

Critical infrastructure protection under imperfect attacker perception

Corresponding author:

Erik Jenelius

Jonas Westin

Åke J. Holmgren

July 12, 2009

Abstract

This paper considers the problem of allocating finite resources among the elements of an infrastructure system in order to protect it from antagonistic attacks. Previous studies have assumed that the attacker has complete information about the system and the utilities associated with attacks on each element. In reality, it is likely that the attacker's information about the system is not as precise as the defender's information, due to geographical separation from the system, secrecy and surveillance, complex system properties etc. As a result, the attacker's actions may not be those anticipated under the assumption of complete information. In this paper we present a modeling framework that incorporates imperfect attacker perception by introducing random observation errors in a previously studied baseline model. Within this more robust framework, we analyze how the attacker's perceptive ability affects attack probabilities and the defender's disutility and optimal resource allocation. We demonstrate e.g. that a less perceptive attacker may cause greater disutility for the defender, that increasing the investment in a poorly chosen element can increase the expected disutility even in a zero-sum situation, and that the elements which should be protected may differ significantly from the baseline model. The policy implications of the analysis are discussed.

KEY WORDS: Attack, Protection, Information, Perception, Uncertainty

1 Introduction

1.1 Antagonistic threats and game theory

Critical infrastructures are technical systems utilized to distribute energy, information, water, goods and people, and are of the utmost importance for the quality of everyday life. A major disturbance in the flow of services provided

by the critical infrastructures can constitute a severe strain on business, government and society in general. A typical situation when analyzing such risks is that there are few data of disturbances with severe consequences (the low probability high consequence problem). Nevertheless, three principal risk analysis approaches can be discerned: (i) ordinary statistical analysis of empirical data (accidents, incidents, disturbances etc.); (ii) mathematical modeling in combination with empirical element data; and (iii) expert judgments (collected through more or less formalized methods, i.e. interviews, surveys, group discussions etc.) [11].

This paper emphasizes the security aspects of the risk analysis of large technical systems, more specifically the threat from qualified antagonists. This is a broad category of threats that spans from insiders and saboteurs, to crime syndicates and transnational terrorist organizations. The purpose of an attack can be to cause severe damage to a technical system in an attempt to disable important functions of a society. However, the goal can also be to make a symbolic demonstration, or to cause a large enough consequence in order to achieve a psychological effect such as a spread of fear and anxiety.

Intentional attacks are different from random failures in the sense that the antagonist intentionally chooses the time and place for the attack. Furthermore, the measures applied to protect a system will, most likely, affect the antagonist's course of action. Changes in how the defender perceives that the opponent will act will, in turn, affect how the defense is allocated, which once more can affect the antagonist's behavior, and so on. Hence, there is a strategic interaction between the attacker and the defender, and studies of antagonistic threats embrace, as many authors before have pointed out, a "game" situation rather than a static decision situation.

A number of papers have studied various aspects of protecting potential targets against antagonistic attacks. Underlying most if not all of the studies is the assumption, going back to [15], of a rational and informed attacker. That is, given a set of alternatives such as different potential targets, the attacker associates a certain utility with each alternative for any protective measures taken by the defender, and will choose an alternative that yields the highest utility.

One branch of research focuses on how a single defender should allocate protective measures among targets to minimize the losses due to subsequent attacks. For instance, [5] analyze a problem where an attacker chooses a given number of facilities to disable in order to maximize the transportation costs between the remaining facilities. Prior to the attack, a defender chooses a given number of facilities to fortify in order to minimize the same objective function. Similar models are studied by [4]. [16, 18, 2, 10] assume that an attacker chooses one target to attack, possibly randomizing, and the probability that the attack is successful is determined by the resources previously invested in the target by a defender. [16, 18] show that given a limited resource budget, the optimal resource allocation for the defender under quite general assumptions is to minimize the attacker's maximum expected utility of an attack. [22, 1] model the decision variables of both the defender and the attacker as continuous levels

of effort for each target, which together determine the probability that the target is disabled.

A related line of research has considered situations where multiple agents protect private targets, and the externalities associated with such distributed decision-making [19, 13, 21, 14, 9, 2]. For example, [19] analyze a situation where two countries can invest in protection against a terrorist that will choose to attack either country or not attack at all. They show that the substitution effect, in which an investment by one country increases the probability that the other is attacked, leads to overinvestments if the countries obey their own self-interests.

1.2 Incomplete information and imperfect observations

A number of papers have considered that the defender may be uncertain about the preferences of the attacker. [18, 2] model the uncertainty as a probability distribution across possible attacker types; [2] explicitly frame the problem as a Bayesian game (e.g. [7]). In this case, the problem for the defender entails considering the expected disutility across the attacker types associated with each possible allocation. [21] uses a multinomial logit model (e.g. [20]) to represent the probabilities of a set of targets being attacked considering the defender's uncertainty about the attacker's preferences. [10] evaluate different strategies for protection of elements in electric power networks and handle uncertainty about the nature of the attack using a scenario-based approach.

Less attention has been given to the decision process of the *attacker*. [17] considers a game where the attacker, prior to the defender's resource allocation, is uncertain about the level of vulnerability of one of the targets. After observing the defender's investments, the attacker updates his beliefs about the vulnerability via Bayes' rule taking the rationality of the defender into account; the outcome is a perfect Bayesian equilibrium (e.g. [7]). This model incorporates attacker uncertainty prior to the defender's actions, but still assume that the attacker observes these actions perfectly. It also assumes that the attacker uses quite sophisticated reasoning (specifically, Bayes' rule) to arrive at its response.

It is unlikely in reality that an antagonist could perceive the precise gains associated with attacking a target as accurately as the defender of the system, due to factors such as undisclosed information, surveillance, complex system structure and geographical separation between the antagonist and the target.¹ Therefore, its subsequent actions may not necessarily be those that would yield the highest utility given all the information about the system that the defender possesses. This can be expected to have significant impacts on how the defender should allocate resources to protect the targets. For example, given a limited defense budget, assuming a perfectly perceptive attacker may mean that some elements should be left unprotected since they would be suboptimal alternatives for the attacker. However, if there is a chance that the attacker chooses a

¹The situation where this assumption would be the most accurate is perhaps when the antagonist has access to perfect insider information.

suboptimal alternative, this strategy could be very dangerous if attacks on some of these elements would cause high disutility for the defender (see Section 2.2).²

In this paper we analyze the impacts if the attacker cannot perfectly observe the utilities associated with attacking elements of an infrastructure system. In particular, we examine the implications for the defender’s problem of allocating a limited resource budget in order to protect the system against attack. As baseline we use the model of [18] which assumes complete information and perfect perception on behalf of both actors. We also include a non-attack option for the antagonist. The observation errors are modeled as random variables whose outcomes are not observed by the defender. In effect, the actions chosen by the attacker become probabilistic from the defender’s point of view.

We highlight a number of important implications of imperfect attacker perception that are not present in the baseline model. For example, an unwise defense investment can increase the expected disutility of the defender even in a zero-sum situation. Also, a less perceptive attacker can yield higher expected disutility for the defender even if resources are allocated optimally, which is in contrast to the notion that a perfectly informed attacker represents a worse-case scenario.

The baseline model is described and analyzed in Section 2. In Section 3 we generalize the model by introducing imperfect perceptive ability of the attacker. The properties of this modeling framework are analyzed in Section 4 with a numerical example given in Section 5. In Section 6, we extend the model by combining the attacker’s imperfect observations with the defender’s uncertainty about attacker types. The paper concludes with a discussion in Section 7.

2 Baseline model with perfect observations

2.1 Model formulation

In this section we introduce the baseline model with perfect observations, which is then generalized in Section 3. This model is similar to the game studied in detail by [16, 18], to which we refer the reader for a rigorous treatment. The only essential difference is the presence of the non-attack alternative, which does not change the analysis in any significant way.

We consider a system consisting of n elements, indexed by $i = 1, \dots, n$. Each element represents a component of an infrastructure system that can be defended independently of other elements and that is a potential target for an attack. A defender has a total resource budget c_{total} to distribute among the elements. The resources that the defender allocates to an element are used to

²A related issue is the possibility that the attacker makes mistakes when *executing* its chosen strategy. For example, using network analysis, [12] suggests that one of the three bombings of the London underground on July 7, 2005 may have been intended to hit the station (King’s Cross) immediately succeeding the station (Edgware Road) where the bomb actually exploded. Although this represents a different source of error from that which is the focus of the present paper (physical proximity rather than utility similarity), it provides support to the notion that attackers’ actions can indeed be suboptimal.

strengthen its protection and determine the probability that an attack on the element is successful [19, 2, 18, 17]. We let c_i denote the resources allocated to element i and let $p_i(c_i)$ denote the corresponding probability that an attack against i is successful. For every element, $p_i(c_i)$ is a continuous, positive, decreasing and convex, and hence invertible, function on the interval of feasible investments.

The preferences of the defender are represented as disutilities, where the defender prefers a lower disutility before a higher. These should be interpreted as total assessments of the different aspects of the event (e.g. number of casualties, economic loss, induced fear, etc.). The defender associates a successful attack on element i with disutility $d_i^s > 0$ and a failed attack with disutility 0. The defender's expected disutility of an attack against i is $d_i(c_i) = p_i(c_i)d_i^s$. A non-realized attack is associated with disutility $d_0 < 0$, i.e., the defender prefers to deter an attack completely rather than to cause a realized attack to fail.

An attacker observes the resource allocation and subsequently chooses one of the elements to attack or not to attack the system. The preferences of the attacker are represented as utilities, where higher utilities are preferred over lower. The attacker associates a successful attack on element i with utility $v_i^s > 0$ and a failed attack with utility 0. The attacker's expected utility of an attack on element i is $v_i(c_i) = p_i(c_i)v_i^s$. The non-attack alternative is associated with utility $v_0 > 0$, i.e., the attacker prefers not to attack over a failed attack. This represents the utility of the best possible alternative to attacking the system, which may be to do nothing, to attack another system, or to do something else.

In game-theoretical terms, an allocation of resources \mathbf{c} from the set $C = \{(c_1, c_2, \dots, c_n) \mid c_i \geq 0 \forall i, \sum_i c_i \leq c_{\text{total}}\}$ represents a strategy for the defender. Since the attacker observes the defender's resource allocation, a strategy $\mathbf{a}(\mathbf{c})$ for the attacker in the baseline model is a mapping from C to the set $A = \{(a_0, a_1, \dots, a_n) \mid a_i \in \{0, 1\} \forall i, \sum_i a_i = 1\}$. Thus, for every feasible resource allocation $\mathbf{c} \in C$, $\mathbf{a}(\mathbf{c})$ is a vector \mathbf{a} with element a_i equal to 1 for some i and remaining elements equal to 0. $a_0 = 1$ represents not attacking the system and $a_i = 1, i = 1, \dots, n$, represents attacking element i .

2.2 Analysis

A subgame-perfect equilibrium [7] of the game is a pair of strategies $\hat{\mathbf{c}}, \hat{\mathbf{a}}(\mathbf{c})$ such that $\hat{\mathbf{c}}$ minimizes the defender's expected disutility given that the attacker acts according to $\hat{\mathbf{a}}(\mathbf{c})$, and $\hat{\mathbf{a}}(\mathbf{c})$, for any $\mathbf{c} \in C$, maximizes the attacker's expected utility given \mathbf{c} . Thus, $\hat{\mathbf{c}}$ and $\hat{\mathbf{a}}(\mathbf{c})$ simultaneously satisfy

$$\hat{\mathbf{c}} \in \arg \min_{\mathbf{c}} d_0 \hat{a}_0(\mathbf{c}) + \sum_{j \geq 1} d_j(c_j) \hat{a}_j(\mathbf{c}) \quad (1)$$

$$\text{s.t. } \mathbf{c} \in C, \quad (2)$$

and

$$\hat{\mathbf{a}}(\mathbf{c}) \in \arg \max_{\mathbf{a}} v_0 a_0 + \sum_{j \geq 1} v_j(c_j) a_j \quad (3)$$

$$\text{s.t. } \mathbf{a} \in A. \quad (4)$$

For a given resource allocation \mathbf{c} , let $v_{\max}(\mathbf{c})$ denote the attacker's highest utility across the alternatives, i.e. $v_{\max}(\mathbf{c}) = \max\{v_0, v_1(c_1), \dots, v_n(c_n)\}$, and let $A_{\max}(\mathbf{c})$ denote the set of alternatives with utility $v_{\max}(\mathbf{c})$. Obviously, a rational attacker will only play strategies that involve choosing some alternative in $A_{\max}(\mathbf{c})$. Further, let $d_{\min\max}(\mathbf{c})$ be the defender's lowest disutility among the actions in $A_{\max}(\mathbf{c})$, and let $A_{\min\max}(\mathbf{c})$ be the set of alternatives in $A_{\max}(\mathbf{c})$ with disutility $d_{\min\max}(\mathbf{c})$. [16] shows that in any subgame-perfect equilibrium, the attacker will choose some alternative in $A_{\min\max}(\mathbf{c})$.

Since the attacker will only choose alternatives in $A_{\max}(\mathbf{c})$, increasing the investment on any element i not in $A_{\max}(\mathbf{c})$ will clearly have no effect. Since any investment c_i will reduce $d_i(c_i)$ as well as $v_i(c_i)$, it follows that the defender should allocate the resources in order to reduce $v_{\max}(\mathbf{c})$ as much as possible, i.e. to adopt a min-max strategy. [16] shows that in any subgame-perfect equilibrium, the defender will play the min-max strategy (which is unique). It is easily verified that none of the players has an incentive to change strategy in this case, so that the strategy pair is indeed an equilibrium.³

The baseline model highlights some important aspects of protection against a strategic attacker as opposed to random failures: Rather than investing in elements where the defense can be improved the most efficiently, the defender should invest in the elements that yield the highest utility for the attacker. This principle is robust with regard to the utilities v_i and disutilities d_i of the two actors and hinges only on the fact that an investment in element i will decrease both v_i and d_i (cf. [3, 10, 18]).

However, some of the implications of the model may be questionable and even hazardous to rely on in a real situation. For example, the model predicts that no element with utility less than $v_{\max}(\mathbf{c})$ will be attacked. Thus, given a limited defense budget, these elements should be left unprotected, even if an attack on some of these elements would cause very high disutility for the defender. Also, perfect perception implies that an investment will have no deterring effect on the attacker unless its effect is precisely to make v_0 the uniquely largest utility. Further, the model predicts that the expected disutility of the defender is non-increasing in the investment on any element, so that an additional investment, *ceteris paribus*, will not make the situation worse. All these properties stem from the fact that the attacker observes the utility of every alternative, given the defender's resource allocation, perfectly.

³If the attacker would choose some alternative in $A_{\max}(\mathbf{c})$ but not in $A_{\min\max}(\mathbf{c})$ (given that such an alternative exists), then the defender could reduce its disutility by redistributing resources from elements of less disutility and increase the investment in the attacked element. Since the defender could always achieve a smaller disutility by redistributing a smaller amount of resources, there is no equilibrium in which the attacker plays such a strategy (see [16] for details).

3 Model with imperfect observations

3.1 Formulation of the attacker's problem

We assume now that due to an imperfect ability to assess the defender's protective measures and the outcome of a successful attack, the attacker's observations of the elements are associated with errors. These errors can be associated with the success probabilities p_i as well as the success utilities v_i^s , although we only consider their combined effect. Instead of observing the true utility $v_i = p_i v_i^s$ for each element, the antagonist's observation or best guess based on its available information is u_i . The utility of not attacking, v_0 , is observed without error, since the characteristics of this alternative should be well known to the attacker.

To conform to the standard formulations of extensive-form games used in [16], the baseline model specifies that the attacker's actions are directly based on the defender's strategy, i.e. the resource allocation \mathbf{c} . In reality, it seems unlikely that the attacker would be able to observe much of the invested amounts directly. Rather, the attacker would perceive the manifestations of those investments in terms of the risks and potential benefits of executing an attack. Hence, in the following we assume that the attacker does not observe \mathbf{c} or c_{total} , but bases its strategies on (imperfect) observations of the utilities v_i .⁴ Since there is a one-to-one correspondence between c_i and the utility $v_i = p_i(c_i)v_i^s$ for every element i , the two formulations are equivalent in the baseline model.

The observation errors of the attacker are modeled as outcomes of random variables ε_i , so that the *observed* utility u_i is an outcome of the random variable

$$U_i(c_i) = v_i(c_i)\varepsilon_i, \quad i = 1, \dots, n. \quad (5)$$

We assume that every ε_i is continuous on some interval and strictly positive. The multiplicative error structure means that the variability of the observations increase with the true attack utility, so that large utilities are more difficult to perceive accurately than small. It also means that the observed expected attack utility can never be less than 0, the utility associated with a failed attack, which is reasonable requirement.⁵

Further, we assume that the attacker treats the observed utilities u_i as if they were true utilities, so that a best response for the attacker is to choose any alternative with maximum observed utility (cf. Section 2.2). For a given vector of observed utilities $\mathbf{u} = (u_1, \dots, u_n)$, the problem that the attacker solves is thus (cf. (3))

$$\hat{\mathbf{a}}(\mathbf{u}) \in \arg \max_{\mathbf{a}} v_0 a_0 + \sum_{j \geq 1} u_j a_j \quad (6)$$

$$\text{s.t. } \mathbf{a} \in A. \quad (7)$$

⁴We do not consider the possibility to use c_i as a signal for the true utility v_i (cf. [17]).

⁵Assuming an additive error structure instead would not affect the results from the analysis in any significant way (see also Section 3.3), nor would assuming that also v_0 is observed with error (cf. [21]).

3.2 Formulation of the defender's problem

We consider the situation when the defender knows the probability distributions of the attacker's observations $U_i = v_i \varepsilon_i$ but not the actual outcomes u_i (defender uncertainty is introduced in Section 6). To facilitate the analysis, we further assume that all ε_i are independent and identically distributed (i.i.d.). Hence, there are no systematic differences or correlations in the observation errors between different elements.

From the point of view of the defender, the *ex ante* probability that the attacker chooses action i is denoted $q_i(\mathbf{c})$. Thus, $q_0(\mathbf{c})$ is the probability that the system is not attacked, and $q_i(\mathbf{c})$, $i = 1, \dots, n$, is the probability that element i is attacked. We have⁶

$$q_0(\mathbf{c}) = \Pr[v_0 > U_j(c_j) \forall j] = \Pr\left[v_0 > \max_j U_j(c_j)\right], \quad (8)$$

and

$$\begin{aligned} q_i(\mathbf{c}) &= \Pr[U_i(c_i) > v_0, U_i(c_i) > U_j(c_j) \forall j \neq i] = \\ &= \Pr\left[\max_j U_j(c_j) > v_0, U_i(c_i) = \max_j U_j(c_j)\right], \quad i = 1, \dots, n. \end{aligned} \quad (9)$$

By the i.i.d. assumption, $q_i(\mathbf{c})$ can be decomposed as

$$\begin{aligned} q_i(\mathbf{c}) &= \Pr\left[\max_j U_j(c_j) > v_0\right] \cdot \Pr\left[U_i(c_i) = \max_j U_j(c_j)\right] = \\ &= (1 - q_0(\mathbf{c}))q_{i|A}(\mathbf{c}), \quad i = 1, \dots, n, \end{aligned} \quad (10)$$

where $q_{i|A}(\mathbf{c})$ denotes the probability that element i is attacked conditional on that the system is attacked.

As a reference, let us first consider the situation analyzed by [10] and [16] where the attacker does not have the non-attack alternative, i.e. conditional on that the system is attacked. In this case, the expected utility of the defender is

$$\bar{d}_A(\mathbf{c}) = \sum_{j \geq 1} d_j(c_j)q_{j|A}(\mathbf{c}). \quad (11)$$

With the non-attack alternative present, the defender's expected disutility is (cf. (1))

$$\begin{aligned} \bar{d}(\mathbf{c}) &= d_0 q_0(\mathbf{c}) + \sum_{j \geq 1} d_j(c_j)q_j(\mathbf{c}) = \\ &= d_0 + \sum_{j \geq 1} (d_j(c_j) - d_0)q_j(\mathbf{c}), \end{aligned} \quad (12)$$

⁶Since the observation errors ε_i are continuous random variables, the probability that two alternatives yield the same observed utility is zero, $\Pr[U_i(c_i) = U_j(c_j)] = \Pr[U_i(c_i) = v_0] = 0 \forall i, j \neq i$. As a result, we do not need to consider what the attacker's strategy would be in cases of multiple best responses (cf. Section 2.2).

and the defender's problem can be written as

$$\min_{\mathbf{c}} \bar{d}(\mathbf{c}) \quad (13)$$

$$\text{s.t. } \mathbf{c} \in C. \quad (14)$$

3.3 Distribution of observation errors

In the following analysis, we assume that every observation error ε_i is distributed according to the Fréchet distribution with shape parameter $\lambda \in (0, \infty)$. The cumulative distribution function of ε_i is thus

$$F_i(x) = \Pr[\varepsilon_i \leq x] = \exp(-x^{-\lambda}), \quad x \in (0, \infty), \quad i = 1, \dots, n. \quad (15)$$

The main advantage of adopting the Fréchet distribution is that attack probabilities can be expressed on closed form. This is unlike e.g. the Log-normal distribution, which is otherwise similar in shape. Since very little can be known about an attacker's observation errors in practice, we believe that the analytical merits of the Fréchet distribution justifies its use. The principal results we obtain should be valid for many other probability distributions with support on the positive real line.⁷

The variance of ε_i is strictly decreasing in the shape parameter λ , which is used throughout this paper to represent the perceptive ability of the attacker. Values close to 0 mean that the observation errors completely dominate the true utilities, so that the attacker chooses an action largely at random. Large values of λ , on the other hand, mean that the observed utilities are close to the true utilities, so that the actions of the attacker tend to those predicted from the baseline model. Figure 1 shows the density function $F'_i(x)$ of ε_i for different λ .

With the Fréchet distribution, it can be shown that, for given resource allocation \mathbf{c} and attacker's perception λ , the conditional attack probability of element i is

$$q_{i|A}(\mathbf{c}, \lambda) = \frac{v_i(c_i)^\lambda}{\sum_{j \geq 1} v_j(c_j)^\lambda}, \quad i = 1, \dots, n, \quad (16)$$

and that the non-attack probability is

$$q_0(\mathbf{c}, \lambda) = \exp\left(-\sum_{j \geq 1} \left(\frac{v_j(c_j)}{v_0}\right)^\lambda\right), \quad (17)$$

⁷Since the attack probabilities depend only on the relative differences between the observed utilities, we may take the logarithm of all utilities, which gives $\log U_i = \log v_i + \log \varepsilon_i$, $i = 1, \dots, n$. Thus, we get an equivalent additive model where it can be shown that the error term $\log \varepsilon_i$ is Gumbel distributed with scale parameter λ . This model formulation, known as a multinomial logit model, is commonly and successfully applied in many areas involving demand modeling, see e.g. [20].

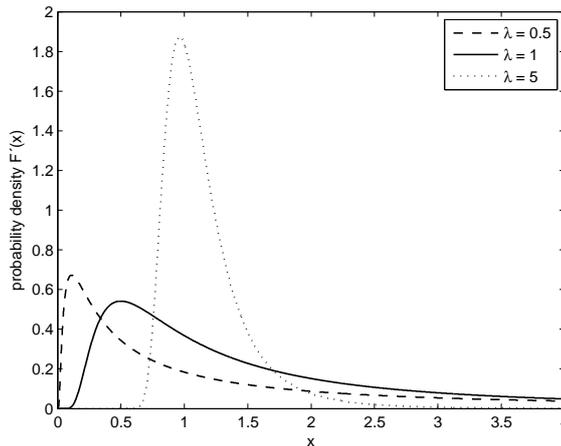


Figure 1: The probability density $F'(x)$ of the Fréchet distribution for different values of the scale parameter λ .

Hence, we obtain the unconditional probability q_i that element i is attacked as

$$q_i(\mathbf{c}, \lambda) = \left(1 - \exp \left(- \sum_{j \geq 1} \left(\frac{v_j(c_j)}{v_0} \right)^\lambda \right) \right) \frac{v_i(c_i)^\lambda}{\sum_{j \geq 1} v_j(c_j)^\lambda}, \quad i = 1, \dots, n. \quad (18)$$

One can verify that $v_i > v_j$ for any elements i, j implies that $q_i > q_j$, and that an increase in v_i implies that q_i increases, i.e. $\partial q_i / \partial v_i > 0$.

4 Properties of the model

4.1 Effects of attacker perception

4.1.1 Attack probabilities

We first examine the effects of the attacker's level of perception λ on attack probabilities. Given a small increase in the attacker's perception λ while keeping the resource allocation fixed, the change in the conditional attack probability $q_{i|A}$ of an arbitrary element i is

$$\frac{\partial q_{i|A}}{\partial \lambda} = \ln \left(\frac{v_i}{\bar{v}_A} \right) q_{i|A}, \quad i = 1, \dots, n, \quad (19)$$

where \bar{v}_A is the geometric mean utility conditional on an attack,

$$\bar{v}_A = \prod_{j \geq 1} v_j^{q_{j|A}}. \quad (20)$$

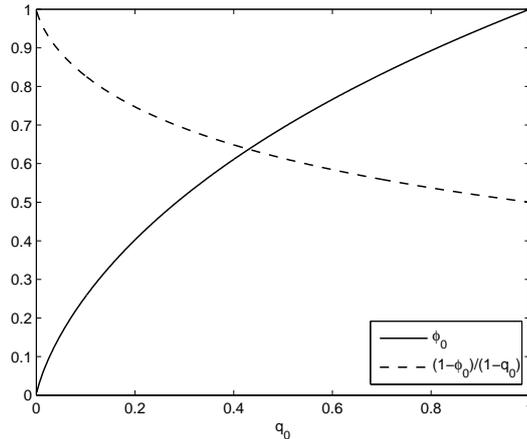


Figure 2: The deterrence factor ϕ_0 and the expression $(1 - \phi_0)/(1 - q_0)$ as functions of the non-attack probability q_0 .

Thus, the conditional attack probability will increase or decrease depending on whether the attack utility v_i is above or below the average \bar{v}_A . This represents the increased ability to perceive the relative utilities of the elements, so that an attack will with increasing probability be targeted towards the most attractive elements.

For small λ , $q_{i|A}$ will approach $1/n$ for every element so that the target will be chosen completely at random. For large λ , $q_{i|A}$ will approach $1/n_{\max}$ if v_i equals $v_{A\max} = \max_{j \geq 1} v_j$, where n_{\max} is the number of elements with utility $v_{A\max}$, or 0 otherwise. That is, the target will be chosen randomly among the elements with the largest utility.

Meanwhile, the change in the non-attack probability q_0 is

$$\frac{\partial q_0}{\partial \lambda} = \ln \left(\frac{v_0}{\bar{v}_A} \right) (1 - q_0) \phi_0, \quad (21)$$

where ϕ_0 , which we will call the *deterrence factor*, is defined as

$$\phi_0 = -\frac{q_0}{1 - q_0} \ln q_0. \quad (22)$$

The deterrence factor ϕ_0 serves as an adjustment of q_0 due to the fact that the utility of not attacking the system, v_0 , is observed without error unlike the other utilities. Figure 2 shows how ϕ_0 and $(1 - \phi_0)/(1 - q_0)$, which occurs in several formulas below, vary with q_0 . Note that ϕ_0 , through q_0 , is a function of the resource allocation \mathbf{c} .

Returning to (21), we see that the change in q_0 is positive or negative depending on whether the non-attack utility v_0 is larger or smaller than the geometric mean utility of attacking, \bar{v}_A . This represents the attacker's increased ability

to perceive the utility of attacking in relation to not attacking the system. For small λ , q_0 will approach $\exp(-n)$, which is small when n is large. For large λ , q_0 will tend to 1 if $v_0 > v_{\text{Amax}}$, to $\exp(-n_{\text{max}})$ if $v_0 = v_{\text{Amax}}$, or to 0 if $v_0 < v_{\text{Amax}}$.

All in all, the change in the unconditional attack probability q_i can be written as

$$\frac{\partial q_i}{\partial \lambda} = \left(\ln \left(\frac{v_i}{\bar{v}_A} \right) + \ln \left(\frac{\bar{v}_A}{v_0} \right) \phi_0 \right) q_i, \quad i = 1, \dots, n. \quad (23)$$

The attack probability q_i will certainly increase if v_i is larger than the average \bar{v}_A and \bar{v}_A is larger v_0 , and it will certainly decrease if the opposite relations both hold. When the two terms are of different signs, the relative magnitudes of the utilities determine whether q_i increases or decreases. Note that the average \bar{v}_A changes with λ as well, which makes the behavior of q_i over a larger range of λ non-trivial.

4.1.2 Defender's disutility

If we first consider the case without the non-attack alternative, the change in the defender's expected disutility \bar{d}_A given a small change in λ is

$$\frac{\partial \bar{d}_A}{\partial \lambda} = \sum_{j \geq 1} d_j \ln \left(\frac{v_j}{\bar{v}_A} \right) q_{j|A} \quad (24)$$

In the extreme case that $v_i = d_i$ for all elements, i.e. a zero-sum situation where the actors have completely opposite preferences, (24) gives that the expected disutility \bar{d}_A will always increase. This is intuitive, since a more perceptive attacker is more likely to attack the most attractive elements, which are also the most valuable for the defender. More generally, \bar{d}_A will increase if the defender's disutilities d_i and attacker's utilities v_i are sufficiently similar across the elements, so that d_i typically is large when $v_i > \bar{v}_A$ and small when $v_i < \bar{v}_A$.

With the non-attack alternative available, the change in the defender's expected disutility \bar{d} is

$$\frac{\partial \bar{d}}{\partial \lambda} = \sum_{j \geq 1} (d_j - d_0) \left(\ln \left(\frac{v_j}{\bar{v}_A} \right) + \ln \left(\frac{\bar{v}_A}{v_0} \right) \phi_0 \right) q_j. \quad (25)$$

The behavior is now more complex due to the introduction of v_0 . However, we can say that \bar{d} will increase if \bar{v}_A is sufficiently larger than v_0 , and/or if d_i and v_i are sufficiently similar across the elements. Correspondingly, \bar{d} will decrease if \bar{v}_A is sufficiently smaller than v_0 , so that the attacker is deterred, and/or if d_i and v_i are sufficiently dissimilar, so that the attack is diverted to less valuable elements.

The results show that a less perceptive attacker can cause higher disutility for the defender. This means that a perfectly perceptive attacker need not represent

a “worst-case” attack scenario against the system. Such an interpretation only holds if the valuation of the defender and the attacker are sufficiently similar.

4.2 Effects of defense resources

4.2.1 Attack probabilities

When the investment c_i in element i is increased while keeping all other resources and the attacker’s level of perception λ fixed, the conditional attack probability $q_{i|A}$ for i decreases. For every other element j , however, the conditional attack probability $q_{j|A}$ increases:

$$\begin{aligned}\frac{\partial q_{i|A}}{\partial c_i} &= \lambda \frac{p'_i}{p_i} q_{i|A} (1 - q_{i|A}) < 0, \\ \frac{\partial q_{j|A}}{\partial c_i} &= -\lambda \frac{p'_i}{p_i} q_{i|A} q_{j|A} > 0, \\ & i, j = 1, \dots, n, \quad i \neq j, \quad (26)\end{aligned}$$

This represents the substitution effect among the elements of increasing the protection [19, 21, 14, 9]. It can be seen that the strength of this effect depends on the attacker’s perception λ , the relative marginal effectiveness of the investment in reducing the success probability, p'_i/p_i , and the current conditional attack probabilities $q_{i|A}$ and $q_{j|A}$.

For the system as a whole there is a deterrence effect, since the probability that the system is not attacked, q_0 , increases:

$$\frac{\partial q_0}{\partial c_i} = -\lambda \frac{p'_i}{p_i} q_i \phi_0 > 0, \quad i = 1, \dots, n, \quad (27)$$

When increasing c_i , both substitution and deterrence work in favor of reducing the unconditional probability that element i is attacked. For any other element j , the substitution effect will to some extent be counterbalanced by the deterrence effect. Overall, however, the probability that j is attacked increases,

$$\begin{aligned}\frac{\partial q_i}{\partial c_i} &= \lambda \frac{p'_i}{p_i} q_i \left(1 - \frac{1 - \phi_0}{1 - q_0} q_i\right) < 0, \\ \frac{\partial q_j}{\partial c_i} &= -\lambda \frac{p'_i}{p_i} q_i q_j \frac{1 - \phi_0}{1 - q_0} > 0, \\ & i, j = 1, \dots, n, \quad i \neq j. \quad (28)\end{aligned}$$

The deterrence and substitution effects of an investment are consequences of the attacker’s observation errors; in the baseline model, a small investment will have no effect on the attacker’s actions unless it changes which alternative yields the highest utility.

4.2.2 Defender's disutility

Without the non-attack alternative, the effect on the expected disutility \bar{d}_A of a small increase of c_i is

$$\frac{\partial \bar{d}_A}{\partial c_i} = \frac{p'_i}{p_i} (d_i + \lambda (d_i - \bar{d}_A)) q_{i|A},$$

$$i = 1, \dots, n. \quad (29)$$

For sufficiently low values of λ , the outer parenthesis will be dominated by d_i , which is always positive so that \bar{d}_A will decrease. The decrease will be the largest at elements where the marginal effectiveness of defense investments is the highest, a well-known result for random attacks (cf. [3, 10]).

For large λ , the term $d_i - \bar{d}_A$ will dominate the parenthesis, and \bar{d}_A will decrease or increase depending on whether d_i is larger or smaller than \bar{d}_A . Since \bar{d}_A is a weighted average of all d_i 's, we know that for any λ , increasing the defense resources will decrease the expected disutility \bar{d}_A at least for elements with the largest utility. This, in turn, means that it is always optimal to allocate all the available resources c_{total} .

On the other hand, we know that there in general are some elements for which $d_i < \bar{d}_A$, so that for sufficiently large λ an increased investment in such an element will increase the expected disutility. Hence, spending more resources on an element can actually make the situation worse, and the defender must be careful when allocating the defense.

With the non-attack alternative present, the effect on the expected disutility \bar{d} of a small increase of c_i is

$$\frac{\partial \bar{d}}{\partial c_i} = \frac{p'_i}{p_i} (d_i + \lambda (d_i - d_0 - (1 - \phi_0) (\bar{d}_A - d_0))) q_i,$$

$$i = 1, \dots, n. \quad (30)$$

Compared with (29), the main addition is the introduction of the deterrence factor ϕ_0 , which works in favor of reducing the expected disutility. Thus, it still holds that there is at least one element where an investment will reduce \bar{d} , but there may be other elements for which \bar{d} will increase.

4.3 Optimal resource allocation

The first-order necessary optimality conditions for the defender's problem (13) require that the marginal change in expected disutility (30) should be negative and equal for any elements with investments $c_i > 0$, and greater or equal to this for elements with zero investments.

Since $\bar{d}(\mathbf{c}, \lambda)$ for a given λ is a continuous function on the closed, bounded domain defined by C , the defender's problem has an optimal solution. It can be shown that $\bar{d}(\mathbf{c}, \lambda)$ is convex for sufficiently small λ , so that there is a unique optimal point. For larger λ , $\bar{d}(\mathbf{c}, \lambda)$ is not convex and a local minimum need not

be a global minimum. Numerical experiments that we have performed, however, suggest that it is common with only one local minimum within C (always on the boundary), which must then be the unique optimal point.

To find the optimal resource allocation when $\bar{d}(\mathbf{c}, \lambda)$ is non-convex we may use the algorithm presented by [8]. In short, the algorithm involves following a path of resource allocations from the global minimum of a convex problem with small λ to the global minimum, if such exists, of the non-convex problem with the actual λ .

5 Numerical example

5.1 Setting

In the following, some of the general results from Section 4 are illustrated with a small example system consisting of $n = 3$ elements. We study a case where the defender's and the attacker's valuations of the elements differ (i.e., not a zero-sum situation). More precisely, we assume that $d_1^s = 0.2$, $d_2^s = 0.45$ and $d_3^s = 1$, while $v_1^s = 1$, $v_2^s = 0.45$ and $v_3^s = 0.2$. Their valuations of the non-attack alternative are $d_0 = -0.3$ and $v_0 = 0.3$, respectively.

Further, we assume that the elements are equally easy to protect, i.e. that $p_1(c) = p_2(c) = p_3(c)$ for any c . This is done so that conclusions can be drawn more easily, but the calculations would not become more complicated if the p_i 's were different. Specifically, we assume that

$$p_i(c_i) = \frac{1}{1 + c_i}, \quad i = 1, 2, 3. \quad (31)$$

Hence, for any c , $i < j$ implies that $d_i(c) < d_j(c)$, $d'_i(c) > d'_j(c)$ and $v_i(c) > v_j(c)$. This functional form was chosen mainly for its simplicity; the principal results do not change much when other forms (e.g., $p_i(c_i) = \exp(-\alpha c_i)$) are used.

5.2 Influence of attacker's perception

First we study how the optimal resource allocation varies with the antagonist's level of perception λ . We assume here that the size of the defense budget is $c_{\text{total}} = 1$. The results are presented in Figure 3. To compute optimal resource allocations for a range of λ , we used the optimal point from the last λ as starting point for the next, slightly larger λ .

For small λ , it is optimal to spend resources on elements where the defender's local disutility can be reduced the most efficiently, since every element has about the same probability of being attacked. In this case, the budget permits that both element 2 and 3 are defended, but no resources are spent on element 1.

As λ increases, the attack probabilities are shifted towards the elements with the largest attack utility, as shown to the right in Figure 3. It therefore becomes increasingly important to reduce the largest attack utility, even if this may be expensive. For intermediate λ , the behavior of the optimal allocation is quite

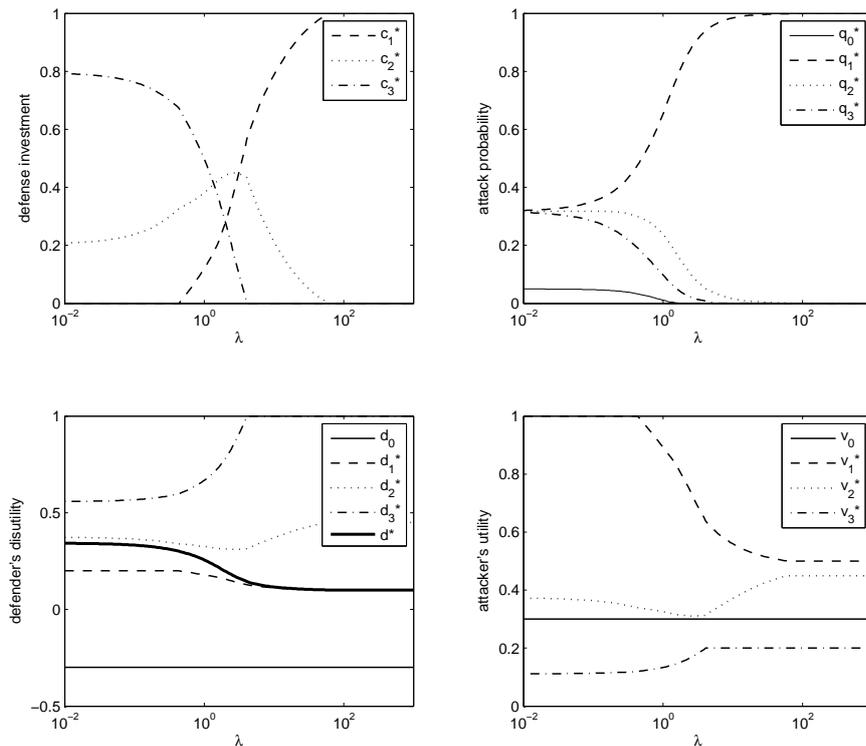


Figure 3: Numerical example (Section 5.2): Optimal resource allocation for different levels of attacker perception λ (shown on logarithmic scale). In the legends, stars indicate values evaluated at the optimal resource allocation. Note that the optimal resource allocation varies greatly with λ (top left), and that the expected disutility decreases with λ (bottom left, thick line).

complex with the investment on element 2 first increasing and then decreasing; in a small interval, element 2 should get the largest investment. For large λ , the priority order among the elements is reversed: Element 1, which has the highest success utility for the attacker, receives all resources while both element 2 and 3 are left undefended. This is also the optimal resource allocation in the baseline model. Note that the budget is not large enough to reduce the attack utilities below the non-attack utility $v_0 = 0.3$, so the non-attack probability q_0 tends to zero.

As shown to the bottom left, the defender's expected disutility \bar{d} (thick line) decreases with λ . This is because the defender and the attacker value the elements in reversed orders, so that a perceptive attacker attacks elements causing less disutility for the defender (cf. (25)).

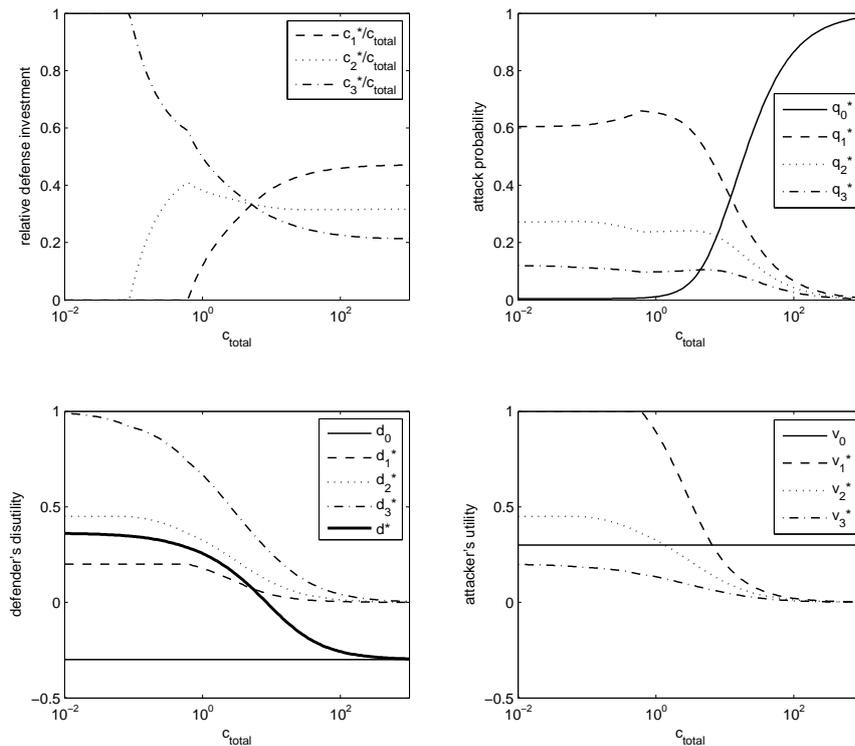


Figure 4: Numerical example (Section 5.3): Optimal resource allocation for different defense budgets c_{total} (shown on logarithmic scale). In the legends, stars indicate values evaluated at the optimal resource allocation. Note that the relative resource distribution varies greatly with c_{total} (top left), and that the deterrence effect becomes significant for large c_{total} (top/bottom right).

5.3 Influence of resource budget

Figure 4 shows how the optimal resource allocation depends on the size of the defense budget c_{total} . We assume here that $\lambda = 1$ which, in relation to the scale of the attack utilities, represents a relatively unperceptive antagonist. Note that the point $\lambda = 1$, $c_{\text{total}} = 1$ is represented in both Figure 3 and 4.

Since the attacker's perception is relatively low, the optimal resource allocation depends much on how efficiently the disutility can be reduced for each element. For small budgets it is optimal to spend all resources on element 3, which also receives the largest investment for completely random attacks as shown in Figure 3.

As the budget increases, it becomes affordable and worthwhile to distribute resources to the other elements as well, due in part to the diminishing return to investment and in part to the non-randomness of the attacker. For intermediate-sized budgets (c_{total} around 1), the figure reveals that the expected disutility

\bar{d} decreases more due to the reduction in elementwise disutility than the deterrence effect, since the non-attack probability q_0 remains low. For large budgets, however, the defender is able to reduce all attack utilities well below the non-attack utility v_0 . The attacker perceives this and the non-attack probability q_0 increases towards 1. As a result, the defender’s expected disutility \bar{d} tends to the non-attack disutility $d_0 = -0.3$.

6 Representing defender uncertainty

In order not to obscure the main ideas of the paper, we have assumed that the defender has complete information about the attacker’s characteristics, except for the actual outcomes u_i of the attacker’s observed utilities. It is straightforward to extend the analysis to situations where the defender is uncertain about the preferences and/or the perceptive ability of the attacker. Uncertainty about preferences can stem from limited knowledge about the attacker’s motives and available resources, while uncertainty about the perceptive ability can stem from limited knowledge about the attacker’s available information, e.g. whether or not the attacker has access to insider information about the technical system.

If the defender’s uncertainty can be represented as a probability distribution across possible attacker characteristics or “types”, as is done in [18, 2], we can model the situation as a Bayesian game (see e.g. [7]). In this representation, the game between the defender and the attacker is preceded by a draw by nature in which the type of the attacker is determined. The outcome of this draw is observed by the attacker but not by the defender.

In our setting, an attacker type θ corresponds to a joint specification of a non-attack utility $v_{0,\theta}$, a success utility $v_{i,\theta}^s$ for every element i and a level of perception λ_θ . For a given attacker type θ and resource allocation \mathbf{c} , the probability that the attacker chooses alternative i , $q_i(\mathbf{c}, \theta)$, and the expected disutility of the defender, $\bar{d}(\mathbf{c}, \theta)$, are given by equations (17), (18) and (12) respectively as before.

As in [18, 2] we assume here that the set of possible attacker types is finite. The probability that the attacker is of type θ is then denoted π_θ and it holds that $\sum_\theta \pi_\theta = 1$. Considering the defender’s uncertainty regarding the attacker type, the overall expected disutility is then (cf. (12))

$$\bar{d}(\mathbf{c}) = \sum_\theta \pi_\theta \bar{d}(\mathbf{c}, \theta) = d_0 + \sum_{j \geq 1} (d_j(c_j) - d_0) \sum_\theta \pi_\theta q_j(\mathbf{c}, \theta). \quad (32)$$

We see that the consideration of uncertainty regarding attacker type involves taking the weighted arithmetic mean across the attacker types of each attack probability. The results for the case with a continuous distribution of attacker types are analogous.⁸

⁸In Section 3.3 we noted that the proposed attack probability model for a given attacker type is equivalent to a multinomial logit model. Introducing randomness in the parameters makes the model equivalent to a mixed logit model, which is also commonly used in demand modeling (e.g. [20]).

7 Conclusion

In the paper, we have considered the problem of allocating finite resources for protecting a system against antagonistic attacks when the attacker's observations are imperfect. The rationale of this analysis is that factors such as secrecy, opacity, surveillance and remoteness make it unlikely that the attacker is capable of predicting its true risks and benefits of attacking different elements of the system. Hence, a protection strategy that is based on the assumption of perfect observations could be ineffective and, in the worst cases, even counter-productive.

The proposed modelling framework extends a previous game-theoretic model with complete information by introducing random observation errors on behalf of the attacker. With the model, we have found that the optimal allocation of resources can vary significantly depending on the attacker's perceptive ability. Further, we have showed that spending more resources on an element is not necessarily better, since this may redirect the attack to more critical elements causing the expected disutility for the overall system to rise. However, if the defense is distributed optimally among elements, it is optimal to spend all defense resources; in other words, the optimal expected disutility is a decreasing function of the defense budget.

The model incorporates both a substitution effect, in that increasing the protection of one element will make attacks on the other elements more likely, and a deterrence effect, in that increasing the defense will make the system as a whole less likely to be attacked, possibly making some other system a more likely target. Thus, when conducting modeling for security investments, just as always when we study the critical infrastructures, the systems definition and delimitation becomes very important. For public decision-making, the system might be all the infrastructures in society, and the issue of externalities then has implications for how we distribute defense resources between different infrastructures. Private infrastructure operators can also be forced to compete in the security arms race. Therefore, as has been pointed out earlier in e.g. [2], the issue of security investments is a mix between a public and private goods, which will involve both decentralized and centralized decision-making on the resource allocation.

Naturally, all defensive measures need not be focused on protecting the elements in the system. On the contrary, many authors emphasize the need for other policy options. In many situations it might be suitable to spend more resources on intelligence and defense resources that are less local. [21] stress the need for fighting terrorism at its source, and [6] argue that it might be more effective to try to increase the opportunity cost of terrorism, making it a less attractive approach. In our model this would correspond to uniformly reducing the attack utility of every element, or equivalently increasing the utility of not attacking.

We have deliberately kept the model of the antagonist's actions simple to avoid more technical discussions and instead focus on the generic aspects of this modeling approach. There are more general choice models that can be worth

studying at a latter stage. For example, it is assumed here that the antagonist attacks at most one element of the system. This can be generalised so that the choice set of the antagonist consists of groups of elements of any size. In that case it may be assumed that the attacker's utility estimates for overlapping groups are correlated, and a more refined choice model may be appropriate. Another area for further research is to combine the present approach with that of [22], where the actions of the antagonist consists of continuous levels of attack effort for each target.

In this paper we have not attempted to explicitly formulate the utilities and disutilities of the attacker and defender, respectively. In practice, a possible approach to this problem would be to use a combination of expert judgments and what empirical data that exists. In a future paper, we intend to investigate the consequences of optimizing the defense against the "wrong" set of beliefs. An earlier study of this kind in [10] showed that the consequences of such mistakes can be substantial. Examples to analyze would include underestimating the level of knowledge that the antagonist has about the system and misjudging his willingness to take risks in order to reach his objectives.

Acknowledgements

The paper has benefited greatly from discussions with Lars-Göran Mattsson. The first author wishes to thank the Swedish Governmental Agency for Innovation Systems (Vinnova) for the financial support.

References

- [1] W. I. Al Mannai and T. G. Lewis. A general defender-attacker risk model for networks. *Journal of Risk Finance*, 9(3):244–261, 2008.
- [2] Vicki Bier, Santiago Oliveros, and Larry Samuelson. Choosing what to protect: Strategic defensive allocation against an unknown attacker. *Journal of Public Economic Theory*, 9(4):563–587, 2007.
- [3] Vicki M. Bier, Aniruddha Nagaraj, and Vinod Abhichandani. Protection of simple series and parallel systems with components of different values. *Reliability Engineering & System Safety*, 87(3):315–323, 2005.
- [4] Gerald Brown, Matthew Carlyle, Javier Salmerón, and Kevin Wood. Defending critical infrastructure. *Interfaces*, 36:530–544, 2006.
- [5] Richard L. Church and Maria Paola Scaparra. Protecting critical assets: The r -interdiction median problem with fortification. *Geographical Analysis*, 39:129–146, 2007.
- [6] Bruno S. Frey and Simon Luechinger. How to fight terrorism: Alternatives to deterrence. *Defense and Peace Economics*, 14(4):237–249, 2003.

- [7] Drew Fudenberg and Jean Tirole. *Game Theory*. MIT Press, 1991.
- [8] Ward Hanson and Kipp Martin. Optimizing multinomial logit profit functions. *Management Science*, 42(7):992–1003, 1996.
- [9] Kjell Hausken. Income, interdependence, and substitution effects affecting incentives for security investments. *Journal of Accounting and Public Policy*, 25:629–665, 2006.
- [10] Åke J. Holmgren, Erik Jenelius, and Jonas Westin. Evaluating strategies for defending electric power networks against antagonistic attacks. *IEEE Transactions on Power Systems*, 22(1):76–84, 2007.
- [11] Åke J. Holmgren and Torbjörn Thedéen. Risk analysis. In Göran Grimvall, Åke J. Holmgren, P. Jacobsson, and T. Thedéen, editors, *Risk in Technological Systems*. Springer, 2009. Forthcoming.
- [12] Ferenc Jordán. Predicting target selection by terrorists: A network analysis of the 2005 London underground attacks. *International Journal of Critical Infrastructures*, 4(1–2):206–214, 2008.
- [13] H. Kunreuther and G. Heal. Interdependent security. *Journal of Risk and Uncertainty*, 26:231–249, 2003.
- [14] Darius Lakdawalla and George Zanjani. Insurance, self-protection, and the economics of terrorism. *Journal of Public Economics*, 89:1891–1905, 2005.
- [15] William M. Landes. An economic study of U.S. aircraft highjacking, 1961–1975. *Journal of Law and Economics*, 21(1):1–31, 1978.
- [16] Robert Powell. Defending against strategic terrorists over the long run. Manuscript, Department of Political Science, U.C. Berkeley, 2006.
- [17] Robert Powell. Allocating defensive resources with private information about vulnerability. *American Political Science Review*, 101(4):799–809, 2007.
- [18] Robert Powell. Defending against terrorist attacks with limited resources. *American Political Science Review*, 101(3):527–541, 2007.
- [19] Todd Sandler and Harvey E. Lapan. The calculus of dissent: An analysis of terrorists’ choice of targets. *Synthese*, 76:245–261, 1988.
- [20] Kenneth Train. *Discrete Choice Methods With Simulation*. Cambridge University Press, 2003.
- [21] Manuel Trajtenberg. Defense R&D in the anti-terrorist era. *Defense and Peace Economics*, 17(3):177–199, 2006.
- [22] Jun Zhuang and Vicki M. Bier. Balancing terrorism and natural disasters – defensive strategy with endogenous attacker effort. *Operations Research*, 55(5):976–991, 2007.