Secure Team Composition to Thwart Insider Threats and Cyberespionage

ARON LASZKA, Pennsylvania State University
BENJAMIN JOHNSON, University of California, Berkeley
PASCAL SCHÖTTLE, University of Münster
JENS GROSSKLAGS, Pennsylvania State University
RAINER BÖHME, University of Münster

We develop a formal non-deterministic game model for secure team composition to counter cyber-espionage and to protect organizational secrets against an attacker who tries to sidestep technical security mechanisms by offering a bribe to a project team member. The game captures the adversarial interaction between the attacker and the project manager who has a secret she wants to protect but must share with a team of individuals selected from within her organization. Our interdisciplinary work is important in the face of the multi-pronged approaches utilized by well-motivated attackers to circumvent the fortifications of otherwise well-defended targets.

Categories and Subject Descriptors: K.6.5 [**Management of Computing and Information Systems**]: Security and Protection

General Terms: Security, Management, Economics

Additional Key Words and Phrases: Insider threat, cyberespionage, game theory, human factor, management of information security, access control

ACM Reference Format:
Aron Laszka, Benjamin Johnson, Pascal Schöttle, Jens Grossklags, and Rainer Böhme, 2014. Secure Team Composition to Thwart Insider Threats and Cyberespionage. *ACM Trans. Internet Technol.* 0, 0, Article 0 (2014), 22 pages.
DOI:http://dx.doi.org/10.1145/0000000.0000000

1. INTRODUCTION

Nowadays the success of economic and political organizations depends less on the conventional factors capital and labor, but increasingly on making the right strategic decisions. These decisions require careful preparation, often involving large numbers of staff, and for a variety of reasons their outcomes must be protected until the official release. At the same time, the proliferation of information technology makes organiza-

Acknowledgments: This work was supported by the Penn State Institute for Cyber-Science. The first author's research visit was supported by the Campus Hungary Program. The second author acknowledges support from the National Science Foundation under ITR award CCF-0424422 (TRUST). The third author's research visit was supported under Visiting Scientists Grant N62909-13-1-V029 by the Office of Naval Research (ONR).

Author's addresses: Aron Laszka, Laboratory of Cryptography and System Security (CrySyS), Budapest University of Technology and Economics; Benjamin Johnson, School of Information, University of California, Berkeley; Pascal Schöttle and Rainer Böhme, Department of Information Systems, University of Münster; Jens Grossklags, College of Information Sciences and Technology, Pennsylvania State University.

DOI:http://dx.doi.org/10.1145/0000000.0000000

tions more transparent and vulnerable to information leaks, turning effective protection of organizational secrets, including intellectual property, into a hard problem.

Providing effective access control in organizations has been referred 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 [Anderson 2008]. 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 critical systems and sensitive data [Saltzer and Schroeder 1975; Sandhu and Samarati 1994]. Beyond the technical level, the management literature distinguishes four thrusts of organizational measures to prevent information leaks: deterrence, prevention, detection, and remedies [Straub and Welke 1998]. Our work addresses prevention on a non-technical level by studying the composition of the teams who prepare or work with organizational secrets. In particular, we study how to compose teams in such a way that the risk of information leakage is minimized.

This research is motivated by the steady rise of cyber-espionage activities, in particular 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" [Finn 2013]. 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" [Randazzo et al. 2005].

The interpretation of such statements is not always fully conclusive because insider threat tends to be a catch-all term for attacks that involve privileged users [Schultz 2002]. However, it is important to distinguish between intrinsically motivated insider threats – the disgruntled employee – and those with external cause. Arguably the latter are more relevant in the described scenarios. They include passive observation of insiders' actions, which might reveal secrets if an employee's observable behavior is correlated with the secret state he knows about, as well as active variants ranging from deception (social engineering), coercion, extortion, to bribery; depending on the level of voluntary cooperation of the target person. Evidently, the employees of an organization differ in their ability to successfully reject advances by an attacker.

Turning a trusted employee into a spy provides a number of benefits for an outside attacker. First, a security compromise by an insider might be harder to detect than external network-based attacks, which might leave traces identifiable for forensics teams. 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 article, we study a two-player stochastic game for modeling secure team composition to add resilience against insider threats with external cause. 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 give necessary conditions on both players' best-response strategies and on the Nash equilibria of the game. We find that Alice's equilibrium 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.

This article extends previous work [Laszka et al. 2013] addressing team composition, and offers a number of new contributions. We have expanded our framework to more broadly consider previous work on insider threats. We show that a manager's mixed strategy can be efficiently computed from our simplified representation of the strategy (Theorem 3.1). We exhibit a computable Nash equilibrium (Theorem 4.7), and prove the uniqueness of the manager's non-perfectly-secure equilibrium strategy (Theorem 4.8). Finally, we prove uniqueness of the attacker's equilibrium payoff (Corollary 4.9). With these new contributions, we continue the discussion of the composition of project teams as a formal and critical dimension of a comprehensive corporate security policy.

The remainder of the article 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, including additional theoretical analysis and numerical illustrations. We discuss our results and provide concluding remarks in Section 6.

2. BACKGROUND AND RELATED WORK

2.1. Studies on Insider Threats and Cyber-espionage

Over the last years, several reports have 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 [Moore et al. 2011]. 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, CERT identified several indicators that preceded either industrial espionage or sabotage, and thus could give hints if an employee might be vulnerable to being bribed [Band et al. 2006]. In our research, we do not explicitly model behavioral and motivational factors that influence the trustworthiness of an employee (see also [Schultz 2002] for an overview). 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) [FBI 2013], 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 [Federal Bureau of Investigation 2013]." 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 [Federal Bureau of Investigation 2013]."

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 € [Corporate Trust (Business Risk & Crisis Mgmt. GmbH) 2012]. In that 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 [Anderson et al. 2013].

2.2. Related Work

This article touches several different research areas. One directly connected area is the organization of firms under weak intellectual property rights. For example, in [Rønde 2001], 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.

Proposals for deterrence strategies to prevent misuse of computing resources are complementary to our work [D'Arcy et al. 2009]. These strategies may include security education and training, awareness programs, and computer monitoring. However, the effectiveness of such approaches against sophisticated insider threats is a cause for concern. A report from the intelligence community on insider threats therefore highlighted the importance of monitoring by suggesting that researchers should "focus on detection, not prevention" when fighting insider threats [Brackney and Anderson 2004]. The apparent lack of ability to focus on prevention might partly rest on the lack of appropriate models and methods (beyond the basic strategies outlined above). Our research on secure composition of teams addresses this problem space.

Also complementary to our work are models and other approaches to exhaustively find ways for an insider attacker to gain access to a specific resource (see, for example, [Chinchani et al. 2005]). These models typically assume that a willing insider is already in place and the obstacle is merely how to extract information from the organization. Our work is focused on preventing an outside attacker to successfully "turn" an insider who has knowledge of a business secret or intellectual property (and does not necessarily need to breach sophisticated access control systems to leak information to the outsider).

There are many additional research directions covering the subject of insider threats, including game theory [Liu et al. 2008] and trust models [Colwill 2009], 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. We respond to the call for research that looks "beyond information technology to the organization's overall business processes" to prevent insider threats from causing substantial harm [Cappelli et al. 2009].

3. MODEL DEFINITION

In this section, we describe a two-player, simultaneous, non-deterministic game which models the team composition scenario. First, we describe the general context and environment of the game and introduce the two players. Then, we define these players' pure strategies and the payoffs resulting from the pure-strategy choices. Finally, we introduce notation to represent mixed strategies and express the players' expected payoffs in terms of this notation. Figure 1 illustrates our game setup.

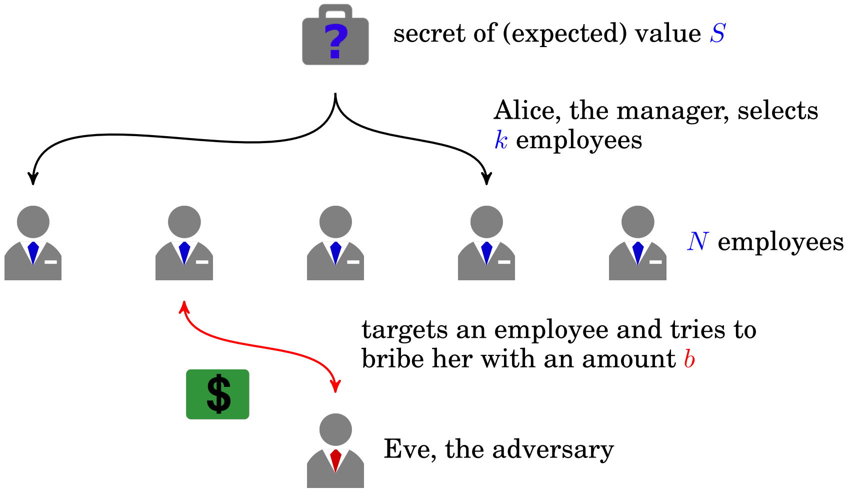

?
secret of (expected) value S
Alice, the manager, selects k employees
N employees
targets an employee and tries to bribe her with an amount b
$
Eve, the adversary

Fig. 1. Illustration for our model with $N = 5$ and $k = 2$.

3.1. Environment

In our model, an organization with a secret of high value has N employees who are qualified to work on projects that require knowledge of the secret. The organization must share the secret with at least k employees in order to operate. The employees have varying levels of trustworthiness, which can only be estimated. For a given employee i, this uncertain trustworthiness level is modeled by a random variable T_i, whose distribution $\mathcal{T}_i$ is known to all players.[1] We explicitly disregard other constraints on team composition and assume that all aspects of the trustworthiness of an employee are captured by the random variable T_i. If $T_i = t_i$, then employee i will reveal what she knows whenever she is bribed[2] with an amount greater than or equal to t_i, but she will never reveal the secret if she is bribed with an amount less than t_i. We use the standard cumulative distribution function 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 named Alice and Eve. Alice is the project manager of the organization, 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

[1] Our game is complete information only in the sense that all players know its parameters. However, as we will see, the players' payoffs are non-deterministic for a given strategy profile, due to the randomness of the trustworthiness levels.

[2] For brevity, we use this bribe interpretation for our formal analysis. Our results generalize to cases where trustworthiness measures the susceptibility to eavesdropping, deception, coercion, extortion, or other forms of social engineering that impose a variable cost on the attacker.

organization, and this secret has (expected) value S, which is assumed to be known to both players.[3] Alice needs to share this secret with k of her N qualified employees to ensure the operation of the company.

Eve is a spy from either inside or outside of the organization. Eve wants to learn the secret and has the resources to bribe or eavesdrop on one of Alice's employees. If she eavesdrops on an employee, the trustworthiness level of the employee can be interpreted as a measure of the difficulty of eavesdropping on that employee.

Note that, since our game is simultaneous, Eve does not know which employees are on the team, and Alice does not know which employee is bribed.

3.3. Pure-Strategy Sets

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

Eve's pure strategy choice is to select one employee[4] and an amount to bribe. Formally, she chooses a pair (i, b) consisting of an employee 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 receives the value of the secret S minus the amount of the bribe b, and Alice loses the value of the secret S. In all other cases, Eve loses the amount of the bribe b, and Alice loses nothing. The payoffs for the different scenarios are summarized by Table I. Recall that each T_i is a random variable, and the players only know its distribution $\mathcal{T}_i$.

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

Strategy profile	Payoff for	
and outcome	Alice	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 player's mixed strategy is a distribution over the set of her pure strategies. For Alice, the canonical representation of her mixed-strategy space is a finite probability distribution over 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 expected payoffs for both players can be determined by representations of the mixed-strategy spaces that are simpler than the canonical ones. In the following subsections, we introduce these representations and use them to express the players' expected payoffs.

3.5.1. Mixed Strategy for Alice. In the canonical representation of Alice's mixed strategy, we would let α_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 payoffs for

[3] Our analysis does not require the secret to be equally valuable to both players. Furthermore, our analysis does not change if we assume the secret's value is a random variable with expectation S.

[4] The restriction of Eve's choice to bribe only one employee can be motivated by the adversary's incentive to keep her operation covert and, thus, to minimize the number of interactions with employees. For example, unsuccessful bribes may be reported by the employee, making the success of repeated attempts unlikely.

any mixed-strategy profile depend only on the probabilities of Alice sharing the secret with each employee. More specifically, the probability of Eve learning the secret is

$$\sum_{i\in\{1,\ldots,N\}} \Pr[\text{Alice shares the secret with } i]\cdot \Pr[\text{Eve succeeds with her bribe} \mid \text{Eve targets } i]\cdot \Pr[\text{Eve targets } i]\ . \tag{2}$$

In other words, Alice's strategy choice influences the payoffs only through the probabilities of sharing with each employee, not the actual distribution over the subsets. Since several different mixed strategies might induce the same marginal probabilities for the employees, we gain simplicity by restricting our attention to these marginal sharing probabilities. Our goal here is to describe the space of these marginal probabilities.

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

$$a_i = \sum_{I:\ i\in I} \alpha_I. \tag{3}$$

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

$$\sum_{i=1}^{N} a_i = k. \tag{4}$$

It is easy to see that, for any mixed strategy of Alice, the vector of marginal probabilities $\boldsymbol{a}$ satisfies $0 \le a_i \le 1$ for every i, and $\sum_{i=1}^{N} a_i = k$. However, it remains to show that this is also true vice versa; that is, to show that, for any vector $\boldsymbol{a}$ of N probabilities whose sum is k, there exists a mixed strategy for Alice whose vector of marginal probabilities is $\boldsymbol{a}$. The following theorem shows that this is indeed true.

THEOREM 3.1. *For any vector of probabilities $\boldsymbol{a}$ that satisfies $\sum_i a_i = k$, there exists a mixed strategy $\boldsymbol{\alpha}$ for Alice such that, for every i, the probability of sharing the secret with employee i is a_i. Furthermore, there is such a mixed strategy whose support consists of at most N sets.*

PROOF. Our proof is constructive, and it is based on the following algorithm.

(1) For every k-subset I, let $\alpha_I = 0$.
(2) Let I be a k-subset consisting of the positions with the k highest a_i (if there are multiple such subsets, select an arbitrary one).
(3) Let p be the maximum value subject to
   - for every $i \in I$, $a_i - p \ge 0$ and
   - for every $i \notin I$, a_i satisfies the MaxProb constraint (for the definition of this constraint, see below).
(4) Increase α_I by p and, for every $i \in I$, decrease a_i by p.
(5) If there is an $a_i > 0$, then continue from Step 2.

Now, we introduce the MaxProb constraint. First, notice that a non-negative vector $\boldsymbol{a}$ has to satisfy two necessary constraints to be a mixed strategy over k-subsets: $\sum_i a_i = k$ and, for every i, $a_i \le 1$. It is easy to see that a vector cannot be a mixed strategy over k-subsets if it violates one of the constraints. Similarly, at any step of the algorithm's execution, it has to hold that $a_i \le k'$ for every i, where $k' = \sum_i a_i/k$. From this, we can formulate the MaxProb constraint as $p \le \sum_j a_j/k - a_i$. Finally, we call a vector $\boldsymbol{a}$ proper if, for every i, $a_i \ge 0$ and $a_i \le k'$. Obviously, we have that the input vector is proper.

Next, we prove the correctness of the algorithm. First, it is easy to see that the vector $\boldsymbol{a}$ stays non-negative (first constraint of Step 3). Second, we can show that the vector $\boldsymbol{a}$ stays proper. Every element $i \in I$ is decreased by p, but the sum is decreased by $k \cdot p$; thus, if the elements of I satisfied $a_i \leq \sum_j a_j / k$ before the decrease, they still satisfy it after the decrease. As for the non-elements $i \notin I$, the MaxProb constraint ensures that the vector stays proper. Third, it is easy to see that if a vector is proper and non-zero, then it has at least k positive elements (as no element can be higher than the sum over k). Fourth, it can be shown that if there are k positive elements, then the maximum p of Step 3 has to be positive (as there are at most k elements for which the equality $a_i = \sum_j a_j / k$ holds; hence, $p = \sum_j a_j / k - a_i$ does not hold for $p = 0$ and $i \notin I$).

Note that, at this point, we already have that the algorithm starts with a proper non-zero vector, it decreases the elements (possibly an infinite number of times) keeping the vector proper and non-negative, and finally decreases the last k positive elements to zero at once. It remains to show that the algorithm terminates after a finite number of iterations. However, we can do much better than that. Let M be the set of elements i for which the equality $a_i = \sum_j a_j / k$ holds (i.e., the set of maximal elements), let Z be the set of zero elements, and let O be the set of elements neither in M nor in Z. First, if an element belongs to Z, then it obviously remains there after a decrease. Second, if an element belongs to M, then it remains there after a decrease (as any element of M has to be a member of I). Third, in every iteration, at least one element of O is moved to either M or Z (as one of the constraints of Step 3 has to be an equality for at least one element for the maximum p). Fourth, $|O| \leq N$ trivially. Therefore, there are at most N iterations, as we remove an element from the set O in every iteration and $|O|$ is at most N initially. Notice that this also implies that the cardinality of the resulting distribution's support (the number of k-subsets with non-zero probability) is also at most N.

Finally, we have to show that the resulting $\boldsymbol{\alpha}$ is indeed a distribution, but this is very easy. First, $\sum_I \alpha_I = 1$, as $\sum_i a_i = k$ initially and we decrease it by $k \cdot p$ when we assign p probability to one of the subsets. Second, for every i, $\sum_{I \ni i} \alpha_I = a_i$, as we increase the probability of a containing subset by p when we decrease the value of a_i by p. $\square$

Note that we not only proved the existence of a mixed strategy, but also devised an algorithm for finding a simple one. This is important from a practical point of view, as we will establish results in Section 4 on best-response and equilibrium strategies based on the marginal probabilities representation. The above algorithm can be used in practice to find a feasible mixed strategy.

3.5.2. 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 the employee Eve has chosen to bribe. Random variable B takes values in $\mathbb{R}_{\geq 0}$, and represents the amount of the bribe.

Similarly 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

$$e_i = \Pr[Y = i]. \tag{5}$$

Since Eve always chooses exactly one employee, we have

$$\sum_{i=1}^{N} e_i = 1. \tag{6}$$

To describe a distribution over bribes, we employ the notation

$$F_B(b) = \Pr[B \leq b], \tag{7}$$

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

$$F_{B_i}(b) = \Pr[B_i \leq b] = \Pr[B \leq b | Y = i]. \tag{8}$$

In what follows, we will represent Eve's mixed strategies as pairs $(\boldsymbol{e}, \boldsymbol{\mathcal{B}})$, 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 $\boldsymbol{a}$ and Eve plays a mixed strategy represented by $(\boldsymbol{e}, \boldsymbol{\mathcal{B}})$, then the expected payoff for Alice is

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

and the expected payoff for Eve is

$$S \cdot \sum_{i=1}^{N} (a_i \cdot e_i \cdot \Pr[T_i \leq B_i]) - \sum_{i=1}^{N} e_i \cdot E[B_i], \tag{10}$$

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 our game's Nash equilibria. We begin with giving necessary conditions on Alice's and Eve's best-response strategies. Then, we use these conditions to constrain the players' strategies in an equilibrium. Finally, based on these constraints, we provide results on the existence and uniqueness of the equilibrium strategies and payoffs.

4.1. Best-Response Strategies

4.1.1. 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 that minimizes $\sum_{i \in I} e_i \cdot \Pr[T_i \leq B_i]$. However, as the set of k employees minimizing the probability of the secret being disclosed may be non-unique, Alice's best response can be a mixed strategy $\boldsymbol{a}$ whose support consists of more than k employees. This notion is formalized by the following lemma.

LEMMA 4.1. *Given Eve's mixed strategy $(\boldsymbol{e}, \boldsymbol{\mathcal{B}})$, any best-reponse strategy for Alice has to satisfy the following properties.*

- *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. □

4.1.2. 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,

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

See Figure 2 for an example satisfying $\mathrm{MaxUE}(\mathcal{T}_i, a_i) > 0$.

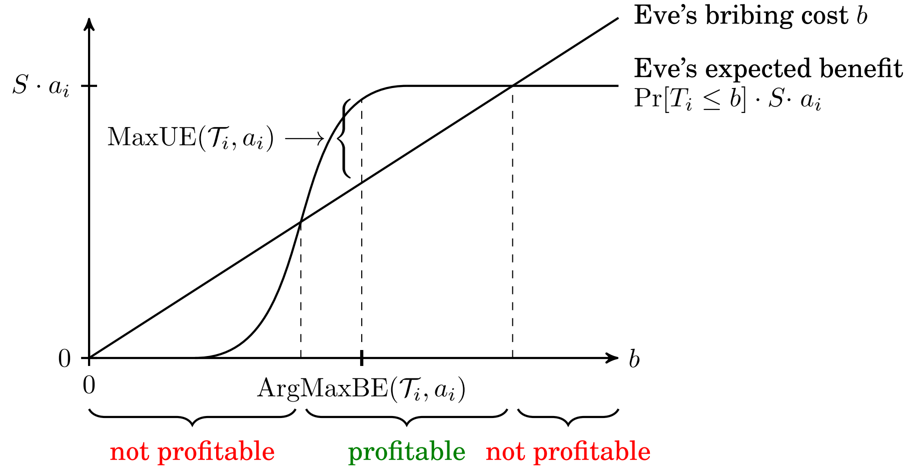


Fig. 2. Results of Eve targeting employee i as a function of her bribe amount b.

LEMMA 4.2. *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*) *if* $\mathrm{MaxUE}(\mathcal{T}_i, z) > 0$ *for some z, then* $\mathrm{MaxUE}(\mathcal{T}_i, x)$ *is strictly increasing in x on* $(z, 1]$*,*
(*4*) $\mathrm{MaxUE}(\mathcal{T}_i, x)$ *is uniformly continuous in x.*

PROOF.

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

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

(4) Finally, to show uniform continuity in x, 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, $\mathrm{MaxUE}(\mathcal{T}_i, y) = y \cdot S \cdot \Pr[T_i \leq b_y] - b_y$. Using the previous result that $\mathrm{MaxUE}(\mathcal{T}_i, y)$ is increasing, we have

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

So $\mathrm{MaxUE}(\mathcal{T}_i, x)$ 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 $\mathrm{ArgMaxBE}(\mathcal{T}_i, x)$ 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,

$$\mathrm{ArgMaxBE}(\mathcal{T}_i, a_i) = \underset{b \in \mathbb{R}_{\geq 0}}{\mathrm{argmax}} \left(a_i \cdot S \cdot \Pr[T_i \leq b] - b \right). \tag{12}$$

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

LEMMA 4.3. *Given Alice's mixed strategy* $\mathbf{a}$*, Eve's best response selects an employee* i *with the largest* $\mathrm{MaxUE}(\mathcal{T}_i, a_i)$ *over all* $i \in \{1, \ldots, N\}$*, and then chooses a bribe value* b *from* $\mathrm{ArgMaxBE}(\mathcal{T}_i, a_i)$*. 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 (10), (11), and (12). □

4.2. Strategies in Nash Equilibria

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

4.2.1. 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 $\mathrm{MaxUE}(\mathcal{T}_i, a_i)$ 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 4.4. *In any Nash equilibrium, Alice's strategy satisfies the following constraints.*

(*1*) *For any pair of employees i and j, if $a_i, a_j < 1$, then $\mathrm{MaxUE}(\mathcal{T}_i, a_i) = \mathrm{MaxUE}(\mathcal{T}_j, a_j)$.*
(*2*) *For any pair of employees i and j, if $a_j < a_i = 1$, then $\mathrm{MaxUE}(\mathcal{T}_i, a_i) \leq \mathrm{MaxUE}(\mathcal{T}_j, a_j)$.*

PROOF. Let $\boldsymbol{a}, (\boldsymbol{e}, \boldsymbol{\mathcal{B}})$ 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 $\mathrm{MaxUE}(\mathcal{T}_i, a_i) \neq \mathrm{MaxUE}(\mathcal{T}_j, a_j)$. We can assume without loss of generality that $\mathrm{MaxUE}(\mathcal{T}_i, a_i) < \mathrm{MaxUE}(\mathcal{T}_j, a_j)$. Then, $\mathrm{MaxUE}(\mathcal{T}_j, a_j) > 0$, which (from Lemma 4.2.1) implies that $a_j > 0$. From Lemma 4.3, 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 $\mathrm{MaxUE}(\mathcal{T}_i, a_i) > \mathrm{MaxUE}(\mathcal{T}_j, a_j)$. Then, $\mathrm{MaxUE}(\mathcal{T}_i, a_i) > 0$, which (based on Lemma 4.2) implies that $a_i > 0$. Consequently, we have (from Lemma 4.3) 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 4.4 that Alice's equilibrium strategy $\boldsymbol{a}$ 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 $\mathrm{MaxUE}(\mathcal{T}_i, a_i)$. Furthermore, these quantities do not depend on Eve's strategy, a fact on which we will rely when computing an equilibrium.

From Theorem 4.4, we also have that:

COROLLARY 4.5. *In any Nash equilibrium,*

— *Alice is either perfectly secure, that is, Eve has no strategy against her with a positive payoff, or else Alice shares the secret with every employee with a non-zero probability. Formally, either $\mathrm{MaxUE}(\mathcal{T}_i, a_i) = 0$ 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 4.3. The former says that Alice shares the secret with every employee with a non-zero probability (when she cannot be secure), while Lemma 4.3 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 expected payoff.

4.2.2. Eve's Strategy in an Equilibrium. In this section, we build on the constraints on Alice's equilibrium strategies presented in Theorem 4.4 to describe Eve's strategy in an equilibrium. In the previous paragraph, we argued that, in an equilibrium, Eve equalizes the probability of targeting and successfully bribing over the set of employees that maximize her payoff.

This notion is made formal by the following theorem.

THEOREM 4.6. *In a Nash equilibrium, if* $a_i, a_j < 1$ *for a pair of employees* i *and* j, *then* $e_i \cdot \Pr[T_i \leq B_i] = e_j \cdot \Pr[T_j \leq B_j]$.

PROOF. Let $\boldsymbol{a}, (\boldsymbol{e}, \boldsymbol{\mathcal{B}})$ 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 $e_i \cdot \Pr[T_i \leq B_i]$ is non-uniform over the set of employees with whom Alice does not always share the secret. Furthermore, let I_{max} be the set of employees i for which $e_i \cdot \Pr[T_i \leq B_i]$ is maximal.

First, suppose 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 4.1). 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 $e_i \cdot \Pr[T_i \leq B_i]$ 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 4.1). Consequently, the only employees i for which $a_i < 1$ holds are the employees in I_{max}. But this contradicts that $e_i \cdot \Pr[T_i \leq B_i]$ is non-uniform since all employees in I_{max} have the same maximal $e_i \cdot \Pr[T_i \leq B_i]$. □

4.3. Existence and Multiplicity of Equilibrium Strategies and Payoffs

In the previous subsections, we have formulated constraints on the equilibria of the game. Here, we provide existence and uniqueness results on the equilibrium strategies and payoffs.

We begin with showing the existence of an equilibrium strategy profile.

THEOREM 4.7. *The game always has at least one Nash equilibrium.*

PROOF. Our proof is constructive, that is, we show the existence of an equilibrium strategy profile by providing an algorithm for computing one. Based on Theorems 4.4 and 4.6, we devise the following algorithm.

(1) *Find an equilibrium strategy* $\boldsymbol{a}^*$ *for Alice*:
We begin with finding a mixed-strategy $\boldsymbol{a}^*$ that satisfies Theorem 4.4. Since we have from Lemma 4.2 that every $\mathrm{MaxUE}(\mathcal{T}_i, a_i)$ is increasing and uniformly continuous in a_i, there always exists a solution $\boldsymbol{a}^*$ satisfying the constraints of Theorem 4.4.[5]

[5] Note that, since $\mathrm{MaxUE}(\mathcal{T}_i, a_i)$ is not strictly increasing, the solution might not be unique. We deal with uniqueness in the subsequent theorem.

(2) *Find an equilibrium strategy* $(\boldsymbol{e}^*, \boldsymbol{\mathcal{B}}^*)$ *for Eve*:
We continue with finding a mixed-strategy $(\boldsymbol{e}^*, \boldsymbol{\mathcal{B}}^*)$ that satisfies both Lemma 4.3 and Theorem 4.6. Let $MaxUE^* = \max_i \text{MaxUE}(\mathcal{T}_i, a_i^*)$ and let I^* be the set of employees for whom the maximum is attained. If $MaxUE^* = 0$, then there is no strategy with a positive expected payoff for Eve, so we let $B_i^* \equiv 0$ for every i (and $\boldsymbol{e}^*$ can be an arbitrary distribution). Otherwise, we find a strategy which gives Eve an expected payoff of $MaxUE^*$ and which ensures that Alice will not deviate from her strategy as follows.
   (a) For every $i \notin I^*$, we let $e_i^* = 0$.
   (b) For every $i \in I^*$, we choose an arbitrary bribe value from $\text{ArgMaxBE}(\mathcal{T}_i, a_i^*)$ and let B_i^* always take this value. Finally, we let

$$e_i^* = \frac{\frac{1}{\Pr[T_i \leq B_i^*]}}{\sum_j \frac{1}{\Pr[T_j \leq B_j^*]}}. \tag{13}$$

We now prove that the mixed-strategy profile $\boldsymbol{a}^*, (\boldsymbol{e}^*, \boldsymbol{\mathcal{B}}^*)$ forms an equilibrium, i.e., that both strategies are best responses to each other. If $MaxUE^* = 0$, then we have the claim readily. Thus, we assume that $MaxUE^* > 0$. First, it is easy to see that Eve's strategy is indeed a best-response, as she targets the employees which give her maximal expected payoff and bribes them with optimal bribe values.

Second, we show that Alice's strategy is a best-response. Observe that, in Step (2b), we have chosen a distribution for Eve so that the probability of targeting and successfully bribing is uniform over employees in I^* and 0 for employees not in I^*. Since $\boldsymbol{a}^*$ satisfies Theorem 4.4, we also have that Alice shares the secret with a probability less than one with employees in I^* and with a probability of one with employees not in I^*. As Alice's best response is to share the secret with those employees whose probabilities of being targeted and successfully bribed are the lowest, we have that $\boldsymbol{a}^*$ is indeed a best response. □

The algorithm presented above proves that the game always has at least one equilibrium; however, it can also be used to compute an equilibrium strategy profile in practice. In this case, the challenge lies in finding a strategy $\boldsymbol{a}^*$ that satisfies the constraints given by Theorem 4.4 in the first step. This challenge is easily reducible to a constrained multidimensional optimization problem, for which there are many well-known numerical approximation methods. Note that we are only concerned with approximate solutions because, in practice, all of our model's initial parameters, such as the trustworthiness level distributions, would need to be approximated.

Our last set of general results provide criteria for uniqueness of Alice's equilibrium strategy and Eve's payoff.

THEOREM 4.8. *If Alice has no perfectly secure strategy, then her equilibrium strategy is unique.*

PROOF. For the sake of contradiction, suppose that the claim of the theorem does not hold, that is, there exist two distinct equilibrium strategies $\boldsymbol{a}'$ and $\boldsymbol{a}''$. From Theorem 4.4, we have that Eve's maximum payoff MaxUE for targeting an employee is uniform over the employees with whom Alice does not certainly share the secret. We let these uniform maximum payoffs for the strategies $\boldsymbol{a}'$ and $\boldsymbol{a}''$ be u' and u'', respectively. Note that, since Alice has no perfectly secure strategy, we have $u' > 0$ and $u'' > 0$.

First, suppose that $u' = u''$. Since $\text{MaxUE}(\mathcal{T}_i, a_i)$ is strictly increasing, there exists only one a_i for each i such that $\text{MaxUE}(\mathcal{T}_i, a_i) = u'$. Thus, we have $a_i' = a_i''$ for the employees who have a sharing probability lower than 1. On the other hand, the set of employees who have a sharing probability of 1 has to be equal for the two strategies, since an employee i for whom $\mathrm{MaxUE}(\mathcal{T}_i, a_i) < u' = u''$ is always in this set. Thus, we have $a'_i = a''_i$ for every employee. However, this leads to a contradiction with the initial supposition that $\boldsymbol{a}' \neq \boldsymbol{a}''$.

Second, suppose that $u' > u''$. For every employee i with $a''_i = 1$, we have $a'_i = 1$, as $\mathrm{MaxUE}(\mathcal{T}_i, a'_i) < u' < u''$. Thus, we have $a'_i = a''_i$ for these employees. On the other hand, for every employee i with $a''_i < 1$ (i.e., $\mathrm{MaxUE}(\mathcal{T}_i, a''_i) = u''$), we either have $a'_i = 1$ or $\mathrm{MaxUE}(\mathcal{T}_i, a'_i) = u'$. In the second case, we have $a'_i > a''_i$ since $\mathrm{MaxUE}(\mathcal{T}_i, a_i)$ is strictly increasing. Thus, we have $a_i > a''_i$ for these employees. However, this leads to the contradiction $1 = \sum_i a' > \sum_i a'' = 1$.

Finally, the case $u' < u''$ leads to a contradiction for the same reasons as the previous case. Therefore, the claim of the theorem has to hold. □

COROLLARY 4.9. *Eve's equilibrium payoff is always unique.*

PROOF. If Alice has only perfectly secure strategies, then Eve's equilibrium payoff is 0. On the other hand, if Alice has no perfectly secure strategy, then we have from Theorem 4.8 that her strategy and, hence, Eve's payoff has to be unique. Thus, it remains to show that non-secure and secure equilibrium strategies for Alice cannot exist at the same time.

For the sake of contradiction, suppose that this is not true, that is, there exist a non-secure and a secure strategy, denoted by $\boldsymbol{a}'$ and $\boldsymbol{a}''$. First, for every employee with $a'_i = 1$, we obviously have $a'_i \geq a''_i$. Second, for every employee with $a'_i < 1$, we have that $\mathrm{MaxUE}(\mathcal{T}_i, a'_i) > 0$ (otherwise, $\boldsymbol{a}'$ would a perfectly secure strategy as Eve's uniform maximum payoff would be 0). Since $\mathrm{MaxUE}(\mathcal{T}_i, a_i)$ is strictly increasing at a'_i, this implies that $a'_i > a''_i$ for these employees. However, this leads to the contradiction $1 = \sum_i a' > \sum_i a'' = 1$. □

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. Recall from Section 3 that we assume both players know the employees' trustworthiness level distributions, which in this case entails knowing the values of l_i and h_i for every i.

5.1. Analysis

We begin our analysis by computing Eve's optimal bribe values for a given mixed strategy $\boldsymbol{a}$ of Alice.

LEMMA 5.1. *Eve's optimal bribe values are*

$$\mathrm{ArgMaxBE}(\mathcal{T}_i, a_i) = \begin{cases} \{0\} & \textit{if } a_i < \frac{h_i}{S} \\ \{0, h_i\} & \textit{if } a_i = \frac{h_i}{S} \\ \{h_i\} & \textit{otherwise.} \end{cases} \tag{14}$$

The proof of the lemma can be found in Appendix A.

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

THEOREM 5.2. *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.*

The proof of the theorem can be found in Appendix B.

5.2. Numerical Illustrations

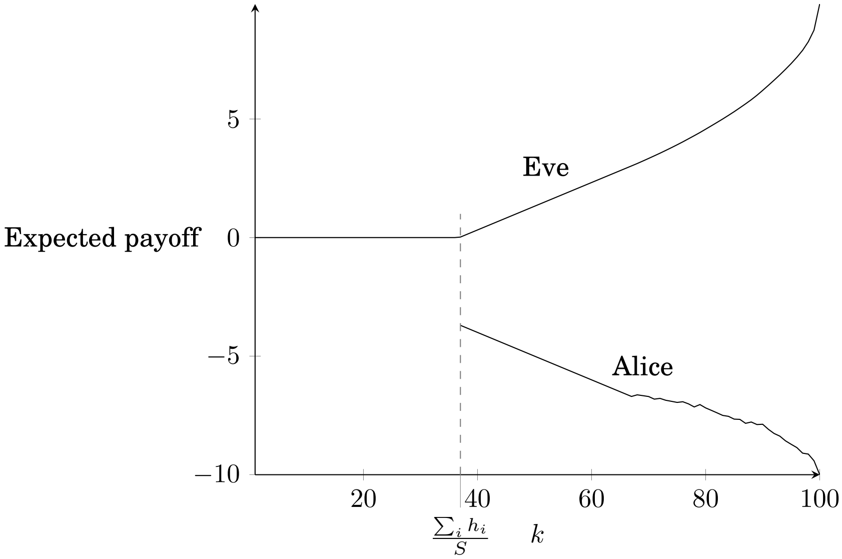


Fig. 3. The players' equilibrium payoffs as functions of the number of employees k that have to know the secret. The total number of employees is $N = 100$, the value of the secret is assumed to be $S = 10$, and the trustworthiness level of each employee i is assumed to be a random variable of the uniform distribution $\mathcal{U}(l_i, h_i)$. For this example, each parameter h_i was drawn from the set $(0, 7)$ uniformly at random. Once a parameter h_i has been drawn, its value is known to both players (recall from Section 3 that we assume both players to know the trustworthiness level distributions).

We next provide numerical illustrations for the special case in which the trustworthiness levels of the employees are modeled by independent uniform random variables T_i with parameters l_i and h_i.

Figure 3 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 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.

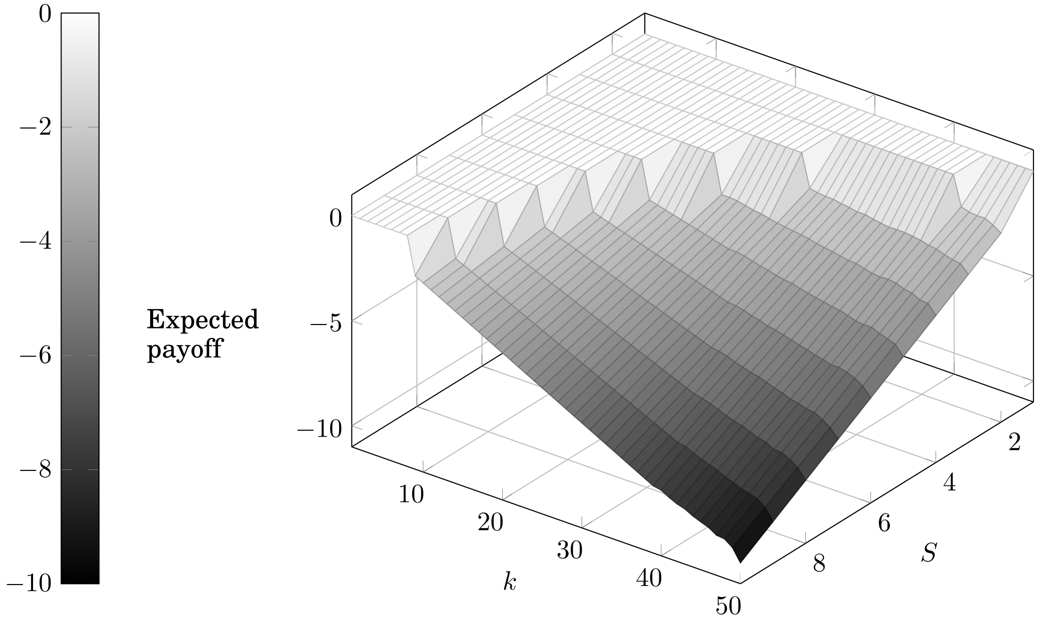
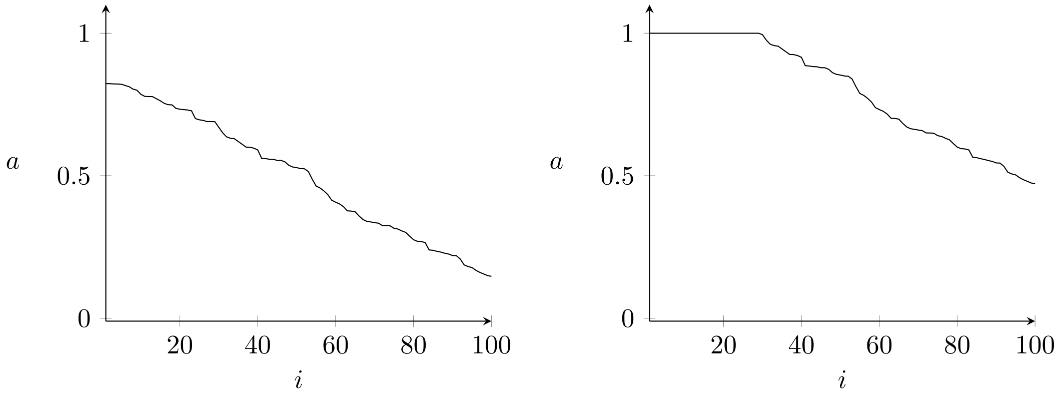


Fig. 4. Alice's equilibrium payoff for all combinations of $1 \leq k \leq 50$ and $1 \leq S \leq 10$. The parameters for this figure were generated in the same way as for Figure 3, but with $N = 50$.

Figure 4 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 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 5 shows Alice's equilibrium strategies for two different values of k. Figure 5(a) shows a case where k is small enough such that Alice does not assign probability 1 to any of her employees, while Figure 5(b) depicts a case where several employees get to know the secret with certainty. Figure 6 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 4.4. Furthermore, the figure also confirms our analytical result that no a_i can be 0.



(a) $k = 50$

(b) $k = 80$

Fig. 5. 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. Again, the parameters for this figure were generated in the same way as for Figure 3.

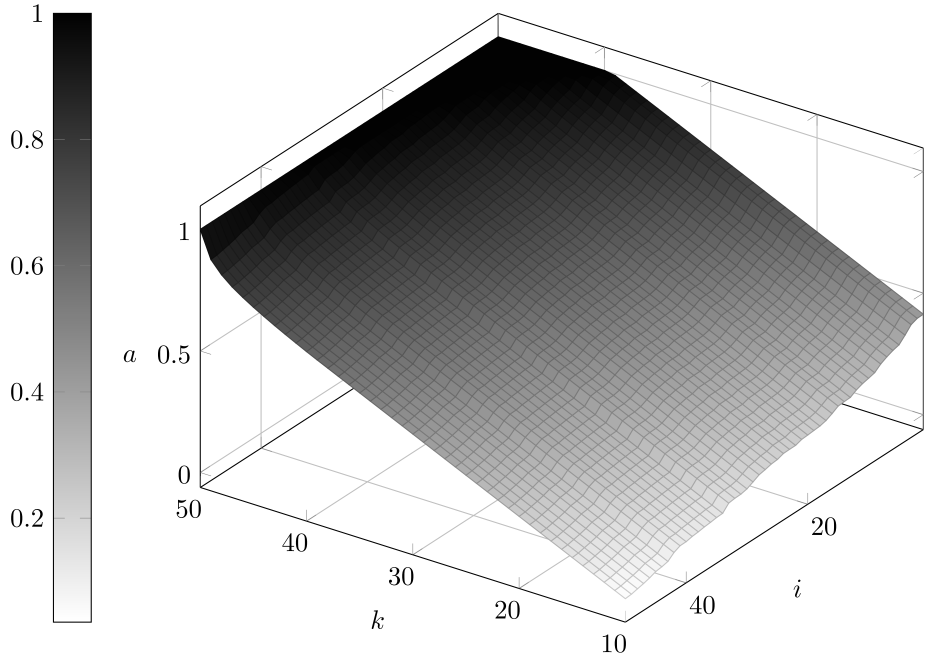


Fig. 6. Alice's equilibrium strategies for $\frac{\sum_i h_i}{S} < k \leq 50$. The parameters for this figure were generated in the same way as for Figure 3, but with $N = 50$. Again, the employees are sorted in decreasing order based on their h_i values.

6. DISCUSSION & CONCLUDING REMARKS

In this article, we present and analyze a game-theoretic model for studying the decision making of a project manager who wants to maximize the protection of organizational secrets.[6] Motivated in part by known behavioral methods of assessing trustworthiness [Munshi et al. 2012], 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 (expected) value of the organizational secret [Bontis 2001].

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 advances. 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 one. Again, this contradicts the naïve assumption that an adversary will always attack 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.[7] 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 she bribes each employee with the minimal cost 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.

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 interactions with employees. 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 article, 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.

REFERENCES

Ross Anderson. 2008. *Security engineering - A guide to building dependable distributed systems (2nd Ed.)*. Wiley.

Ross Anderson, Chris Barton, Rainer Böhme, Richard Clayton, Michel van Eeten, Michael Levi, Tyler Moore, and Stefan Savage. 2013. Measuring the Cost of Cybercrime. In *The Economics of Information Security and Privacy*, Rainer Böhme (Ed.). Springer, Berlin Heidelberg, 265–300.

Stephen Band, Dawn Cappelli, Lynn Fischer, Andrew Moore, Eric Shaw, and Randall Trzeciak. 2006. *Comparing insider IT sabotage and espionage: A model-based analysis*. Technical Report CMU/SEI-2006-TR-026. Carnegie Mellon University.

Nick Bontis. 2001. Assessing knowledge assets: A review of the models used to measure intellectual capital. *International Journal of Management Reviews* 3, 1 (2001), 41–60.

[6] The game was first introduced in a conference version of this research [Laszka et al. 2013].

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

Richard Brackney and Robert Anderson. 2004. Understanding the Insider Threat: Proceedings of a March 2004 Workshop. RAND Corporation, Santa Monica, CA.

Dawn Cappelli, Andrew Moore, Randall Trzeciak, and Timothy Shimeall. 2009. *Common Sense Guide to Prevention and Detection of Insider Threats 3rd Edition – Version 3.1*. Technical Report. Carnegie Mellon University, CERT.

Ramkumar Chinchani, Anusha Iyer, Hung Ngo, and Shambhu Upadhyaya. 2005. Towards a theory of insider threat assessment. In *Proceedings of the 2005 International Conference on Dependable Systems and Networks (DSN)*. 108–117.

Carl Colwill. 2009. Human factors in information security: The insider threat – Who can you trust these days? *Information Security Technical Report* 14, 4 (2009), 186 – 196.

Corporate Trust (Business Risk & Crisis Mgmt. GmbH). 2012. Studie: Industriespionage 2012 - Aktuelle Risiken für die deutsche Wirtschaft durch Cyberwar. (2012).

John D'Arcy, Anat Hovav, and Dennis Galletta. 2009. User Awareness of Security Countermeasures and Its Impact on Information Systems Misuse: A Deterrence Approach. *Information Systems Research* 20, 1 (March 2009), 79–98.

FBI. April 2013. The insider threat. http://www.fbi.gov/about-us/investigate/counterintelligence/insider_threat_brochure.

Federal Bureau of Investigation. 2013. Economic Espionage. http://www.fbi.gov/about-us/investigate/counterintelligence/economic-espionage. (2013).

Peter Finn. 2013. Chinese citizen sentenced in military data-theft case. http://articles.washingtonpost.com/2013-03-25/world/38006926_1_development-of-military-technologies-information-and-technologies-chinese-citizen, *Washington Post* (March 2013).

Aron Laszka, Benjamin Johnson, Pascal Schöttle, Jens Grossklags, and Rainer Böhme. 2013. Managing the Weakest Link: A Game-Theoretic Approach for the Mitigation of Insider Threats. In *Proceedings of the 18th European Symposium on Research in Computer Security (ESORICS) (Lecture Notes in Computer Science)*, Jason Crampton, Sushil Jajodia, and Keith Mayes (Eds.), Vol. 8134. Springer, Berlin Heidelberg, 273–290.

Debin Liu, XiaoFeng Wang, and L. Jean Camp. 2008. Game Theoretic Modeling and Analysis of Insider Threats. *International Journal of Critical Infrastructure Protection* 1 (Dec. 2008), 75–80. DOI: http://dx.doi.org/j.ijcip.2008.08.001

Andrew Moore, Dawn Cappelli, Thomas Caron, Eric Shaw, Derrick Spooner, and Randall Trzeciak. 2011. A Preliminary Model of Insider Theft of Intellectual Property. *Journal of Wireless Mobile Networks, Ubiquitous Computing, and Dependable Applications* 2, 1 (March 2011), 28–49.

Asmaa Munshi, Peter Dell, and Helen Armstrong. 2012. Insider Threat Behavior Factors: A comparison of theory with reported incidents. In *Proceedings of the 45th Hawaii International Conference on System Sciences (HICSS)*. 2402–2411.

Marisa Randazzo, Michelle Keeney, Eileen Kowalski, Dawn Cappelli, and Andrew Moore. 2005. *Insider threat study: Illicit cyber activity in the banking and finance sector*. Technical Report CMU/SEI-2004-TR-021. Carnegie Mellon University.

Thomas Rønde. 2001. Trade secrets and information sharing. *Journal of Economics and Management Strategy* 10, 3 (2001), 391–417.

Jerome Saltzer and Michael Schroeder. 1975. The protection of information in computer systems. *Proceedings of the IEEE* 63, 9 (1975), 1278–1308.

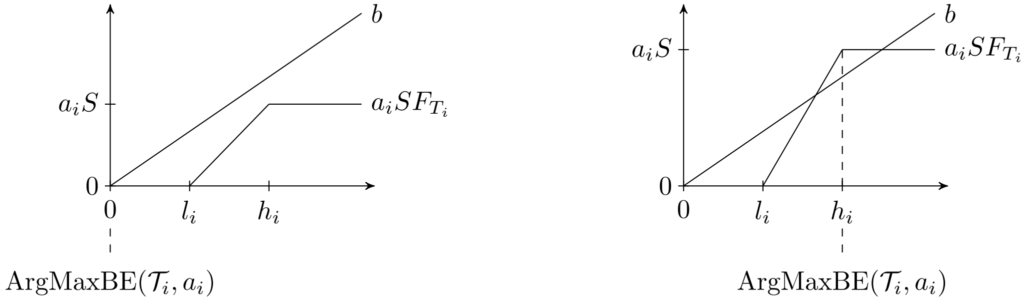
Ravi Sandhu and Pierangela Samarati. 1994. Access control: Principle and practice. *IEEE Communications Magazine* 32, 9 (1994), 40–48.

Eugene E. Schultz. 2002. A Framework For Understanding and Predicting Insider Attacks. In *Proceedings of Compsec*. London, UK, 526–531.

Detmar W. Straub and Richard J. Welke. 1998. Coping with Systems Risk: Security Planning Models for Management Decision Making. *MIS Quarterly* 22, 4 (1998), 441–469.

A. PROOF OF LEMMA 5.1

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



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

Fig. 7. Illustration of the proof of Lemma 5.1.

when targeting employee i is

$$S \cdot a_i \cdot \frac{b - l_i}{h_i - l_i} - b \ . \tag{15}$$

See Figure 7 for an illustration. When $h_i > S \cdot a_i$ (Figure 7(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. On the other hand, when $h_i < S \cdot a_i$ (Figure 7(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. □

B. PROOF OF THEOREM 5.2

PROOF. Let $\boldsymbol{a}, (\boldsymbol{e}, \boldsymbol{\mathcal{B}})$ 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, $\text{MaxUE}(\mathcal{T}_i, a_i) > \text{MaxUE}(\mathcal{T}_j, a_j)$ and, from Lemma 4.3, we have that $e_j = 0$. Furthermore, from Theorems 4.4 and 4.6, 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 4.3, 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 5.1.

— $k > \frac{\sum_i h_i}{S}$: First, it is easy to see that, for any strategy $\boldsymbol{a}$, there has to be at least one i such that $a_i > \frac{h_i}{S}$, which implies $\text{MaxUE}(\mathcal{T}_i, a_i) > 0$. 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 $\boldsymbol{a}$, 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 4.3. 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 $\boldsymbol{a}$ and $\boldsymbol{e}$ 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 $\boldsymbol{a}$ and $\boldsymbol{e}$ 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 5.1. □