Modeling Content-Adaptive Steganography with Detection Costs as a Quasi-Zero-Sum Game

Aron Laszka, *Member, IEEE,* Ádám Máté Földes, *Member, IEEE,*

***Abstract*—Recently, content-adaptive steganography was modeled by Johnson et al. as a stochastic, two-player, zero-sum game between a steganographer and a steganalyst [1]. To model economically rational steganalysts, we generalize this model by introducing a non-uniform *cost of steganalysis*. We characterize the Nash equilibria of our game based on the theory of *blocking games* [2], a class of quasi-zero-sum games, which were previously used to study the attack-resilience of systems and networks. Finally, we provide efficiently computable linear programs for finding an equilibrium. To the best of our knowledge, our paper is not only the first one to solve our generalized model, but it is also the first one to solve the original model for every possible combination of the parameter values.**

***Index Terms*—game theory, content-adaptive steganography, economics of security, information hiding.**

I. Introduction

STEGANOGRAPHY is the practice and study of techniques for hiding messages into cover media in such a way that the very existence of the messages is concealed [3], [4], [5]. Even though steganography resembles cryptographic encryption in many aspects, they are fundamentally different: the latter uses messages that are meant to be undecipherable to anyone except the intended recipient, while the former uses messages that are meant to be "invisible" to anyone except the recipient. The advantage of steganography over cryptography[1] is that, ideally, its practitioners can communicate without raising suspicion, even if their communication channel is being observed. This can be useful in many situations, for example, in countries where encryption itself is illegal.

A considerable portion of the literature on steganography (e.g., [6]) discusses specific steganographic and steganalytic methods for hiding and revealing hidden messages in specific cover media (e.g., JPEG images, MP3 audio files). However, in order to assess and quantify the security of a general class of steganographic algorithms, models must abstract away from the specifics of the carrier medium. These abstract models can be used to quantify steganographic capacity, regardless of the specifics of the employed algorithms. One of the most important common ideas in many recent algorithms is the concept of content-adaptive steganography. It is based on the observation that cover objects are usually heterogeneous in the sense that different parts have varying predictability (e.g., noisy parts of images are harder to predict). In its most basic form, which is called naïve content-adaptive embedding, the steganographer always embeds into the most unpredictable parts since changes are harder to detect there.

Several models – including ours – make use of the game theory nomenclature, and describe information hiding as a game between a steganographer (the defender) and a steganalyst (the attacker). Game-theoretic models allow the steganalyst to employ a strategy that takes the anticipated actions of the stenographer into consideration, and vice versa. The steganalyst's goal is either to detect the presence of a hidden message, or to introduce noise into the covert communication in order to disrupt it [7]. In this paper, we focus on the first problem and, consequently, assume that the attacker is passive, i.e., she does not manipulate the communication, only observes it.

A passive attacker model is described by Orsdemir et al. in [8], where the steganalyst uses a statistical classifier to distinguish between benign cover objects and objects with embedded information. Both the steganographer and the steganalyst are assumed to be able to choose a sophisticated or a naïve strategy – in other words, to anticipate or not to anticipate the other party's efforts. In [8], it is shown that a Nash equilibrium does not exist for pure strategies, but it does exist for mixed strategies, i.e., when both parties select their pure strategies at random according to distributions chosen by them beforehand.

In Ker's model [9], the steganographer is assumed to be in the possession of a set of cover media, and she is free to distribute the information to be hidden between the media arbitrarily. She transmits the – potentially information-bearing – media, and the steganalyst bases her decision on the pool of collected media and a detection threshold value. In [9], it is shown that, somewhat surprisingly, the steganographer's best strategy is either to concentrate all information into one cover or to distribute it evenly between all covers. However, a Nash equilibrium is shown to exist if the steganographer uses a mixed strategy (i.e., if she selects her parameter randomly according to a distribution chosen by her beforehand), but the steganalyst's detection threshold is constant; it is also conjectured that this is the only Nash equilibrium.

Schöttle and Böhme discuss strategic content-adaptive steganography in [10]. In their model, the steganographer first hides her message at some position in the cover, and the steganalyst then inspects a chosen position to detect the presence of the message. To model content-adaptive steganography, the steganographer uses a heterogeneity metric to compute

Submitted October 29, revised December 1, 2013.

A. Laszka (aron@laszka.hu) is with the Laboratory of Cryptography and System Security (CrySyS Lab), Department of Networked Systems and Services, Budapest University of Technology and Economics.

Á. M. Földes (foldesa@gmail.com) was with the Department of Networked Systems and Services, Budapest University of Technology and Economics, at the time of writing the paper.

[1]Note that, in practice, these approaches can and should be combined: the message is first encrypted, and then the ciphertext is embedded using some steganographic scheme.

the probability of hiding at a given position. The authors demonstrate that there exists a unique Nash equilibrium, and also show that strategic adaptive steganography is always more secure than naïvely choosing hiding positions starting with the most heterogeneous and then going to the less heterogeneous.

Johnson et al. describe a model of content-adaptive steganography, in which a steganalyst tries to infer the positions where information was hidden by the steganographer [1]. The information uses k hiding positions out of n, and the steganographer is assumed to select the actual positions according to a predictability metric. The steganalyst guesses the probable cover value for a certain position based on previously obtained information, and compares it to the observed value. The authors show that the game has a unique Nash equilibrium and propose formulas for computing the players' equilibrium strategies. The model is generalized in [11], where the steganalyst is allowed to obtain information about every bit position; however, the authors characterize the equilibrium strategies only for the special case of hiding a single bit in cover media of two bits.

Schöttle et al. study a variant of the above model, in which the steganographer chooses whether to embed in a bit position independently of the other positions, with the constraint that the expected number of embedded bits has to be k [12]. The authors show that the steganalyst's best-response strategy can be expressed as a linear aggregation threshold formula, similar to those used in practical steganalysis.

The main contributions of our paper are the following:

- We generalize the model of Johnson et al. [1] by introducing a non-uniform cost of steganalysis.
- We provide a solution based on the theory of blocking games [2] for our general model. To the best of our knowledge, our solution is the first one in the literature on steganography that is based on this class of games. Furthermore, our solution is the first one to solve the model of [1] for every possible combination of the parameter values.
- We show that in the special case of embedding one bit and zero cost of steganalysis, our solution is equivalent to that of Johnson et al. [1].

The remainder of this paper is organized as follows. First, we introduce the model of content-adaptive steganography that is used in this paper in Section II. In Section III, we summarize those results from the theory of blocking games that are essential to our analysis. In Section IV, we show how steganography can be modeled as blocking game, and provide an efficient solution for the game in Section V. In Section VI, we discuss the implications of our results. Finally, we provide concluding remarks in Section VII.

Notations: Vectors are assumed to be column vectors and denoted by bold lowercase letters (e.g., $\boldsymbol{x}$). Vectors of ones and zeros are denoted by $\mathbf{1}$ and $\mathbf{0}$, respectively (their sizes are not indicated, as they are never ambiguous in this paper). Matrices are denoted by bold uppercase letters (e.g., $\boldsymbol{\Lambda}$). The prime sign is used to denote transposition (e.g., $\boldsymbol{x}'$ or $\boldsymbol{\Lambda}'$). Elements of vectors are referred to using subindices (e.g., $\boldsymbol{x} = [x_0, \dots, x_{n-1}]'$). As an example to using these notations, consider the equality $\mathbf{1}'\boldsymbol{x} = \sum_i x_i$, which we will use repeatedly throughout this paper.

II. The Steganography Game

In this section, we first summarize the game-theoretic model of Johnson et al. [1], to which we will refer as the *basic steganography game*, and briefly discuss its previously proposed solution in Subsection II-A. Then, we generalize the model by introducing a *cost of steganalysis* in Subsection II-B.

The strategic interactions between the steganographer, whom we will call Alice, and the steganalyst, whom we will call Eve, are modeled as a two-player, zero-sum, one-shot game. In a nutshell, the game is played as follows. First, Alice embeds a k-bit hidden message into a randomly drawn cover object. Then, an unbiased coin is flipped by Nature to decide whether the original cover object or Alice's stego object is sent through the communication channel, which is being observed by Eve. Finally, Eve chooses one bit position to query, and uses side information about the most likely value of that bit to decide whether the observed object is cover or stego. If Eve's decision is right, she wins and receives a payoff of 1, while Alice receives a payoff of -1 (i.e., she loses an amount of 1). On the other hand, if Eve's decision is wrong, then vice versa (Eve receives a payoff of -1, while Alice receives 1). For an illustration of the entire steganographic system, see Figure 1 (notice that – for the sake of completeness – this figure also includes the recipient of the message, who is not part of the game).

It is assumed that both players try to maximize their respective expected payoffs (or, equivalently, minimize their expected loss). For a summary of the players' payoffs depending on what has actually been sent through the channel and what Eve's decision is, see Table I. Note that element-wise positive affine transformations on the players' payoff matrices do not change the equilibria of the game; thus, even though the payoff values might seem unrealistic for certain situations, the results apply to a wider range. Also note that this model differs from the standard model of steganography by allowing Eve to query some side information directly from the cover (for an elaborate discussion on this assumption, we refer the reader to [1]).

TABLE I
PAYOFFS FOR EVE AND ALICE

		Reality	
		cover	stego
Eve's decision	cover	$(1,-1)$	$(-1,1)$
	stego	$(-1,1)$	$(1,-1)$

Now, we discuss the model in more detail. Cover and stego objects – which can represent digital images, audio files, etc. – are assumed to consist of n bits; hence, each object is a vector $\boldsymbol{x} \in \{0,1\}^n$.[2] Cover objects are drawn by Nature from a random source $\boldsymbol{X} = [X_0, \dots, X_{n-1}]'$, which is a vector of n independent Bernoulli random variables[3]. The probability that X_i takes its more likely value is given by the function

[2]In practice, embedding is usually restricted to a subset of the bits; for example, to the least significant bits in a bitmap image. In this case, we can simply ignore all other bits and consider this restricted set to be our vector.

[3]A Bernoulli random variable's support is the set $\{0,1\}$.

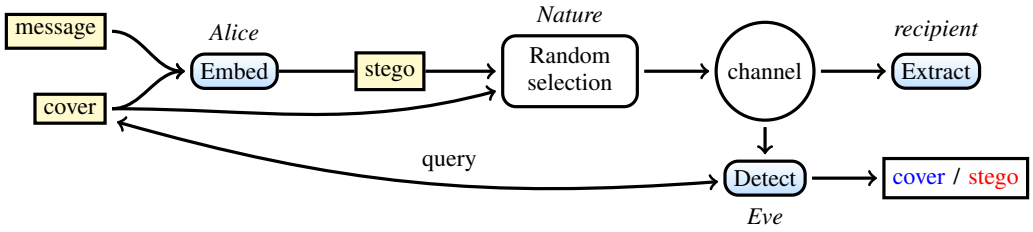

Fig. 1. Block diagram of the steganographic system.

$f(i) : \{0, \dots, n-1\} \mapsto \left[\frac{1}{2}, 1\right]$. To model content-adaptive steganography, both players are assumed to know f. Without any loss of generality, let $f(i) = P(X_i = 1)$ for the remainder of this paper.[4]

Alice embeds her k-bit message into a randomly drawn cover object $\boldsymbol{x}$ by flipping the values of $\boldsymbol{x}$ at k different positions. As she is free to choose the embedding positions, her pure strategies are k-subsets of the set of positions $\{0, \dots, n-1\}$, and her pure-strategy set is the set of all k-subsets, which will be denoted by $\mathcal{S}$. Since always embedding in the same set of positions is almost never optimal, we allow Alice to use a *mixed strategy*. When using a mixed strategy, Alice first chooses a distribution $\boldsymbol{\alpha}$ over her pure-strategy set $\mathcal{S}$ (i.e., a vector $\boldsymbol{\alpha} \in \mathbb{R}_{\geq 0}^{|\mathcal{S}|}$ satisfying $\mathbf{1}'\boldsymbol{\alpha} = 1$). Then, she embeds into a k-subset randomly chosen according to the distribution $\boldsymbol{\alpha}$.

Unfortunately, Alice's mixed-strategy space can be very complex, as her pure-strategy set is exponential in size. More specifically, the size of her pure-strategy set is the number of all k-subsets of the set of n positions, which is equal to $\binom{n}{k}$. Hence, the natural representations of her mixed strategies are vectors of length $\binom{n}{k}$. Consequently, a simpler representation of her mixed-strategy space, which has the right payoff-equivalence-class properties, was proposed in [1]. Let a_i be the probability that Alice embeds in position i. The probability of embedding in position i is equal to the sum probability of all the k-subsets that contain i; formally,

$$a_i = \sum_{S \ni i} \alpha_S \ , \tag{1}$$

where $S \ni i$ means that the sum is over all k-subsets which contain i. It can be shown that Alice's expected payoff is the same for any two mixed strategies that have the same projection $\boldsymbol{a}$. Furthermore, for every non-negative vector $\boldsymbol{a}$ that satisfies $\mathbf{1}'\boldsymbol{a} = k$ (i.e., k bits are embedded) and $\boldsymbol{a} \leq \mathbf{1}$ (i.e., probabilities can never be higher than 1), there exists a mixed strategy whose projection is $\boldsymbol{a}$ (for a constructive proof, see Appendix A). Hence, the set of these vectors can be used to represent Alice's mixed-strategy space.

Since Eve's task would be trivial if only stego objects were transmitted through the communication channel, Nature randomly selects – obviously unknown to Eve – either the original cover object or the stego object, which contains Alice's message. The probability of sending the original cover object is $\mathcal{P}_0$, while the probability of sending the stego object is $\mathcal{P}_1 = 1 - \mathcal{P}_0$. Following the convention of [13], which requires cover and stego objects to be equally likely, it is assumed that $\mathcal{P}_0 = \mathcal{P}_1 = \frac{1}{2}$.

Eve's strategy space is rather complex as her pure strategies consist of two steps: first, selecting a position to query and, then, deciding if the observed object is cover or stego. However, once a position has been chosen and its value has been observed[5], Eve's optimal decision becomes trivial. For $\mathcal{P}_0 = \mathcal{P}_1 = \frac{1}{2}$, it can be shown that Eve's optimal decision rule $\text{Decision}(x_i)$ is

$$\text{Decision}(x_i) = \begin{cases} \text{cover} & \text{if } x_i = 1 \ , \\ \text{stego} & \text{if } x_i = 0 \ . \end{cases} \tag{2}$$

Thus, the only strategic choice Eve has to make is to pick one position, as her decision will then be trivial based on the above optimal decision rule. Consequently, Eve's pure-strategy set can be simplified to the set of positions $\{0, \dots, n-1\}$, and her mixed strategies can be represented by distributions over the set of positions (i.e., a mixed strategy is a vector $\boldsymbol{\beta} \in \mathbb{R}_{\geq 0}^{n}$ satisfying $\mathbf{1}'\boldsymbol{\beta} = 1$).

A. A Solution for the Basic Steganography Game

Johnson et al. proposed a solution for – what we call here – the basic steganography game in [1]. In this paper, we show that this solution covers the special case of $k = 1$ (i.e., hiding only a single bit), but does not always work in the more general case $k \geq 1$ (i.e., hiding an arbitrary number of bits). Here, we briefly discuss the main ideas of this solution (for a more detailed discussion, we refer the interested reader to [1]) and show examples where the solution does not work.

Let *Eve's local advantage* at position i be the product of the probability a_i that Alice embeds in position i and the value $f(i) - \frac{1}{2}$. It can be shown that Eve's local advantage is proportional to her expected payoff for querying a given position (hence the name). Consequently, in a best response, Eve always queries the position (or positions) where her local advantage attains its maximum. Since, Alice's loss is equal to Eve's payoff in the basic game, Alice tries to minimize the maximum of Eve's local advantage.

Now, assume that there exists a mixed strategy for Alice such that Eve's local advantage is uniform over the positions. Then, it can be shown that this uniform local advantage

[4]Note that the convention $f(i) = P(X_i = 1)$ is indeed without loss of generality, as we can easily swap the definitions of 0 and 1 in the cover source.

[5]More precisely, Eve observes whether the value is the more likely one or the other. However, since the more likely value is assumed to be 1 to simplify our notations, Eve simply observes whether it is 1 or 0.

strategy is Alice's unique optimal strategy. For the sake of contradiction, suppose that this is not true, that is, there exists an optimal strategy $\boldsymbol{a}^*$ where Eve's local advantage is not uniform. Let I be the set of positions where the local advantage attains its maximum, and j be a position where it attains its minimum. However, this leads to a contradiction as Alice could decrease the maximum of Eve's local advantage by decreasing a_i^* for every $i \in I$ and increasing a_j^* at the same time. Thus, this uniformity constraint is indeed necessary. Finally, it can also be shown that the constraint is sufficient as well.

However, in the case of $k > 1$, the existence of a strategy satisfying the uniform local advantage constraint is not guaranteed. As a simple example, let $n = 3$, $k = 2$, $f(0) = \frac{5}{8}$, and $f(1) = f(2) = \frac{7}{8}$. Then, even if Alice hides in position 0 with probability 1, the local advantage $\frac{1}{8}$ at position 0 is still less than the average local advantage $\frac{3}{16}$ at the other positions; thus, a strategy with uniform local advantage cannot exist.

More generally, we can show that there exist an infinite number of counterexamples.

Lemma 1. *Let $k = 2$, $f(0) = \frac{1}{2} + \varepsilon$, and $f(i) = 1 - \varepsilon$ for every $i > 0$, where ε is an arbitrary number in $(0, \frac{1}{2})$. Then, no strategy satisfying the uniform local advantage constraint can exist if $n < \frac{1}{2\varepsilon}$.*

Notice that the threshold for n grows without bound as ε approaches zero.

Proof. Assume that $n < \frac{1}{2\varepsilon}$. Then, $\varepsilon < \frac{1}{2n}$.

To prove that no strategy with a uniform local advantage can exist, we now show that the local advantage at position 0 is always less than the average. First, consider the extreme case of $a_0 = 1$. In this case, the local advantage at position 0 is the highest possible value, which is $1 \cdot \left(\frac{1}{2} + \varepsilon - \frac{1}{2}\right) = \varepsilon$. Since the sum of the probabilities of the remaining positions is 1 and f is uniform $1 - \varepsilon$ over them, the sum of their local advantages is $1 \cdot \left(1 - \varepsilon - \frac{1}{2}\right) = \frac{1}{2} - \varepsilon$. Hence, the average local advantage over all positions is

$$\frac{\varepsilon + \frac{1}{2} - \varepsilon}{n} = \frac{1}{2n} \ . \tag{3}$$

By combining this with $\varepsilon < \frac{1}{2n}$, we have that the local advantage at position 0 is strictly less than the average. Finally, it is obvious that in every strategy that assigns a probability smaller than 1 to position 0, the local advantage at position 0 is even smaller compared to the average. □

B. Cost of Steganalysis

In the basic steganography game, the steganalyst is interested solely in minimizing her decision error, without any regard to the cost of her operation. In other words, the basic steganography game assumes that the steganalyst acts as if she bears zero cost. However, in practice, the cost of steganalysis is non-zero: operation and maintenance costs of the system performing steganalysis, cost of acquiring side information, cost of implementing detection algorithms against new steganographic techniques, etc. These costs might seem negligible at first compared to the payoff for a successful detection, but as the probability of detection decreases and the size of the steganalytic system increases, the cost of steganalysis can exceed the expected payoff. Furthermore, the cost of steganalysis might be non-uniform over the set of positions; for example, the cost of acquiring side information might be different for the EXIF and the image data of a JPEG file. Consequently, an economically rational player 1) might use a querying strategy which differs from the optimal strategy of the zero-cost case or 2) might decide not to perform steganalysis at all if it is economically infeasible. To model an economically rational steganalyst, in this subsection, we generalize the basic steganography game by introducing a cost of steganalysis.

We assume that querying and predicting a given position i requires some effort or expenditure from Eve in the amount of $\mu_i \in \mathbb{R}_{\geq 0}$, where μ_i can depend on the position i. Thus, her payoff is $1 - \mu_i$ when she makes the right decision and $-1 - \mu_i$ when she does not. Note that

- the cost has to be paid by Eve in advance, regardless of whether she will be successful in detecting Alice or not.
- The cost does not affect Alice's payoff directly. However, it might affect her strategy indirectly through changing Eve's optimal strategy.
- The cost can depend only on which position is chosen, but not on the value of the bit at the chosen position.
- The basic steganography game is the special case $\boldsymbol{\mu} = \mathbf{0}$.

III. Blocking Games

Blocking games are *"quasi"-zero-sum*[6] games that model the strategic interactions between a defender, who requires a set of resources to perform her task, and an adversary, who is capable of carrying out availability (or denial-of-service) attacks against the resources [2]. The first blocking game was proposed by Gueye et al. in [14] to study the problem of designing attack-resilient network topologies. The general concept of blocking games was introduced in [2] to allow studying a wider range of security and availability problems. In this section, we summarize the results of [2], on which our analysis is built. Note that we will use the blocking game terminology instead of the steganographic throughout this section (e.g., defender, adversary, and resources instead of steganographer, steganalyst, and bit positions), and we will connect the two in the next section.

A blocking game is a one-shot, two-player game between a defender and an adversary. The defender has a non-empty set of resources E available to her. To perform her task, she has to select a collection of resources $S \subseteq E$; however, she cannot choose any collection of resources, only those that are feasible for her task. This non-empty set of feasible collections is denoted by $\mathcal{S} = \{S_1, \ldots, S_N\}$, where each $S_i \subseteq E$; hence, the defender's pure-strategy set is $\mathcal{S}$. Meanwhile, the adversary targets a resource $e \in E$ to be attacked in order to disrupt the task of the defender; hence, the adversary's pure-strategy set is E. To successfully carry out her attack against resource e, the adversary has to spend μ_e, which is called the cost of attack.

[6]The reason for calling these games quasi-zero-sum will soon be discussed.

Since sufficiently high costs can make all attacks unprofitable for the adversary, she also has the option of not attacking.

The players' payoffs are determined by a loss function $\lambda(S,e) : \mathcal{S} \times E \mapsto \mathbb{R}_{\geq 0}$. When the defender selects collection S and the adversary targets resource e, the defender's payoff is $-\lambda(S,e)$ (in other words, her loss is $\lambda(S,e)$) and the adversary's payoff is $\lambda(S,e) - \mu_e$ (i.e., the loss caused to the defender minus the cost of the attack). It is often assumed that there is no loss when the adversary targets a resource that is not used by the defender; in other words, the value of the loss function is zero when $e \notin S$. Notice that the game would be zero sum if there were no attack costs (i.e., if $\boldsymbol{\mu} = 0$), since the sum of the players' payoffs would be $-\lambda(S,e) + \lambda(S,e) = 0$. The model generalizes zero-sum games by adding an extra term to one player's payoff; hence, we call the resulting game quasi-zero-sum. These games are more amenable to theoretical analysis than general, non-zero-sum (or – equivalently – non-constant-sum) games.

As a simple example to illustrate blocking games, consider a local area network that is attacked by a strategic adversary. The defender (i.e., the network operator) has to maintain loop-free connectivity between the network nodes using the set of available network links E. For this, she selects a spanning tree $S \subseteq E$ as the communications infrastructure; hence, the set of feasible collections $\mathcal{S}$ is the set of all spanning trees. A selected spanning tree can be implemented in practice, for example, as the forwarding table entries of the network switches. Meanwhile, the adversary targets a link $e \in E$ and pays the cost μ_e of attacking it. Attacking a link can be implemented in practice, for example, as physical destruction. Finally, the loss is $\lambda(S,e) = 1$ if the adversary manages to disconnect the network (i.e., if $e \in S$), and $\lambda(S,e) = 0$ otherwise (i.e., if $e \notin S$).

Since pure strategies are almost never optimal in blocking games, the players are allowed to employ mixed strategies. The defender can choose a distribution $\boldsymbol{\alpha}$ over the set of feasible collections $\mathcal{S}$, while the adversary can choose a distribution $\boldsymbol{\beta}$ over the set of resources E. Then, the selected collection S and the targeted resource e are drawn randomly from the chosen distributions $\boldsymbol{\alpha}$ and $\boldsymbol{\beta}$.

The goal of blocking game analysis is to 1) compute the adversary's equilibrium payoff and to 2) find optimal defender and adversarial strategies. The characterization of the Nash equilibria of blocking games established in [2] builds on the theory of blocking pairs of polyhedra (BPP). Here, we introduce the concepts of BPP that are essential for understanding this characterization, and refer the interested reader to [15] for a more detailed discussion. The *polyhedron* $P_{\boldsymbol{\Lambda}}$ of a nonnegative $N \times m$ matrix $\boldsymbol{\Lambda}$ is defined as the vector sum of the convex hull of the rows $\boldsymbol{\lambda}_1, \ldots, \boldsymbol{\lambda}_N$ of $\boldsymbol{\Lambda}$ and the nonnegative orthant; formally, $P_{\boldsymbol{\Lambda}} = \text{conv.hull}(\boldsymbol{\lambda}_1, \ldots, \boldsymbol{\lambda}_N) + \mathbb{R}^m_{\geq 0}$. In other words, the polyhedron $P_{\boldsymbol{\Lambda}}$ consists of vectors which are the sums of a convex linear combination of the rows of $\boldsymbol{\Lambda}$ and a non-negative vector. The *blocker* $bl(P_{\boldsymbol{\Lambda}})$ of $P_{\boldsymbol{\Lambda}}$ is defined as

$$bl(P_{\boldsymbol{\Lambda}}) = \left\{ \boldsymbol{y} \in \mathbb{R}^m_{\geq 0} \;\middle|\; \forall \boldsymbol{x} \in P_{\boldsymbol{\Lambda}} : \; \boldsymbol{x}'\boldsymbol{y} \geq 1 \right\} \; . \tag{4}$$

Alternatively, the blocker $bl(P_{\boldsymbol{\Lambda}})$ can also be defined as the set of vectors that "block" every row of $\boldsymbol{\Lambda}$; formally, $bl(P_{\boldsymbol{\Lambda}}) = \{\boldsymbol{y} \in \mathbb{R}^m_{\geq 0} \; : \; \boldsymbol{\Lambda}\boldsymbol{y} \geq \mathbf{1}\}$. Note that the blocker of a polyhedron itself is also a polyhedron.

Now, let $\boldsymbol{\Lambda}$ be the loss (i.e., negative payoff) matrix of the defender; formally, $\Lambda_{S,e} = \lambda(S,e)$. Using the notation introduced above, $P_{\boldsymbol{\Lambda}}$ is the polyhedron associated with $\boldsymbol{\Lambda}$, and $bl(P_{\boldsymbol{\Lambda}})$ is the blocker of $P_{\boldsymbol{\Lambda}}$. Before characterizing the equilibria of the blocking game using its blocker $bl(P_{\boldsymbol{\Lambda}})$, we have to introduce a few more concepts. First, let $\Omega = \{\boldsymbol{\omega}_1, \ldots, \boldsymbol{\omega}_K\}$ be the set of the extreme points of the blocker $bl(P_{\boldsymbol{\Lambda}})$. For a vector $\boldsymbol{y} \in bl(P_{\boldsymbol{\Lambda}})$, let the quantity $\theta(\boldsymbol{y})$ be

$$\theta(\boldsymbol{y}) = \frac{1}{\boldsymbol{y}'\mathbf{1}} \left(1 - \boldsymbol{y}'\boldsymbol{\mu}\right) \; , \tag{5}$$

and let $\theta_{max} = \max_{\boldsymbol{y} \in bl(P_{\boldsymbol{\Lambda}})} \theta(\boldsymbol{y})$. It was shown that the maximum θ_{max} is attained at an extreme point (or at some extreme points) of the blocker; that is, $\max_{\boldsymbol{y} \in bl(P_{\boldsymbol{\Lambda}})} \theta(\boldsymbol{y}) = \max_{\boldsymbol{\omega} \in \Omega} \theta(\boldsymbol{\omega})$. Finally, let Ω_{max} denote the set of extreme points for which the maximum is attained; formally, $\Omega_{max} = \{\boldsymbol{\omega} \in \Omega \mid \theta(\boldsymbol{\omega}) = \theta_{max}\}$.

Theorem 1 (Gueye [2])**.** *For the general blocking game, the following always hold.*

1) *If $\theta_{max} \leq 0$, then not attacking is always optimal for the adversary.*
2) *If $\theta_{max} \geq 0$, then for every probability distribution $\boldsymbol{\gamma}$ over Ω_{max}, the adversary's strategy $\boldsymbol{\beta}$ defined by*

$$\beta_e = \sum_{\boldsymbol{\omega} \in \Omega_{max}} \gamma_{\boldsymbol{\omega}} \frac{\omega_e}{\boldsymbol{\omega}'\mathbf{1}} \tag{6}$$

is in Nash equilibrium with any strategy $\boldsymbol{\alpha}$ of the defender that satisfies the following properties:

$$\begin{cases} \sum_{S \in \mathcal{S}} \alpha_S \lambda(S,e) - \mu_e = \theta_{max}, & \forall e \in E \text{ s. t. } \beta_e > 0 \; , \\ \sum_{S \in \mathcal{S}} \alpha_S \lambda(S,e) - \mu_e \leq \theta_{max}, & \forall e \in E \; . \end{cases}$$

Furthermore, there exists at least one such strategy $\boldsymbol{\alpha}$. The corresponding payoffs are θ_{max} for the adversary and $\sum_{\boldsymbol{\omega} \in \Omega_{max}} \frac{\gamma_{\boldsymbol{\omega}}}{\boldsymbol{\omega}'\mathbf{1}}$ for the defender.
3) *If $\boldsymbol{\mu} = \mathbf{0}$, every Nash equilibrium pair of strategies is of the above type.*

For the proof of the theorem, see [2].

Recall that the goal of blocking game analysis is to compute the adversary's equilibrium payoff and a pair of equilibrium strategies. If the defender's payoff matrix $\boldsymbol{\Lambda}$ is explicitly given, this problem can easily be formulated as a linear program, which can be solved efficiently. However, in most models, the input of the computational problem is not the payoff matrix itself, but some implicit definition of it. For example, in the simple network blocking game used above as an illustration, the input of the problem is the network graph, and even the set of feasible collections is only implicitly given as the set of spanning trees. This can lead to a very challenging computational problem, as the number of feasible collections can be exponential in the size of the input. For example, the number of spanning trees in a complete graph of only 60 network nodes is $60^{58} \approx 1.36 \times 10^{103}$, which is several orders of magnitude larger than the number of atoms in the observable universe.

Unfortunately, there is no general algorithm for solving a blocking game in polynomial time given such an implicit definition of the feasible collections. However, for a number of models, the game can be solved efficiently using various tricks (for examples of polynomial-time solutions, see [14], [16]). In Section V, we will show that this is possible for the steganography game as well.

IV. Modeling Steganography as a Blocking Game

In this section, we show that the steganography game can be formulated as a blocking game, and provide a characterization of the game's blocker $bl(P_{\mathbf{\Lambda}})$. This formulation will allow us to solve the game not only for the general case (arbitrary k) of the basic steganography game (zero cost of steganalysis $\boldsymbol{\mu} = \mathbf{0}$), but also for our generalized model (arbitrary cost of steganalysis $\boldsymbol{\mu} \geq \mathbf{0}$).

First, to simplify our formulas, we introduce the *bias function*

$$\tilde{f}(i) = 2f(i) - 1 \ . \tag{7}$$

Since $f(i)$ was defined as the probability of the more likely outcome of bit i, the bias function $\tilde{f}$ can be interpreted as the predictability of bit i. If $\tilde{f}(i) = 0$, then $f(i) = \frac{1}{2}$; hence, bit i is an unbiased coin flip. On the other hand, if $\tilde{f}(i) = 1$, then $f(i) = 1$, which means that bit i is completely deterministic (always takes its more likely value).

Before we can formulate steganography as a blocking game, we have to prove the following lemma.

Lemma 2. *If Alice embeds in subset S and Eve queries position i, Alice's expected payoff is*

$$\begin{cases} 0 & \text{if } \ i \notin S \ , \\ -\tilde{f}(i) & \text{if } \ i \in S \ . \end{cases} \tag{8}$$

Proof. We prove the two cases separately.

- $i \notin S$: Recall that Eve's optimal decision rule is to guess cover when bit i takes its more likely value ($x_i = 1$), and to guess stego when it takes its less likely value ($x_i = 0$). Consequently, when a cover object is transmitted, Eve's decision will be right iff bit $x_i = 1$; thus, her chance of winning is $f(i)$ in this case. On the other hand, when a stego object is transmitted, Eve's decision will be right iff $x_i = 0$; thus, her chance of winning is $1 - f(i)$ in this case. By combining the two cases, we have that the probability of Eve winning is
$$\frac{1}{2}f(i) + \frac{1}{2}\left(1 - f(i)\right) = \frac{1}{2} \ . \tag{9}$$
Since Alice's payoff is -1 if Eve's decision is right and 1 if it is not, Alice's payoff is $\frac{1}{2} \cdot (-1) + \frac{1}{2} \cdot 1 = 0$.
- $i \in S$: When a cover object is transmitted, the probability that Eve's decision will be right is $f(i)$ for the same reasons as in the previous case. However, when a stego object is transmitted, bit i has been flipped by Alice. Consequently, Eve will make the right decision iff the bit had taken its more likely value before being flipped; thus, her probability of winning is $f(i)$. By combining these two cases, we have that the probability of Eve's decision being right is
$$\frac{1}{2}f(i) + \frac{1}{2}f(i) = f(i) \ . \tag{10}$$
Therefore, Alice's payoff is $f(i) \cdot (-1) + (1 - f(i)) \cdot 1 = 1 - 2f(i) = -\tilde{f}(i)$.

□

We can now formulate steganography as a blocking game as follows.

- First, let the *set of resources* E be the *set of bits* $\{0, \ldots, n-1\}$.
- Let the role of the *defender* be played by *Alice*, the steganographer. Let *selecting a collection* S of the resources represent *embedding into the subset* S of bits. Since Alice always embeds into k bits, the set of feasible resource collections $\mathcal{S}$ is the set of all k-subsets.
- Let the role of the *adversary* be played by *Eve*, the steganalyst. Let *targeting one of the resources* represent *querying the corresponding bit* (and deciding whether she sees a stego or a cover object).
- Finally, let the *cost of attack* μ_i be the *cost of steganalysis* introduced in Subsection II-B.

In contrast to conventional blocking games, the payoff for a given pure-strategy profile (S, i) in the steganography game is a random variable, not a constant value. However, since both players try to maximize their expected payoffs, we can define the value of the loss function $\lambda(S, i)$ using the expected payoffs for a given strategy profile (S, i). Thus, based on Lemma 2, we have that the loss function of the steganography game is

$$\lambda(S, i) = \begin{cases} 0 & \text{if } \ i \notin S \ , \\ \tilde{f}(i) & \text{if } \ i \in S \ . \end{cases} \tag{11}$$

To apply Theorem 1, we have to characterize the blocker $bl(P_{\mathbf{\Lambda}})$ of the steganography game.

Theorem 2. *The blocker $bl(P_{\mathbf{\Lambda}})$ of the steganography game can be characterized as*

$$bl(P_{\mathbf{\Lambda}}) = \Big\{ \boldsymbol{y} \in \mathbb{R}^n_{\geq 0} \ \Big| \ \exists K \in \mathbb{R}_{\geq 0}, \boldsymbol{z} \in \mathbb{R}^n_{\geq 0} : \\ kK - \mathbf{1}'\boldsymbol{z} \geq 1 \ \wedge \ \forall i \left(K \leq z_i + \tilde{f}(i) y_i \right) \Big\} \ . \tag{12}$$

Proof. We prove Equation (12) in two steps.

- Righ-hand side (RHS) of Equation (12) $\subseteq bl(P_{\mathbf{\Lambda}})$: We have to show that every element of the RHS of Equation (12) is also an element of the blocker $bl(P_{\mathbf{\Lambda}})$ (i.e., "blocks" every vector in the polyhedron $P_{\mathbf{\Lambda}}$). Consider an arbitrary element $\boldsymbol{y}$ of the RHS of Equation (12). Since every vector of the polyhedron $P_{\mathbf{\Lambda}}$ is a linear combination of the rows of $\mathbf{\Lambda}$ (plus a non-negative vector), it suffices to show that $\boldsymbol{y}$ "blocks" every row of $\mathbf{\Lambda}$ to prove that $\boldsymbol{y}$ "blocks" every vector of the polyhedron $P_{\mathbf{\Lambda}}$. Formally, it suffices to show that, for every row $\boldsymbol{\lambda}_S$ of $\mathbf{\Lambda}$, it holds that $\boldsymbol{\lambda}'_S \boldsymbol{y} \geq 1$. Equivalently, we have to show that $\min_{\boldsymbol{\lambda}_S} \boldsymbol{\lambda}'_S \boldsymbol{y} \geq 1$.
Now, consider an arbitrary row $\boldsymbol{\lambda}_S$ of $\mathbf{\Lambda}$. Let the vector $\boldsymbol{a} \in \{0, 1\}^n$ be such that $a_i = 0$ if the ith element of

$\boldsymbol{\lambda}_S$ is zero, and $a_i = 1$ otherwise (i.e., $\boldsymbol{a}$ is an "indicator vector" of the non-zero elements of $\boldsymbol{\lambda}_S$). By definition, we have that the ith element of $\boldsymbol{\lambda}_S$ is $\tilde{f}(i)$ if $i \in S$, and it is 0 otherwise. Consequently, since S is a k-subset, it holds that $\mathbf{1}'\boldsymbol{a} = k$.

Now, because there exists such a vector $\boldsymbol{a}$ for every row $\boldsymbol{\lambda}_S$, we have that $\min_{\boldsymbol{\lambda}_S} \boldsymbol{\lambda}_S'\boldsymbol{y}$ is greater than or equal to the value of the following integer linear program:

$$\text{Minimize } \sum_i a_i \tilde{f}(i) y_i \tag{13}$$

subject to

$$\mathbf{1}'\boldsymbol{a} = k \ , \tag{14}$$

where $\boldsymbol{a} \in \{0,1\}^n$. Thus, it suffices to show that the value of this integer program is at least 1.

By relaxing some constraints of a minimization problem, its value can only decrease, but never increase. Consequently, the value of the following relaxed linear program is a lower bound of the above program:

$$\text{Minimize } \sum_i a_i \tilde{f}(i) y_i \tag{15}$$

subject to

$$\mathbf{1}'\boldsymbol{a} \geq k \tag{16}$$

$$\boldsymbol{a} \leq \mathbf{1} \ , \tag{17}$$

where $\boldsymbol{a} \in \mathbb{R}^n_{\geq 0}$ (notice that the variables are non-integer in the relaxed program).

Finally, the dual of the above relaxed linear program is the following:

$$\text{Maximize } kK - \mathbf{1}'\boldsymbol{z} \tag{18}$$

subject to

$$\forall i: \ K \leq z_i + \tilde{f}(i) y_i \ , \tag{19}$$

where $K \in \mathbb{R}_{\geq 0}$ and $\boldsymbol{z} \in \mathbb{R}^n_{\geq 0}$. As $\boldsymbol{y}$ satisfies the RHS of Equation (12), there exists a $K \in \mathbb{R}_{\geq 0}$ and $\boldsymbol{z} \in \mathbb{R}^n_{\geq 0}$ that satisfy $\forall i \left(K \leq z_i + \tilde{f}(i) y_i\right)$ and $kK - \mathbf{1}'\boldsymbol{z} \geq 1$. Since there exists a solution $k, \boldsymbol{z}$ for which the objective function attains 1, the value of the above dual program and, hence, all former linear programs is at least 1. Therefore, any element $\boldsymbol{y}$ of the RHS is also an element of the blocker, and RHS of Equation (12) $\subseteq bl(P_{\boldsymbol{\Lambda}})$ has to hold.

- $bl(P_{\boldsymbol{\Lambda}}) \subseteq$ RHS of Equation (12): We have to show that every element of the blocker $bl(P_{\boldsymbol{\Lambda}})$ is also an element of the RHS of Equation (12). For the sake of contradiction, suppose that the claim $bl(P_{\boldsymbol{\Lambda}}) \subseteq$ RHS of Equation (12) does not hold. In other words, suppose that there is an element $\boldsymbol{y} \in bl(P_{\boldsymbol{\Lambda}})$ for which no $K \in \mathbb{R}_{\geq 0}$ and $\boldsymbol{z} \in \mathbb{R}^n_{\geq 0}$ can exist that satisfy the constraints of the RHS of Equation (12) with $\boldsymbol{y}$. Then, consider the following linear program:

$$\text{Maximize } kK - \mathbf{1}'\boldsymbol{z} \tag{20}$$

subject to

$$\forall i: \ K \leq z_i + \tilde{f}(i) y_i \ , \tag{21}$$

where $K \in \mathbb{R}_{\geq 0}$ and $\boldsymbol{z} \in \mathbb{R}^n_{\geq 0}$. It is easy to see that the value of the above linear program has to be less than 1. If this were not true, then a solution (i.e., some K and $\boldsymbol{z}$) which attained a value of 1 would exist. But this solution would satisfy the constraints of the RHS of Equation (12) with $\boldsymbol{y}$, which would lead to a contradiction with our initial supposition. Thus, the value of the above linear program has to be strictly less than 1.

Next, consider the dual of the above linear program:

$$\text{Minimize } \sum_i a_i \tilde{f}(i) y_i \tag{22}$$

subject to

$$\mathbf{1}'\boldsymbol{a} \geq k \tag{23}$$

$$\boldsymbol{a} \leq \mathbf{1} \ , \tag{24}$$

where $\boldsymbol{a} \in \mathbb{R}^n_{\geq 0}$. Since the value of this linear program is less than 1, we have that $\sum_i a_i \tilde{f}(i) y_i < 1$ for every optimal solution $\boldsymbol{a}$.

Now, let $\boldsymbol{a}^*$ be an optimal solution for which $\mathbf{1}'\boldsymbol{a}^* = k$ holds (since the dual is a minimization problem and the sign of $\boldsymbol{a}$ is positive in the objective function, there always exists at least one such solution). Notice that $\boldsymbol{a}^*$ is a mixed strategy for Alice (see Section II). Finally, let $\boldsymbol{\alpha}$ be a distribution corresponding to $\boldsymbol{a}^*$; that is, let $\boldsymbol{\alpha}$ be such that $a_i^* = \sum_{S \ni i} \alpha_S$. Then, the vector $\sum_S \alpha_S \boldsymbol{\lambda}_S$ is an element of the polyhedron $P_{\boldsymbol{\Lambda}}$ by definition, but it is not blocked by $\boldsymbol{y}$, since $\left(\sum_S \alpha_S \boldsymbol{\lambda}_S\right)' \boldsymbol{y} = \sum_i a_i \tilde{f}(i) y_i < 1$. However, this leads to contradiction with our initial supposition that $\boldsymbol{y} \in bl(P_{\boldsymbol{\Lambda}})$. Therefore, the claim $bl(P_{\boldsymbol{\Lambda}}) \subseteq$ RHS of Equation (12) has to hold.

Finally, from RHS of Eq. (12) $\subseteq bl(P_{\boldsymbol{\Lambda}})$ and $bl(P_{\boldsymbol{\Lambda}}) \subseteq$ RHS of Eq. (12), it follows readily that there has to be an equality between these two sets. □

V. Solving the Game Efficiently

In the previous section, we gave a characterization of the blocker of the steganography game. In theory, this characterization combined with Theorem 1 can be used to compute an equilibrium for a given instance $(n, \boldsymbol{f})$ of the game. However, performing this computation in polynomial time, which is the criterion for feasibly computable according to the Cobham-Edmonds thesis, is not straightforward due to the exponential size of Alice's strategy set.

First, applying Theorem 1 directly by finding the maximum of the function $\theta(\boldsymbol{\omega})$ over the set of all extreme points $\boldsymbol{\omega}$ is not feasible, as the number of extreme points to be enumerated is generally exponential. Second, solving the game in the more conventional way of finding optimal strategies using linear programs is also infeasible. Even though linear programs can be solved in polynomial time, in the case of the steganography game, the input of the problem is not the payoff matrix, but the description of the cover source $(n, \boldsymbol{f})$. As Alice's strategy set is exponential in the size of this input and it has to appear

either as a set of constraints or as a set of variables, the size of the resulting linear program is also exponential. Consequently, the running time of this approach is generally exponential.

As an example of how infeasible this is in practice, consider the problem of hiding messages of $k = 20$ bits into covers of $n = 200$ bits. The number of possible embeddings and, hence, the number of variables (or constraints) is $\binom{n}{k} = \binom{200}{20} \approx 1.6 \times 10^{27}$. Solving linear programs of this size would be a very challenging problem to say the least.

Therefore, in this section, we provide linear programs of polynomial size for efficiently computing 1) Eve's equilibrium payoff and 2) a pair of equilibrium strategies for Alice and Eve. To formulate our linear programs, we build on both the general description of the equilibria of blocking games (Theorem 1) and our characterization of the blocker of the steganography game (Theorem 2).

Theorem 3. *Given an instance $(n, \boldsymbol{f})$ of the steganography game, Eve's equilibrium payoff and a pair of equilibrium strategies $(\boldsymbol{a} \in \mathbb{R}^n_{\geq 0}, \boldsymbol{\beta} \in \mathbb{R}^n_{\geq 0})$ for Alice and Eve can be computed in polynomial time.*

Proof. We begin with computing Eve's equilibrium payoff. According to Theorem 1, her equilibrium payoff is $\theta_{max} = \max_{\boldsymbol{y} \in bl(P_{\boldsymbol{\Lambda}})} \theta(\boldsymbol{y})$. Our goal is to formulate this problem as a linear program of polynomial size. First, observe that we already have the constraints of our linear program from Theorem 2, which characterizes $bl(P_{\boldsymbol{\Lambda}})$ using a set of linear constraints:

$$\boldsymbol{y} \in bl(P_{\boldsymbol{\Lambda}}) \tag{25}$$

iff

$$kK - \boldsymbol{z}'\mathbf{1} \geq 1 \tag{26}$$

$$\forall i: \; K \leq z_i + \tilde{f}(i) y_i \; , \tag{27}$$

where $K \in \mathbb{R}_{\geq 0}$ and $\boldsymbol{z} \in \mathbb{R}^n_{\geq 0}$.

Unfortunately, the desired objective function $\theta(\boldsymbol{y}) = \frac{1}{\mathbf{1}'\boldsymbol{y}} (1 - \boldsymbol{\mu}'\boldsymbol{y})$ cannot be expressed as a linear function in $\boldsymbol{y}$ because of the division by $\mathbf{1}'\boldsymbol{y}$. Therefore, we have to "scale" our variables. First, we introduce a new variable ϕ, which will be equal to $\frac{1}{\mathbf{1}'\boldsymbol{y}}$. Then, we divide the existing variables (K, $\boldsymbol{z}$, and $\boldsymbol{y}$) and inequalities by $\mathbf{1}'\boldsymbol{y}$; that is, we multiply them by ϕ. We will denote the scaled versions of K and $\boldsymbol{z}$ with the same letters, but we will denote $\frac{\boldsymbol{y}}{\mathbf{1}'\boldsymbol{y}}$ by $\boldsymbol{\beta}$.[7] Using these scaled variables, our problem can be formulated as the following linear program:

$$\text{Maximize } \phi - \boldsymbol{\mu}'\boldsymbol{\beta} \tag{28}$$

subject to

$$\mathbf{1}'\boldsymbol{\beta} = 1 \tag{29}$$

$$kK - \boldsymbol{z}'\mathbf{1} \geq \phi \tag{30}$$

$$\forall i: \; K \leq z_i + \tilde{f}(i)\beta_i \; , \tag{31}$$

where $K, \phi \in \mathbb{R}_{\geq 0}$ and $\boldsymbol{\beta}, \boldsymbol{z} \in \mathbb{R}^n_{\geq 0}$. First, observe that, as expected, the objective function is $\phi - \boldsymbol{\mu}'\boldsymbol{\beta} = \frac{1}{\mathbf{1}'\boldsymbol{y}} - \boldsymbol{\mu}'\left(\frac{1}{\mathbf{1}'\boldsymbol{y}}\boldsymbol{y}\right) = \frac{1}{\mathbf{1}'\boldsymbol{y}}(1 - \boldsymbol{\mu}'\boldsymbol{y}) = \theta(\boldsymbol{y})$. Second, notice that the constant 1 in the first constraint is replaced by ϕ due to the scaling. Finally, notice that we have to introduce a new constraint to ensure that $\mathbf{1}'\boldsymbol{\beta} = \mathbf{1}'\frac{\boldsymbol{y}}{\mathbf{1}'\boldsymbol{y}} = 1$ holds.

Let $\boldsymbol{\beta}^*$, ϕ^*, K^*, and $\boldsymbol{z}^*$ be an optimal solution to the above linear program. Since the linear program has a polynomial number of variables and constraints in the size of the input, we can compute Eve's equilibrium payoff $\theta_{max} = \phi^* - \boldsymbol{\mu}'\boldsymbol{\beta}^*$ efficiently using any standard linear program solver.

Next, we show how to find a pair of equilibrium strategies. Since an optimal solution $\boldsymbol{\beta}^*$ is a non-negative vector of length n that sums up to 1, it can be interpreted as a probability distribution over the set of positions. We now show that $\boldsymbol{\beta}^*$ is actually an equilibrium strategy for Eve. If Eve employs $\boldsymbol{\beta}^*$ as her strategy, then Alice's expected loss for embedding in position i is $\tilde{f}\beta_i$. Thus, Alice's best response is a mixed strategy $\boldsymbol{a}$ that minimizes $\sum_i \tilde{f}(i)\beta_i a_i$. We can formulate the problem of finding a best response as the following linear program:

$$\text{Minimize } \sum_i \tilde{f}(i)\beta^*_i a_i \tag{32}$$

subject to

$$\mathbf{1}'\boldsymbol{a} = k \tag{33}$$

$$\boldsymbol{a} \leq \mathbf{1} \; , \tag{34}$$

where $\boldsymbol{a} \in \mathbb{R}^n_{\geq 0}$. The dual of the above linear program is:

$$\text{Maximize } kK - \mathbf{1}'\boldsymbol{z} \tag{35}$$

subject to

$$\forall i: \; K \leq z_i + \tilde{f}(i)\beta^*_i \; , \tag{36}$$

where $K \in \mathbb{R}_{\geq 0}$ and $\boldsymbol{z} \in \mathbb{R}^n_{\geq 0}$. Since the objective function for the solution K^* and $\boldsymbol{z}^*$ is ϕ^*, the value of the linear program is at least ϕ^*, which means that Alice's loss is at least ϕ^* regardless of her strategy. Thus, if Eve uses the strategy $\boldsymbol{\beta}^*$, her payoff is at least the equilibrium payoff $\phi^* - \boldsymbol{\mu}'\boldsymbol{\beta}^*$ regardless of Alice's strategy.

Finally, we show how to find an equilibrium strategy for Alice. Consider the following linear program:

$$\text{Minimize } L \tag{37}$$

subject to

$$\mathbf{1}'\boldsymbol{a} = k \tag{38}$$

$$\boldsymbol{a} \leq \mathbf{1} \tag{39}$$

$$\forall i: \; \tilde{f}(i) a_i \leq L + \mu_i \; , \tag{40}$$

where $L \in \mathbb{R}_{\geq 0}$ and $\boldsymbol{a} \in \mathbb{R}^n_{\geq 0}$. Let $\boldsymbol{a}^*$ be an optimal solution to the above linear program. Using linear programming duality in the same way as before, it can be shown that if Alice employs $\boldsymbol{a}^*$ as her strategy, her loss is at most ϕ_{opt} regardless of Eve's strategy. Consequently, as the number of variables and constraints in the above program is polynomial in the size of the input, we can compute an equilibrium strategy $\boldsymbol{a}^*$ for Alice in polynomial time. □

[7] The reason for denoting the scaled version of $\boldsymbol{y}$ with $\boldsymbol{\beta}$ will be revealed soon.

A. Special Case: Basic Steganography Game

Johnson et al. proposed a solution in [1] for – what we call here – the basic steganography game, which is the special case of $\boldsymbol{\mu} = \mathbf{0}$ (i.e., zero cost of steganalysis). In Subsection II-A, we have shown that this solution does not always work for arbitrary $k \geq 1$. Here, we show that for the special case of $k = 1$, this solution is equivalent to ours.[8]

Recall from Subsection II-A that the solution proposed by Johnson et al. is based on the idea of uniform local advantage. From this criterion, they derive closed-form formulas for both players' equilibrium strategies. Quite surprisingly, the formulas for Alice's and Eve's equilibrium strategies are almost identical (and identical for $k = 1$):

$$a_i = \frac{\frac{k}{\tilde{f}(i)}}{\sum_j \frac{1}{\tilde{f}(j)}} \quad \text{and} \quad \beta_i = \frac{\frac{1}{\tilde{f}(i)}}{\sum_j \frac{1}{\tilde{f}(j)}} \ . \tag{41}$$

Now, we solve our linear programs for the special case of $k = 1$ and $\boldsymbol{\mu} = \mathbf{0}$. By substituting k for 1 and $\boldsymbol{\mu}$ for $\mathbf{0}$, the linear program for finding an equilibrium strategy for Eve simplifies to:

$$\text{Maximize } \phi \tag{42}$$

subject to

$$\mathbf{1}'\boldsymbol{\beta} = 1 \tag{43}$$

$$K - \boldsymbol{z}'\mathbf{1} \geq \phi \tag{44}$$

$$\forall i: \ K \leq z_i + \tilde{f}(i)\beta_i \ , \tag{45}$$

where $K, \phi \in \mathbb{R}_{\geq 0}$ and $\boldsymbol{\beta}, \boldsymbol{z} \in \mathbb{R}^n_{\geq 0}$. Notice that, in an optimal solution, it always holds that $\phi = K - \boldsymbol{z}'\mathbf{1}$. Thus, we can eliminate ϕ and reformulate the program as:

$$\text{Maximize } K - \boldsymbol{z}'\mathbf{1} \tag{46}$$

subject to

$$\mathbf{1}'\boldsymbol{\beta} = 1 \tag{47}$$

$$\forall i: \ K \leq z_i + \tilde{f}(i)\beta_i \ , \tag{48}$$

where $K \in \mathbb{R}_{\geq 0}$ and $\boldsymbol{\beta}, \boldsymbol{z} \in \mathbb{R}^n_{\geq 0}$. Next, we show that there always exists an optimal solution where $\boldsymbol{z} = \mathbf{0}$. Let K^* and $\boldsymbol{z}^*$ be an optimal solution (with some $\boldsymbol{\beta}^*$). Then, consider the solution $\hat{K} = K^* - \max_i z_i^*$ and $\hat{\boldsymbol{z}} = \mathbf{0}$ (with the same $\boldsymbol{\beta}^*$). First, it is easy to see that the value of the objective function for $\hat{K}$ and $\hat{\boldsymbol{z}}$ is at least as high as for K^* and $\boldsymbol{z}^*$ (strictly higher if there are multiple non-zero elements in $\boldsymbol{z}^*$). Second, the solution $\hat{K}$ and $\hat{\boldsymbol{z}}$ (with $\boldsymbol{\beta}^*$) does not violate the constraints, as we decrease the left-hand side of Equation (48) by at least as much as its right-hand side. Thus, there always exists an optimal solution to the linear program where $\boldsymbol{z} = \mathbf{0}$. Consequently, we can reformulate the program as:

$$\text{Maximize } K \tag{49}$$

subject to

$$\mathbf{1}'\boldsymbol{\beta} = 1 \tag{50}$$

$$\forall i: \ K \leq \tilde{f}(i)\beta_i \ , \tag{51}$$

where $K \in \mathbb{R}_{\geq 0}$ and $\boldsymbol{\beta} \in \mathbb{R}^n_{\geq 0}$. Finally, the above linear program is obviously equivalent to the problem of maximizing $\min_i \tilde{f}(i)\beta_i$ subject to $\mathbf{1}'\boldsymbol{\beta} = 1$.

We now show that the optimal solution to the above maximization problem is the same as the solution to Equation (41). First, in an optimal solution, $\tilde{f}(i)\beta_i$ has to be uniform over the positions. Otherwise, decreasing β_i where $\tilde{f}(i)\beta_i$ attains its maximum and increasing β_i where $\tilde{f}(i)\beta_i$ attains its minimum would increase the value of the objective function $\min_i \tilde{f}(i)\beta_i$ while still satisfying the constraint $\mathbf{1}'\boldsymbol{\beta} = 1$. Thus, we have that $\tilde{f}(0)\beta_0 = \ldots = \tilde{f}(n-1)\beta_{n-1}$. Consequently, there has to exist a constant $C \in \mathbb{R}_{\geq 0}$ such that, for every i, we have $\beta_i = \frac{C}{\tilde{f}(i)}$ and $\sum_i \beta_i = 1$. It is easy to see that $C = \frac{1}{\sum_i \frac{1}{\tilde{f}(i)}}$ is the only solution satisfying these constraints. Therefore, the optimal solution to our linear program is equal to the solution of Equation (41). Moreover, using a similar argument, it can be shown that Alice's equilibrium strategy is also equal to the solution of Equation (41).

However, in the more general case of $k > 1$, the solution of Equation (41) may differ from the solutions of our linear programs. In these cases, the former solution assigns probabilities that are greater than 1 to some of the positions. For examples of such instances, see Subsection II-A.

VI. Discussion

The solution proposed by Johnson et al. has an interesting implication. Observe that the "shapes" of the players' optimal strategies do not change as the number of embedded bits k increases, only the steganographer's strategy is "scaled up" linearly. Consequently, steganalyst's payoff (i.e., success rate) increases linearly with the number of embedded bits [1]. This rule deviates from the square root law of steganographic capacity, which predicts asymptotically quadratic advantage, even for homogeneous covers [17]. However, based on our solution proposed in this paper, we can refine the above rule as follows. First, for smaller numbers of embedded bits (that is, when our solution coincides with that of [1]), the steganalyst's success rate increases linearly indeed. For larger numbers, however, the success rate can increase superlinearly, as the steganographer has to deviate from the strategy described by Equation (41) (since she can longer find a strategy satisfying the uniform local advantage constraint).

Our generalization introducing a potentially non-uniform and non-zero cost of steganalysis has two important implications. First, if $\theta_{max} < 0$, then not performing steganalysis is the only optimal strategy for the attacker. In other words, if the expected cost of steganalysis (see Section II-B) is always higher than the expected reward for successful detections, an economically rational attacker should choose not to perform steganalysis. While this might seem counter-intuitive at first, the operation and maintenance costs can actually be very high when looking for a "needle in a haystack", and the would-be steganalyst should look for other means of catching the steganographer.

Second, if the cost of one (or some) of the positions is relatively high compared to the costs of the other positions, then it can be optimal for the steganalyst to never query that (or

[8]Note that, in this subsection, we restrict ourselves to $\boldsymbol{\mu} = \mathbf{0}$ (i.e., the basic steganography game), since the solution of [1] was devised for this model.

those) position(s) but query others. As a very simple example, consider the game of hiding $k = 2$ bits in covers of $n = 3$ bits, where the predictability is uniform $\tilde{f} \equiv \frac{1}{2}$, and the costs are $\mu_0 = 1$ and $\mu_1 = \mu_2 = 0$. By solving our linear program, we can compute that the steganalyst's mixed-strategy is $\beta = \left[0, \frac{1}{2}, \frac{1}{2}\right]'$. In other words, she should never query the first position.

VII. Conclusions and Future Work

In this paper, we have proposed a generalization of the basic steganography game [1]: to model an economically rational steganalyst, we have introduced a potentially non-uniform and non-zero cost of steganalysis. We have proposed a novel solution for the generalized model based on the theory of blocking games, and have shown how an equilibrium can be computed efficiently using linear programs. Finally, we have compared our solution with the solution of Johnson et al. [1].

Our research can be extended into multiple directions. The limitation of the discussed model (and solution) is the assumption that the steganalyst can query only a single position. This constraint can be justified by arguing that the steganalyst is budget- or resource-constrained. However, as the cost of steganalysis may be non-uniform, even a budget-constrained steganalyst might be able to query multiple low-cost positions instead of one high-cost position. Thus, the model (and its solution) could be extended to the general case of querying multiple positions.

References

[1] B. Johnson, P. Schöttle, and R. Böhme, "Where to hide the bits?" in *Proceedings of the 3rd Conference on Decision and Game Theory for Security (GameSec)*, 2012, pp. 1–17.

[2] A. Gueye, "A game theoretical approach to communication security," Ph.D. dissertation, University of California, Berkeley, Electrical Engineering and Computer Sciences, March 2011.

[3] S. Katzenbeisser and F. A. Petitcolas, Eds., *Information hiding techniques for steganography and digital watermarking*. Artech House, 2000.

[4] I. Cox, M. Miller, J. Bloom, J. Fridrich, and T. Kalker, *Digital watermarking and steganography*. Morgan Kaufmann, 2007.

[5] A. D. Ker, P. Bas, R. Böhme, R. Cogranne, S. Craver, T. Filler, J. Fridrich, and T. Pevnỳ, "Moving steganography and steganalysis from the laboratory into the real world," in *Proceedings of the 1st ACM Workshop on Information Hiding and Multimedia Security (IH&MMSEC)*, 2013, pp. 45–58.

[6] A. Westfeld and A. Pfitzmann, "Attacks on steganographic systems," in *Proceedings of the 3rd International Workshop on Information Hiding*, 1999, pp. 61–76.

[7] J. M. Ettinger, "Steganalysis and game equilibria," in *Proceedings of the 2nd International Workshop on Information Hiding*, 1998, pp. 319–328.

[8] A. Orsdemir, H. O. Altun, G. Sharma, and M. F. Bocko, "Steganalysis aware steganography: Statistical indistinguishability despite high distortion," in *Security, Forensics, Steganography, and Watermarking of Multimedia Contents X*, vol. 6819. SPIE, 2008.

[9] A. D. Ker, "Batch steganography and the threshold game," in *Security, Steganography, and Watermarking of of Multimedia Contents IX*, vol. 6505. SPIE, 2007, pp. 401–413.

[10] P. Schöttle and R. Böhme, "A game-theoretic approach to content-adaptive steganography," in *Proceedings of the 14th Information Hiding Conference*, 2012, pp. 125–141.

[11] B. Johnson, P. Schöttle, A. Laszka, J. Grossklags, and R. Böhme, "Bitspotting: Detecting optimal adaptive steganography," in *Proceedings of the 12th International Workshop on Digital-Forensics and Watermarking (IWDW)*, 2013.

[12] P. Schöttle, A. Laszka, B. Johnson, J. Grossklags, and R. Böhme, "A game-theoretic analysis of content-adaptive steganography with independent embedding," in *Proceedings of the 21st European Signal Proceedings Conference (EUSIPCO)*, 2013.

[13] S. Katzenbeisser and F. Petitcolas, "Defining security in steganographic systems," in *Security and Watermarking of Multimedia Contents IV*, vol. 4675. SPIE, April 2002, pp. 50–56.

[14] A. Gueye, J. Walrand, and V. Anantharam, "Design of network topology in an adversarial environment," in *Proceedings of the 1st Conference on Decision and Game Theory for Security (GameSec)*, 2010, pp. 1–20.

[15] D. R. Fulkerson, "Blocking and anti-blocking pairs of polyhedra," *Mathematical Programming*, vol. 1, no. 1, pp. 168–194, 1971.

[16] A. Laszka, D. Szeszlér, and L. Buttyán, "Game-theoretic robustness of many-to-one networks," in *Proceedings of the 3rd International Conference on Game Theory for Networks (GameNets)*, 2012, pp. 88–98.

[17] A. D. Ker, T. Pevnỳ, J. Kodovskỳ, and J. Fridrich, "The square root law of steganographic capacity," in *Proceedings of the 10th ACM Workshop on Multimedia and Security*, 2008, pp. 107–116.

Appendix

A. Algorithm for Computing a Mixed-Strategy from Embedding Probabilities

Theorem 4. *For every vector $\boldsymbol{a}$ of embedding probabilities satisfying $\sum_i a_i = k$, there exists a mixed strategy $\boldsymbol{\alpha}$ for Alice (i.e., a distribution over k-subsets) such that the projection of $\boldsymbol{\alpha}$ is $\boldsymbol{a}$.*

We provide a constructive proof, which is based on a polynomial-time algorithm. It is noteworthy that the mixed strategy output by the algorithm is a fairly simple one: its support consist of at most n sets.

Proof. We prove the theorem by providing an algorithm that can compute a mixed strategy $\boldsymbol{\alpha}$ from any vector of probabilities $\boldsymbol{a}$ satisfying $\sum_i a_i = k$.

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 $i \in I$, $a_i - p \geq 0$ 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, $a_i \leq 1$. 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 $a_i \leq k'$ for every i, where $k' = \sum_i a_i / k$. From this, we can formulate the MaxProb constraint as $p \leq \sum_j a_j / k - a_i$. Finally, we call a vector $\boldsymbol{a}$ proper if, for every i, $a_i \geq 0$ and $a_i \leq k'$. 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. □

Aron Laszka received the B.Sc. degree with honors and M.Sc. degree with honors in computer engineering from the Budapest University of Technology and Economics, Hungary, in 2009 and 2011. He is currently pursuing the Ph.D. degree in computer science under the supervision of Prof. Levente Buttyán, with an expected graduation date of early 2014. He has been a graduate student member of the HSN Laboratory since 2011.

In 2013, he was a visiting research scholar at the Pennsylvania State University, under the supervision of Prof. Jens Grossklags. His research interests include studying the robustness of network topologies, game-theoretic analysis of steganography, modeling the mitigation of covert compromises using games of timing, and security of interdependent information systems.

Ádám M. Földes received his MSc in Technical Informatics from the Budapest University of Technology and Economics, Hungary, in 2009. He wrote his master's thesis about the development of an application that leverages steganographic techniques to hide information in a pool of files on a file system. Between 2009 and 2013, he was a PhD student at the Dept. of Telecommunications of the same university with the focus area of privacy-enhancing technologies. He is currently working for the R&D department of Ericsson, Inc. in Silicon Valley. Ádám has been an editor of the International PET Portal and Blog since 2008.