems that can degrade or distort the situational awareness capability of the adversary by spreading misinformation inside their intelligence systems with either a tactical objective to obscure specific activities that may be under surveillance or to achieve a strategic surprise in preparation for an attack. It can also involve subversion of physical devices and processes that are guided or operated by information systems, such as manipulation of weapons guidance systems to cause them to fire off-target. Targets can also include civilian critical infrastructure resources, such as electric grid, stock market and other financial databases, water filtration plans and others. The Stuxnet worm, discovered in June 2010, is believed to be the first publicly known nation-state sponsored integrity cyberattack, which is speculated to target the Iranian uranium enrichment program for subtle and long-term sabotage with the goal of destroying Iran's centrifuges by covertly making them spin at faster frequencies than they had been designed to do[8]. It is quite clear that these types of integrity cyberattacks pose a severe danger to advanced nation-states, whose economies, critical infrastructure and military systems are dependent on information systems, as they can be used to remotely wreak havoc on financial, energy, food, water, and transportation infrastructure sectors, as well as degrade abilities of advanced militaries to collect and analyze reliable battlefield intelligence and even execute kinetic operations. Deterrence of these types of attacks must be a priority for any effective cyberdeterrence policies.

VI. AVAILABILITY ATTACKS

Availability attacks are those that attempt to bring information systems offline in order to shutdown or destroy critical physical or virtual processes or prevent access to information. Long-lasting attacks can cause devastating damage to the economy, such as those that cause prolonged electricity or communication network blackouts. Short duration attacks that are surgically targeted at intelligence collection and analysis capability can blind a nation's ability to see an immediate strategic conventional or broader cyber threat by denying defenders access to vital situational awareness data or intelligence resources. Thus, just as with integrity attacks, availability threats can, under certain circumstances, present a serious national security danger and must be deterrable in a broader deterrence strategy.

VII. DETERRENCE IN CYBERSPACE

To achieve deterrence against integrity and availability cyberattacks, according to the deterrence theory, requires that a nation-state first build a credible threat detection capability that will protect its ability to counter-strike. While advanced cyber threats can use advanced obfuscation and polymorphic techniques to avoid detection for a prolonged period of time (as Stuxnet had done, avoiding all public detection for at least 12 months since its earliest proven sighting in the wild in 2009), the chances of them avoiding discovery permanently are quite low as their sabotage or other destructive activities will likely bring attention to themselves sooner or later.

Attribution is another problem that presents a challenge to the detection capability. It is very difficult and, often, impossible to accurately and quickly attribute a cyberattack, once it is discovered, to a specific adversary through technical means alone. The anonymity of the Internet easily allows an attacker to lay a false trail and hide behind a myriad of intermediately hop-points or proxy actors. The use of traditional non-cyber intelligence resources, including HUMINT and SIGINT, can help with that goal but they also cannot provide a reasonable level of assurance required for credible deterrence that the attacker will be identified. Despite this challenge, it is conceivable that deterrence will be effective even without accurate and timely attribution. For one, the required level of attribution required for a counter-strike is directly proportional to the degree of criticism a nation-state is prepared to endure in the international and domestic courts of public opinion and is greatly influenced by the destructiveness of the original threat and a cui bono analysis of the attack objectives. The threshold is not proof beyond reasonable doubt in the court of law but sufficient mix of suspicion and evidence to justify the retaliatory strike to the plurality of domestic and international audiences. For instance, strategic context of international relations at the time at which a cyber attack may take place can offer strong clues as to its origins[9].

One of the other major challenges one faces in applying lessons learned from traditional strategic deterrence theories to cyberspace is timely detection of the attack. Due to the nature of cyber weapons, cyber offensive capability can be developed, tested and pre-deployed offline without any credible means for detecting it. Unlike physical weapons, there is no missile silo, mobile launch facility or submarines to observe and monitor for early-warning detection. The attacks themselves may propagate literally with a speed of light when deployed through fiber optic networks, but even on slower copper networks, the speed of fully automated cyber ordinance will still be faster than what a human can reasonably evaluate and respond to. Deployment of defensive measures with automated retaliation capability, on the other hand, presents too high of a risk for misfire or targeting of an innocent party due to attribution challenges.

Accurately and timely determining the intent of the attack, once it has been discovered, is yet another problem. If international norms that dictate

proportionality of response prevent retaliation to attacks on confidentiality, or espionage, determining whether the goal of an intrusion is to attack confidentiality, integrity or availability of information system becomes nearly just as critical as detecting the attack itself. The process of intrusion classification can be very challenging for a defender, at least in its initial stages, as the tactics, techniques and procedures (TTPs) may be identical for all types of attacks. For instance, attacks on confidentiality, integrity and availability may all begin with an intrusion exploiting a vulnerability in an externally connected information system, followed by malware deployment, which provides the adversary with full remote access capabilities to the internal network. Until that remote access capability is leveraged for an integrity or availability attack, it may be impossible to know through technical means the true purpose of the intrusion. This uncertainty further adds to the complexity of establishing an early-warning detection system that can provide sufficient time for the appropriate Command and Control (C2) decision makers to evaluate the information and launch a counter-strike.

Advanced defensive tactics, technologies and highly trained personnel will contribute to the shrinking of the detection and classification gap. Separation of defensive and offensive resources, such as storage of offensive cyberweapons in offline locations which are less vulnerable to virtual targeting and distributing the retaliatory information systems and networks across wide virtual and physical space will help to build credible resilience to the counter-strike force. This can reduce the reliance on rapid detection and classification of inbound attack by providing the means for the decision makers to retaliate even after suffering a devastating first strike, minimizing the chance that the adversary can count on taking out all of the counter-strike assets in a single attack.

Second, is the need to preserve a rapid C2 decision-making and execution of a counter-strike option when facing a devastating cyber attack. This can be accomplished by preserving the resiliency and integrity of command chain communications by instituting or preserving offline communications channels that are less likely to be impacted by cyber attacks, such as dedicated traditional secure POTS (plain old telephone service) lines and encrypted radio and satellite communications that are physically separated from virtual networks which can carry attack codes.

Third, the counter-strike itself must be capable of instituting devastating damage on the attacker's own virtual and physical infrastructure to make the first-strike prohibitively expensive. Limited public demonstrations of cyber offensive capabilities can serve a useful purpose in alerting potential opponents to what they may face should they decide to attack. However, this part of the deterrence equation presents the biggest challenge to developed nation-states with advanced cyber defensive and offensive capabilities but who face developing nation-state adversaries with dangerous offensive cyber weapons but are themselves not reliant on cyberspace for their national economic or military interests. It is hard to cause

prohibitively devastating damage on your opponent through cyber means alone if his vital infrastructure is completely disconnected from the network. This problem presents a serious conundrum to policy makers, who face the unappealing choice of rising up the escalatory ladder and retaliating with conventional or perhaps even nuclear weapons in response to a cyber-only attack, in the process risking violations of international norms of proportional response, or absorbing the attack without a response and looking weak to their enemies, friends and populations alike. Yet, while this is a significant unresolved policy problem today, it is reasonable to expect that its consequences will lessen with time, as more and more developing countries rapidly increase their reliance on cyberspace in order to reap the economic, efficiency and force-multiplier benefits it affords.

Lastly, political leaders must demonstrate the credible will to issue a cyber counter-strike in response to a highly damaging integrity or availability attack of national security consequence. In today's world of complete ambiguity with regards to cyber-offense that can create much uncertainty in the minds of potential opponents, this can be best accomplished with a declaratory policy that defines, even if in opaque terms that provide sufficient room for decision makers to maneuver, the red lines that will trigger a counter-strike or even a preemptive first-strike in response to a credible and imminent threat.

VIII. CONCLUSION

This paper has argued that an effective cyberdeterrence strategy must focus on consequential integrity and availability cyber attacks and deter them through a declaratory policy that establishes red lines which may trigger a counter-strike against all identifiable responsible parties. It is also advantageous to provide public limited demonstration of offensive capabilities, spend resources on increasing threat detection and resiliency of both defensive and offensive networks, increase off-line redundancies for C2 communications and enhance HUMINT and SIGINT intelligence collection and analysis efforts to focus on cyber threat actors, their capabilities and intentions. This strategy can credibly influence nation-state threat actors who themselves exhibit serious vulnerabilities to cyber attacks from launching a first strike.

IX. FUTURE EXPLORATION

This paper did not explore several areas that need to be covered by a comprehensive cyberspace strategic deterrence doctrine and that should be explored in future works. These areas include the challenge of deterring non-state actors, such as terrorist and criminal groups, a problem that has not been adequately solved in neither the physical nor the virtual world. Another aspect that should be examined in the future is whether tactical as opposed to strategic cyberdeterrence is achievable against attacks that don't rise to the level of strategic impact.

ACKNOWLEDGMENT

The author would like to acknowledge Greg Conti, Adam Meyers, Jose Nazario, Jason Shepherd, and Jeff Stambolsky for their valuable and thoughtful contributions to the ideas expressed in this paper.

REFERENCES

[1] Joint Publication (JP) 1-02, *Department of Defense Dictionary of Military and Associated Terms*, 8 November 2010 (as amended through 31 January 2011), p107, http://www.dtic.mil/doctrine/new_pubs/jp1_02.pdf

[2] R. Rozoff, *U.S. Cyber Command: Waging War in the World's Fifth Battlespace* (Montreal: Centre for Research on Globalization, May 27 2010), http://www.globalresearch.ca/index.php?context=va&aid=19360

[3] R. Powell, *Nuclear Deterrence Theory: The Search for Credibility* (Cambridge: Cambridge University Press, 1990), p8

[4] A. Cohen, *Israel and the Bomb* (Columbia University Press, 1998)

[5] Department of Defense Directive 8500.01E, "Information Assurance (IA)", October 24, 2002 (certified current as of April 23, 2007), p3, http://www.dtic.mil/whs/directives/corres/pdf/850001p.pdf

[6] P. Knightley, *The second oldest profession: spies and spying in the twentieth century* (Pimlico, 2003)

[7] E. Patterson, *Just War Thinking: Morality and Pragmatism in the Struggle against Contemporary Threats* (Lexington Books, 2007), p63-69

[8] D. Albright, P. Brannan, and C. Walrond, "Did Stuxnet Take Out 1,000 Centrifuges at the Natanz Enrichment Plant? Preliminary Assessment," *Institute for Science and International Security (ISIS)*, http://isis-online.org/uploads/isis-reports/documents/stuxnet_FEP_22Dec2010.pdf

[9] E. Sterner, "Retaliatory Deterrence in Cyberspace," *Strategic Studies Quarterly*, Spring 2011