Optimal Investment in Cyber Attack and Resilience: A Dynamic Differential Game

Alexander Alexeev
School of Public and Environmental Affairs
Bloomington, Indiana

Eric Jardine
Assistant Professor Political Science
Virginia Polytechnic Institute and State University, in Blacksburg, Virginia

Kerry Krutilla
Associate Professor
School of Public and Environmental Affairs
Bloomington, Indiana

Abstract

In this article, we develop a differential game to assess optimal investment in cyber measures. The model is based on an augmented contest success function in which efforts to influence an endogenous probability of attack reflect a combination of resource commitments this period and the state of knowledge. The state of knowledge decays with exogenous technical advance, but increases as a function of resource commitments this period. The model is solved, and steady-state solutions for optimal cyber investment as a function of changes in the models parameters assessed.

1. Introduction

It is sometimes argued that a full-blown cyberwar between state actors will not take place (Rid 2013). However, the short history of digital engagements shows that lower-intensity governmental conflicts are becoming a regular occurrence (Healey 2012; Clarke and Knake 2010; Kaplan 2016; Valeriano and Maness 2015; Stiennon 2015). Well known examples include the Russian distributed denial of service (DDoS) attacks on Estonia in 2007 and Georgia in 2009, the deployment of the US-Israeli Stuxnet virus that destroyed Iranian nuclear centrifuges at Natanz in 2010 (Zetter 2014), and the successful attacks on the Ukrainian power grid in 2015, 2016, and now 2017.

Moreover, while firms, particularly in the financial sector, are often the target of data breaches, government is the fourth most popular target, with over 20% of the data breaches recorded between 2005 and 2017, as well as the third highest number of compromised records, at 14%. Over this period, governments were breached some 743 times and had some 183,668,599 records compromised, according to data from Privacy Rights Clearinghouse. As was the case with the hack of the US Office of Personnel Management (OPM), many of these breached files include highly sensitive information such as social insurance numbers.

Concerns about the effects of cyberattacks have stimulated governments to invest in their cyber arsenals. These investments have taken many forms, ranging from the highly technical to regulatory development, human capital generation, and diplomatic initiatives. As the recent Shadow Broker leaks make clear, the US government (alongside many other governments) is deeply involved in the purchase and retention of so-called zero-day vulnerabilities for which there is no technological defense (Cox 2016). The Federal Bureau of Investigation's (FBI) purchase of a zero-day exploit to gain access to the iPhone of one of the San Bernardino terrorists is a classic example, although other US government agencies such as the DEA and, especially the NSA, are also deeply involved in buying up these highly valued software defect (Cox 2017; Hampson and Jardine 2016).

The US government is also working to develop domestic regulatory frameworks to help protect critical national infrastructure (NIST 2017), while simultaneously investing heavily in infrastructure modernization through initiatives such as Information Technology Modernization Fund (The White House 2016). The Department of Homeland Security hosts annual Cybersecurity Awareness Months, with the aim of developing higher levels of human capital among the general population. Additionally, some governmental cybersecurity actions play out on the international stage, where diplomatic efforts in the United Nations Group of Governmental Experts (UNGGE) have led to a list of normative principles that would put a leash on governmental use of cyberweapons (United Nations 2015).

Notwithstanding growing attention to cybersecurity concerns and increasing public resource commitments to offensive and defensive measures, resources are limited and government confront the economic challenge of balancing the gains from cyber investments against their opportunity costs. This trade-off has a probabilistic dimension; resource commitments in cyber measures do not yield certain results, and outcomes are also affected by the reaction of rivals to a country's behavior. A probabilistic game theoretic formulation is the method to model these

interactions. This article extends the one-period game theoretic model in Alexeev and Krutilla (2015) to the more realistic setting of a repeated rivalry between governments. In the expanded model explored here, resource commitments to cyber actions this period increase the stock of knowledge, and the probability of a successful attack is a function both of the stock of accumulated knowledge from past actions and behavior in the present.

We start in the next section with a review of the literature on optimal investment in cybersecurity. The following section then describes the model developed in this research. The next section presents the solution for steady state resources devoted to cyber defense and attack as a function of the model's parameters. The final section of the article offers conclusions and recommendations for future research.

2. Background

The literature on optimal investment in cyber measures is relatively limited. The benchmark model is by Gorden and Loeb (2002). It is based on a one period model in which firms are risk neutral, and within this context, examines optimal investments a firm should make to defend against cyber attacks. The model assumes that the probability of a cyber attack is exogenous, but firms can unilaterally reduce their vulnerability through investment in security measures. Given this set-up, the model demonstrates that firms should not invest more than 1/e (≈ 37%) of the expected cost of a data breach (Gordon and Loeb 2002). This result turns out to be robust across a more general range of functional forms (Baryshnikov 2012).

However, the 37% result does not hold up in a one-period game formulation of Alexeev and Krutilla (2015). In this model, the probability of a successful attack endogenously depends on the resources rivalrous governments devote to attack and defense. The model allows for asymmetric valuations of gains and losses by the rivals, and relative differences in the efficacy of their resource committments. In this setting, optimal investments can be significantly greater or less than 37% of expected damages.

Extending the Gorden and Loeb (2002) model to risk averse firms, a model by Huang, Hu and Behara 2008 shows that there is a minimal data breach cost below which the optimal level of investment in cybersecurity protections drops to zero. As the potential cost of a breach goes up optimal investment in cybersecurity increases, but the value never exceeds the total cost of the incident.

A rare empirical study shows that the cost of most data breaches tend to be roughly commensurate with a firm's IT security budget. For example, making an assumption about the fraction of budget devoted to IT security, 77% of data breach costs are within +/- 10 million dollars of the firm's IT security budget, while fully 50% of incidents fall within +/- 1 million dollars (Romanosky 2016, 13).

This small literature suggests several research gaps in study of optimal cyber investment. Most significantly, single-period models convey limited information. The value from investments in cyber measures in the current period does not necessarily fall exclusively in the period and some

cyber security measures, such as staff training in cybersecurity digital hygiene, might not pay out much immediately but could continue to pay dividends well into the future. With past investments paying future dividends, it is possible that the safety of a cyber system could increase even as the annual rate of investment decreases.

There is also the possibility of additional learning over time. Some sort of information accretion process, such as Bayes Rule, could be used to form a better judgement about risk levels with experience (Cavusoglu, Mishra and Raghunathan 2004). Organizations can also come to learn more about their adversaries over time with repeated conflictual interactions. The so-called "attribution problem" is an example. (Tsagourias 2012; Rid and Buchanan 2015). Through a series of technical steps, careful attackers can obfuscate their identity, motive and location, making deterrence of attacks via the threat of credible punishment more difficult. Effective attribution leverages both the particular details of a specific attack, but also historical details from past attacks. Attack methods, idiosyncrasies in the code and target types across multiple attacks can be combined with forensic details on the current assault to produce a more complete picture of who launched a particular attack.

A prime example is the hack of the Democratic National Committee (DNC) by Russian operatives in the lead up to the 2016 presidential election. DHS and the FBI released a joint analysis report in December of 2016 presenting the technical details involved in the hack of the DNC (NCCIC/FBI 2016). One of the telling features of the report is the focus on how advanced persistent threats 28 and 29, as they were technically known, had entered the system on multiple occasions. As the report put it, "Both groups have historically targeted government organizations, think tanks, universities, and corporations around the world." (Ibid., 2). This pattern of historical interaction breeds a familiarity with an adversary's toolkit and social engineering approaches. This familiarity, in turn, makes defense easier, potentially necessitating less investment, but it cannot be captured in a single period model.

The model developed in this article focuses on the first of the issues mentioned: the fact that effects of investments in cyber measures this period can have dynamic multi-period effects. To our knowledge, this is the first application of a differential game model exploring optimal investment with cyber knowledge accumulating over time.

2. Model

The model extends the contest success function approach of Alexeev and Krutilla (2015) to the multi-period setting. There are two competing governments, an "attacker" and a "defender", with the objectives:

$$\max_{R_A} \int_0^\infty \left(U_A(R_A, K_A, R_D, K_D)\right) e^{-\eta t} dt \qquad (1)$$

$$\max_{R_D} \int_0^\infty \left(U_D(R_A, K_A, R_D, K_D)\right) e^{-\eta t} dt \qquad (2)$$

and constraints:

$$\dot{K}_A = -\delta K_A + \nu_A R_A \tag{3}$$

$$\dot{K}_D = -\delta K_D + \nu_D R_D \tag{4}$$

The variables and parameters of this model are defined in Table 1. The expression for the attacker and defender's utilities in (1) and (2) are represented as expected net payoffs, as follows:

$$U_A = G_A P(\bullet) - R_A \tag{5}$$

$$U_D = G_D P(\bullet) + R_D \tag{6}$$

where G_A and G_D are the utility gains to attacker and losses to defender respectively from a successful attack, $P(\bullet)$ is the probability of a successful attack, and R_A and R_D are the flow resources committed this period to attack and defense respectively.

Note that K_A and K_B represent the state of knowledge of the attacker/defender this period, and that equations (3) and (4) show the rate of change of the state of knowledge, $\dot{K}_A$ and $\dot{K}_D$, as a function of two influences. The first, the parameter $\delta \in [0,1]$, shows the decay rate of the state of knowledge. This parameter is assumed to be driven by global, exogenous technical advance in cyber security and offensive capabilities. This technical advance reduces the effectiveness of the existing state of knowledge at rate δ over time, all else constant. On the other hand, this period's effort in attack (R_A) and defense (R_D) have enduring effects on the state of knowledge. The fractional parameters, $\nu_A \in [0,1]$ and $\nu_D \in [0,1]$, show what part of this period's efforts have effects on the state of knowledge lasting beyond the period. To summarize $\dot{K}_A$ and $\dot{K}_D$ are a function of the rate of depreciation and new investment, with the latter being some part of this period's efforts in attack and defense.

Following Alexeev and Krutilla (2015), the probability, $P(\bullet)$, of a successful attack is represented as:

$$P = \left[\frac{E_A + \omega}{E_A + \sigma E_D + 2\omega} \right] \tag{7}$$

where (E_A) is "Effective Effort in Attack" and (E_D) is "Effective Effort in Defense", σ is the relative technical efficiency of "Effective Effort in Attack" compared to "Effective Effort in Defense," ω is a noise parameter allowing for a degree of bounded rationality. To fix ideas, let $\sigma = 1$ and $\omega = 0$, and notice that, under these conditions, if $E_A = E_D$, $P = .5$. Again under the same assumptions, if $E_A = 2E_D$, then $P = .66$,whereas, if $.5E_A = E_D$, $P = .33$. In short, when "Effective Effort in Attack" is greater than "Effective Effort in Defense", the probability of an effective attack is greater than 50%, and vice versa -- under the default parameter settings.

The σ parameter represents differences in the efficiency of effective effort. If $\sigma > 1$, the defender's effort reduces the probability of a successful attack more than the attacker's effort increases it, and vice versa. In fact, there is some literature suggesting attackers have an inherent advantage, implying that $\sigma < 1$ might be relatively typical.

Finally, notice that as ω goes from zero to infinity, P will go to 50% whatever the rivals do. This parameter reflects noise in the sense that rivals do not respond with perfect sensitivity to each other's actions when ω assumes a non-zero value.

We now depart from Alexeev and Krutilla (2015) in defining "Effective Effort" as composites of two variables:

$$E_A = R_A^{\alpha_A} K_A^{1-\alpha_A}, \qquad E_D = R_D^{\alpha_D} K_D^{1-\alpha_D} \tag{8}$$

The parameters α_A and α_D are the share parameters for "Effective Effort" arising from resource commitments this period on the part of the attacker and defender respectively, while $1-\alpha_A$ and $1-\alpha_D$ are the share parameters for "Effective Effort" arising from the attacker's and defender's cumulated stock of knowledge. Entering (8) into (7) gives the complete expression for the probability of an effective attack:

$$P = \left[\frac{R_A^{\alpha_A} K_A^{1-\alpha_A} + \omega}{R_A^{\alpha_A} K_A^{1-\alpha_A} + \sigma R_D^{\alpha_D} K_D^{1-\alpha_D} + 2\omega} \right] \tag{9}$$

Using all of the information discussed, the current value Hamiltonians for the differential game are:

$$H_A = \psi \left[\frac{R_A^{\alpha_A} K_A^{1-\alpha_A} + \omega}{R_A^{\alpha_A} K_A^{1-\alpha_A} + \sigma R_D^{\alpha_D} K_D^{1-\alpha_D} + 2\omega} \right] - R_A + \lambda_1 \left(-\delta K_A + \nu_A R_A \right) + \lambda_2 \left(-\delta K_D + \nu_D R_D \right) \tag{10}$$

$$H_D = -\left[\frac{R_A^{\alpha_A} K_A^{1-\alpha_A} + \omega}{R_A^{\alpha_A} K_A^{1-\alpha_A} + \sigma R_D^{\alpha_D} K_D^{1-\alpha_D} + 2\omega} \right] - R_D + \lambda_3 \left(-\delta K_D + \nu_D R_D \right) + \lambda_4 \left(-\delta K_A + \nu_A R_A \right) \tag{11}$$

In (11), note that G_D is normalized to 1, and $\psi \equiv \frac{G_A}{G_D}$ in (10). $G_D = 1$ might be thought of as the economic loss to the defender from a successful attack, while ψ is the relative value of a successful attack for the attacker compared to the economic costs that the attack imposes.

4. Results

We focus on the way the steady-state solutions respond to the parametric variations. The dynamic transition paths are of less interest, reflecting arbitrary variations in initial conditions. The model does not have analytical solution, so numerical simulation is used.

The parametric variations considered in this preliminary analysis are shown in Table 2. The corresponding results for steady-state resource commitments in attack and defense are shown in Table 3.

The first simulation varies the relative effectiveness of resource commitments in attack and defense, with permutation SA1 showing the case that the defender's resource commitments are relatively more effective, and SA2 showing greater efficiency of the attacker's efforts. Interestingly, these asymmetries result in a symmetric decline for both the attacker and defender in resource commitments from the base case, suggesting that disparities in relatively efficiency in resource use can reduce resource equilibrium commitments.[1] The logic of this reality can be seen in the limiting case where one party or the other's resource commitments are totally effective, and the other party's are completely ineffective. For example, if the defenders efforts were 100% effective, there would be no point for the attacker to waste resources in attacking. From the cyber security perspective, the policy implication would be increasing the effectiveness of defensive efforts would both reduce the probability of successful attacks and reduce the resources needed to deter them.

Comparing the resource commitments as a ratio of expected damages -- the common metric used in the literature – gives a higher fraction of expenditure when deterrence is relatively effective (.58 for SA1) than when it isn't (.29 for SA2). Although the ratio of cost to expected damages is an intuitively logical metric, some policy relevant insight (as discussed above) is lost if the absolute comparison is not also made.

Turning to variation in the share parameters for "Effective Effort" arising from resource commitments this period, SA3 shows the case where this period's effort has a relatively low impact on Effective Effort (.2), while SA4 shows the case where the impact is relatively high (.8). Steady-State resource commitments decline in the first instance relative to the base case, and increase in the second case. However, the probability of attack does not change.

The next simulation pair (SA5) and (SA6) compares the impact of different valuations on the part of the attacker per unit of economic damages a successful attack causes to the defender. In SA5, the attacker's valuation is twice the economic damages caused; in SA6 it is half. For SA5, the asymmetry increases the resources the attacker devotes to attack from .22 to .39, and decreases the defenders recourse commitment from .22 to .19. For SA6, the resources the attacker devotes drop from .22 to .10; the defenders resource commitments drop to .19 as before. It is interesting that the relative valuation asymmetry in either direction reduces the resources a defender rationally commits to defense.

[1] A similar result was observed in Krutilla and Alexeev (2012) in another game theory model using a contest success function.

The final comparison assesses the effect of changing the fraction of resource commitments this period that affects the state of state of knowledge beyond one period. In SA7, the parameter is reduced from the base case of .5 to .25; in SA8, the parameter is raised to .75. The surprise from this comparison is that it has no effect on the base case level of resource commitments, or change in probability of attack. It does significantly affect the steady state accumulation of knowledge (not shown in Table 3). We are now exploring the reasons for this result.

These results are 1quite preliminary and we will be continuing to assess the effects of changing the other parameters, as well other asymmetries between attackers and defenders. An additional next step is to assess whether the literature suggests parameter settings that are consist with empirical studies.

5. Conclusions

This research develops a differential game formulation to study temporal optimal investments in cyber measures. The literature on optimal cyber investment is relatively small, and our model represents a significant extension. Among other attributes, it distinguishes between the effects of resource commitments in influencing attacks in current versus future periods, allows for asymmetries in the effectiveness of cyber measures between attackers and defenders, and shows the impact of asymmetric valuations by attackers and defenders of the damages cyber attacks cause. The model allows for sensitivity analysis of a number of policy-relevant parameters.

The study is quite preliminary with much simulation work in progress. Additionally, we are reviewing the literature to calibrate parameters to reflect empirically reasonable cases.

Table 1: Variable and Parameter Definitions

Variables	**Definitions**
U_A / U_D	utility from attack/defense
R_A / R_D	flow resource commitment this period to attack/defense
K_A / K_D	state of knowledge in attack/defense at time t
$\dot{K}_A / \dot{K}_D$	instantaneous change in state of knowledge at time t
Parameters	
η	discount rate
δ	state of knowledge depreciation
ν_A / ν_D	fraction of flow resources to attack/defense this period that increases the state of knowledge beyond one period

Table 2: Parameter Values in Sensitivity Analyses

Value of Parameters in One Way Sensitivity Analyses	Parameters in Contest Success Function			Relative damage valuation of attacker to defender	Equation of Motion Parameters		Discount Rate
Sensitivity Analyses	relative efficiency (σ)	production share parameter ($\alpha_A = \alpha_D$)	bounded rationality (ω)	$\psi = \frac{G_A}{G_B}$	state of knowledge depreciation (δ)	fraction of flow resources to attack/defense this period that increases the state of knowledge beyond one period (ν_A, ν_D)	η
Base Case	1	.6	0	1	.1	.5	.05
SA 1	2	.6	0	1	.1	.5	.05
SA 2	.5	.6	0	1	.1	.5	.05
SA 3	1	.2	0	1	.1	.5	.05
SA 4	1	.8	0	1	.1	.5	.05
SA 5	1	.6	0	2	.1	.5	.05
SA 6	1	.6	0	.5	.1	.5	.05
SA 7	1	.6	0	1	.1	.25	.05
SA 8	1	.6	0	1	.1	.75	.05

Table 3: Steady State Resource Commitments

	Attacker		Defender		Prob
	Resources as a Fraction of Damages	**Resources as a Fraction of Expected Damages**	**Resources as a Fraction of Damages**	**Resources as a Fraction of Expected Damages**	
Base Case	0.22	0.43	0.22	0.43	0.50
SA 1	0.19	0.58	0.19	0.58	0.33
SA 2	0.19	0.29	0.19	0.29	0.67
SA 3	0.18	0.37	0.18	0.37	0.50
SA 4	0.23	0.47	0.23	0.47	0.50
SA 5	0.39	0.58	0.19	0.29	0.67
SA 6	0.10	0.29	0.19	0.58	0.33
SA 7	0.22	0.43	0.22	0.43	0.50
SA 8	0.22	0.43	0.22	0.43	0.50

References

Aleexev, Alexander and Kerry Krutilla, 2015. "Cyber-Attack as a Contest Game." In A.Kott (ed.) Proceedings of the NATO IST-128 Workshop: Assessing Mission Impact of Cyberattacks. US Army Research Laboratory, ARL-SR-0349, pp. 66-74, 2015. www.arl.army.mil/www/default.cfm?technical_report=7602.

Anderson, Ross and Tyler Moore, 2006, "The Economics of Information Security," *Science*, 314, 610-613. Accessed at: http://tylermoore.ens.utulsa.edu/science-econ.pdf

Baryshnikov, Yuliy, 2012, "IT Security Investment and Gordon-Loeb's 1/3 Rule," Workshop on the Economics of Information Security (WEIS 2012). Accessed at: http://www.econinfosec.org/archive/weis2012/papers/Baryshnikov_WEIS2012.pdf

Cavusoglu, Huseyin, Birendra Mishra and Srinivasan Raghunathan, 2004, "A Model for Evaluating IT Security Investment," *Communications of the ACM*, vol. 47, no. 7, 87-92. Accessed at: http://utd.edu/~huseyin/paper/investment.pdf

Clarke and Knake, 2010, *Cyberwar: The Next Threat to National Security and What to Do about It*. New York: Harper Collins.

Cox, Joseph, 2016, "A Brief Interview with The Shadow Brokers, The Hackers Selling NSA Exploits," *Motherboard*. Accessed at: https://motherboard.vice.com/en_us/article/a-brief-interview-with-the-shadow-brokers-the-hackers-selling-nsa-exploits

Cox, Joseph, 2017, "Here's a DEA Invoice for Zero-Day Exploits." *Motherboard*. Accessed at: https://motherboard.vice.com/en_us/article/heres-a-dea-invoice-for-zero-day-exploits

Gordon, Lawrence A. and Martin P. Loeb, 2002, "The Economics of Information Security Investment," *ACM Transactions on Information and Systems Security*, vol. 5, no. 4, 438-457.

Hampson, Fen and Eric Jardine, 2016, *Look Who's Watching: Surveillance, Treachery and Trust Online*. Waterloo: Centre for International Governance Innovation Press.

Healy, Jason, ed., 2012, *A Fierce Domain: Conflict in Cyberspace, 1996-2012. New York: Atlantic Council*.

Huang, C. Derrick, Qing Hu and Ravi S. Behara, 2008, "An Economic Analysis of the Optimal Information Security Investment in the Case of a Risk-Averse Firm," *Int, J. Production Economics*, vol. 114, 793-804.

Kaplan, Fred, 2016, *Dark Territory: The Secret History of Cyber War*. New York: Simon & Schuster.

Krutilla, Kerry and Alexander Alexeev, 2012, "The Normative Implications of Political Decision-Making for Benefit-Cost Analysis." *Journal of Benefit-Cost Analysis*, 3(2): Article 2, 2012

NCCIC/FBI, 2016, "GRIZZLY STEPPE – Russian Malicious Cyber Activity." Accessed at: https://www.us-cert.gov/sites/default/files/publications/JAR_16-20296A_GRIZZLY%20STEPPE-2016-1229.pdf

NIST, 2017, "Framework for Improving Critical Infrastructure Cybersecurity." Accessed at: https://www.nist.gov/sites/default/files/documents////draft-cybersecurity-framework-v1.11.pdf

Rid, Thomas, 2013, *Cyberwar Will Not Take Place*. New York: Oxford University Press.

Rid, Thomas and Ben Buchanan, 2015, "Attributing Cyber Attacks," *Journal of Strategic Studies*, vol. 38, no. 1-2, 4-37. http://dx.doi.org/10.1080/01402390.2014.977382

Roberts, Daniel, 2016, "Tom Ridge: Cyber attacks are now worse than physical attacks." *Yahoo! Finance*. Accessed online at: http://finance.yahoo.com/news/tom-ridge-cybersecurity-attacks-are-now-worse-than-physical-attacks-170426390.html?soc_src=social-sh&soc_trk=tw

Romanosky, Sasha, 2016, "Examining the Cost and Causes of Cyber Incidents." *Journal of Cybersecurity*, Vol. 2, no. 2, 1-11. DOI: https://doi.org/10.1093/cybsec/tyw001

Stiennon, Richard, 2015, *There Will be Cyberwar: How the Move to Network-Centric War Fighting Has Set the Stage for Cyberwar*. New York: IT-Harvest Press.

Tsagourias, Nicholas, 2012, "Cyber attacks, self-defence and the problem of attribution." Journal of Conflict and Security Law, vol. 17, no. 2, 229-244. https://doi.org/10.1093/jcsl/krs019

United Nations, 2015, "70/237. Developments in the field of information and telecommunications in the context of international security," Resolution adopted by the General Assembly on 23 December 2015. Available at: https://unoda-web.s3-accelerate.amazonaws.com/wp-content/uploads/2016/01/A-RES-70-237-Information-Security.pdf

Valeriano, Brandon and Ryan C. Maness, 2015, *Cyberwar Versus Cyber Realities: Cyber Conflict in the International System*. New York: Oxford University Press.

White House, 2016, "FACT SHEET: Cybersecurity National Action Plan." Accessed at: https://obamawhitehouse.archives.gov/the-press-office/2016/02/09/fact-sheet-cybersecurity-national-action-plan

Zetter, Kim, 2014, *Count Down to Zero Day: Stuxnet and the Launch of the World's First Digital Weapon*. New York: Broadway Books.