Cyber Deterrence: A Wargaming Approach

Dr. Abderrahmane Sokri
DRDC – Ottawa Research Centre
CANADA

abderrahmane.sokri@drdc-rddc.gc.ca

ABSTRACT

Cyber risk is one of the most serious challenges the NATO nations are currently facing. Its impact can immediately or gradually harm a nation's safety and reputation. Deterrence strategies may be used to protect defenders from cyber threats. Deterrence dissuades would-be attackers from committing unwanted behavior by manipulating their cost-benefit analysis. Deterrence in the cyber domain is very complex and faces many enduring problems. The most challenging of them is the attribution dilemma. An analytical wargame is designed in this paper to show that cyber deterrence may be possible if conducted properly. A metric derived from data gathered in-game is employed to characterize the effectiveness of different cyber deterrence strategies. This paper builds a bridge between game theory and wargaming and shows that wargaming reasoning is well-suited to cyber defense problems.

1. INTRODUCTION

The fundamental idea of deterrence theory is to prevent adversaries from taking unwanted actions by influencing their cost-benefit analysis. Deterrence theory falls under the economic theory of utility. It asserts that challengers evaluate their expected benefit and the cost that the defender may impose before taking any action. If the cost outweighs the benefit, they are likely to be deterred (Morgan, 2003; Taipale, 2010; Wilner, 2017; Brantly, 2018). Morgan (2003) distinguished deterrence theory from deterrence strategies. The author states that the strategies vary in how they operationalize the six key elements of the theory (i.e., severe conflict, rationality, retaliatory threat, unacceptable damage, credibility, deterrence stability).

Two types of deterrence are generally used in the physical world: deterrence by punishment and deterrence by denial. Deterrence by punishment uses tit for tat or equivalent retaliation strategy to increase the aggressor's perceived cost. Deterrence by denial uses impenetrability strategies to reduce the aggressor's perceived benefits. But few papers have empirically evaluated the complexity of applying deterrence theory to cyber conflict (Wilner, 2019). To the best of my knowledge, a wargaming approach has never been used for cyber deterrence. Because of many unique characteristics of cyberspace, communicating coercive threats (deterrence by punishment) may not be credible (Chen and Dinerman, 2016; Moisan and Gonzalez, 2017; Sokri, 2020b). These characteristics can be divided into three main categories: (1) The nature of cyber weapons, (2) the multitude of actors in cyberspace, and (3) the attribution dilemma.

1.1 The nature of cyber weapons

Cyber weapons are effective, cheap and can be launched from anywhere at any time. They can be used not only against virtual and physical targets but also in psychological warfare (Rustici, 2011; Chen, 2017). Cyber weapons present two main paradoxes: (1) They are subject to time-decay: When the exploited vulnerability is detected by the defender, the weapon becomes useless. (2) Their usage may enhance the target's defence: When the vulnerability is identified by the defender, the attacked target becomes upgraded (Sokri, 2020a; Podins and Czosseck, 2012).

1.2 The multitude of actors in cyberspace

In digital space, defenders may face various state and non-state actors from multiple locations. Their numbers are mutch higher than those of conventional conflicts due to (1) the low cost of entry to cyberspace and (2) the absence of any agreed upon concept of cyberspace sovereignty (Fischerkeller and Harknett, 2017). These actors support their interests and exercise their influence through highly sophisticated and dynamic threats. They can be rational or irrational (individuals or machines) and their objectives are inherently conflicting (Fischerkeller and Harknett, 2017; Sokri, 2020a).

1.3 The attribution dilemma

The credibility of any deterrence depends on the information captured by attribution. Attribution is the determination of the identity or the location of an attacker (Wheeler, 2003; Sokri, 2020b; Robinson et al., 2015). This identity can be digital (e.g., an Internet Protocol (IP) address) or physical (e.g., a geographical address) (Guan and Zhang, 2010). The blame attribution can be used to prevent future attacks and improve defensive techniques (Nicholson et al., 2012). In digital space, attacks go beyond all geographic and political boundaries and determining who to blame for them may be time-consuming and very challenging (Wilner, 2017; Sokri, 2018).

To the best of my knowledge, this study is the first to use a wargame to demonstrate that cyber deterrence may be useful, if conducted properly. The paper builds a bridge between game theory and wargaming. It is organized into four sections. Following the introduction, section 2 provides a brief review of literature on the applicability of deterrence theory in cyberspace. Section 3 presents a wargaming approach to evaluate the impact of cyber deterrence. Future research directions are recommended in section 4.

2. LITERATURE REVIEW

Cyber risk is one of the main complex and challenging issues the NATO nations are currently facing. Deterrence strategies may be used to protect defenders from cyber threats. Since the seminal book by Morgan on deterrence theory, there has been a growing body of literature debating how deterrence can be applied in cyberspace. This literature can be divided into three main categories of publications.

2.1 The first category

The first category asserts that cyber deterrence is inherently weak. This literature generally cites attribution as a reason against implementing the strategy (Bordelon, 2017). Iasiello (2014), for example, indicated that many challenging problems can inhibit quick and accurate attribution processes. These problems include the time and effort it takes to collect and analyze the used attack method and the high probability of misattribution. Stevens (2012) compared cyber deterrence with nuclear deterrence using the six conditions presented in Morgan (2003) and found that cyber deterrence fails to satisfy any of them. Bordelon (2017) maintained that retaliation can be difficult to accomplish and the likelihood of escalation is high, if a state acts against a lone individual in another sovereign country. Clark and Landau (2011) argued that the Internet was not designed with the goal of deterrence in mind. The authors concluded that several changes in thinking should be conducted to tackle the cyber deterrence problem. These changes may include a new Internet designed differently.

2.2 The second category

The second category of publications argues that deterrence by denial can resolve the problem of attribution. Unlike Deterrence by punishment which relies on knowing the attacker, deterrence by denial shows impenetrability to threaten failure (Wilner, 2017; Bordelon, 2017). Policy makers are increasingly gravitating towards deterrence by denial (Taipale, 2010, Sokri 2020b). The defender can reduce the probability of a

successful attack by investing in information security. Cavusoglu et al. (2008) concluded that by revealing its security investment strategy in a sequential game, deterrence is more effective than in simultaneous games. Sokri (2020b) used a sequential game theoretic approach, with a disclosure mechanism, to show how a deterrence strategy can be formulated in cyberspace.

2.3 The third category

The third category asserts that cyber conflicts have their own characteristics that are not necessarily similar to those of conventional conflicts. Therefore, new innovative cyber deterrence frameworks (or theories) have to be developed for this unique domain (Chen, 2017). Rosenzweig (2010), for example, showed that there is a dichotomy between current cyber deterrence strategies. They analyse either deterrence by denial or deterrence by punishment. Chen (2017) examined unique ways of implementing deterrence in cyber warfare. The author improved Rosenzweig's classification and suggested a framework that can be uniquely applied to enhance cyber defense.

3. A WARGAMING APPROACH

This section sets up a wargaming method for testing the effectiveness of deterrence by denial in cyber space. It presents its underlying theoretical foundation, a possible scenario, and a possible execution. To the best of our knowledge, a wargaming approach has never been used for cyber deterrence. This is a first suggestion.

3.1 A possible scenario

The scenario is based on a game theoretic model developed by the author on the same topic. In this application, we consider a security game between an attacker *a* (the Red Team) and a defender *d* (the Blue Team) in a cyber infrastructure system.

Let $T = \{t_1, t_2, \dots, t_n\}$ be a set of *n* targets at risk of being attacked (e.g., vulnerabilities in Internet-connected systems) and $c(t_i),\ i = 1,\ 2,\ \dots n$, be the defender's cost if the target t_i is successfully attacked.

Let $S = \{s_1, s_2, \dots, s_m\}$ be a set of resources to cover the targets (e.g., firewalls, inspection procedures) and $c(s_i),\ i = 1,\ 2,\ \dots m$, be the defender's cost associated with s_i.

Let $S = \{a_1, a_2, \dots, a_l\}$ be a set of l types of attacks to attack the targets (e.g., Sokri, 2020b), $d(a_i),\ i = 1,\ 2,\ \dots l$, the attacker's time to prepare the attack a_i, and $p(a_{ij}),\ i = 1,\ 2,\ \dots l$, the probabilities of a successful attack on the target t_j using a_i.

3.2 A possible execution

The objective of the Red Team is to conduct the maximum number of successful attacks in the minimum time possible. The objective of the Blue Team is to cover the maximum number of targets with the minimum total cost. At each turn, the defender publicly releases her/his level of investment. The attacker reacts with a certain level of willingness-to-attack for each target. Simultaneous games (i.e., Myopic approaches) and sequential games (i.e., Non-myopic approaches) will be played and their results will be compared (Cavusoglu et al., 2008; Sokri, 2020b). The expected effort to compromise each target can be expressed in terms of time. At the end, a correlation coefficient will capture the potential correlation between the defender investment level and the expected effort to be exerted by the attacker.

Bivariate correlation analysis is one of the most useful methods for determining the strength and direction of the probable relationship between two variables (Sokri and Solomon, 2014). The closer the value is to -1 or +1, the stronger is the relationship.

- A negative value would denote negative linear correlation. When the defender investment is high, the attacker's effort should be relatively low, and deterrence by denial would be effective.
- A positive value would denote positive linear correlation. The defender investment would have the opposite of the desired effect by increasing the attacker's willingness-to-attack.
- A value close to 0 would denote a very weak linear relationship. The defender investment would have a small impact on the attacker's willingness-to-attack and deterrence by denial would be useless.

4. CONCLUSION

In this paper, we designed the first analytical wargame to test the effectiveness of deterrence by denial in cyberspace. A correlation coefficient derived from data gathered in-game is used to measure the strength of the relationship between the defender investment and the attacker effort. Further efforts will be undertaken to explore possible extensions of this method. These extensions include (but are not limited to):

- the application of this theoretic wargame to a real-world cyber-security problem using real-life parameters;
- assessing the risk to the defender of a disclosure strategy;
- including deception mechanisms to enhance security;
- developing models to deal with bounded rationality of human adversaries;
- developing models where the defender faces multiple attackers.

REFERENCES

[1] Morgan, P. (2003). Deterrence Now, Cambridge University Press, Cambridge, UK.

[2] Taipale, K.A. (2010). Cyber-deterrence. Law, Policy and Technology: Cyberterrorism, Information, Warfare, Digital and Internet Immobilization, IGI Global.

[3] Wilner, A. (2017). Cyber deterrence and critical-infrastructure protection: Expectation, application, and limitation. Comparative Strategy, Vol. 36, Issue 4, pp. 309-318.

[4] Brantly, A.F. (2018). The cyber deterrence problem. 10th International Conference on Cyber Conflict (CyCon), IEEE, pp. 31-54.

[5] Wilner, A. S. (2019). US cyber deterrence: Practice guiding theory. Journal of Strategic Studies, 1-36.

[6] Chen, J., and Dinerman, A. (2016, July). On cyber dominance in modern warfare. In European Conference on Cyber Warfare and Security (p. 52). Academic Conferences International Limited.

[7] Moisan, F. and Gonzalpez, C. (2017). Security under Uncertainty: Adaptive Attackers Are More Challenging to Human Defenders than Random Attackers. Frontiers in Psychology, Vol. 8:982.

[8] Sokri, A. (2020a). Game Theory and Cyber Defense. In Games in Management Science (pp. 335-352). Springer, Cham.

[9] Rustici, R. (2011). Cyberweapons: Leveling the International Playing Field, Parameters, Volume 41, No. 3, Pages 32-42.

[10] Chen, J. (2017). Deterrence and its implementation in cyber warfare. In International Conference on Cyber Warfare and Security (p. 83). Academic Conferences International Limited.

[11] Podins K and Czosseck C. (2012). A Vulnerability-Based Model of Cyber Weapons and its Implications for Cyber Conflict. International Journal of Cyber Warfare and Terrorism, 2 (1), p. 14–26

[12] Fischerkeller, M. P., and Harknett, R. J. (2017). Deterrence is not a credible strategy for cyberspace. Orbis, 61(3), 381-393.

[13] Wheeler DA and Larsen GN. Techniques for Cyber Attack Attribution. Institute for Defense Analysis, IDA Paper P-3792, 2003

[14] Robinson M, Jones K, and Janicke H. (2015) Cyber Warfare: Issues and Challenges, Computer and Security, Vol. 49, p. 70-94.

[15] Guan Y and Zhang L. (2010). Network Forensics. In: Managing Information Security, Vacca, J.R. (eds) Syngress, p. 197-212.

[16] Nicholson A, Watson T, Norris p, Duffy A, and Isbell R. (2012). A Taxonomy of Technical Attribution Techniques for Cyber Attacks. Proceedings of the 11th European Conference on Information Warfare and Security. Filiol, E. and Erra, R. (eds). Laval, France, p. 188-197.

[17] Sokri, A. (2018). Optimal Resource Allocation in Cyber-Security: A Game Theoretic Approach. Procedia computer science, 134, 283-288.

[18] Bordelon, E. (2017). Approaching Cyber Warfare: Geopolitics, Deterrence, and International Law.

[19] Iasiello, E. (2014). Is cyber deterrence an illusory course of action?. Journal of Strategic Security, 7(1), 54-67.

[20] Stevens, T. (2012) “A Cyberwar of Ideas? Deterrence and Norms in Cyberspace”, Contemporary Security Policy, Volume 33, No. 1, Pages 148-170.

[21] Clark, D. and Landau, S. (2011). Untangling Attribution, Harvard Law School, National Security Journal, 1–30.

[22] Cavusoglu, H., Raghunathan, S., and Yue, W.T. (2008). Decision-Theoretic and Game-Theoretic Approaches to IT Security Investment. Journal of Management Information Systems, Vol. 25:2, pp. 281-304.

[23] Sokri, A. (2020b). Deterrence in Cyberspace: A Game Theoretic Approach

[24] Rosenzweig, P. (2010). The Organization of the United States Government and Private Sector for Achieving Cyber Deterrence, Proceedings of a Workshop on Deterring Cyber Attacks: Informing Strategies and Developing Options for U.S. Policy, Pages 245-269. National Academies Press.

[25] Sokri, A. and Solomon, B. (2014) “Cost Risk Analysis and Contingency for the NGFC”, DRDC CORA TM 2013-224.