Optimal Decision Making in Interdependent Network Security

Alan Nochenson
Information Systems
Carnegie Mellon University
Pittsburgh, PA 15213
anochenson@gmail.com

April 26, 2012

Contents

Acknowlegements

This work could not have been done without my supervisor Larry Heimann. He has been there from the beginning with ideas, suggestions, and guidance on this thesis and otherwise. I would also like to thank my friends and family for listening to my ideas, and I'd especially like to thank Naomi for editing parts of this work. Additionally I'd like to thank the Information Systems Department at CMU and the faculty thereof for their continued support throughout my time here. I'd also like to thank the Dietrich College of Humanities & Social Sciences for the opportunity to write this thesis, and Carnegie Mellon University for fostering a great environment to produce it in.

Abstract

Although people are frequently urged to protect the machines they use and oversee, the fact remains that the decision to invest in protection software is far from universal. To better understand this decision, we formulate two models of interdependent network security. In the first, there is a system administrator responsible for a network of size n against attackers attempting to penetrate the network and infect the machines with viruses or other exploits. Through analysis of this interdependent network security scenario, we conclude that the decision to buy protection is dependent upon a number of factors including external and internal vulnerabilities, the types and likelihoods of different amounts of loss, the degree of autonomy of the attacker, and others.

The second model looks at network security from a game-theoretic point of view. Through the formulation and examination of increasingly complex scenarios, we formulate a model for utility-based security decisions for an individual in a network of individuals. We look at the decision for one person buy security software for herself and to buy security software in the context of two or more people. By modeling security as a public good, we examine externalities that players impose upon each other. We then examine Olson's theory of groups [13] in a network security context to evaluate the effect of network size on optimal decision-making. Network topologies are also discussed to investigate the limitations of the common game-theoretic interdependent security models. We conclude that these models work well for small to medium-sized networks with fairly uniform topologies. Through analysis of these two models, we propose methodologies for decision-making that are simple to understand and applicable to many other interdependent security scenarios.

Chapter 1: Introduction

Network security is a topic that can be looked at from a variety of angles. The pure computer science approach focuses on topics such as protocol examination and verification, design of intrusion detection systems, and the exploitation of cryptographic primitives. The pure social science approach focuses on effects of security policies, privacy v. convenience trade-offs, and monetary cost-benefit analysis, among others. Recently, there has been a convergence of these two approaches - certain researchers have become interested in a more interdisciplinary view of security. Rather than isolating and analyzing a specific vulnerability or looking at the legality of a privacy issue, we are concerned with the more economic side of security.

This work focuses on modeling network security as an *interdependent* problem, and analyzing mathematically these situations[1]. The term interdependent refers to the expanded scope of any decision made within a network - my choices affect my neighbors' and vice-versa. As an example, without knowing it, a user may be sharing his media library with everyone around him. This provides other users in his network with some benefit and some risk. If a file of his is malicious, it can infect someone else who tries to access it remotely. However, if he had invested in adequate security measures, that malicious file would have been caught before it had time to spread. Kunreuther and Heal [11] refer to this as an interdependent security problem. Interdependency refers to the external effects that decisions have. In the media library scenario above, a person who does not invest in security is negatively affecting all others who are networked with him.

The structure of this thesis is as follows: Chapter 2 discusses related work in interdependent decision-making and game theory applied to security, Chapter 3 discusses the case of a network administrator purchasing protection for her network, Chapter 4 discusses multiple users each making security decisions in the context of a network of other individuals and Chapter 5 concludes.

[1]This thesis is based on our papers that are in a preprint database [9], accepted to a conference [8] and under review for a conference [12], respectively.

Chapter 2: Related work

Kunreuther and Heal [11] analyze a variety of interdependent security scenarios. These scenarios are framed as problems where there are some number of users, companies, etc. that are attempting to minimize their own personal loss (or maximize their gain). The interdependency refers to the types of threats that each of the organizations is facing. In their key example, they look at a malicious person carrying an explosive device onto a specific airline. The defenders in this scenario are airlines that each do not want a bomb to go off on their airline . They each face the risk that a suspect will carry a bomb from outside onto their airline (an external threat) as well as the risk that a suspect will carry a bomb from another airline onto theirs (an internal threat). These internal threats are expressed as negative externalities that entities factor into their decisions, which are based strictly on utility. We then look at behavior under various conditions and derive Nash Equilibria for a variety of scenarios, including a section on network security.

Heal and Kunreuther [7] follow up with a revised model that accounts for heterogeneous populations. The heterogeneity refers to the agents in this model having different risks and costs associated with decisions that they make. They examine conditions under which systemic changes to security decisions can tip from one choice to another (from some level of protection to defection).

Grossklags, et al. [5] also use game theory to look at different security problems and classify these problems into five canonical security games: *total effort*, *weakest-link*, *best shot*, and *weakest target (with and without mitigation)*. Each game uses the notion of protection levels and insurance levels and differs in how they are measured. For instance, in the *total effort* security game, the protection level of an individual is the mean protection level of all entities in the system. They look at Nash Equilibria for the various games as well as tipping points. They also discuss strategies to mitigate free-riders, people who will not invest in security due to the form of the game (e.g. if one player in the weakest-link game will never invest, then there is no incentive for anyone else to invest in security either.)

Grossklags and Johnson [6] follow up by examining uncertainty in one of their security games, the *weakest link* game. This game is characterized by a tightly coupled interdependent network where the utility of any player is equal to that of the least-well-protected player in the network.

They then go on to propose a revised utility model for this game, dependent upon the player's level of expertise. They also examine the role of information levels in the same game and discuss the role information plays on the different types of players. They conclude that payoffs are no more than 18% better with complete versus incomplete information in their particular formulation.

Johnson, et al. [10] continue looking into interdependent games, this time using a simpler model of a corporate LAN. They derive utility functions per player in homogeneous and heterogeneous situations based on a number of factors. They forgo entirely the notion of insurance and look at internal and external protection levels. Even if a player protects, she still has the chance to be compromised from within the local network. They go on to discuss uncertainty in certain modeling parameters, and they use probability distributions to clear up some uncertainty before computing Nash equilibria for the homogeneous and heterogeneous cases. They conclude that uncertainty in certain parameters does not significantly affect where the equilibria tend to lie, and that homogeneity in some parameters does not affect some conclusions in their model.

Professor Larry Heimann and I have built on this model [10] by examining the case of a single-decision maker (e.g. a network administrator) and by introducing the concept of a loss profile, a statistical distribution representing the amount of loss incurred by an entity once infected [8]. In the next chapter we will go on to show the effects of these profiles on the administrator's decisions. In the following chapter we will examine the decisions of multiple people in the context of each other, and examine how network topology and size (and behavior in these various situations) affects optimal decision-making.

Chapter 3: The Effect of Loss Profiles in a Single-User Interdependent Security Scenario

In this chapter, we formulate a model of interdependent network security where there is a system administrator responsible for a network of size n against autonomous attackers attempting to penetrate the network and infect the machines with viruses or other exploits. We introduce the concept of a loss profile, which encapsulates the idea of variable loss due to infection. Through the application of a simple loss profile to this interdependent network security scenario, we conclude that the decision is dependent upon a number of factors including external and internal vulnerabilities, the types and likelihoods of different amounts of loss, and the interaction of all of these effects. Through this analysis, we form a model for centralized decision-making that is simple to understand and applicable to many other interdependent security scenarios.

3.1 Model

The model that we will investigate here is that of a corporate network. There is a single network administrator that is responsible for mandating security policies for n users in the network. The policy we are investigating is whether or not the administrator should mandate universal protection within his network[1]. In order to evaluate this decision, we need to look at some more specifics of the scenario.

Each computer in this network is connected to an external network and is connected to every other computer in the network. The external connection (e.g. to the Internet) poses some risk of infection from attackers and viruses. These attackers have their individual preferences on which computers they want to attack. In the language of game theory, each attacker is playing a purely

[1] This scenario is a common one, where users in the network must follow corporate policies.

mixed strategy (exogenous to this discussion and decided prior to it). Each attacker will attack some machines with (possibly) nonzero probabilities. If any attacker attacks an unsecured machine, he is sure to compromise it[2]. Once a machine is compromised, it cannot be compromised again. For our purposes, we can aggregate all possible attackers playing mixed strategies into one attacker that will attack machine i with (possibly zero) probability p. If an unsecured machine is attacked, it is surely Due to the interdependent nature of the network, once a computer is infected, it may spread its infection to other machines. The probability that machine i becomes infected and infects machine j is q. The internal infection parameter q and the external one p are both assumed to be the same for all machines in the network, respectively. That is, all p's equal each other and all q's equal each other. The internal infection probability q includes the situation where i does not become infected from an outside source, therefore $q < p$.

This scenario has a number of stages, which are performed in the following order:

1. The administrator mandates total protection or total defection.
2. Outside attackers compromise some network machines.
3. Infected machines compromise some other network machines.

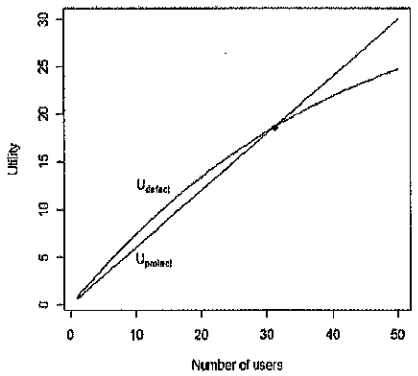
Each machine in the network has a value, which we will refer to as A_i. Accordingly, the entire network (before the scenario begins) has a value of $\sum_{i=1}^{n} A_i$. If the administrator chooses to mandate protection, then he will pay a cost c for each machine in his network (e.g. for virus protection licenses). Therefore, the loss he incurs for protecting all machines is cn. By mandating protection, the value of his network becomes:

$$U_{protect} = -cn + \sum_{i=1}^{n} A_i$$

By paying the protection costs, the administrator ensures that every machine is protected from the outside, and consequently from the inside (since a precondition to internal infection is a machine to have been infected from the outside). If the administrator chooses to mandate a policy of defection, he risks external and internal infection across all machines. In the tradition of other research in this area ([5] [6] [10] [11]), we will assume that the parties involved (namely the network administrator) are perfectly rational, utility-maximizing individuals. Therefore, the administrator will mandate protection when his expected utility from doing so is greater than the alternative.

[2]This is not completely realistic, but serves to simplify the model a good deal. See Bier, et. al. [2] for a model (in a related domain) that includes attacks that may fail.

Figure 3.1: Utilities for protecting and defecting intersecting at a tipping point



3.1.1 Loss profiles

Up until this point, we have assumed that all machines, if infected, will incur a total loss with probability 1. In order to more closely model the variable nature of loss due to infection, we introduce the concept of a *loss profile* to quantify these possible differences. Formally, a loss profile is a discrete probability distribution characterized by a probability mass function $L_i : [0, 1] \to [0, 1]$, from percentage of machine i's loss to probability of occurrence (e.g. $L_3(0.1)$ is the chance that machine 3 loses 10% of its value A_3, given that machine 3 is infected). The expected utility maintained by machine i, if infected, is therefore $A_i l_i$, where l_i is a randomly drawn percentage from L_i.

There are many factors that may contribute to the characterization of a computer's loss profile. A machine holding a lot of sensitive data may be considered a total loss if any of it gets out (meaning $L_i(1) \approx 1$ for this particular machine). On the opposite end of the spectrum is a machine with a casual user (and no sensitive data), which will only be considered a total loss if rendered unusable, which is not very likely. To better explain the concept of loss profiles, we will look at a bimodal loss profile, which is characterized by:

$$L_i(x) = \begin{cases} p_l, & \text{if } x = 1 \\ 1 - p_l, & \text{if } x = 0 \\ 0 & \text{otherwise} \end{cases}$$

The parameter p_l is the probability that a compromised machine incurs a total loss. This loss profile, while surely a simplification, is enough to represent users who will either incur a total loss, or no loss (for this reason, a bimodal loss profile can be thought of as all-or-nothing).

The bimodal loss profile is a simplification of the multimodal loss profile, where there are M discrete amounts of possible loss, that each occur with some probability. If we let $X = \{x_i : i \in [1, M] \text{ and } 0 \leq x_i \leq 1\}$ be the set of possible losses, then an multimodal loss profile is characterized by:

$$L_i(x) = \begin{cases} p_x, & \text{if } x \in X \\ 0 & \text{otherwise} \end{cases}$$

In this loss profile, p_x is the probability that x percent of machine i's value will be lost upon infection[3]. The multimodal loss profile is applicable to many situations. From the standpoint of perceived loss, a machine may fit an multimodal loss profile if its user has M distinct problems that may occur if infected (e.g. confidential document leaked, password file stolen, machine rendered unusable, etc.). While these types of loss profiles are not all-encompassing, they will help us to analyze the administrator's decision making in the following sections.

3.2 Analysis

We now have sufficient information to calculate the expected utility maintained by the network if the administrator mandates defection. This utility is as follows:

$$U_{defect} = (1-p)(1-q)^{n-1} * \sum_{i=1}^{n} A_i l_i$$

And from before:

$$U_{protect} = -cn + \sum_{i=1}^{n} A_i$$

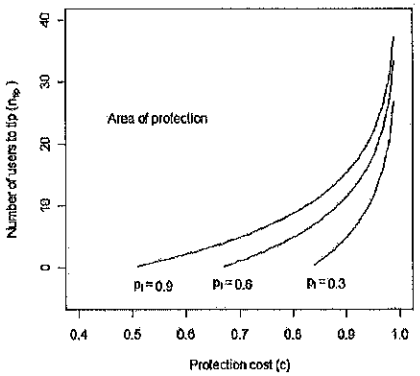
The administrator will make a purely utility-based decision and will mandate defection when $U_{defect} > U_{protect}$.

3.2.1 Simplification

For the purpose of our analysis, we will assume the administrator knows that all machines have the same values ($A_1 = \cdots = A_n = 1$), and that loss profiles are bimodal and the same for all machines (l_i is still different for every machine, but is drawn from the same distribution $L(x)$). While this may be an oversimplification, in a real scenario the administrator will not know the composition

[3]Since $L_i(x_i)$ is a probability distribution, $\sum_{i=1}^{M} L_i(x) = 1$

Figure 3.2: Protection costs vs. tipping points (q=0.1, p=0.5)



of his network, and would possibly assume an "average" loss profile for each machine, which would aggregate to a roughly correct picture of the population, assuming the initial estimate is reasonable. The assumption of a single type of loss profile captures this situation.

We will also assume fixed values of p and q, the outside and inside infection probabilities respectively. These are just as hard to estimate, and that analysis is beyond the scope of this work[4].

3.2.2 Application of bimodal loss profile

Now that we have simplified the model, we can apply it to a population of machines with bimodal loss profiles. Though each machine has its own parameter p_l, since each parameter is drawn from the same distribution the expected percent loss of a randomly selected machine is p_l (Expected percent loss $= 1(p_l) + 0(1 - p_l) = p_l$). With this in mind, the administrator's expected utilities are:

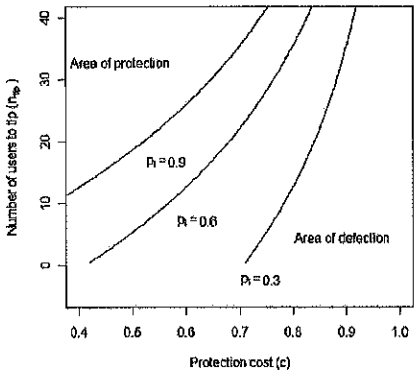
$$U = \begin{cases} np_l(1-p)(1-q)^{n-1}, & \text{if he mandates defection} \\ n(1-c) & \text{if he mandates protection} \end{cases}$$

When the expected utility for mandating equals that for mandating defection, we call that value of n, the number of players, a *tipping point* represented by n_{tip}. If $n > n_{tip}$, then the administrator will choose to mandate protection. Equating the parts of the above piecewise function yields that tipping points are calculated by:

$$n_{tip} = 1 + log_{1-q}\frac{1-c}{p_l(1-p)}$$

[4]For a thorough analysis of estimating these parameters see [10].

Figure 3.3: Protection costs vs. tipping points (q=0.03, p=0.05)



As previously stated, p and q are some fixed values, which the administrator knows. He also knows the average loss profile parameter p_l for the population. The only right-hand-side quantity which has not yet been discussed is the protection cost c. The protection cost is the most obvious factor that contributes to the administrator making a decision. Even if he had much less knowledge about his network, the administrator would not buy grossly expensive protection. Therefore, imagine that the administrator is approached by a salesperson willing to sell him protection at a per-machine cost of c. He looks at the size of his network, then buys and mandates protection if he has more machines than n_{tip} for this per-machine protection cost. Keeping this in mind, some example scenarios are shown in Figures 3.2 and 3.3.

3.2.3 General insights

We can draw a number of general insights from the tipping point equation and figure above. The function for tipping points is continuous. It does not have full support, leading us to conclude that for a given network, an administrator will always do better by mandating protection where the function is non-positive. Due to the constraints that we put on the problem ($A_i = 1 \; \forall i \in [1, n]$) , the function has a vertical asymptote at $c = 1$. If we had not imposed this restriction, there would be a vertical asymptote at:

$$c = \frac{\sum_{i=1}^{n} A_i}{n} = A_{average}$$

As the protection cost approaches the average value of a machine in his network, the administrator is more compelled to mandate defection. The administrator should mandate defection when protection costs more than the average value of a network machine. As the protection cost approaches zero, the administrator is more compelled to mandate protection.

3.2.4 Specific insights

The role of the infection parameters

For the specific values of p and q in Figure 3.2, the *Area of Protection* where the administrator should definitely mandate protection, is much larger than the *Area of Defection* (for protection costs between zero and one, the area of protection is at least 7 times larger than the area of defection).

For the specific values of p and q in Figure 3.3, the area of protection is significantly smaller than in Figure 3.2. From this, we can conclude that the infection parameters p and q play a significant role in the administrator's decision to mandate protection, holding all other parameters constant.

Another notable result regarding the infection parameters concerns their ratio q/p. For some distinct machines i, j:

$$\begin{aligned} q &= P[(i \text{ becomes infected}) \text{ and } (i \text{ infects } j)] \\ p &= P[i \text{ becomes infected}] \\ q/p &= P[(i \text{ infects } j) \mid (i \text{ has been infected already})] \end{aligned}$$

The effect of q/p is clearly shown between the two figures. In Figure 3.2, $q/p = 0.2$ whereas in Figure 3.3 it is 0.6. This variation in this ratio is partially responsible for the difference in the appearances of the graphs. The ratio quantifies the events of third stage of the scenario, where machines infect one another. By this point, all machines have either been infected from the outside or not. This third stage will not occur unless at least one machine has been compromised from the outside. This ratio represents the chance that one machine will infect another, given it is infected. It is of note that in a real-world version of this scenario, the administrator would be able to reduce both the chance of outside infections p (through the parameter c in our model) and the chance of inside infection q (through a parameter which is exogenous to this model). This is an area for further study.

The role of the loss profile parameter

Since the loss profile parameter p_l dictates the average degree of loss given infection, it should have an effect on the size of the area of protection. The result which is counterintuitive is that an increase in p_l towards 1 (which would mean that an infected machine is completely compromised) makes the decision to mandate protection less enticing. This can be seen somewhat in Figure 3.2,

and much more so in Figure 3.3. The intuitive explanation for this is that as it becomes harder to mitigate loss on a machine, protection becomes less worthwhile.

The infection parameters' effect on the loss profile parameter's effect

Since the space between pairs of tipping point lines is different between the figures, the infection parameters affect the degree to which the loss profile parameter affects the decision. This is intuitive, since if there is a small chance of infection (from inside or outside), then it matters less how much loss a machine incurs if infected.

Another noteworthy result is that any decrease in the infection parameters will cause a "left shift" in the tipping point curve, holding the loss profile parameter constant. At the lower average level of infection, the administrator is not as enticed to mandate protection, and will therefore defect more often.

3.3 Conclusions

From the proceeding section, it is obvious that this model captures some very real world effects in the problem of an administrator setting security policies. Sometimes the administrator has an easy decision based upon protection costs alone (protect when cost is low and defect when it is high). This decision is heavily influenced by how exposed machines in the network are - it makes less sense to mandate protection when there is a very low chance of infection.

The largest contribution of this work is the introduction and analysis of loss profiles. Other work [10] has drawn similar conclusions regarding the infection parameters, but did not address the unequal losses possibly incurred by district machines. However, Grossklags, et. al. [5] posed distinct security games that had their own conditions for loss, which solves a similar but not identical problem. This work shows how the even the simplest bimodal loss profile drastically affects the size of area of protection, and therefore the overall decision to mandate protection. The degree of this effect is either mitigated or enhanced by the magnitude of the infection parameters - when a machine is likely to become infected, the amount of loss incurred upon infection becomes more relevant.

The results here apply not only to the computer security scenario, but to many situations that can be modeled as interdependent security problems[5]. An immediately relevant scenario is airline security, where an "infection" means a prohibited device or person getting through airline security, and a bimodal loss profile would have a very high loss profile parameter (if a bomb gets onto a plane, it will do a lot of damage to the airline as well as the industry). In that scenario, the relative

[5]See [11] for some additional examples of these types of problems.

values of differently sized airlines and planes would factor heavily into the analysis. Here we have proposed a model that can be applied to a number of different situations, keeping in mind which parameters are more important based on the risks of the scenario.

3.4 Future Research

This chapter introduced and performed analysis using the concept of a loss profile. There is more research to be done in this area, specifically applying and analyzing a variety of loss profiles in similar scenarios. Moreover, in this chapter we have treated the population as homogeneous, with all the machines sharing the same loss profile. In a more heterogeneous population where machines had different loss profiles, the results could be influenced by the distribution of various loss profile distributions and their respective parameters among the population of individuals.

In a related manner, this chapter has a single decision-maker, i.e., the system administrator, who could act on behalf of the entire network. If we allowed for individuals within the network to act on their own, we introduce a whole new level of complexity.

Another possible area for expansion is making the attacker a player in an attacker-defender game. It is possible to let the attacker choose his strategies, rather than be a conglomeration of various individuals playing mixed strategies[6]. Additionally, we discussed how an real-world administrator could not only reduce the probability of outside infection, but that of internal infection as well. One possible strategy for reducing the chance of inside infection is changing the network topology by removing or changing connections. These changes come at some cost - by removing a link between two machines, the administrator makes file transfers between these machines slower or impossible. This opportunity cost needs to be accurately weighed against the benefits of performing the change in topology. Along with the possibility of an intelligent attacker, this may help us better quantify some of the effects that may arise in these types of scenarios.

[6]See [2] for a similar scenario without any interdependencies.

Chapter 4: Multi-user Interdependent Security Decisions

In this chapter, we address some of the concerns in the previous chapter. Namely, we formulate a similar model for decision-making, but this time regard each individual as a decision-maker rather than having a single administrator make the decisions. We show how interdependency impacts decision-making for individuals in a number of increasingly complex situations. We build from a very simple example a single user purchasing virus protection software to a complex multi-person computer network scenario, explaining new complications as they arise. However, making decisions in these scenarios involves much more than mathematical calculations. We look at Olson's [13] theory of groups, and how various factors including group size and homogeneity affect the acquisition of public goods, such as security. We also discuss network topologies and how they affect these situations.

This chapter begins by building a model for utility-based decision making assuming that people are perfectly rational. It then builds from a network with no externalities to one with many externalities, and shows the effects of these externalities. We then examine discusses how network size and topology can affect how players behave in these scenarios. Then, it looks to undermine the assumption of perfectly rational players by examining Olson's theory of groups [13] in this context.

4.1 Connectivity and security models

4.1.1 The case for virus protection

Let us begin by looking at the actions of Alice, a perfectly rational person. Alice has a personal computer that she uses often, though it has no security software installed. Alice decides to analyze a scenario where she can buy a full-featured protection mechanism. This ideal mechanism, if purchased and installed, will completely remove the threat caused by outside infections and let her

Figure 4.1: Single-user protection decision examples

Value (A)	**Infection probability (p)**	**Protection cost (C)**	**Expected loss (Ap)**	**Decision**
10	0.5	0.5	5	*Protect*
10	0.5	9	5	*Defect*
10	0.5	5	5	*Tie*
10	0.7	5	7	*Protect*

browse the Internet safely[1]. However, this full protection comes at a cost C, which captures not only the monetary cost of purchasing the package, but the time to install it, etc. She can either protect and pay C, or defect. Let us call the worth of Alice's computer A (as perceived by her). Let us also say that without protection, Alice's machine will be infected and suffer a total loss with a probability p. Her utility can be represented as:

$$U_i = \begin{cases} A - Ap, & \text{if Alice defects} \\ A - C & \text{if Alice protects} \end{cases}$$

Alice will buy protection if the cost of protection (C) is less than her expected loss (Ap). The cost of protection needs to be small enough to mitigate the risk caused by going unprotected. Alice should buy protection when it has a low cost (close to zero or much less than her machine's value). Accordingly, she should not buy protection when its cost is close to her current value. When the probability of infection is high protection is more optimal assuming mid-range values for the other parameters. When neither the probability of infection nor the cost of protection is extreme either solution may be optimal.

4.1.2 Two players

Right before Alice makes her decision about buying protection, Alice's younger brother Bob declares that he is considering connecting to the Internet, and wishes to share files he downloads (legally, of course) with his sister Alice. He solicits Alice's advice on if he should buy protection, and will follow her advice if he has an incentive to do so. She decides to approach the decision using the model from before. She adds in another player Bob, and another term for the "badness" that they

[1]While no security software is perfect, the recommendations yielded from examining this scenario can serve as boundary cases (e.g. if Alice does not find it worthwhile to buy a perfect security package, she will certainly not buy an imperfect one).

might impose upon one another. We will define $X_{Bob \to Alice} \in [0, 1]$ as that "badness", the expected negative externality that Bob imposes on Alice[2]. This affects A_{Alice} by scaling it negatively. Let subscripts identify which person, Alice or Bob, we are discussing. We can discuss Alice's expected utility as:

$$U_i = \begin{cases} A_{Alice}(X_{Bob \to Alice})(1 - p_{Alice}), & \text{if Alice defects} \\ A_{Alice}(X_{Bob \to Alice}) - C, & \text{if Alice protects} \end{cases}$$

Bob's expected utility can be defined by replacing all "Alice" with "Bob" and vice-versa in the equation above. Without loss of generality, we will discuss Bob's effect on Alice from here onwards. Let us define the probability $q_{Bob \to Alice}$ as the chance that Bob infects Alice, regardless of if he is infected or not (e.g. $q_{Bob \to Alice} = P$[Bob becomes infected and spreads the infection to Alice]). This presumes that becoming infected and spreading an infection are independent events. Therefore, $q_{Bob \to Alice} < p_{Bob}$, since Bob must be infected to infect Alice. While there is certainly more to say on the subject, for the current discussion, we will simply assume $q_{Bob \to Alice}$ is a function of p_{Bob} (see [10] for a way to estimate this parameter and [8] for discussion on its effect). As before, if Alice contracts the infection, she will suffer a total loss. It is therefore expected that Bob imposes the following externality upon Alice (which is one if he protects, since he can't infect Alice if he isn't able to be infected):

$$X_{Bob \to Alice} = 1 - q_{Bob \to Alice}$$

We can think of $X_{Bob \to Alice}$ as Bob removing a portion of Alice's initial value. There are four possible situations that can result from combinations of Alice and Bob protecting and defecting. Our goal is to find which situations are Nash equilibria and when they are. A Nash equilibrium is a situation when every player cannot benefit for switching his choice in the contexts of everyone else's choice.

Previous analysis [11] shows that Nash equilibria are only possible when both players choose the same strategy. Accordingly, Bob and Alice should both protect when the cost of protecting mitigates both the risk of internal and external infection. The same basic conclusions hold from above. They should both protect when $C < Ap$ (assuming their values, protection costs, and infection risks are the same).

In non-symbolic terms, the protection strategy is a Nash equilibrium when expected loss (due to internal and external infection) is smaller than protection cost. This is consistent with the single-user case, where a protection cost much smaller than the currently value dictates protection.

[2]The negative externality imposed between two players can manifest itself in a variety of ways, including one player spreading infections to another and an attacker compromising one machine in order to reach another. We will use the term infect loosely here to refer to any of these scenarios.

Figure 4.2: Normal form representation of the Alice-Bob situation

		Bob	
		Protect	Defect
Alice	Protect	$(A_{Alice} - C_{Alice}, A_{Bob} - C_{Bob})$	$(A_{Alice}(1 - q_{Bob \to Alice}) - C_{Alice}, A_{Bob}(1 - p_{Bob}))$
	Defect	$(A_{Alice}(1 - p_{Alice}), A_{Bob}(1 - q_{Alice \to Bob}) - C_{Bob})$	$(A_{Alice}(1 - q_{Bob \to Alice})(1 - p_{Alice}), A_{Bob}(1 - q_{Alice \to Bob})(1 - p_{Bob}))$

4.1.3 Generalizing to N people

Alice suddenly got a new job as a systems administrator for corporation X. They have just decided to overhaul their current corporate network, and ask Alice to lead the redesign. In discussions with management, Alice says that investing in protection may not be worthwhile right now. They are intrigued by the possibility of saving some money, and ask Alice to research into this option. If she can show them definitively that leaving the system unprotected can save them a good deal of money, they will suggest that employees follow her recommendation. She begins to use the same model as before, adding in a new player for each employee of the company. The employees in the company are referred to by their id numbers, which range from 1 to N inclusive, where N is the number of employees in the company. Taking this into consideration, the utility for player i can be expressed as:

$$U_i = \begin{cases} A_i(X_{-i})(1 - p_i), & \text{if player } i \text{ defects} \\ A_i(X_{-i}) - C, & \text{if player } i \text{ protects} \end{cases}$$

In this equation, X_{-i} is the net externality imposed on player i from all other players in the network. The portion of the negative externality imposed from a player who chooses to protect is nearly zero (assuming scenarios such as A infects B infects C are rare). Let us call k the number of players who defect ($k \leq N$). Also, let us call the chance that any player j infects player i ($q_{j \to i}$) the same as any other chance, which we will refer to as q. With these considerations in mind, we can quantify the net externality imposed on player i, as follows:

$$\begin{aligned} X_{-i} &= \prod_{j \neq i} X_{j \to i} \\ &= \prod_{\substack{j \neq i \text{ and} \\ j \text{ defects}}} 1 - q_{j \to i} \\ &= \prod_{k} 1 - q_{j \to i} \\ &= (1-q)^k \end{aligned}$$

In order to further simplify things, let us assume that every individual has the same chance of outside infection, which we will call p. We can therefore refer to player i's utility as:

$$U_i = \begin{cases} A_i(1-p)(1-q)^k, & \text{if player } i \text{ defects} \\ A_i(1-q)^k - C, & \text{if player } i \text{ protects} \end{cases}$$

This model is consistent with others that have been previously analyzed ([5], [10], and [11]). It has been generalized to fit various types of infections [5] and reduced to investigate effects on users with a uniform amount to lose [10]. We will defer equilibrium analysis of this scenario and its variants to those works.

4.2 The impact of size and topology

Once Alice made her recommendation to the company, she realized that there may be other things she needs to consider in recommending decisions. Namely, she wants to examine how the size of her network and the value of users' machines within in impact their decisions. She also wants to examine how various network topologies impact these decisions. We will discuss these in the following sections.

4.2.1 Player value and network size

The model above has been analyzed almost exclusively where all player values (A_i's) are equal and every player has roughly the same amount to lose. As discussed in [8] and above, different people can be characterized by different loss profiles represented by statistical distributions with different parameters. These profiles relate the value of a person's machine to the likelihood of infections. Imagine two people that have the same amount of data on their computers, but one has periodic remote backups set up while the other does not. The person with backups is likely to lose a small amount of his recent work to an infection, but very unlikely to lose all of it. This contrasts with

the other person, who has a high chance of a total loss if infected. These players have different loss profiles.

From another perspective, Olson discusses how groups composed of members of unequal values will be more likely to provide for themselves a public good [13]. In network security scenarios, this means that if one person has a lot to lose (a large A_i), they have more of an incentive to buy protection. The decision of that more valuable person to buy protection is largely independent of other players in the network. This suggests the existence of equilibria with heterogeneous strategies in situations of unequal player wealth (all A_i's not equal). This is due to a scenario where the most valuable player can pay for the protection of others in his network (to protect his own assets) or one where he can convince others in the network to invest. Therefore, smaller groups or groups with more pronounced heterogeneity are more likely to "mobilize" and end up fully protected.

Olson also discusses how on average larger groups tend to be less productive than smaller ones [13]. The same effect applies to networks; larger networks are much less likely to end up with total protection than smaller ones, where one person may be able to logically convince all of the others. Additionally, Heal and Kunreuther [7] discuss how attackers will focus on less-well-protected targets. Therefore, assuming that there is a malicious adversary actively attacking a network, the probabilities of being infected from the outside is dependent upon how many people have already protected. The people who are "targeted" are now more likely to become infected and spread a virus to other people in the network. Clearly, network size and the proportion of people currently protected influence internal infection probabilities.

4.2.2 Network topology

While network size clearly has an impact on these types of decisions, it is also important to look at the connectivity that exists among members of the network. Different network topologies capture these connections. Topology is closely related to the uniformity previously examined in the probability of infection from an inside source (q). Without investigating topology, Johnson, et al. [10] drew the internal infection parameter from a distribution over $[0, p]$ to quantify differences in these probabilities. In the following paragraphs, we will examine how network topologies affect these probabilities and the decision to buy protection as a whole.

Anderson suggests that "different [network] topologies have different robustness properties with respect to various attacks [1]." Ganesh, et. al. examine with a high degree of mathematical rigor the role of various network topologies on the spread of infections [4]. They examine many configurations, such as a star-shaped network (where all but one node have only one connection, to the central node). They show that as the number of network participants (N) increases, the probability that the entire network becomes infected decreases towards zero, given that an infection

could die out. Ignoring the central node (which is identified by index 0), which will almost certainly become infected if any other node does, every node can only become infected from the central node ($q_{i \to j} = 0 \; \forall \; i \neq 0, j \neq 0$).

Therefore, the probability that a node infects another node is strongly related to its position within the star topology. It also follows that the infection of a leaf node after the initial one does not affect the chance of any other leaf node becoming infected. Also, let us suppose that a virus lives for a certain lifetime in a certain machine, and afterwards the machine is immune to infection. Given this, the presence of more network participants decreases the chance a specific node will be infected. Knowing this information, players in a large star-shaped network are more likely to free-ride (not buy protection).

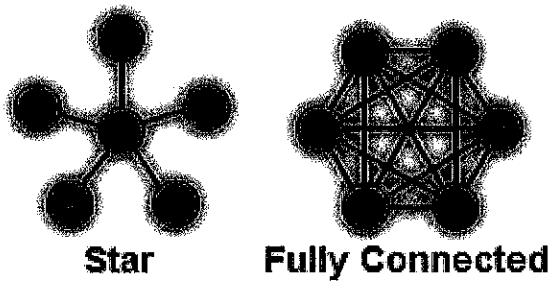
Christin, et. al. [3] suggest that incomplete information plays an important role in choice to buy protection. Using the example of a star-shaped network from above, we can clearly see this - if a non-central player knows that the central node has implemented some form of internal protection, then he is effectively rid of the threat from other nodes[3]. If the central node does not have this special protection or if users are not aware of it, then they gain no information about the probability of internal infection (except for its source). If we extend the concept of internal protection to leaf nodes, then knowledge of network topology is very important. For example, if Alice knows she is a leaf of a star-shaped network, she will want to invest in a mechanism to shield herself from the central node. This defense would be impossible if the player was not aware that her network is star-shaped.

Another common network topology is the fully-connected network. In this configuration each node is connected directly to each other node. This type of network represents a situation where machines have a physical or logical link to all other machines in the network. Since the number of connections in this type of network grows quadratically with the number of nodes in the network, this type of network is only practical when the network is small and or direction connection is valued heavily (e.g. when quick file transfers are necessary).

The incentives for individual protection from the outside or from individual nodes are much worse in the case of an all-connected network. For example, if everyone is a given network is protected and one new machine enters without protecting every other node is now vulnerable due to the link to the entrant. As we discussed for star-shaped networks, it is rational to protect against threats from the node that can inflict upon you the most risk. In the star-shaped network, only the center node poses a reasonable risk to a leaf node. In a fully-connected network, for an agent to prudently protect herself she must protect the links between her machine and every unsecured machine (as well as the link to the outside).

[3] Internal protection refers to a program such as a firewall which can screen or block connections between machines.

Figure 4.3: Two common network topologies



When an unsecured machine (or group of machines) enters the network, due to the increased number of links each agent needs to secure, the rational decision may be for all agents to not protect against the entrant(s). In the situation we have discussed here (where protection is absolute), there are multiple protection schemes that are stable, assuming risks are greater than protection costs. One of these schemes is every node protecting from only the outside. In this configuration, no node pays a useless internal protection cost and every node has no chance of being infected. Other protection configurations for different network topologies will be examined in a further work.

This analysis can be similarly applied to other network topologies, which suggests that topology plays a very important role in interdependent security models. Pal and Hui [14] do some of this analysis with regards to the role that topological uncertainty can make in these types of decisions. They do not however discuss the role that placement within a network of a given shape can play in decision-making. For homogeneous topologies a reliable mathematical model for protection decisions can be formed, but for mixed topologies which are more common in larger corporate networks, the process will be much more complicated. This is certainly an area for further study.

4.3 Conclusions

We discussed the case for buying protection software for a one-computer network. In this situation, protection should be purchased when the cost of protection is smaller than the potential loss from going unprotected. In the sections immediately following, we expanded the model to include two or more players that all affect each other. We introduced the concept of a negative externality, which is an important factor in a player's utility-based decision-making in a multi-player network. This concept, which was shown by Kunreuther and Heal [11] is implicitly used in Johnson, et. al. [10]. This chapter explains the derivation of this model from a network security standpoint.

We then discuss some additional considerations that need to be made when making these decisions and show their importance. We look at network size and player value to show that proposed models may not be sufficient for large or homogeneous networks. We also discuss the role of network topology using the star-shaped and fully-connected topologies as examples and show how topology differences play an important role in security decisions.

In future research, we intend to examine more network topologies mathematically to further examine the some of the conclusions above. It would be interesting to examine networks with random topologies as well as networks where people have various amounts of information about the size of the network (e.g. Alice is 75% sure there are at least 50 people in her network). Each of these is an area for further research.

Chapter 5: Conclusions

This thesis has examined a common problem - securing a computer network - from two distinct perspectives. We first looked at a situation where one person can mandate others to follow her security decisions. We formulated a model for her to make a decision based on a variety of salient information she may have about her network. In the process, we introduced a previously unknown concept called a loss profile. Loss profiles can be used to quantify the variable loss a machine may incur when it is infected (e.g. how likely is a user to lose a large project or be denied access to her own machine). We introduced the bimodal loss profile and its generalized cousin the multimodal loss profile and analyzed how these profiles affect decision making. To summarize, loss profiles affect the decision because a machine that is likely to incur near total loss upon being infected it is less enticing to protect.

After examining the case of a single administrator, we then looked at the case of users in a network being able to make their own decisions. We started from a single user making her own decision, which is simply the administrator-deciding model ignoring loss profiles and having a single user. From there, we built in other users an examined the ways in which they affect each other. By examining these negative externalities, we were able to analyze how a single user should analyze the purchasing of protection software in the context of other users in the network.

We then introduced more complexity into the game-theoretic model by looking at network size and topology. We concluded that both of these are significant factors that have been largely ignored by most work up to this point. We discussed how introducing internal protection mechanisms in various network topologies (star-shaped, fully-connected) can affect how users protect against each other and the outside world. While looking at network size, we examined how groups of various size tend to behave according to Olson's [13] theory of group behavior. Through this analysis, we concluded that the models we have proposed are not sufficient for groups of all sizes.

Bibliography

[1] R. Anderson and T. Moore, "The economics of information security," *Science*, vol. 314, no. 5799, pp. 610–613, 2006.

[2] V. M. Bier, "Choosing what to protect: Strategic defensive allocation against an unknown attacker," *Risk Analysis*, vol. 27, no. 3, pp. 607–620, 2007.

[3] N. Christin, J. Grossklags, and J. Chuang, "Near rationality and competitive equilibria in networked systems," in *Proceedings of the ACM SIGCOMM workshop on Practice and theory of incentives in networked systems*, ser. PINS '04. New York, NY, USA: ACM, 2004, pp. 213–219.

[4] A. Ganesh, L. Massouli, and D. Towsley, "The effect of network topology on the spread of epidemics," in *IN IEEE INFOCOM*, 2005, pp. 1455–1466.

[5] J. Grossklags, N. Christin, and J. Chuang, "Security and insurance management in networks with heterogeneous agents," in *Proceedings of the 9th ACM conference on Electronic commerce*, ser. EC '08. New York, NY, USA: ACM, 2008, pp. 160–169.

[6] J. Grossklags and B. Johnson, "Uncertainty in the weakest-link security game," in *Proceedings of the First ICST international conference on Game Theory for Networks*, ser. GameNets'09. Piscataway, NJ, USA: IEEE Press, 2009, pp. 673–682.

[7] G. Heal and H. Kunreuther, "You only die once: Managing discrete interdependent risks," in *Columbia Business School and Wharton Risk Management and Decision Processes*, 2002.

[8] C. F. L. Heimann and A. Nochenson, "The effects of loss profiles in interdependent network security," in *The World Congress on Internet Security (WorldCIS)*, 2012, to appear.

[9] C. F. L. Heimann and A. Nochenson, "Identifying Tipping Points in a Decision-Theoretic Model of Network Security," *ArXiv e-prints*, Mar. 2012.

[10] B. Johnson, J. Grossklags, N. Christin, and J. Chuang, "Uncertainty in interdependent security games," in *Decision and Game Theory for Security*, ser. Lecture Notes in Computer Science, T. Alpcan, L. Buttyn, and J. Baras, Eds. Springer Berlin / Heidelberg, 2010, vol. 6442, pp. 234–244.

[11] H. Kunreuther and G. Heal, "Interdependent security," *Journal of Risk and Uncertainty*, vol. 26, pp. 231–249, 2003.

[12] A. Nochenson and C. F. L. Heimann, "Optimal security investments in networks of varying size and topology," in *International Workshop on Socio-Technical Aspects in Security and Trust*, 2012, Under review.

[13] M. Olson, *The Logic of Collective Action: Public Goods and the Theory of Groups*, ser. Harvard Economic Studies. Harvard University Press, 1965, no. v. 124.

[14] R. Pal and P. Hui, "Modeling internet security investments: Tackling topological information uncertainty," in *Decision and Game Theory for Security*, ser. Lecture Notes in Computer Science, J. Baras, J. Katz, and E. Altman, Eds. Springer Berlin / Heidelberg, 2011, vol. 7037, pp. 239–257.

Last updated: April 26, 2012