Cyber Deterrence by Engagement and Surprise

By Jim Chen

The conventional deterrence strategies of denial and punishment do not factor in the unique characteristics of the man-made cyber domain. This domain needs a new and holistic deterrence strategy that involves prompt and direct cyber responses that are sudden, dynamic, stealthy, and random so that adversaries can be defeated mentally and virtually. This article offers such an approach that I refer to as "deterrence by engagement and surprise."

Deterrence

Released in January 2017, Department of Defense Joint Publication 3–0 defines deterrence as "the prevention of action by the existence of a credible threat of unacceptable counteraction and/or belief that the cost of action outweighs the perceived benefits."[1] To make it effective, deterrence should depend on capability, credibility, and communication:

- capability helps to destroy what the adversary values most highly, thus making the cost of an attack exceed the benefit that an adversary could gain;
- credibility can be achieved via the demonstration of the willingness to use capability;
- communication requires capability, the willingness to use capability, and that credible consequences be made known to an adversary.

Simply put, deterrence is a coercive approach used for the purpose of avoiding a war or preventing the escalation of a war. It is used as a strategy to help achieve goals, and varied means can be adopted and diverse capabilities can be used to support such a strategy.

Our current deterrence strategies are heavily influenced by the nuclear and conventional deterrence models—deterrence by denial and deterrence by punishment. Strategist Herman Kahn held that defensive capabilities should be greatly enhanced to limit damage caused by an adversary, so that retaliation by the adversary can be countered, and a credible and real threat can be generated against the adversary during a conflict. In this sense, the capability to defend oneself for survival is a key element. This approach lays the foundation for deterrence by

Dr. Jim Q. Chen is a professor of Cybersecurity Studies in the College of Information and Cyberspace at National Defense University.

denial, which intends to scare an adversary away by denying his ability to inflict sufficient harm to justify the risk of retaliation.

Strategist Thomas Schelling, however, argued for the deterring effect of uncertainty in a stable balance of terror. He used uncertainties as the magic of threats since an adversary may fear irrationality or accident. As well explained by former Deputy Assistant Secretary of Defense Keith Payne, stable deterrence, which provides reliable, predictable, and mutual deterrence, "could be orchestrated to proceed from mutual prudence born of mutual vulnerability."[2] It is a strategy of having the other party be ultimately "persuaded to exercise self-control" because of the irreversible and disastrous consequences that may ensue without self-control. Payne retains, during the Cold War, the basic ingredients of this theory were the U.S. capability to threaten nuclear retaliation against the Soviet Union as well as the vulnerability of U.S. society to Soviet nuclear attack.[3] In this sense, uncertainties are involved in the outcome of this strategy as one does not directly control an adversary, who makes decisions on how to act and what to do. This approach lays the foundation for deterrence by punishment.

In the cyber domain, deterrence by punishment does not work well owing to the complexities of attribution and the challenges of stealth operations. To have a measure in place, deterrence by denial brings in responses from diplomatic, military, economic, political, legal, ethical, and other instruments of national power. If it is well prescribed, this approach can make an adversary feel the pressure and pain from multiple domains, thereby deterring further action in the cyber domain. However, this approach requires a well-orchestrated and near-perfect collaboration from all relevant domains—something that is difficult to achieve within a short period.[4]

The current DOD cyber strategy calls for a holistic approach, asserting that the deterrence of cyberattacks against U.S. interests will only be achieved through "the total of U.S. action, including declaratory policy, substantial indications and warning capabilities, defensive posture, effective response procedures, and the overall resiliency of U.S networks and systems."[5] If there is a gap in collaboration, however, the effectiveness of deterrence is immediately reduced.

A lack of a deterrence theory or a framework that accounts for the unique challenges in the cyber domain may account for the present ineffectiveness of cyber deterrence. The next question is how best to develop such a theory or framework to be effective in the cyber sphere?

Missing Components

An intensive study of deterrence indicates it is neither strictly an offensive nor defensive approach, despite a close relation to both. Offense and defense are focused on external factors while deterrence requires a near-simultaneous focus on both external and internal factors.

- The external factor reflects the unambiguous exhibition of power that serves as an enormous threat to the other side. This power projection is supported by unmatchable capabilities in number, volume, quantity, quality, size, and other relevant components.
- The internal factor reflects the intimidation truly felt by the other side. This overwhelming state is accompanied by the feeling of being exhausted, helpless, and defenseless. This can help to convince adversaries of the potential damage and failure that they are going to receive if they continue what they are doing. This psychological state could be reached through a number of factors, to include surprise. If surprise is so strong that it leads to a shock, intimidation may ensue.

Depending on the context, deterrence might have a closer relation to offense or defense. Offense,

deterrence, and defense can be launched at nuclear force level, at physical force level, at cyber level, and at the diplomatic and economic level.[6] If offense, defense, and deterrence strategies are inserted into each level, a revised representation of levels can be generated:

- Nuclear force: Nuclear weapons can be used in an offensive operation and for nuclear deterrence. Missile defense systems such as the Terminal High Altitude Area Defense (THAAD) systems can be used for defense.
- Conventional physical force: In a small-scale conflict, automatic weapons can be used in an offensive operation or as physical deterrence. In this event, body armors such as bulletproof vests can be used for defense.
- Cyber: Cyber weapons such as denial-of-service tools can be used in offensive operations. However, they are not effective for cyber deterrence, as they are less violent than other means of deterrence such as nuclear weapons. Firewalls, intrusion detection systems, intrusion prevention systems, anti-malware tools are used for defense.
- Diplomatic and economic: Measures such as sanctions can be used in offense or for economic deterrence. Improving diplomatic and economic relations with third-party countries and adjusting internal markets are measures that can be used for defense.

Offense may restrictively be applied at the cyber level. However, there is no unique and effective deterrence at the cyber level.

Unique Characteristics of the Cyber Domain

Current cyber deterrence approaches are polarized, either focused on deterrence by punishment or on deterrence by denial. These approaches do not factor in the unique characteristics of the man-made cyber domain, which resembles a black box. Someone who uses a network connection and runs an operating system or perhaps an application, has no concept of how networks are connected, what codes are required for the operating systems, and what codes are executed for the application. Codes are run and processed at low levels while human machine interface occurs at a high level, supporting anonymity. When this anonymity is used in defense, it is privacy protection. When this is used in offense and in deterrence, it becomes stealth operations.

Given stealth, surprise can be generated at the user end; stealth maneuvers can be launched; and intelligence can be collected covertly, even with meta-data. Cyber feature sets, which include intelligence collection, stealth maneuvers, and surprise effect, can serve as force multipliers and eventually lead to military dominance if they are integrated appropriately into conventional military capabilities.[7] An examination of retaliation in the cyber domain reveals five unique features:

- Targeting is not an easy task, as attribution in cyberspace may require substantial time and effort. The delay in attribution affects deterrence by punishment more than deterrence by denial, as the former requires a target be accurately identified prior to any retaliatory response.
- Cyber weapons are not as severe as nuclear weapons or other physical weapons. There is no virtual massive destructive weapon like a nuclear weapon in the cyber domain currently, even though critical infrastructure might be targeted in an attack. In this sense, cyber retaliation is relatively limited in scale and capacity.
- Uncertainty is required for deterrence by punishment. It does not matter whether it is used in the physical world or in cyberspace.
- Retaliation is expected to be executed within a short period of time, especially in the cyber domain.

- Cyber weapons can generate unique effects that nuclear weapons or other physical weapons cannot generate. Likewise, they are good at generating surprise effects in the virtual environment, or in a combination of the virtual and physical environments.

Deterrence by Engagement and Surprise

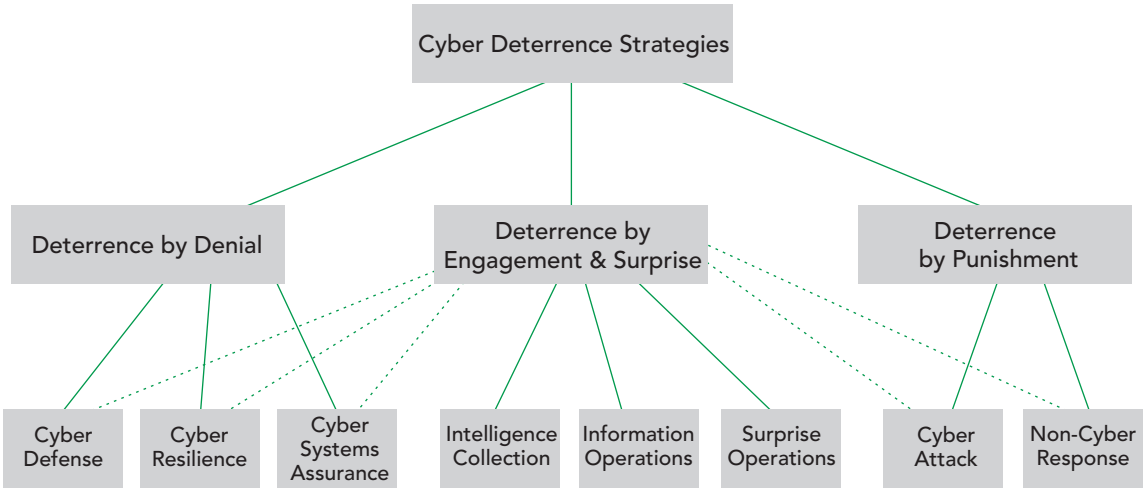
Deterrence by engagement and surprise offers the depth and flexibility to support sudden, dynamic, and random changes initiated by different contexts. Empowered by artificial intelligence (AI) and machine learning, this deterrence strategy is able to effectively and efficiently support intelligence collection, information operations, and surprise operations.

Intelligence Collection

Utilizing various intelligent sensors in varied parts of networks, collections engage the devices used by an attacker in revealing their true identity overtly and covertly via multiple channels and methods right after the devices that an attacker uses make abnormal requests. Such engagement, supported by machine learning, contributes to accurate attribution. It can eventually lead to precise and prompt targeting. Intelligent sensors can collect relevant information whenever necessary and feed it into machine learning algorithms. They take advantage of the fact that no hacker can control every single device along a transmission route in the internet environment. This makes it possible for such sensors to record the Medium Access Control (MAC) address and the Internet Protocol (IP) address of both the sender and the recipient in any leg of transmission. If the information of the previous leg is unknown, an engagement is initiated to chat with the device, such as a router, a switch, a proxy device, or a host device, to find out the relevant information. This capability can be built with the ability-to-learn algorithms powered by AI.

Artificial intelligence also makes it possible for a cyber weapon to mutate its appearance or even rewrite itself completely based on the context of when it is executed. In this sense, it is perpetually changing its behavior. In addition, different phases of maneuvers can be initiated from different parts of the world, thus confusing an adversary in finding out who sent out the responses. The dynamics built here help to create a stealth environment for cyber maneuvers.

FIGURE 1: Deterrence by Engagement and Surprise.



Information Operations

Advances in AI are able to drive change in information superiority. The capabilities for the collection and analysis of data as well as capabilities for the creation and manipulation of data can be dramatically improved. Disinformation and misinformation can appear persuasive. Meanwhile, "AI-enhanced forgery of audio and video media is rapidly improving in quality and decreasing in cost."[8] Likewise, AI can further improve electronic warfare (EW), computer network operations (CNO), psychological operations (PSYOP), military deception (MILDEC), and operations security (OPSEC), thus enhancing information-related capabilities (IRCs) "to gain advantages in the information environment" and "to influence, disrupt, corrupt, or usurp the decision-making of adversaries and potential adversaries while protecting our own."[9] All of these capabilities can be used to confuse and frighten adversaries.

Surprise Operations

Stealth maneuvers generate unexpected actions from various angles, aspects, directions, and locations, be it physical or virtual. The virtual munitions are loaded with varied payload. They range from audio warnings to light-weighted offensive operations. The virtual munitions are dynamically utilized based on contexts. A dynamic defense posture is thus created. This can successfully take an adversary by surprise psychologically, disabling his/her willingness to fight further or to continue the attack.[10]

AI systems can randomly select locations for launching surprise operations, which makes it difficult for an attacker to determine where the countermeasures are actually coming from. AI systems can also generate various responses, such as a surprise warning message, audio sound, web page, video clip, or anything that can warn or scare an attacker individually via different media. The purpose of surprise responses is to make an attacker realize the fact that he is under close surveillance and is responsible for what he is doing. This takes away the advantage of a covert cyberattack. Unless it is in an overt conflict, the attacker will withdraw from the attack in most cases unless he willing to receive the punishment. Meanwhile, evidence collection for digital forensics gets started. Determinant of the situation, a cyber offensive operation can also be launched as a retaliatory measure if it is legal and necessary. In addition, relevant diplomatic, legal, economic, and military measures can be taken.

This approach to deterrence can help foster a state of mind that decisively influences the decisionmaking calculus of the adversary who sees the intolerable consequence of aggressive action and who starts to fear such consequences.[11] Ultimately, this new approach is able to generate significant impact virtually, psychologically, morally, and physically.

Virtual

Virtual impact is achieved via intelligent responses from autonomous computer systems, supervised by humans if needed. Responses are usually at machine speed, avoiding any unnecessary delay. They are either defensive or offensive, or both, based on the specific situation, even though they are less severe. These responses are also precise as they are pointing exactly toward perpetrators with the help of collected intelligence. With respect to functionality, they can reject illegitimate requests, disable services, generate alerts, call in additional defensive forces, log abnormal events, back-track to find out the identity of the device that makes the initial request or even the individual who uses that device to make the initial request.

Psychological

This is achieved through surprise responses that range from a warning utilizing text, image, voice, or video messages on relevant devices including the initial device once discovered. These unexpected responses are manipulated by AI algorithms. Clearly

displayed as an unambiguous exhibition of power via disparate capabilities, the responses are used to scare adversaries. When they suddenly realize that there are some unknown but powerful capabilities possessed by the opposing force, adversaries will reconsider the continuation of their attacks as they are not certain about the consequences of their attack actions. In this way, their cyber aggression can be dissuaded.

Moral

The moral impact is achieved via surprise responses that remind the user of the relevant devices of the moral and legal responsibilities they have in cyberspace.

Physical

The physical impact is achieved via intelligent systems under the close supervision of humans. It can cause disruption or destruction of a physical system.

Advantages of Deterrence by Engagement and Surprise

This new, holistic approach can successfully handle the challenge in a manner that deterrence by denial and deterrence by punishment cannot—it fills the deterrence gap. Engagement and surprise can lead to accurate attribution and precise targeting. It can also help to build a strategic buffer zone in the cyber domain and also help to eliminate the delay in responses as a whole. It applies not just to state actors but also nonstate actors and can help to avoid unnecessary escalation of conflict while providing prompt, dynamic, flexible, expandable, and effective retaliatory responses. This game changing capability offers at least nine advantages:

- It bridges the deterrence gap, thus enriching the theory and forming a holistic approach for which new deterrence mechanisms can be developed.
- Capability is exhibited in a unique way without delay and, during this process, credibility is enhanced through an effective display.
- The approach also addresses the unique characteristics of the cyber domain, so that responses can be generated at the cyber level thereby helping to avoid escalation.
- When contexts change, deterrence strategies can easily move upward or downward along the ladder of deterrence theory, which creates strategic depth.
- Prompt and direct responses are possible without conflict, be it virtual or physical. Warnings can carry several messages to include: close surveillance is on; further intrusion may escalate the situation; self-defense is initiated, and corresponding retaliatory responses will be generated.
- It applies Schelling's magic of threat—i.e. uncertainty in a new environment—thus adding new meaning to this old trick.
- With sudden, dynamic, stealthy, and random changes, deterrence by engagement and surprise is able to catch an adversary by surprise, thus defeating an adversary virtually, psychologically, morally, and physically.
- This new approach can also be applied to the physical world.
- Furthermore, the approach supports accurate attribution and precise targeting, which can support evidence collection for digital forensic investigation.

Conclusion

The cyber domain needs a new and holistic deterrence strategy that involves prompt and direct cyber responses that are sudden, dynamic, stealthy, and random so that adversaries can be defeated mentally and virtually. Deterrence by engagement and surprise is such a deterrence strategy. It takes advantage

of the unique characteristics of cyber conflicts and creates a strategic buffer zone that makes it possible to dynamically select countermeasures based on specific contexts in addition to its support for intelligence collection, information operations, and surprise operations. Empowered by AI and machine learning, this deterrence approach is capable of exercising deterrence with virtual, psychological, moral, and physical aspects in an integrated way, thus leveraging cyber power (i.e. information power) together with diplomatic, military, economic, political, and legal power when dealing with challenges in the cyber domain. PRISM

Notes

[1] Joint Publication 3-0, *Joint Operations*, (Washington DC: The Joint Staff, August 11, 2011).

[2] Keith Payne, *The Great American Gamble: Deterrence Theory and Practice from the Cold War to the Twenty-First Century*, (National Institute Press, 2008).

[3] Ibid.

[4] Current cyber deterrence strategists generally align with deterrence by punishment, deterrence by denial, or both. For more information on these please see: Tim Stevens, "A Cyberwar of Ideas? Deterrence and Norms in Cyberspace," *Contemporary Security Policy*, vol. 33, no. 1, (2012), 148–70; Patrick Morgan, *Deterrence Now*, (Cambridge, UK: Cambridge University Press, 2003); Tim Stevens, "A Cyberwar of Ideas? Deterrence and Norms in Cyberspace," *Contemporary Security Policy*, vol. 33, no. 1, (2012), 148–70. Eric Sterner, "Deterrence in Cyberspace: Yes, No, Maybe," *Returning to Fundamentals: Deterrence and U.S. National Security in the 21st Century*, (Washington DC: George C. Marshall Institute, 2011), 27; Frank Cilluffo, Sharon Cardash, and George Salmoiraghi, "A Blueprint for Cyber Deterrence: Building Stability Through Strength," *Military and Strategic Affairs*, vol. 4, no. 3, (2012), 3–23; Martin Libicki, *Cyberdeterrence and Cyberwar*, (RAND Corporation, Project Air Force, 2009), 97.

[5] U.S. Department of Defense, *The Department of Defense Cyber Strategy*, (Washington DC: U.S. Department of Defense, May 2015).

[6] The levels of severity are based on the levels of belligerence as outlined by Martin Libicki in *Cyberdeterrence and Cyberwar*, (RAND Corporation, Project Air Force, 2009), 97

[7] For more information on this please see: Jim Chen and Alan Dinerman, "On Cyber Dominance in Modern Warfare," *Proceedings of the 15th European Conference on Cyber Warfare and Security*, (Reading, UK: Academic Conferences & Publishing International (ACPI) Limited, 2016), 52–7.

[8] See Greg Allen and Taniel Chan, "Artificial Intelligence and National Security," (Cambridge, MA: Belfer Center for Science and International Affairs, Harvard Kennedy School, July 2017), 2, available at <https://www.belfercenter.org/publication/artificial-intelligence-and-national-security>.

[9] Joint Publication 3–13, *Information Operations*, (Washington DC: The Joint Staff, November 20, 2014).

[10] This is reminiscent to what Scott Beidleman explores in his work on "Defining and Deterring Cyber War," (Carlisle Barracks, PA: U.S. Army War College, 2009).

[11] See: Martin Libicki, *Cyberdeterrence and Cyberwar*, (RAND Corporation, Project Air Force, 2009), 8.