Managing the Weakest Link

A Game-Theoretic Approach for the Mitigation of Insider Threats

Aron Laszka[ab], Benjamin Johnson[ca], Pascal Schöttle[ad], Jens Grossklags[a], and Rainer Böhme[d]

[a]College of Information Sciences and Technology, Pennsylvania State University, USA
[b]Department of Networked Systems and Services, Budapest University of Technology and Economics, Hungary
[c]Department of Mathematics, University of California, Berkeley, USA
[d]Department of Information Systems, University of Münster, Germany

Abstract. We introduce a two-player stochastic game for modeling secure team selection to add resilience against insider threats. A project manager, Alice, has a secret she wants to protect but must share with a team of individuals selected from within her organization; while an adversary, Eve, wants to learn this secret by bribing one potential team member. Eve does not know which individuals will be chosen by Alice, but both players have information about the bribeability of each potential team member. Specifically, the amount required to successfully bribe each such individual is given by a random variable with a known distribution but an unknown realization.

We characterize best-response strategies for both players, and give necessary conditions for determining the game's equilibria. We find that Alice's best strategy involves minimizing the information available to Eve about the team composition. In particular, she should select each potential team member with a non-zero probability, unless she has a perfectly secure strategy. In the special case where the bribeability of each employee is given by a uniformly-distributed random variable, the equilibria can be divided into two outcomes – either Alice is perfectly secure, or her protection is based only on the randomness of her selection.

Keywords: Insider Threats, Cyberespionage, Game Theory, Computer Security, Access Control

1 Introduction

Providing effective access control in organizations has been refered to as the "traditional center of gravity of computer security" since it is a melting pot for human factors, systems engineering and formal computer science approaches [1]. Over the last decades, a large number of important contributions have been made to address various technical challenges to the problem of access control for important systems and sensitive data [18, 19].

This body of research is motivated in equal parts by the threat of malicious attackers from the outside and potential abuse by legitimate system users. Anderson further distinguishes between those situations in which insiders exploit

technical vulnerabilities of a system in opportunistic ways, and other situations in which employees abuse the trust placed in them [1]. In our work, we address the latter dimension of the problem space.

Data theft by trusted employees covers a significant share of insider attacks. For example, a CERT investigation of 23 attacks showed that "in 78% of the incidents, the insiders were authorized users with active computer accounts at the time of the incident. In 43% of the cases, the insider used his or her own username and password to carry out the incident" [16].

These attacks are occasionally attributed to disgruntled employees and are said to be primarily destructive in nature. However, the steady rise of cyber-espionage activities strongly motivates the threat scenario of employees stealing information for monetary rewards. A recent article summarized publicly-known United States legal data from the past four years and stated that "nearly 100 individual or corporate defendants have been charged by the Justice Department with stealing trade secrets or classified information" [10]. The article just considered theft benefiting one particular foreign nation. Therefore, it is reasonable to assume that the data merely represents the tip of the proverbial iceberg.

Turning a trusted employee into a spy provides a number of benefits for an outside attacker. First, a security compromise by an insider might not be discoverable in comparison to external network-based attacks that might leave traces identifiable for expert forensics teams. The result is that a corporation cannot adequately plan and respond to evidence of a stolen trade secret. Second, an insider can point the attacker towards particularly valuable secrets by identifying the so-to-speak needle in the haystack. Given the accelerating data growth within corporations it makes sense to assume that attackers are also suffering from information overload as a result of their successful but unguided network penetrations. Third, an insider can help the attacker interpret the stolen data through complementary communications that do not have to take place at the work location. Lastly, having an insider conduct the attack might be the only feasibly way for an attacker to circumvent the defenses of particularly well-defended targets such as military and intelligence services, i.e., the attacker makes use of the human as the weakest link.

In this paper, we develop a formal model in which an attacker sidesteps technical security mechanisms by offering a bribe to one member of a project team who works with sensitive data or business secrets. By applying game-theoretic tools, we derive optimal strategies for the defender and attacker, respectively, and provide numerical results to illustrate and explain our findings.

With our work, we intend to start a discussion about considering the composition of project teams as a formal and critical dimension of a comprehensive corporate security policy.

The remainder of the paper is structured as follows: Section 2 provides the background for our research and considers related work. In Section 3, we define the basic properties of our model. The conditions for Nash equilibria are given in Section 4. Section 5 instantiates our model with explicit distributions, and

numerical illustrations of the derived solutions are given in Section 6. We discuss our results and provide concluding remarks in Section 7.

2 Background and Related Work

2.1 Studies on Insider Threats and Cyber-espionage

Over the last several years, much research has been published in the area of insider threats, using different models and loss figures. For example, Carnegie Mellon University's CERT has published several reports concerning the field of insider threats, and industrial and economic espionage. Their 2011 report identifies two different models of espionage [13]. Motivating for our scenario is the so-called *Ambitious Leader Model*, where a leader (either from the inside or the outside of the organization), tries to convince (other) employees to follow her and to divulge secrets. In an earlier work, the institute identified several indicators that preceded either industrial espionage or sabotage, and thus could give hints if an employee might be vulnerable to being bribed [3]. In our research, we do not explicitly model behavioral and motivational factors that influence the trustworthiness of an employee. Instead, we assume that the defender has an indicator available to measure the *level of trustworthiness*.

The awareness of this threat is represented, for example, by a brochure published by the Federal Bureau of Investigation (FBI) [8], that lists:

> "A domestic or foreign business competitor or foreign government intent on illegally acquiring a company's proprietary information and trade secrets may wish to place a spy into a company in order to gain access to non-public information. *Alternatively, they may try to recruit an existing employee to do the same thing.*"

Additionally, the FBI "estimates that every year billions of U.S. dollars are lost to foreign and domestic competitors who deliberately target economic intelligence in flourishing U.S. industries and technologies [9]." The FBI further lists the following recommended activities for organizations: "Implement a proactive plan for safeguarding trade secrets, and confine intellectual knowledge on a need-to-know basis [9]."

Another example from Germany includes a 2012 report which identifies the loss for the German industry caused by industrial espionage to be around 4.2 billion € [6]. In this study, over 70% of these losses were caused by members of their own organization, through a combination of giving away intellectual property (47.8%) and failing to disclose their knowledge due to social factors (22.7%). Note that these numbers might be unreliable and interest-driven, as highlighted in [2].

2.2 Related Work

This paper touches several different research areas. The struggle between hiders of information and seekers of information is ubiquitous in the study of steganography, the field from which our idea originated [11]. This inspiration arose from

exploring the plight of a steganographer who wishes to hide k bits in a binary cover sequence of length n, and a steganalyst who wishes to detect whether the sequence has been modified. That model differs significantly from our model here, as the authors assume an equal a priori probability of modified and unmodified sequences, and the function that measures the predictability of sequence positions is part of the model as a parameter.

Another area that is directly connected to the situation we model is the organization of firms under weak intellectual property rights. For example, in [17], the author considers a situation in which a monopolist may distribute intellectual property across two employees. There is also a competitor who might hire one of these two to gain access to the intellectual property. The author models this situation as a leader–follower game, and derives equilibria.

There are many additional research directions covering the subject of insider threats, including deterrence theory [7], game theory [12] and trust models [5], which are all tangent to our model. But, to the best of our knowledge, none of the published models gives directions for a project manager on how to staff a team, that has to know a specific intellectual property, while being aware that an attacker might try to bribe one of his personnel.

3 Model Definition

In this section, we describe a two-player, non-zero-sum, non-deterministic game which models the team composition scenario. First, we describe the general context and environment of the game. Next, we introduce the game's players. Then we define these players' pure strategies, and the payoffs resulting from these simple choices. Finally, we introduce notation to represent mixed strategies and express the players' expected payoffs in terms of this notation.

3.1 Environment

In our model, an organization with a secret of high value has N employees who are qualified to operate on projects that require knowing the secret. The organization must share the secret with at least k employees in order to operate. The employees have varying levels of trustworthiness. For a given employee i, this trustworthiness level is given by a random variable T_i whose distribution $\mathcal{T}_i$ is known. We explicitly disregard other constraints on team building and assume that all aspects of the trustworthiness of an employee can be captured by the random variable T_i. If $T_i = t_i$, then employee i will reveal her known secrets whenever she is bribed by an amount at least t_i, but she will not reveal the secret if she is bribed by an amount less than t_i. We use the standard notation

$$F_{T_i}(b) = \Pr[T_i \leq b] \tag{1}$$

to denote the probability that the trustworthiness level of employee i is at most b.

3.2 Players

The players in our game are Alice and Eve. Alice is an organization's project manager who is responsible for selecting a team of qualified employees to work on a confidential project. The project requires each team member to know a secret of the organization, and this secret has a value S. Alice needs to share this secret with k of her N qualified employees. Eve is a spy from either inside or outside of the organization. Eve wants to know the secret and has the resources to bribe or eavesdrop on one of Alice's employees. If Eve eavesdrops, the trustworthiness level of an employee can be interpreted as a measure of difficulty for Eve to eavesdrop on that employee. Note that Eve does not know which employees are on the team.

3.3 Strategy Sets

Alice's pure strategy choice is to select a subset of her N employees with whom to share the secret. Formally, she chooses a size-k subset I of $\{1, \ldots, N\}$.

Eve's pure strategy choice is to select one employee and an amount to bribe. Formally, she chooses a pair (i, b) consisting of an index $i \in \{1, \ldots, N\}$ and a bribe value $b \in \mathbb{R}_{\geq 0}$.

3.4 Payoffs

Suppose that Alice plays a pure strategy I, and Eve plays a pure strategy (i, b). If $i \in I$ and $T_i \leq b$, then Eve wins the value of the secret minus the amount of the bribe, and Alice loses the value of the secret. In all other cases, Eve loses the amount of the bribe, and Alice loses nothing.

Table 1. Payoffs for Alice and Eve for the strategy profile $I, (i, b)$

Strategy profile and outcome	Payoff for Alice	Payoff for Eve
$i \in I$ and $T_i \leq b$	$-S$	$S - b$
$i \notin I$ or $T_i > b$	0	$-b$

3.5 Representation of Mixed Strategies

A mixed strategy is a distribution over pure strategies. For Alice, the canonical representation of her mixed strategy space is a finite probability distribution on the set of size-k subsets of $\{1, \ldots, N\}$. For Eve, the canonical representation of her mixed strategy space is a continuous probability distribution over the set $\{1, \ldots, N\} \times \mathbb{R}_{\geq 0}$. Because of the structure of the game, the payoff for both players is determined by simpler representations of the strategy spaces than the canonical ones, and we proceed to describe these representations next.

Mixed Strategy for Alice In the canonical representation of Alice's mixed strategy, we would let a_I denote the probability that she recruits the members of the size-k set I into the project team. However, since Eve can bribe only one employee, the payoff for any mixed strategy depends only on the probabilities of sharing the secret with each employee. Since several different mixed strategies might induce the same projection onto employee probabilities, we gain simplicity by restricting our attention to these projections.

By overloading notation, for each $i = 1, \ldots, N$, we let a_i denote the probability that Alice shares the secret with employee i. Formally,

$$a_i = \sum_{I : i \in I} a_I. \tag{2}$$

The requirement that Alice has to share the secret with k employees induces the notational constraint

$$\sum_{i=1}^{N} a_i = k. \tag{3}$$

Furthermore, it can be shown easily that, for any sequence $\langle a_i \rangle$ of N probabilities whose sum is k, there exists a mixed strategy for Alice whose projection is $\langle a_i \rangle$. Consequently, we will represent Alice's mixed strategies by such sequences for the remainder of this paper.

Mixed Strategy for Eve To represent Eve's mixed strategies, which are distributions over the set $\{1, \ldots, N\} \times \mathbb{R}_{\geq 0}$, we introduce two random variables, Y and B. Random variable Y takes values in $\{1, \ldots, N\}$, and it represents which employee Eve has chosen to bribe. Random variable B takes values in $\mathbb{R}_{\geq 0}$, and represents the amount of the bribe.

Overloading notation in a way that is similar to what we did for Alice, for each $i = 1, \ldots, N$, we define e_i to be the probability that Eve bribes employee i, so that we have

$$e_i = \Pr[Y = i]. \tag{4}$$

Since Eve always chooses exactly one employee, we have

$$\sum_{i=1}^{N} e_i = 1. \tag{5}$$

To describe a distribution over bribes, we sometimes use the notation

$$F_B(b) = \Pr[B \leq b], \tag{6}$$

which gives the probability that the value of the bribe chosen by Eve is at most b. It is also useful to describe the conditional distributions over bribes focused on a particular employee i. For each $i = 1, \ldots, N$, let B_i be the random variable whose range is the set of all possible bribes to player i, and whose distribution $\mathcal{B}_i$ is defined by

$$F_{B_i}(b) = \Pr[B_i \leq b] = \Pr[B \leq b | Y = i]. \tag{7}$$

In what follows, we will represent Eve's mixed strategies as pairs $(\langle e_i \rangle, \langle \mathcal{B}_i \rangle)$, where each e_i is the probability that Eve bribes the employee i, and each $\mathcal{B}_i$ is a distribution over bribe values, conditioned on the assumption that Eve chooses to bribe employee i.

3.6 Payoffs for Mixed Strategies

In order to use the simplified mixed-strategy representation defined above, we have to express the players' expected payoffs in terms of these representations. If Alice plays a mixed strategy represented by $\langle a_i \rangle$ and Eve plays a mixed strategy represented by $(\langle e_i \rangle, \langle \mathcal{B}_i \rangle)$, then the expected payoff for Alice is

$$-S \cdot \sum_{i=1}^{N} a_i \cdot e_i \cdot \Pr[T_i \leq B_i] \tag{8}$$

and the expected payoff for Eve is

$$S \cdot \sum_{i=1}^{N} (a_i \cdot e_i \cdot \Pr[T_i \leq B_i]) - \sum_{i=1}^{N} e_i \cdot E[B_i], \tag{9}$$

where $E[B_i]$ denotes the expected value of B_i under the distribution $\mathcal{B}_i$.

4 Analytical Results

Our goal in this section is to derive analytical results on the structure of the Nash equilibria of the game. We begin by characterizing Alice's and Eve's best-response strategies. Then, we use these characterizations to constrain Alice's and Eve's strategies in an equilibrium. Finally, based on these constraints, we formulate an algorithm for computing an equilibrium.

4.1 Best-Response Strategies

Alice's Best Response For a fixed strategy of Eve, Alice's best response minimizes the probability of the secret being compromised. Since the probability of employee i being targeted and successfully bribed is $e_i \cdot \Pr[T_i < B_i]$, Alice has to choose a set I of k employees to minimize $\sum_{i \in I} e_i \cdot \Pr[T_i \leq B_i]$. However, as the set of k employees minimizing the probability of the secret being disclosed can be non-unique, Alice's best response can be a mixed strategy $\langle a_i \rangle$ whose support consists of more than k employees. This notion is formalized by the following lemma:

Lemma 1. *Given Eve's mixed strategy $(\langle e_i \rangle, \langle \mathcal{B}_i \rangle)$, Alice's best response can be characterized as follows:*

- *For any employee i, if there are at least $N - k$ employees whose probabilities of being targeted and successfully bribed are strictly greater than that of i, then $a_i = 1$.*

– *For any employee i, if there are at least k employees whose probabilities of being targeted and successfully bribed are strictly less than that of i, then $a_i = 0$.*

Proof. First, for any employee i, if there are at least $N - k$ employees whose probabilities of sharing the secret are strictly greater than that of i, then i is a member of every size-k subset of employees that minimizes the probability of the secret being disclosed. Thus, in any best response, Alice always shares the secret with this employee i.

Second, for any employee i, if there are at least k employees whose probabilities of sharing he secret are strictly less than that of i, then i cannot be a member of any k-subset that minimizes the probability of the secret being disclosed. Thus, i cannot be in the support of any mixed strategy that is a best response for Alice. □

Eve's Best Response Suppose that Alice is playing a mixed strategy where a_i is the probability that she shares the secret with employee i. We define $\mathrm{MaxUE}(\mathcal{T}_i, a_i)$ to be the maximum payoff that Eve can attain from targeting employee i. Formally,

$$\mathrm{MaxUE}(\mathcal{T}_i, a_i) = \max_{b \in \mathbb{R}_{\geq 0}} \left(a_i \cdot S \cdot \Pr[T_i \leq b] - b\right). \tag{10}$$

Lemma 2. *For any employee i and trustworthiness distribution $\mathcal{T}_i$, Eve's maximum payoff $\mathrm{MaxUE}(\mathcal{T}_i, a_i)$ as a function of Alice's secret-sharing probability a_i has the following properties:*

1. $\mathrm{MaxUE}(\mathcal{T}_i, 0) = 0$,
2. $\mathrm{MaxUE}(\mathcal{T}_i, x)$ *is increasing in* x,
3. $\mathrm{MaxUE}(\mathcal{T}_i, x)$ *is uniformly continuous in* x.

Proof.

1. First, it is clear that the maximum of $\max_{b \in \mathbb{R}_{\geq 0}}(-b)$ is attained at $b = 0$.
2. To show that the function is increasing in x, let $x, y \in [0, 1]$ with $x < y$. Let b_x be a bribe value at which the maximum payoff is attained for secret-sharing probability x, that is, $\mathrm{MaxUE}(\mathcal{T}_i, x) = x \cdot S \cdot \Pr[T_i \leq b_x] - b_x$. Then, we have

$$\begin{aligned} \mathrm{MaxUE}(\mathcal{T}_i, y) &\geq y \cdot S \cdot \Pr[T_i \leq b_x] - b_x \\ &\geq x \cdot S \cdot \Pr[T_i \leq b_x] - b_x \\ &= \mathrm{MaxUE}(\mathcal{T}_i, x). \end{aligned}$$

3. Finally, to show uniform continuity, let $x, y \in [0, 1]$ with $x < y$, and let b_y be a bribe value at which the maximum payoff is attained for secret-sharing

probability y, that is, $\text{MaxUE}(\mathcal{T}_i, y) = y \cdot S \cdot \Pr[T_i \leq b_y] - b_y$. Using the previous result that $\text{MaxUE}(\mathcal{T}_i, y)$ is increasing, we have

$$\begin{aligned} 0 &< \text{MaxUE}(\mathcal{T}_i, y) - \text{MaxUE}(\mathcal{T}_i, x) \\ &\leq (y \cdot S \cdot \Pr[T_i \leq b_y] - b_y) - (x \cdot S \cdot \Pr[T_i \leq b_y] - b_y) \\ &= (y - x) \cdot S \cdot \Pr[T_i \leq b_y] \\ &\leq (y - x) \cdot S. \end{aligned}$$

So $\text{MaxUE}(\mathcal{T}_i, x)$ satisfies a Lipschitz condition in the variable x with Lipschitz constant S; and hence, it is uniformly continuous. □

For a given employee, it is possible for more than one bribe value to give Eve the maximal payoff. We define $\text{ArgMaxBE}(\mathcal{T}_i, x)$ to be the set of bribes that give Eve her maximum payoff for employee i, which is a function of the employee's trustworthiness level distribution and the probability of receiving the secret from Alice. Formally,

$$\text{ArgMaxBE}(\mathcal{T}_i, a_i) = \underset{b \in \mathbb{R}_{\geq 0}}{\text{argmax}} \left(a_i \cdot S \cdot \Pr[T_i \leq b] - b\right). \tag{11}$$

Using this notation, we may define constraints on Eve's best response strategy as follows.

Lemma 3. *Given any strategy $\langle a_i \rangle$ for Alice, Eve's best response selects an employee i with the largest $\text{MaxUE}(\mathcal{T}_i, a_i)$ over all $i \in \{1, \ldots, N\}$, and then chooses a bribe value b from $\text{ArgMaxBE}(\mathcal{T}_i, a_i)$. If there are multiple pairs (i, b) satisfying these constraints, then Eve may choose any distribution whose support is a subset of these payoff-maximizing pure strategies.*

Proof. Follows readily from Equations (9), (10), and (11). □

4.2 Nash Equilibria

Above, we introduced constraints on best-response strategies. In the following subsection, we introduce additional constraints on equilibrium strategies.

Alice's Strategy in an Equilibrium It is generally in Alice's interest to minimize the maximum attainable payoff for Eve, as this generally (but, since the game is non-zero sum, not necessarily) minimizes her loss. We know that Eve's best response is always to choose an employee (or a set of employees) which will maximize $\text{MaxUE}(\mathcal{T}_i, a_i)$ over i. Therefore, in an equilibrium, Alice's strategy should try to equalize these quantities, subject to the constraints that her sharing probabilities cannot exceed 1 and that they sum to k.

This notion is made formal in the following theorem:

Theorem 1. *In any Nash equilibrium,*

1. *if $a_i, a_j < 1$, then $\text{MaxUE}(\mathcal{T}_i, a_i) = \text{MaxUE}(\mathcal{T}_j, a_j)$, and*

2. *if* $a_j < a_i = 1$, *then* $\text{MaxUE}(\mathcal{T}_i, a_i) \leq \text{MaxUE}(\mathcal{T}_j, a_j)$.

Proof. Let $\langle a_i \rangle, (\langle e_i \rangle, \langle \mathcal{B}_i \rangle)$ be Alice's and Eve's mixed strategies and assume that this strategy profile is a Nash equilibrium.

1. For the sake of contradiction, suppose that $a_i, a_j < 1$ and it holds that $\text{MaxUE}(\mathcal{T}_i, a_i) \neq \text{MaxUE}(\mathcal{T}_j, a_j)$. We can assume without loss of generality that $\text{MaxUE}(\mathcal{T}_i, a_i) < \text{MaxUE}(\mathcal{T}_j, a_j)$. Then, $\text{MaxUE}(\mathcal{T}_j, a_j) > 0$, which (from Lemma 2.1) implies that $a_j > 0$. From Lemma 3, we have that the support of Eve's best-response mixed strategy does not include i. Thus, Alice may strictly increase a_i towards 1, and strictly decrease every other non-zero component of her strategy for employees other than i, while still satisfying the constraint $\sum_m a_m = k$. By decreasing her secret-sharing probability on every employee that Eve might bribe, Alice necessarily decreases the total probability of Eve learning the secret. Therefore, Alice can improve her expected payoff by changing her strategy, which contradicts the equilibrium condition.
2. For the sake of contradiction, suppose that $a_j < a_i = 1$ and that $\text{MaxUE}(\mathcal{T}_i, a_i) > \text{MaxUE}(\mathcal{T}_j, a_j)$. Then, $\text{MaxUE}(\mathcal{T}_i, a_i) > 0$, which (based on Lemma 2) implies that $a_i > 0$. Consequently, we have (from Lemma 3) that the support of Eve's mixed strategy does not include employee j. So Alice may simultaneously increase a_j towards 1 and decrease her non-zero secret-sharing probabilities for employees other than j, all while satisfying the constraint $\sum_m a_m = k$. Again, by decreasing her secret-sharing probability on every employee that Eve might bribe, Alice necessarily decreases the total probability of Eve learning the secret. Hence, this strategy change will increase her expected payoff, contradicting the equilibrium condition. □

It follows from Theorem 1 that Alice's equilibrium strategy $\langle a_i \rangle$ may have some employees with whom she shares the secret with certainty, but for all other employees, her secret-sharing distribution is only constrained by a smoothness constraint on the quantities $\text{MaxUE}(\mathcal{T}_i, a_i)$. Furthermore, these quantities do not depend on Eve's strategy, a fact on which we will rely when computing an equilibrium.

From Theorem 1, we also have that:

Corollary 1. *In any Nash equilibrium,*

- *Alice is either secure, that is, Eve has no strategy against her with a positive payoff, or she shares the secret with every employee with a non-zero probability. Formally, either* $\text{MaxUE}(\mathcal{T}_i, a_i) = 0$ *for every employee* i*, or* $a_i > 0$ *for every employee* i*.*
- *The employees with whom Alice shares the secret with certainty are at most as likely to be targeted by Eve as the other employees, with whom Alice is less likely to share the secret.*

It is interesting to compare the first point of the above corollary with Lemma 3. The former says that Alice shares the secret with every employee with a

non-zero probability (when she cannot be secure), while Lemma 3 says that Alice never shares the secret with an employee if there are at least k employees that have lower probabilities of being targeted and successfully bribed. Since an equilibrium strategy is necessarily a best response, it has to satisfy both constraints. This implies that, in an equilibrium, Eve equalizes the probability of targeting and successfully bribing over the set of employees that maximize her payoff.

Eve's Strategy in an Equilibrium In this section, we build on the characterization of Alice's equilibrium strategies presented in Theorem 1 to characterize Eve's equilibrium strategies. In the previous paragraph, we discussed how Eve equalizes the probability of targeting and successfully bribing over the set of employees that maximize her payoff.

This notion is made formal in the following theorem:

Theorem 2. *In any Nash equilibrium, if* $a_i, a_j < 1$*, then* $e_i \cdot \Pr[T_i \leq B_i] = e_j \cdot \Pr[T_j \leq B_j]$.

Proof. Let $\langle a_i \rangle, (\langle e_i \rangle, \langle \mathcal{B}_i \rangle)$ be Alice's and Eve's mixed strategies and assume that this strategy profile is a Nash equilibrium. For the sake of contradiction, suppose that $\langle e_i \cdot \Pr[T_i \leq B_i] \rangle$ is non-uniform over the set of employees with whom Alice does not always share the secret. Let I_{max} be the set of employees i for which $e_i \cdot \Pr[T_i \leq B_i]$ is maximal.

First, assume that $k \leq N - |I_{max}|$. Then, Alice's best response never shares the secret with the employees in I_{max}, that is, $a_i = 0$ for all $i \in I_{max}$, as there are k strictly better employees (as stated in Lemma 1). Consequently, we have $e_i = 0$ for every $i \in I_{max}$, as Eve's strategy also has to be a best response. But this implies that $e_i \cdot \Pr[T_i \leq B_i] = 0$ for every i such that $a_i < 1$, which contradicts that $\langle e_i \cdot \Pr[T_i \leq B_i] \rangle$ is non-uniform. Thus, it has to hold that $k > N - |I_{max}|$.

From $k > N - |I_{max}|$, we have that Alice's best response always shares the secret with every employee i for which $e_i \cdot \Pr[T_i \leq B_i]$ is not maximal (as stated in Lemma 1). Consequently, the only employees i for which $a_i < 1$ holds are the employees in I_{max}. But this contradicts that $\langle e_i \cdot \Pr[T_i \leq B_i] \rangle$ is non-uniform since all employees in I_{max} have the same maximal $e_i \cdot \Pr[T_i \leq B_i]$. □

Finding an Equilibrium Based on Theorems 1 and 2, we can formulate the following algorithm for finding an equilibrium of the game:

1. Find an equilibrium strategy $\langle a_i^* \rangle$ for Alice:
   We have to find an $\langle a_i^* \rangle$ that satisfies the constraints of Theorem 1. This can be done, for example, using any multidimensional numerical optimization method (e.g., the Nelder-Mead algorithm[15]) by using the sum of the amounts by which each constraining equality is violated as the objective function. Since we have from Lemma 2 that every $\text{MaxUE}(\mathcal{T}_i, a_i)$ is increasing and uniformly continuous in a_i, there always exists a solution $\langle a_i^* \rangle$ satisfying the constraints of Theorem 1. Note that, since $\text{MaxUE}(\mathcal{T}_i, a_i)$ is not strictly increasing, the solution might not be unique.

2. Find an equilibrium strategy $(\langle e_i^* \rangle, \langle \mathcal{B} \rangle)$ for Eve:
   We have to find $(\langle e_i^* \rangle, \langle \mathcal{B} \rangle)$ that satisfies both Lemma 3 and Theorem 2. Let $MaxUE^* = \max_i \mathrm{MaxUE}(\mathcal{T}_i, a_i^*)$ and let I^* be the set of employees for whom the maximum is attained. If $MaxUE^* = 0$, then there is no strategy with positive payoff for Eve, so let $B_i^* \equiv 0$ for every i (and $\langle e^* \rangle$ can be arbitrary). Otherwise:
   (a) For every $i \notin I^*$, let $e_i^* = 0$.
   (b) For every $i \in I^*$, let B_i^* always take some arbitrary but fixed bribe value from $\mathrm{ArgMaxBE}(\mathcal{T}_i, a_i^*)$, and let

$$e_i^* = \frac{\frac{1}{\Pr[T_i \leq B_i^*]}}{\sum_j \frac{1}{\Pr[T_j \leq B_j^*]}}. \tag{12}$$

It can be verified easily that $\langle a_i^* \rangle$ also satisfies Lemma 1. Thus, $\langle a_i^* \rangle$ and $(\langle e_i^* \rangle, \langle \mathcal{B}_i \rangle)$ form an equilibrium.

5 Special Case: Uniform Distributions on Trustworthiness

In this section, we assume that the trustworthiness level of each employee i is generated by a uniform random variable $T_i \sim \mathcal{U}(l_i, h_i)$, $0 < l_i < h_i < S$. In other words, we assume that employee i never reveals the secret for a bribe less than l_i, always reveals it for a bribe more than or equal to h_i, and the probability of revealing it increases linearly between l_i and h_i. Note that we allow a different distribution, i.e., different l_i and h_i, for each employee.

We begin our analysis by computing Eve's optimal bribe values for a given mixed strategy $\langle a_i \rangle$ of Alice.

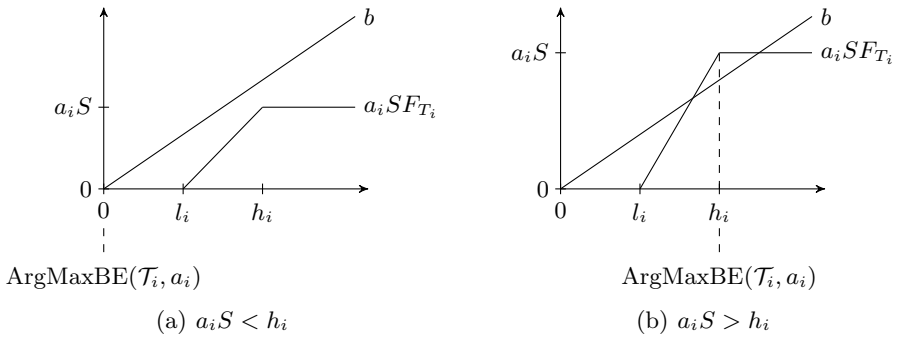
Lemma 4. *Eve's optimal bribe values are*

$$\mathrm{ArgMaxBE}(\mathcal{T}_i, a_i) = \begin{cases} \{0\} & \text{if } a_i < \frac{h_i}{S} \\ \{0, h_i\} & \text{if } a_i = \frac{h_i}{S} \\ \{h_i\} & \text{otherwise.} \end{cases} \tag{13}$$

Proof. First, it is clear that no bribe value in $(0, l_i]$ can be optimal as the probability of successfully bribing is zero in this interval; thus, these bribe values are all dominated by 0. Second, it is clear that no bribe value greater than h_i can be optimal as the probability of successful bribing reaches its maximum at h_i; thus, all values greater than h_i are dominated by h_i. For bribe values in $[l_i, h_i]$, Eve's expected payoff when targeting employee i is

$$S \cdot a_i \cdot \frac{b - l_i}{h_i - l_i} - b \ . \tag{14}$$

See Figure 1 for an illustration. When $h_i > S \cdot a_i$ (Figure 1(a)), we have that $S \cdot a_i \cdot \frac{b-l_i}{h_i-l_i} - b < S \cdot a_i \cdot \frac{b}{h_i} - b < 0$; thus, the only optimal bribe value is 0.



(a) $a_iS < h_i$ (b) $a_iS > h_i$

Fig. 1. Illustration of the proof of Lemma 4.

On the other hand, when $h_i < S \cdot a_i$ (Figure 1(b)), we have that, for a bribe value $b = h_i$, the payoff is $S \cdot a_i \cdot \frac{h_i - l_i}{h_i - l_i} - h_i > 0$. It is also easy to see that the derivative of the expected payoff as a function of b is strictly greater than zero in this case; thus, the only optimal bribe value is h_i. Finally, when $h_i = S \cdot a_i$, we have that, for a bribe value $b = h_i$, the payoff is $S \cdot a_i \cdot \frac{h_i - l_i}{h_i - l_i} - h_i = 0$; thus, both 0 and h_i are optimal. □

For uniform trustworthiness level distributions, the equilibria of the game can be characterized as follows:

Theorem 3. *If the trustworthiness level of each employee is generated according to a uniform distribution $\mathcal{U}(l_i, h_i)$, $0 < l_i < h_i < S$, the equilibria of the game can be characterized as follows:*

- *If $k < \frac{\sum_i h_i}{S}$, then Alice is* perfectly secure*: in any equilibrium, $a_i \leq \frac{h_i}{S}$ for every i, Eve never bribes any of the employees, and both players' payoffs are zero.*
- *If $k = \frac{\sum_i h_i}{S}$, then in any equilibrium of the game, $a_i = \frac{h_i}{S}$ for every i, and Eve's payoff is zero.*
- *If $k > \frac{\sum_i h_i}{S}$, then in any equilibrium of the game, $a_i > \frac{h_i}{S}$ and $B_i \equiv h_i$ for every i, and Eve's payoff is strictly positive while Alice's payoff is strictly negative.*

Proof. Let $\langle a_i \rangle, (\langle e_i \rangle, \langle \mathcal{B}_i \rangle)$ be Alice's and Eve's mixed strategies and assume that this strategy profile is a Nash equilibrium. We prove each case separately:

- $k < \frac{\sum_i h_i}{S}$: For the sake of contradiction, suppose that $a_i > \frac{h_i}{S}$ for some i. Then, there has to be a j such that $a_j < \frac{h_i}{S}$, otherwise $\sum_i a_i = k < \frac{\sum_i h_i}{S}$ would not hold. Consequently, MaxUE($\mathcal{T}_i, a_i$) > MaxUE($\mathcal{T}_j, a_j$) and, from Lemma 3, we have that $e_j = 0$. Furthermore, from Theorems 1 and 2, we also have that $e_i > 0$. Therefore, Alice can increase her payoff by decreasing a_i and increasing a_j, which contradicts the equilibrium condition. Thus, $a_i \leq \frac{h_i}{S}$ has to hold for every i.

Now, for the sake of contradiction, suppose that Eve targets and bribes employee i non-zero probability, that is, $e_i > 0$ and $B_i \not\equiv 0$. Since Eve's strategy has to be a best response, we have that $a_i \geq \frac{h_i}{S}$. Consequently, there has to exist some j satisfying $a_j < \frac{h_i}{S}$. From Lemma 3, we have that $e_j = 0$. Therefore, Alice can increase her payoff by decreasing a_i and increasing a_j, which contradicts the equilibrium condition. Thus, Eve never bribes any of the employees, and it follows immediately that both players' payoffs are zero.

- $k = \frac{\sum_i h_i}{S}$: For the sake of contradiction, suppose that $a_i > \frac{h_i}{S}$ for some i, which implies that there has to be a j such that $a_j < \frac{h_i}{S}$. Then, we can show that this leads to a contradiction using the same argument as in the first paragraph of the previous case. Thus, $a_i = \frac{h_i}{S}$ for every i. The rest follows readily from Lemma 4.
- $k > \frac{\sum_i h_i}{S}$: First, it is easy to see that, for any strategy $\langle a_i \rangle$, there has to be at least one i such that $a_i > \frac{h_i}{S}$, which implies $\mathrm{MaxUE}(\mathcal{T}_i, a_i) > 0$. By using the strategy $e_i = 1$ and some constant bribe value from $\mathrm{ArgMaxBE}(\mathcal{T}_i, a_i)$, Eve can achieve a positive payoff. Consequently, for every strategy $\langle a_i \rangle$, Eve's best response payoff has to be strictly positive. It follows immediately that, in any equilibrium, Eve's payoff is strictly positive while Alice's payoff is strictly negative.
  Now, for the sake of contradiction, assume that $a_i \leq \frac{h_i}{S}$ for some i, which implies $\mathrm{MaxUE}(\mathcal{T}_i, a_i) = 0$. Then, we have that $e_i = 0$ from Lemma 3. Therefore, Alice can increase her payoff (i.e., decrease her loss) by increasing a_i and decreasing every non-zero component of her strategy, which contradicts the equilibrium condition. Thus, $a_i > \frac{h_i}{S}$ has to hold for every i.
  Second, assume indirectly that, for some $\langle a_i \rangle$ and e that form an equilibrium and some i, $a_i < \frac{h_i}{S}$. If $e_i = 0$, then Alice would be able to increase her payoff (i.e., decrease her loss) by simultaneously increasing a_i and decreasing some $a_j > \frac{h_i}{S}$, which would contradict the assumption that $\langle a_i \rangle$ and e form an equilibrium. On the other hand, if $e_i > 0$, then Eve would be able to increase her payoff by simultaneously decreasing e_i and increasing e_j where j is such that $a_j > \frac{h_j}{S}$, which would also lead to a contradiction. Therefore, we have that $a_i \geq \frac{h_i}{S}$ for every i in any equilibrium. Finally, $B_i \equiv h_i$ follows readily from Lemma 4. □

6 Numerical Illustrations

In this section, we provide numerical illustrations for the results derived in the previous section. Thus, throughout this section, we model the trustworthiness levels of the employees as independent uniform random variables T_i with parameters l_i and h_i.

Figure 2 shows both players' equilibrium payoffs as functions of the number of employees k that have to know the secret. First, when k is less than $\frac{\sum_i h_i}{S}$, Alice can choose a secure strategy such that bribing is infeasible for Eve. Thus, both players' payoffs are zero. Second, when k is larger than $\frac{\sum_i h_i}{S}$, but it is low

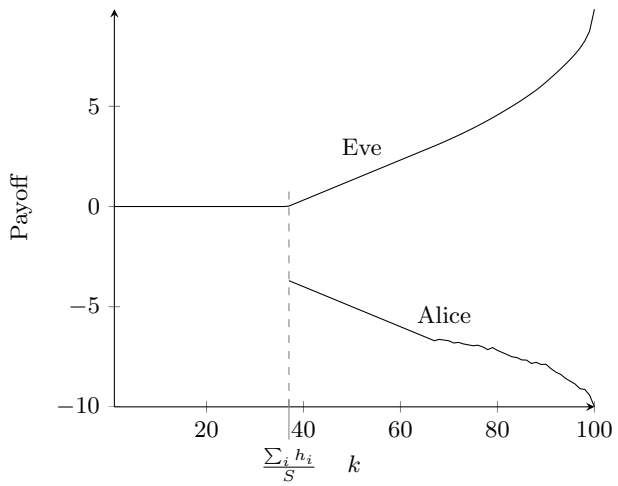


Fig. 2. The players' equilibrium payoffs as functions of the number of employees k that have to know the secret. The total number of employees is $N = 100$, the value of the secret is assumed to be $S = 10$, and the trustworthiness level of each employee i is assumed to be a random variable of the distribution $\mathcal{U}(l_i, h_i)$. For this example, each h_i is drawn from the set $(0, 7)$ uniformly at random.

enough such that $a_i < 1$ for each employee i, Alice distributes $k - \frac{\sum_i h_i}{S}$ evenly among the employees' probabilities. Thus, the probability of compromise and, hence, Alice's loss and Eve's payoff increase linearly with k. It is interesting to note that, while Eve's payoff is a continuous function of k, there is a big drop in Alice's payoff at the point where she can no longer play a secure strategy. This phenomena is caused by the non-zero sum property of our game. Finally, when k is large enough such that Alice assigns probability 1 to some employees, Eve's payoff increases super-linearly, while Alice's loss increases non-monotonically. Although Alice's non-monotonically increasing loss might seem surprising at first, it can be explained easily: as the secret is shared with more and more employees who are more easily bribed (i.e., have lower h_i), Eve can decrease her bribing costs by targeting these employees. This might decrease her success probability, but only by a value that is less than the decrease in her bribing costs. Consequently, sometimes Alice is better off if she shares the secret with more employees than she has to.

Figure 3 shows Alice's payoff (darker values indicate a higher loss) for a wide spectrum of parameter combinations of k and S. The figure clearly shows that, for lower values of S, the area where Alice can play a secure strategy (white plain) is greater than the area for higher values of S. Note that, for most values of S, we can identify the same three regions for k as in the previous figure: for $k < \frac{\sum_i h_i}{S}$, Alice's loss is zero; for $k > \frac{\sum_i h_i}{S}$, Alice's loss first increases linearly with k, but for larger values of k, Alice's loss increases non-monotonically. As

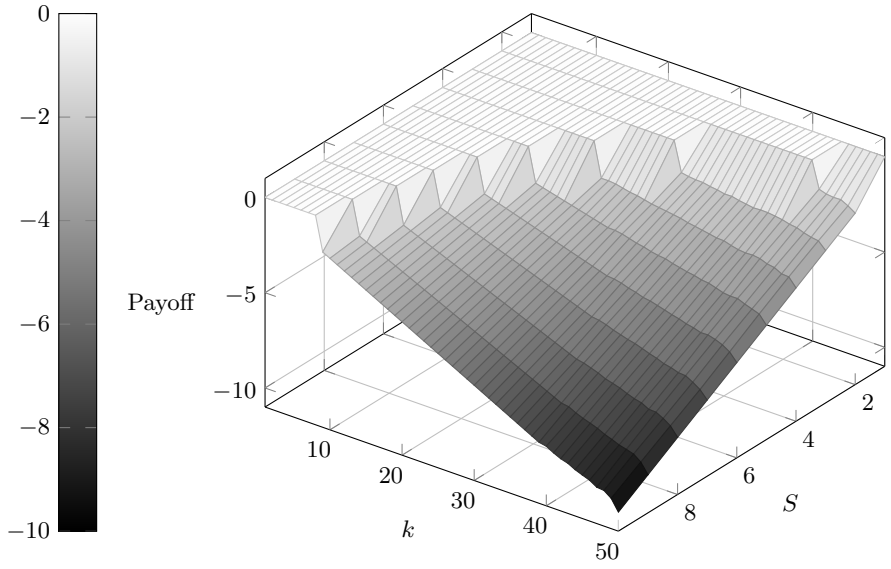


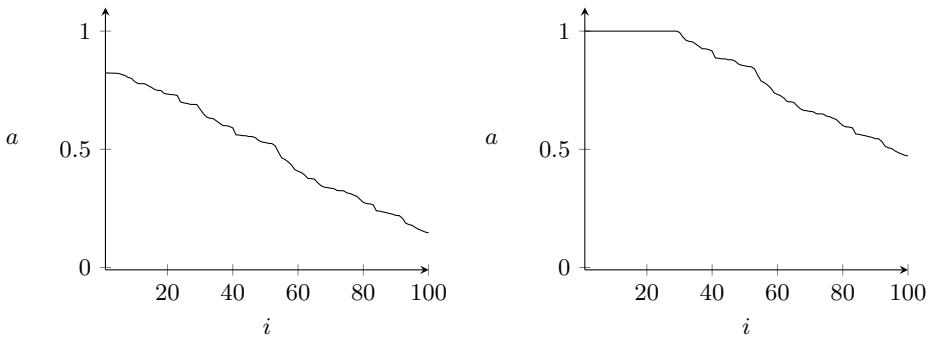
Fig. 3. Alice's equilibrium payoff for all combinations of $1 \leq k \leq 50$ and $1 \leq S \leq 10$. The parameters for this figure were generated in the same way as for Figure 2, but with $N = 50$.

expected, the worst case for Alice is when the number of employees k that have to know the secret is large and the value S of the secret is high.

Figure 4 shows Alice's equilibrium strategies for two different values of k. Figure 4(a) shows a case where k is small enough such that Alice does not assign probability 1 to any of her employees, while Figure 4(b) depicts a case where several employees get to know the secret with certainty. Figure 5 shows her equilibrium strategies for $N = 50$ and $\frac{\sum_i h_i}{S} \leq k \leq 50$. The figure clearly shows that, for all values of k, a_i is a monotonically increasing function of h_i, which can be explained by Theorem 1. Furthermore, the figure also confirms our analytical result that no a_i can be 0.

7 Discussion & Concluding Remarks

In this paper, we introduce a game-theoretic model for studying the decision making of a project manager who wants to maximize the security of an organization's intellectual property. Motivated in part by known behavioral methods of assessing trustworthiness [14], we assume that both the project manager and her adversary know the distribution of a random variable representing the trustworthiness of each employee. Finally, we assume that both players are able to estimate the value of the organization's intellectual property [4].



(a) $k = 50$ (b) $k = 80$

Fig. 4. Alice's equilibrium strategies for (a) $k = 50$ and (b) 80. The total number of employees is 100, the value of the secret is assumed to be $S = 10$, the trustworthiness level of each employee i is assumed to be a random variable of the distribution $\mathcal{U}(l_i, h_i)$, and the employees are sorted in decreasing order based on their h_i values. For this example, each h_i is drawn from the set $(0, 7)$ uniformly at random.

As a result of our analysis, we find that a project manager should select every employee with a non-zero probability, unless there is a secure strategy, where an adversary has no incentives to attack at all. This contradicts the naïve assumption that, to achieve maximal security, only the most trustworthy employees should be selected. The explanation for this is the following: selecting the team members deterministically always gives the adversary the knowledge of which employees to target for bribing. So, by randomizing her strategy, the project manager minimizes the information available to the adversary for planning her attack. It is an even more surprising result that, in an equilibrium, the adversary is at most as likely to target employees that certainly know the secret as those employees that know the secret with a probability less than 1. Again, this contradicts the naïve assumption that an adversary will try to bribe the employees that are the most likely to know the secret.

For the special case of uniform distributions on trustworthiness levels, we find that the game has two distinct outcomes: either the number of team members is small enough, such that the project manager has a perfectly secure strategy, or the security of the secret depends solely on the randomness of selecting the employee with whom it is shared.[1] In the former case, the adversary has no incentives to attack and, consequently, never learns the secret. In the latter case, the adversary always attacks and always bribes the targeted employee with the minimal amount that is never below the employee's trustworthiness level. Thus, if the adversary targeted an employee that actually knows the secret, then it is certainly revealed. The project manager's only possible defense in this case is to randomize the selection of employees.

[1] Note that the probability that an exact equality occurs is negligible in practice.

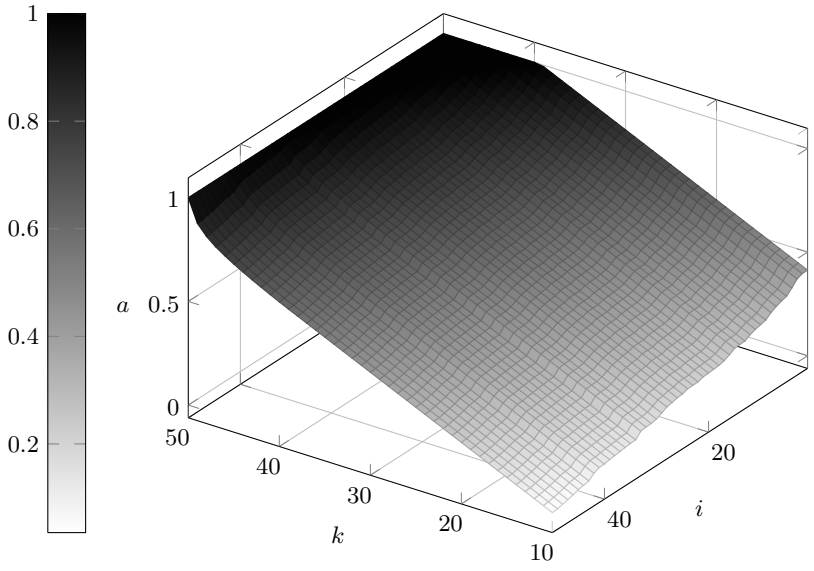


Fig. 5. Alice's equilibrium strategies for $\frac{\sum_i h_i}{S} < k \leq 50$. The parameters for this figure were generated in the same way as for Figure 4, but with $N = 50$. Again, the employees are sorted in decreasing order based on their h_i values.

There are multiple possible directions for future work. First, a limitation of the model is the restriction on the adversary, which constrains her to target only a single employee at a time. This simplification can be motivated by the adversary's incentive to keep her operation covert and, thus, to minimize the number of bribing attempts. However, it would be worthwhile to study the trade-off between the adversary's increased risk of being discovered and the increased probability of learning the secret when she targets multiple employees. As another direction, we want to study our model with specific distributions over trustworthiness levels. In this paper, we provide results for the uniform distribution, which can be well-motivated in practice; however, there are other distributions that can be justified from practical observations: e. g., the beta distribution.

Acknowledgements

We gratefully acknowledge the support of the Penn State Institute for Cyber-Science. The first author would like to thank the Campus Hungary Program for supporting his research visit. The third author would like to thank the Office of Naval Research (ONR) for supporting his research visit under Visiting Scientists Grant N62909-13-1-V029.

References

1. Ross Anderson. *Security engineering - A guide to building dependable distributed systems (2nd Ed.)*. Wiley, 2008.
2. Ross Anderson, Chris Barton, Rainer Böhme, Richard Clayton, Michael van Eeten, Michael Levi, Tyler Moore, and Stefan Savage. Measuring the cost of cybercrime. In *WEIS*, 2012.
3. Stephen Band, Dawn Cappelli, Lynn Fischer, Andrew Moore, Eric Shaw, and Randall Trzeciak. Comparing insider IT sabotage and espionage: A model-based analysis. Technical Report CMU/SEI-2006-TR-026, Carnegie Mellon University, 2006.
4. Nick Bontis. Assessing knowledge assets: A review of the models used to measure intellectual capital. *International Journal of Management Reviews*, 3(1):41–60, 2001.
5. Carl Colwill. Human factors in information security: The insider threat – Who can you trust these days? *Information Security Technical Report*, 14(4):186 – 196, 2009.
6. Corporate Trust (Business Risk & Crisis Mgmt. GmbH). Studie: Industriespionage 2012 - Aktuelle Risiken für die deutsche Wirtschaft durch Cyberwar, 2012.
7. John D'Arcy, Anat Hovav, and Dennis Galletta. User awareness of security countermeasures and its impact on information systems misuse: A deterrence approach. *Information Systems Research*, 20(1):79–98, March 2009.
8. FBI. The insider threat. `http://www.fbi.gov/about-us/investigate/counterintelligence/insider_threat_brochure`, April 2013.
9. Federal Bureau of Investigation. Economic espionage. `http://www.fbi.gov/about-us/investigate/counterintelligence/economic-espionage`.
10. Peter Finn. Chinese citizen sentenced in military data-theft case. *Washington Post*, March 2013.
11. Benjamin Johnson, Pascal Schöttle, and Rainer Böhme. Where to hide the bits? In *GameSec 2012*, volume 7638 of *LNCS*, pages 1–17. Springer, 2012.
12. Debin Liu, XiaoFeng Wang, and L. Jean Camp. Game theoretic modeling and analysis of insider threats. *International Journal of Critical Infrastructure Protection*, 1:75–80, December 2008.
13. Andrew Moore, Dawn Cappelli, Thomas Caron, Eric Shaw, Derrick Spooner, and Randall Trzeciak. A preliminary model of insider theft of intellectual property. *Journal of Wireless Mobile Networks, Ubiquitous Computing, and Dependable Applications*, 2(1):28–49, March 2011.
14. Asmaa Munshi, Peter Dell, and Helen Armstrong. Insider threat behavior factors: A comparison of theory with reported incidents. In *IEEE HICSS 2012*, pages 2402–2411, 2012.
15. John Nelder and Roger Mead. A simplex method for function minimization. *Computer Journal*, 7:308–313, 1965.
16. Marisa Randazzo, Michelle Keeney, Eileen Kowalski, Dawn Cappelli, and Andrew Moore. Insider threat study: Illicit cyber activity in the banking and finance sector. Technical Report CMU/SEI-2004-TR-021, Carnegie Mellon University, June 2005.
17. Thomas Ronde. Trade secrets and information sharing. *Journal of Economics and Management Strategy*, 10:391–417, Fall 2001.
18. Jerome Saltzer and Michael Schroeder. The protection of information in computer systems. *Proceedings of the IEEE*, 63(9):1278–1308, 1975.
19. Ravi Sandhu and Pierangela Samarati. Access control: Principles and practice. *IEEE Communications Magazine*, 32:40–48, 1994.