Game Theory and Adaptive Steganography

Pascal Schöttle and Rainer Böhme

***Abstract*—According to conventional wisdom, content-adaptive embedding offers more steganographic security than random uniform embedding. We scrutinize this view and note that it is barely substantiated in the literature as only recently adaptive steganographic systems are tested against an attacker who anticipates the adaptivity and incorporates this knowledge into her detection strategy. For a better theoretical understanding of strategical embedding and detection, we propose a game-theoretic framework to study adaptive steganography while taking the knowledge of the steganalyst into account. We instantiate the framework with a stylized cover model and study both parties' optimal strategies. The model has a unique equilibrium in mixed strategies, which depends on the heterogeneity of the cover source. We add realism by introducing imperfect recoverability of the adaptivity criterion and prove that naïve adaptive embedding—the strategy implemented in many practical schemes—is only optimal if perfect steganography is possible or if the adaptivity criterion is not recoverable at all. In practice, where steganography is imperfect and adaptivity criteria are partially recoverable, the optimal embedding strategy is between naïve adaptive and random uniform embedding.**

***Index Terms*—Adaptive Steganography, Game Theory, Security**

I. Introduction

STEGANOGRAPHY enables *undetectability*, the protection goal associated with concealing the very existence of a secret message by hiding it in inconspicuous cover data, such as digital media [1]. In a very general sense, cover objects are points in a high-dimensional space, which is partitioned, often key-dependent, into disjoint regions that map to the elements of the hidden message space. A (minimal) steganographic embedding function takes as inputs a message and a key. It outputs a point within the associated region. Steganalysis, the counter-technology, tries to detect hidden messages by *deciding* whether an observed object is "plausible", i. e., if it is drawn from the distribution governing the cover generation process.

Steganography is perfect if the embedding function preserves the cover distribution [2]. This requires knowledge of the distribution or a sampler and computational effort exponential in the size of the message space. However, for empirical covers like digital media, the cover distribution is unknown (and arguably unknowable [3]). In practice, the high-dimensional space is sparsely populated with empirical covers and the hidden message space is too large for *rejection sampling*, a method that draws covers until one is found in the desired region [4]. Therefore, the standard approach in steganography is to take a given cover and move it into the region of the hidden message by slightly modifying its coordinates (e. g., pixel values of an image, samples of an audio file).

 The authors are with the Department of Computer Science, Universität Innsbruck, Austria, e-mail: pascal.schoettle@uibk.ac.at; rainer.boehme@uibk.ac.at.

Simple coding allows the steganographer to partition the high-dimensional space over the message space such that embedding a given message has many possible solutions [5], [6], [7]. *Adaptive embedding* (also known as content-adaptive) increases the steganographic security by selecting a solution that moves the cover along those dimensions of the high-dimensional space that reveal the least information about the fact that a message has been embedded to a potential attacker (called steganalyst).

Because neither the steganographer nor the steganalyst know the cover distribution, both must resort to local models of the unknown joint distribution and make local decisions. This leaves both parties with choices. In adaptive embedding, the steganographer chooses along which dimensions the cover should be moved to the message region. The steganalyst chooses element weights to aggregate local evidence into a global decision. Both choices are clearly interdependent and jointly affect the security of the steganographic communication. Therefore, both choices have to be strategic, i. e., anticipating the opponent's choice. This suggests that adaptive steganography and optimal adaptive steganalysis are best studied in the context of game theory, a well-established framework to model situations in which two (or more) parties act strategically [8].

This article extends our seminal work on adaptive steganography and game theory [9] and makes several contributions. Consistent with recent empirical results [10], [11], the theoretical analysis of the model we propose predicts that adaptive steganography does not improve security against a strategic adversary. In addition, using the solution concept of Nash equilibria, we can identify the *optimal adaptive embedding* strategy, which maximizes the security against detectors that anticipate adaptive embedding. We define heterogeneous cover sources, and show that if they do not allow perfect embedding, this strategy is strictly superior to naïve adaptive and random uniform embedding, commonly used in practice.

Specifically, these results are derived from a universal framework for the theoretical analysis of adaptive steganography. By instantiating this framework, we introduce a stylized model of a cover source. This model captures important characteristics of real covers but is simple enough to obtain closed-form solutions to the resulting game for a fixed local embedding operation and a fixed (locally optimal) detection rule. For the sake of simplicity, earlier models assumed that the steganalyst is capable of perfectly recovering the most likely embedding positions. We relax this assumption by adding the *recovery rate* to our model, which expresses the fraction of embedding positions the steganalyst is able to recover. This brings the game-theoretic models one step closer to reality.

Here is the outline of this article: Section II defines the general game-theoretic framework including terminology and

notation. Section III presents our specific model and proves its relevant properties. Section IV derives an analytical solution under the assumption of perfect recoverability, which is then relaxed in Section V. The main results are discussed with numerical examples in Section VI. We establish relations to prior art in Section VII. Section VIII concludes.

II. Framework

We first define our framework formally and then illustrate it with an example for image steganography in Section II-E. We refer to [12] for a gentle introduction to digital steganography.

A. Notation

We write random variables as upper-case letters, their realizations (and constants) in lower case. Vectors, shorthand for one-dimensional arrays, are typeset boldface $\boldsymbol{x} = (x_0, \dots, x_{n-1})$, with n implicit. Following the notation in [1], superscript (0) in $x_i^{(0)}$ denotes a symbol before embedding and superscript (1) in $x_i^{(1)}$ denotes a symbol after embedding. By extension, superscript $(\bar{a})$ in $x_i^{(\bar{a})}$ means that the symbol has been changed by embedding with probability $\bar{a}$ and $\boldsymbol{x}_{(i)}^{(1)}$ denotes a stego object where only position i has been changed. $\mathcal{P}_0$ is the probability distribution of the cover source. $\mathcal{P}_1$ is the probability distribution of stego objects. $\mathcal{P}_{(x_i)}$ is the probability distribution after embedding only in the i-th element. We use the standard notation for Binomial coefficients, i. e., $\binom{n}{k} := \frac{n!}{k!(n-k)!}$.

B. Definitions

To study adaptive steganography in a general framework, we formally define its key components.

Definition 1 (Cover)**.** *A vector $\boldsymbol{x^{(0)}} = (x_0^{(0)}, \dots, {x_{n-1}}^{(0)})$ of n discrete symbols is called* cover*, if it is a realization of the* cover source $\boldsymbol{X}^{(0)}$ *drawn according to $\mathcal{P}_0$. Every symbol ${x_i}^{(0)}$ of the cover can take values from the cover alphabet $\mathbb{X}$.*

An *embedding function* is a key-dependent mapping of cover $\boldsymbol{x^{(0)}}$ and message to a stego object $\boldsymbol{x^{(1)}}$. To study adaptive embedding, we decompose the embedding function into atomic operations that modify individual cover symbols.

Definition 2 (Embedding Operation)**.** *A function $\mathrm{emb}(\cdot)$ that takes as input a cover symbol $x_i^{(0)}$ and outputs the corresponding symbol $x_i^{(1)}$ with different steganographic semantic is called* embedding operation*.*

Without loss of generality, we assume a one-to-one mapping between cover symbols and bits carrying steganographic semantic. These bits are typically an encrypted and encoded representation of the message.

For a given $\mathcal{P}_0$ and a uniform prior over the key space, $\mathcal{P}_1$ depends on the embedding function. The *Kullback–Leibler divergence* (KLD) between $\mathcal{P}_0$ and $\mathcal{P}_1$ is an information-theoretic measure of steganographic security with regard to undetectability [2]. We leverage this to distinguish between homogeneous and heterogeneous cover sources.

Definition 3 (Homogeneous vs Heterogeneous Cover Source)**.** *Cover source $\boldsymbol{X}^{(0)}$ is called* homogeneous *for a fixed embedding operation, if for every $i, j \in \{0, \dots, n-1\}, i \neq j$, and for any subset of the cover space and the corresponding subsets of the stego space, it holds that $\mathrm{KLD}(\mathcal{P}_0, \mathcal{P}_{(x_i)}) = \mathrm{KLD}(\mathcal{P}_0, \mathcal{P}_{(x_j)})$. Otherwise, $\boldsymbol{X}^{(0)}$ is called* heterogeneous*.*

This definition implies that homogeneous cover sources offer the same security regardless of *where* the embedding changes are made. For typical embedding operations, all i. i. d. and the Markov cover models in [13] are homogeneous cover sources. Because adaptive steganography exploits variations in detectability between embedding positions, we need to model heterogeneous cover sources. In this case, the security impact of changing individual cover symbols may depend on the realization $\boldsymbol{x^{(0)}}$. We define a notion of *suitability* for embedding per position and per cover by decomposing the KLD measure into differences in the likelihood of hypothetical stego objects.

Definition 4 (Suitability)**.** *Position i of cover $\boldsymbol{x}^{(0)}$ is more suitable for embedding than position j, if the stego object $\boldsymbol{x}_{(i)}^{(1)}$ is a more likely realization of the cover distribution $\mathcal{P}_0$ than the stego object $\boldsymbol{x}_{(j)}^{(1)}$, i. e., if $\mathcal{P}_0\left(\boldsymbol{x}_{(i)}^{(1)}\right) > \mathcal{P}_0\left(\boldsymbol{x}_{(j)}^{(1)}\right)$.*

This definition is agnostic about multiple embedding changes appearing together, a common assumption in the literature [14].

Since $\mathcal{P}_0$ is unknown for empirical cover sources, practical adaptive embedding functions use an *adaptivity criterion* to approximate the suitability of individual embedding positions.

Definition 5 (Adaptivity Criterion)**.** *A family of tractable functions, e. g., $\zeta_i : \mathbb{X}^n \times \boldsymbol{\Theta} \to \mathbb{R}$, is called* adaptivity criterion *if it establishes an order of all n embedding positions in a cover $\boldsymbol{x^{(0)}}$ by their approximate suitability. More specifically, $\zeta_i(\boldsymbol{x^{(0)}}, \boldsymbol{\theta}) > \zeta_j(\boldsymbol{x^{(0)}}, \boldsymbol{\theta})$ implies that, to the best of the steganographer's knowledge, position i appears more suitable for embedding than position j.*

Definitions 4 and 5 require some reflection.

Remark 1. *The adaptivity criterion may use side information $\boldsymbol{\theta} \in \boldsymbol{\Theta}$ to improve the quality of the approximation.*

In [6], for example, a steganographic method in the JPEG domain is presented, where $\boldsymbol{\theta}$ stems from the never-compressed image. This enables embedding in coefficients that are close to the boundary of quantization intervals. This side information is neither available to the recipient nor the attacker. The selection channel, a coding technique, and its generalizations ensure that the recipient does not need to know the embedding positions to extract the message [6].

Remark 2. *The mere order relation in Definition 5 ignores quantitative differences in the likelihoods of Definition 4.*

Remark 3. *The assumption of a complete order is a simplification. Some practical schemes establish partial orders and resolve them with random (key-dependent) tie-breaking rules.*

The framework is sufficiently expressive to study canonical strategies. Replacing the order with a quantitative detectability profile [14] or more realistic non-linear distortion functions is

formally straightforward, but depends on detailed knowledge of the specific cover source. The simplifications here allow us to use a handy convention: we write $\boldsymbol{y}^{(\mathbf{0})}$ for a cover $\boldsymbol{x}^{(\mathbf{0})}$ with symbols ordered by *decreasing* suitability for embedding, i. e., $\zeta_{i-1}(\boldsymbol{y}^{(\mathbf{0})}, \boldsymbol{\theta}) \geq \zeta_i(\boldsymbol{y}^{(\mathbf{0})}, \boldsymbol{\theta})$ for $1 \leq i < n-1$. Of course, the stego object is transmitted with its symbols in original order.

However, stego objects $\boldsymbol{x}^{(\mathbf{1})}$ often leak information about the values of ζ to the steganalyst, who can thus infer likely embedding positions and (partially) *recover* the order of $\boldsymbol{y}^{(\mathbf{0})}$. We use the hat notation to express the steganalyst's estimation $\hat{\zeta}$ of the values of ζ. Similarly, let $\hat{\boldsymbol{y}}^{(\mathbf{1})}$ be the stego object with the recovered order of symbols. We say that an adaptivity criterion is *perfectly recoverable* if $\hat{\boldsymbol{y}}^{(\mathbf{1})} = \boldsymbol{y}^{(\mathbf{1})}$. The framework is agnostic about quantifying this information leakage. Deviations from perfect recovery are best studied in the context of specific models (see Section V for an example).

C. The Adaptive Steganography Game

Let *Alice* be the steganographer and *Eve* be the steganalyst. Eve knows the embedding function including its adaptivity criterion. Alice does not know the global cover distribution $\mathcal{P}_0$. Granting Eve access to both global distributions $\mathcal{P}_0$ and $\mathcal{P}_1$ (as suggested by the strictest interpretation of Kerckhoffs' principle for steganography [15]) would enable her to attack with the best-possible detector (although it may be computationally hard). This setup appears unrealistic and is sufficiently studied [16]. Instead, we follow Böhme and Ker who argue that a realistic setup is characterized by incomplete information and computational bounds for all parties [1], [17], [18]. This means that both parties, unaware of the global distributions, must resort to local models based on public knowledge. Deprived of perfect embedding and optimal detection, Alice's best choice of embedding positions may depend on Eve's actions, and vice versa. Game theory helps us to analyze the resulting strategies.

The different entities in our game are: *Nature*, *Alice*, the *Judge*, and *Eve*. Nature is the heterogeneous cover source that emits a cover $\boldsymbol{x}^{(\mathbf{0})}$ of n symbols drawn from $\mathcal{P}_0$. Upon receiving the cover from Nature, Alice changes exactly $k \leq n$ values. She changes position i of the reordered cover $\boldsymbol{y}^{(\mathbf{0})}$ with probability $\bar{a}_i$. (Recall that we abstract from a coding layer. See [14] for a discussion of coding to maximize embedding efficiency for content-adaptive steganography.) The Judge is fair and forwards to Eve with constant probability $\mu = 1/2$ either the cover or the stego object. In the jargon of game theory, Alice and Eve are the strategic players and Nature and the Judge are not strategic. They cause imperfect information in the sense that Alice has little influence on the cover source and Eve does not know what type of object she faces.

When Eve gets either the cover or the stego object, she recovers its order and inspects symbol $\hat{y}_i^{(\bar{a}_i)}$ with probability $\bar{e}_i$. Then, she decides about the type of object. Her disadvantage materializes in the error rates of this decision. These rates quantify steganographic security.

D. Strategies

Game theory distinguishes *pure* and *mixed* strategies. A strategy is pure if a player chooses an action deterministically. By contrast, a mixed strategy is a probability distribution over pure strategies. Alice's strategy space to change k values out of n positions leads to $\binom{n}{k}$ pure strategies. We simplify this by assigning probabilities in mixed strategies to single positions and only look at the projection of the probabilities onto the positions. We define the random binary vector $\boldsymbol{A}$, of which Alice's choice $\boldsymbol{a} = (a_0, \ldots, a_{n-1})$ is a realization, and the random binary vector $\boldsymbol{E}$, of which Eve's choice $\boldsymbol{e} = (e_0, \ldots, e_{n-1})$ is a realization. A value of $a_i = 1$ means that Alice changes $y_i^{(0)}$ for embedding, and $a_i = 0$ means she does not. Similarly, Eve inspects $\hat{y}_i^{(\bar{a}_i)}$ only if $e_i = 1$.

Let $\bar{a}_i = \Pr(A_i = 1)$ and $\bar{e}_i = \Pr(E_i = 1)$ be Alice's and Eve's parameters in mixed strategies, respectively.

The *embedding strategy* is part of the embedding function, besides the embedding operation (Def. 2). We characterize three canonical embedding and three canonical detection strategies.

Definition 6 (Canonical Embedding Strategies)**.**
The steganographer's embedding strategy is called ...

- *(E.i)* naïve adaptive*, if* $\bar{a}_i = 1$ *for* $i \in \{0, \ldots, k-1\}$ *and* $\bar{a}_i = 0$ *otherwise,*
- *(E.ii)* random uniform*, if* $\forall i : \bar{a}_i = k/n$*, and*
- *(E.iii)* optimal adaptive*, if* $\bar{\boldsymbol{a}} = \bar{\boldsymbol{a}}^*$*, a (unique) equilibrium strategy of the adaptive steganography game.*

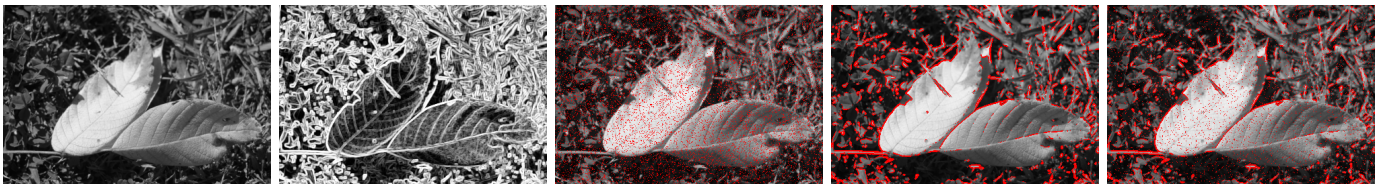
Definition 7 (Canonical Detection Strategies)**.**
The steganalyst's detection strategy is called ...

- *(D.i)* unweighted*, if* $\forall i : \bar{e}_i = k/n$*,*
- *(D.ii)* weighted*, if* $\bar{e}_i = 0$ *for* $i \in \{0, \ldots, n-k-1\}$ *and* $\bar{e}_i = 1$ *otherwise, and*
- *(D.iii)* optimal adaptive*, if* $\bar{\boldsymbol{e}} = \bar{\boldsymbol{e}}^*$*, a (unique) equilibrium strategy of the adaptive steganography game.*

Most practical embedding functions implement random uniform or naïve adaptive embedding. Most steganalysis methods implement unweighted or weighted detection. Observe that weighted detection is blind to naïve adaptive embedding if $k < \frac{n}{2}$, as it puts all weight on the least suitable (i. e., easiest to analyze) positions which are not touched by naïve adaptive embedding. A contribution of this article is to investigate the *optimal adaptive* strategies.

E. Example

To connect the concepts of our framework with simple practical steganography, consider the example of least significant bit (LSB) embedding in the spatial domain (i. e., pixel values) of grayscale images. Suppose the cover source $\boldsymbol{X}^{(\mathbf{0})}$ is a digital camera and let the image in Figure 1 (a) be a *cover* $\boldsymbol{x}^{(\mathbf{0})}$ (Def. 1) drawn from the unknown cover distribution $\mathcal{P}_0$. (The pixel matrix is serialized to a vector.) The *embedding operation* (Def. 2) replaces the LSBs of embedding positions with the encrypted and encoded hidden message bits. A well-known detector of LSB replacement steganography predicts potential cover pixel values from the observed image and aggregates the resulting residuals for further analysis [12]. Obviously, pixels at sharp edges are less predictable and thus more *suitable* for embedding (Def. 4) than pixels in smoother regions. The local variance approximates differences in suitability and serves as



(a) cover (b) adaptivity criterion (c) random uniform (d) naïve adaptive (e) optimal adaptive

Fig. 1. Concepts of our framework by the example of spatial domain image steganography. Red color indicates positions where embedding flips the LSB.

a popular *adaptivity criterion* ζ (Def. 5). Brighter regions in Figure 1 (b) denote higher variance and less risk of detection.

Figure 1 (c) shows the embedding positions of a *random uniform strategy* (Def. 6-E.ii). By contrast, the *naïve adaptive strategy* (Def. 6-E.i) prioritizes positions with high local variance and avoids risky spots, shown in Figure 1 (d). According to empirical measurements [19], the steganalyst can recover more than 98 % of the embedding positions by recomputing the local variance on the stego image. With this knowledge she may concentrate her efforts on likely embedding positions. Using the *optimal adaptive strategy* (Def. 6-E.iii) promises better security against an anticipating steganalyst who can recover the values of the adaptivity criterion. Observe in Figure 1 (e) that some moderately suitable positions are used. This prevents the steganalyst from ignoring these parts of the image and increases steganographic security. The optimal adaptive strategy is an equilibrium of the adaptive steganography game.

III. Specific Model

The simplest model to study adaptive embedding consists of a source of heterogeneous covers of exactly two symbols ($n = 2$), $\boldsymbol{x}^{(0)} = \left(x_0^{(0)}, x_1^{(0)}\right)$, in which Alice makes one embedding change ($k = 1$). To reduce the number of case distinctions, it is convenient to model covers ordered by decreasing suitability $\boldsymbol{y}^{(0)} = \left(y_0^{(0)}, y_1^{(0)}\right)$. By symmetry, this is without loss of generality if we assume perfect recovery. Imperfect recovery can be modeled by flipping the two symbols with probability $1 - r$, where $r \in [0, 1]$ is the *recovery rate*.

A. Two-Symbol Model

The instantiation of $n = 2$ and $k = 1$ simplifies Alice's strategy space $\bar{\boldsymbol{a}}$ to $\bar{a}_0 := \bar{a}$ and $\bar{a}_1 = 1 - \bar{a}$. She embeds with probability $\bar{a}$ into $y_0^{(0)}$ and with probability $1 - \bar{a}$ into $y_1^{(0)}$. A similar simplification works for Eve's strategy space $\bar{\boldsymbol{e}}$. With perfect recoverability, a value of $\bar{e} = 1$ means she inspects $y_0^{(\bar{a})}$, the more suitable symbol, and $\bar{e} = 0$ means she examines $y_1^{(1-\bar{a})}$. More generally, we model Eve's choice such that she can either inspect $\hat{y}_0^{(\bar{a})}$ or $\hat{y}_1^{(1-\bar{a})}$, but not both at the same time. We justify this by the observation that Eve has no knowledge of the global distribution and thus has to use imperfect local rules, thereby discarding some evidence.

The simplifications allow us to draw this instantiation of the adaptive steganography game, as defined in Sect. II-C, in Figure 2. The tree specifies the probabilities for both players' pure strategies in mixed strategies and also incorporates the non-strategic parts: cover source, the Judge's coin flip, and Eve's decision rule.

B. Cover Source

Most digital representations of natural cover sources use positive integers as alphabet $\mathbb{X} := \{0, \dots, 2^\ell - 1\}$. Constant ℓ defines the size of the cover alphabet. To reflect that symbols occur with varying probability, let $f_{t_i}^{(0)} : \mathbb{X} \to [0, 1]$ be a family of probability mass functions (PMFs),

$$f_{t_i}^{(0)}(u) := \Pr(y_i^{(0)} = u) := \frac{(t_i)^u}{d_i}, \tag{1}$$

with parameter $t_i \geq 1$ and normalizing factor $d_i := \frac{1-{t_i}^{2^\ell}}{1-t_i}$. Observe that the probabilities of the values $0, \dots, 2^\ell - 1 \in \mathbb{X}$ are increasing by a constant ratio.[1] In the limit case, $t_i = 1$ creates a uniform distribution (i. e., maximum entropy). The entropy decreases with increasing t_i.

Now extending to $n = 2$ independent cover symbols, we restrict the parameter ranges of t_0 and t_1 to $1 \leq t_0 \leq t_1$. This will allow us to generate homogenous (for $t_0 = t_1$) and heterogenous (for $t_0 < t_1$) covers with ordered suitability. (Corollary 1 in Sect. III-E will prove the very last assertion.)

C. Justification of the Cover Source Model

Although our cover source is very simple and in fact artificial [3], several reasons justify its specific choice.

First, note that the PMF for individual symbols asymptotically converges to (the left half of) a discretized Laplace distribution, which is known to model the marginal distribution of real transform-coded covers reasonably well [20]. The PMF of a mean-free discretized Laplacian distribution with scale parameter p is given by [21]:

$$g_p(u) = \frac{p-1}{p+1} \cdot p^{|u|}, \ p \in (0, 1), \ u \in \mathbb{Z}. \tag{2}$$

We resolve the absolute value function by considering only the left half of the distribution, $u \leq 0$:

$$g_p(u) = \frac{p-1}{p+1} \cdot p^{-u}. \tag{3}$$

As $p < 1$, we substitute $t_i := \frac{1}{p}$ in Equation (1) to obtain

$$f_{\frac{1}{p}}(u) = \frac{\left(\frac{1}{p}\right)^u}{d_i} = \frac{1}{d_i} \cdot p^{-u}. \tag{4}$$

For $t_i = \frac{1}{p}$ fixed, $\mathcal{O}(g_p)$ and $\mathcal{O}(f_{\frac{1}{p}})$ give the asymptotic equivalence in tails as u (and ℓ) go to infinity:

$$\begin{aligned} g_p(u) &\in \mathcal{O}(p^{-u}), \\ f_{\frac{1}{p}}(u) &\in \mathcal{O}(p^{-u}). \end{aligned} \tag{5}$$

[1] This replaces the linear PMF with $\ell = 2$ fixed in our earlier work [9].

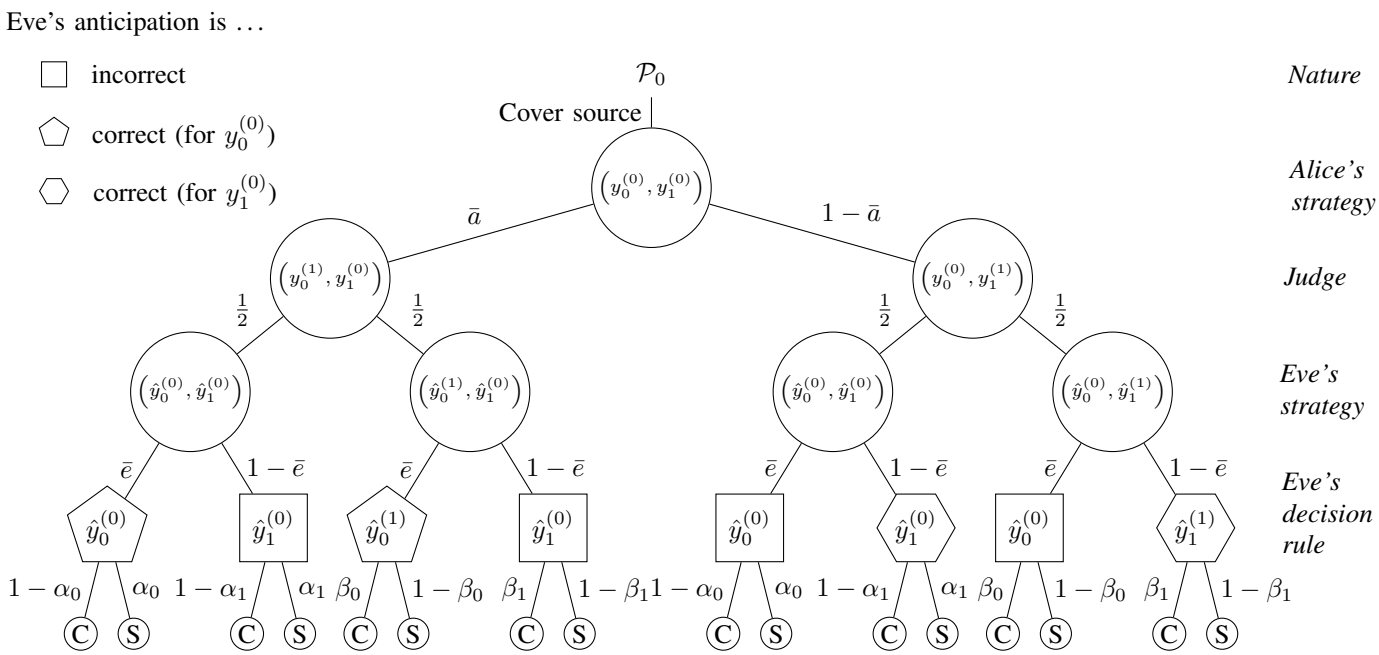


Fig. 2. The adaptive steganography game in the two-symbol model. α (β) is the false positive (negative) rate of Eve's decision rule (C for cover; S for stego).

Second, the restriction to $n = 2$ symbols permits an interpretation of larger heterogenous covers with independent symbols if they can be partitioned into two parts of equal size and suitability. The game is then played simultaneously and independently for each pair of heterogeneous symbols.

Third, the assumption that the ordered symbols $\boldsymbol{y^{(0)}}$ are independent is a common (and possibly realistic) simplification because reordering the cover by the adaptivity criterion likely removes Markov-properties. Of course, this does not prevent Eve from exploiting Markov-properties stemming from the cover in the *unordered* stego object $\boldsymbol{x^{(1)}}$. To resolve this, one may assume that she exhausts this information source when recovering the adaptivity criterion (e. g., local variance).

Fourth, independent cover symbols imply that the entropy of the cover source is the sum of the entropy of its symbols. The entropy of the cover source is an important benchmark quantity. It gives the upper bound for the size of a hidden message which a computationally unconstrained steganographer can embed undetectably. We can easily vary the heterogeneity of the cover source by adjusting t_i while (numerically) enforcing constant entropy. By doing so, entropy and heterogeneity are not confounded and we can isolate the effect of heterogeneity.

Fifth, we will show that our PMF renders the game-theoretic results independent of the size of the cover alphabet ℓ.

D. Embedding Operation and Alice's Strategy

We fix the embedding operation to the popular choice of *least significant bit replacement* (LSBR),

$$\mathrm{emb}(y) := y + (-1)^y \quad \Rightarrow \quad \mathrm{emb}^{-1}(y) = \mathrm{emb}(y). \tag{6}$$

Let $f_{t_i}^{(1)}$ be the family of PMFs resulting from always embedding in $y_i^{(0)}$. Then, for individual values u it holds:

$$f_{t_i}^{(0)}(u) = \Pr(u \mid \text{Cover}) \text{ and } f_{t_i}^{(1)}(u) = \Pr(u \mid \text{Stego}). \tag{7}$$

In the cover model, we can find an analytical expression for $\mathcal{P}_1$ by examining the distribution after embedding in $y_0^{(0)}$ with probability $\bar{a}$ and embedding in $y_1^{(0)}$ with probability $1-\bar{a}$.

As our model is to always change one symbol, it holds that

$$f_{t_i}^{(1)}(u) = f_{t_i}^{(0)}(\mathrm{emb}^{-1}(u)). \tag{8}$$

This yields the following lemma about $f_{t_i}^{(1)}(u)$, the marginal distributions of $\mathcal{P}_1$.

Lemma 1. *The PMF of stego symbols $f_{t_i}^{(1)}(u)$ is*

$$f_{t_i}^{(1)}(u) = f_{t_i}^{(0)}(u) \cdot t_i^{(-1)^u}. \tag{9}$$

Proof: After inserting Eq. (6) into Eq. (8),

$$f_{t_i}^{(1)}(u) = f_{t_i}^{(0)}(\mathrm{emb}^{-1}(u)) = f_{t_i}^{(0)}(u + (-1)^u), \tag{10}$$

we use the definition of Eq. (1) and rearrange,

$$= \frac{{t_i}^{u+(-1)^u}}{d_i} = f_{t_i}^{(0)}(u) \cdot t_i^{(-1)^u}. \tag{11}$$

■

If Alice plays a mixed strategy with parameter $\bar{a}$, the joint distribution $\mathcal{P}_1$ after embedding is a mixture of the kind:

$$\begin{aligned}\mathcal{P}_1(\boldsymbol{y}) &= \Pr(y_0 = u, y_1 = v) \\ &= \bar{a}\left(f_{t_0}^{(1)}(u) \cdot f_{t_1}^{(0)}(v)\right) + (1-\bar{a})\left(f_{t_0}^{(0)}(u) \cdot f_{t_1}^{(1)}(v)\right).\end{aligned} \tag{12}$$

Remark 4. *With our cover model and embedding operation, perfect steganography is only possible if $t_0 = 1$.*

Whenever $t_0 > 1$, some simple algebra shows that $\mathcal{P}_0$ and $\mathcal{P}_1$ differ. Note that this is necessary but not sufficient to rule

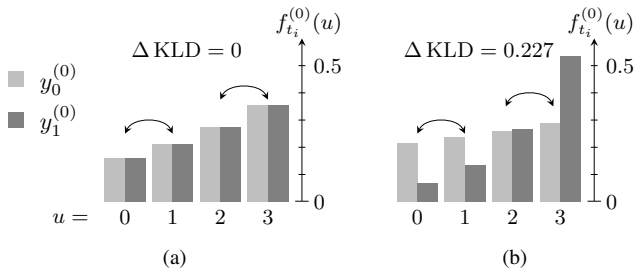


Fig. 3. Example histograms of the cover source for $n = \ell = 2$. Compare the more suitable (brighter bars) to the less suitable (darker bars) position for a: (a) homogeneous ($t_0 = t_1 = 1.3$); (b) heterogeneous ($t_0 = 1.1$, $t_1 = 2$) cover source. The arrows indicate which values are exchanged by the LSBR embedding operation.

out the possibility of perfect steganography. Even if $\mathcal{P}_0$ and $\mathcal{P}_1$ are not the same, the marginal distributions for one symbol (the more suitable one) may be equal.

E. Heterogeneity

Heterogeneity is necessary for adaptive steganography. We discuss how our model can be parametrized for different levels of heterogeneity.

Recall that the definition of heterogeneity (Def. 3) uses the KLD. There is an easy way to calculate it for our model.

Lemma 2. *The Kullback–Leibler divergence between $\mathcal{P}_0$ and $\mathcal{P}_{(y_i)}$ can be calculated as follows:*

$$\mathrm{KLD}(\mathcal{P}_0, \mathcal{P}_{(y_i)}) = \log t_i \cdot \frac{t_i - 1}{t_i + 1}. \tag{13}$$

The proof is given in Appendix A.

As the symbols are independent, the amount of distortion introduced by embedding, as measured by the KLD, only depends on the PMF of the symbol used for embedding.

Corollary 1. *If it holds that $t_0 < t_1$, then $y_0^{(0)}$ is more suitable for embedding than $y_1^{(0)}$.*

Proof: If $t_0 < t_1$, then $\log t_0 \cdot \frac{t_0-1}{t_0+1} < \log t_1 \cdot \frac{t_1-1}{t_1+1}$, hence by Lemma 2: $\mathrm{KLD}(\mathcal{P}_0, \mathcal{P}_{(y_0)}) < \mathrm{KLD}(\mathcal{P}_0, \mathcal{P}_{(y_1)})$. ∎

Remark 5. *The difference in the KLD between (1) changing only the least suitable and (2) changing only the best suitable symbol is a metric to quantify the heterogeneity of a cover source:* $\Delta\,\mathrm{KLD} := \mathrm{KLD}(\mathcal{P}_0, \mathcal{P}_{(y_1)}) - \mathrm{KLD}(\mathcal{P}_0, \mathcal{P}_{(y_0)})$.

Note that this metric depends on the embedding operation, like our notions of heterogeneity and suitability.

The histograms in Figure 3 show examples of two different parameterizations of the cover source with a fixed alphabet of four values ($\ell = 2$). The smaller parameter t_i, the closer is the distribution to a uniform distribution and the less detectable is the embedding operation LSBR (as indicated by the arrows). Figure 3(a) shows a homogeneous cover source. Only for heterogeneous cover sources (Figure 3(b)), Alice can take advantage of adaptively choosing more suitable positions. This advantage increases with the level of heterogeneity.

F. Eve's Decision: Local Optimal Detector

We equip Eve with the locally optimal decision rule, specific to the embedding operation LSBR and the cover source. The rule is not part of Eve's strategy, she follows it deterministically. The rule influences the game-theoretic analysis indirectly by the error rates it induces.

Eve's decision rule $\mathrm{decide}(u)$ between C (for cover) and S (for stego) follows from the *maximum a posteriori* (MAP) estimation [12, for example], and the fairness of the Judge ($\mu = 1/2$).

Lemma 3. *Eve's locally optimal decision rule when examining an individual symbol and finding value u is:*

$$\mathrm{decide}(u) := \begin{cases} \mathrm{S} & : u \equiv 0 \pmod 2 \\ \mathrm{C} & : u \equiv 1 \pmod 2. \end{cases} \tag{14}$$

Proof: MAP estimation minimizes the decision errors by using Bayes' theorem:

$$\hat{q} = \arg\max_q \Pr(q \mid u) = \arg\max_q \Pr(u \mid q) \cdot \Pr(q). \tag{15}$$

With $q \in \{\mathrm{C}, \mathrm{S}\}$, we obtain

$$\hat{q} = \arg\max_q \Pr(u \mid q) \cdot \mu \tag{16}$$

$$\overset{\text{Eq. (7)}}{=} \arg\max \left\{ \mathrm{C} : f_{t_i}^{(0)}(u), \mathrm{S} : f_{t_i}^{(1)}(u) \right\}, \tag{17}$$

now using Lemma 1 and dividing element-wise by $f_{t_i}^{(0)}(u)$,

$$= \arg\max \left\{ \mathrm{C} : 1, \mathrm{S} : t_i^{(-1)^u} \right\}, \tag{18}$$

$$= \begin{cases} \mathrm{S} & : u \equiv 0 \pmod 2 \\ \mathrm{C} & : u \equiv 1 \pmod 2. \end{cases} \tag{19}$$

The last equality follows from the fact that $t_i \geq 1$. If $t_i = 1$, Eve is indifferent, but the rule is still optimal in the sense that she cannot do better than random guessing. ∎

Note that fixing the embedding operation (in Sect. III-D) and this detector generally precludes Alice from embedding at the information-theoretic bound (unless $t_i = 1$), and Eve from using the best-possible detector. This is intentional to reflect the hardness of reaching these goals in practice. These restrictions can be understood as a way to model the players' knowledge and computational constraints while allowing us to still analyze their respective strategies. At the same time, the fact that we use an artificial cover model with tractable globally optimal solutions enables us to benchmark the constrained solutions against the information-theoretic optimum, which would minimize the KLD. This comparison would not be tractable for much richer cover models let alone real covers.

G. Error Rates

As mentioned in Section II-C, Eve's error rates quantify steganographic security. In our model, the error rates depend on the parameters t_i. Let α_i (β_i) be Eve's false positive (false negative) probability when applying decide on $f_{t_i}^{(0)}$ ($f_{t_i}^{(1)}$). We

use Eve's *average error rate* (under equal priors) $\mathrm{AER} := (\alpha_i + \beta_i)/2$ to measure steganographic security in this analysis.

Lemma 4. *If Eve investigates the same position $i \in \{0,1\}$ that Alice has changed for embedding, then*

$$\mathrm{AER} = \frac{1}{t_i + 1}. \tag{20}$$

The proof is given in Appendix B.

Equation (20) is intuitive, as the error probability is $1/2$ (random guessing) for the boundary case $t_i = 1$; uniform i. i. d. where LSBR is undetectable. It is also consistent with Corollary 1 because higher values of t_i imply less suitability for embedding, which leads to a lower AER, and vice versa.

Corollary 2. *The worst case for Eve is Alice choosing $a \in \{0,1\}$ and she herself choosing $e = 1 - a$. In this case, her decision is no better than random guessing, i. e.,* $\mathrm{AER} = 1/2$.

Proof: Recall that $a = a_0, 1 - a = a_1, e = e_0$, and $1 - e = e_1$. If $e = 1 - a$, Eve's decision rule is always applied to symbols drawn from the (marginal) cover distribution. For every symbol $u \in \mathbb{X}$, let bias $b_u \in [0,1]$ be the probability that any probabilistic decision rule (including decide from Lemma 3) returns S (for stego) upon finding value u. Then,

$$\mathrm{AER}\,|_u = \frac{\alpha|_u + \beta|_u}{2} = \frac{b_u + (1 - b_u)}{2} = \frac{1}{2}. \tag{21}$$

$\mathrm{AER}\,|_u$ is independent of u, hence $\mathrm{AER} = 1/2$. ∎

H. Type of Game

We recall the properties of our game to facilitate its classification in the game-theortical literature. Our setup starts as a game with *incomplete information*: the players are uncertain about the cover realization. By introducing Nature and the Judge, we use the Harsanyi transformation [22] to rewrite the game as a game with *imperfect information*, i. e., a Bayesian game. Finally, aggregating the probability distributions of Nature and the Judge to a (frequentist) rate, the AER, transforms the setup to a *simultaneous move game with perfect information*.

IV. Solving the Game

We first derive the pay-off function and then solve the game for Nash equilibria [23]. Throughout this section we assume that Eve can perfectly recover the order of the suitability of the embedding positions; formally: $\hat{\boldsymbol{y}}^{(\mathbf{1})} = \boldsymbol{y}^{(\mathbf{1})}$.

A. Pay-Off

Being agnostic about detailed cost assumptions, we devise a zero-sum game with the AER determining the pay-offs. Zero-sum games are strictly competitive, one player loses what the other wins. Alice wants to perform least detectable steganography, hence she tries to maximize the AER. Eve's goal is to detect as much as possible, hence she tries to minimize the AER. Consequently, Alice's pay-off is her expected AER, and Eve's pay-off is her expected $-\,\mathrm{AER}$. Expectations are taken over realizations of the random variables governed by Nature and the realizations of the players' strategies A and E.

Table I lists all possible states (in rows), the associated AER for two different scenarios (column blocks), and how we obtain it. Note that each row aggregates both possible outcomes of the Judge's coin flip and the AER combines both error rates.

Lemma 5. *The expected* AER *in mixed strategies is*

$$\chi(\bar{a}, \bar{e}) := \frac{1}{t_1 + 1} + \Big(\frac{t_1 - 1}{2(t_1 + 1)}\Big)\bar{a} + \Big(\frac{t_1 - 1}{2(t_1 + 1)}\Big)\bar{e} + \Big(\frac{1 - t_0 t_1}{(t_0 + 1)(t_1 + 1)}\Big)\bar{a}\bar{e} \tag{22}$$

Proof: Figure 2 shows that the nodes of Eve's decision can be classified into three different types (by their shape).

(1) Alice changes $y_0^{(0)}$ and Eve anticipates it (pentagons).
(2) Alice changes $y_1^{(0)}$ and Eve anticipates it (hexagons).
(3) Alice changes $y_i^{(0)}$, but Eve inspects the wrong embedding position (squares in Figure 2).

Table I shows the respective probabilities of occurrence, pay-offs, and justifications in columns 1–5. In combination, this leads to the following expression for $\chi(\bar{a}, \bar{e})$:

$$\chi(\bar{a}, \bar{e}) = \bar{a}\bar{e}\frac{1}{t_0 + 1} + \frac{1}{2}(\bar{a}(1 - \bar{e}) + (1 - \bar{a})\bar{e}) + (1 - \bar{a})(1 - \bar{e})\frac{1}{t_1 + 1}. \tag{23}$$

Equation (22) follows from rearranging Equation (23). ∎

Remark 6. *In the pathological case of $t_0 = t_1 = 1$, i. e., a homogeneous cover source with perfect steganography possible in both symbols, it holds that $\chi(\bar{a}, \bar{e}) = 1/2$. Particularly, $\chi(\bar{a}, \bar{e})$ is independent of $\bar{a}$ and $\bar{e}$. Such situations do not require game theory and are out the scope of this article.*

B. Equilibrium Strategies

Nash equilibria in two-player games are tuples of mixed strategies $(\bar{a}^*, \bar{e}^*)$ such that no player can (strictly) increase her pay-off by unilaterally deviating from her equilibrium strategy [23]. To find a Nash equilibrium we look for a strategy that makes the opponent indifferent, i. e., a strategy where she cannot influence the pay-off by changing her strategy. We find such strategies by taking partial derivatives of the pay-off function, $\chi(\bar{a}, \bar{e})$ with regard to the opponent's strategy and setting them to zero. Then we show that theses strategies indeed constitute a unique equilibrium, which happens to be symmetric.

Theorem 1. *There exists a unique symmetric Nash equilibrium in mixed strategies. In this equilibrium it holds that:*

$$\bar{a}^* = \bar{e}^* = \frac{(1 - t_1)(1 + t_0)}{2(1 - t_0 t_1)}. \tag{24}$$

Proof: The partial derivatives of the pay-off functions are:

$$\frac{\partial \chi(\bar{a}, \bar{e})}{\partial \bar{a}} = \frac{t_1 - 1}{2(t_1 + 1)} + \Big(\frac{1 - t_0 t_1}{(t_0 + 1)(t_1 + 1)}\Big)\bar{e}, \tag{25}$$

$$\frac{\partial - \chi(\bar{a}, \bar{e})}{\partial \bar{e}} = -\frac{t_1 - 1}{2(t_1 + 1)} - \Big(\frac{1 - t_0 t_1}{(t_0 + 1)(t_1 + 1)}\Big)\bar{a}. \tag{26}$$

TABLE I
GAME OUTCOME IN DIFFERENT STATES OF THE WORLD

			Perfect/Correct recovery		Incorrect recovery		
Alice's choice	Eve's choice	Probability	AER	Reason	Reality	AER	Reason
$y_0^{(0)}$	$\hat{y}_0^{(1)}$	$\bar{a}\cdot\bar{e}$	$\frac{1}{t_0+1}$	Lemma 4, $i=0$	$y_1^{(0)}$	$\frac{1}{2}$	Corollary 2
$y_0^{(0)}$	$\hat{y}_1^{(0)}$	$\bar{a}\cdot(1-\bar{e})$	$\frac{1}{2}$	Corollary 2	$y_0^{(1)}$	$\frac{1}{t_0+1}$	Lemma 4, $i=0$
$y_1^{(0)}$	$\hat{y}_0^{(0)}$	$(1-\bar{a})\cdot\bar{e}$	$\frac{1}{2}$	Corollary 2	$y_1^{(1)}$	$\frac{1}{t_1+1}$	Lemma 4, $i=1$
$y_1^{(0)}$	$\hat{y}_1^{(1)}$	$(1-\bar{a})\cdot(1-\bar{e})$	$\frac{1}{t_1+1}$	Lemma 4, $i=1$	$y_0^{(0)}$	$\frac{1}{2}$	Corollary 2

Setting both derivatives to zero yields Equation (24).

To see that $\bar{a}^*$ is an equilibrium strategy, we combine Equations (22) and (24):

$$\begin{aligned}\chi(\bar{a}^*,\bar{e}) = &\frac{1}{t_1+1}+\Big(\frac{t_1-1}{2(t_1+1)}\Big)\cdot\Big(\frac{(1-t_1)(t_0+1)}{2(1-t_0t_1)}\Big)\\ &+\Big(\frac{t_1-1}{2(t_1+1)}\Big)\bar{e}\\ &+\Big(\frac{1-t_0t_1}{(t_0+1)(t_1+1)}\Big)\cdot\Big(\frac{(1-t_1)(t_0+1)}{2(1-t_0t_1)}\Big)\bar{e}. \quad (27)\end{aligned}$$

Considering only the terms containing $\bar{e}$:

$$\bar{e}\cdot\Big(\frac{t_1-1}{2(t_1+1)}+\frac{1-t_1}{2(t_1+1)}\Big)=\bar{e}\cdot 0. \quad (28)$$

As the same holds for $\chi(\bar{a},\bar{e}^*)$, both $\chi(\bar{a}^*,\bar{e})$ and $\chi(\bar{a},\bar{e}^*)$ are independent of the opponent's strategy. Thus, $\forall\bar{a},\bar{e}\in[0,1]:\ \chi(\bar{a}^*,\bar{e}^*)=\chi(\bar{a}^*,\bar{e})=\chi(\bar{a},\bar{e}^*)$, and thus $(\bar{a}^*,\bar{e}^*)$ is a Nash equilibrium.

A quick check that no combination of pure strategies is a Nash equilibrium (for $t_0>1$) establishes the uniqueness of $(\bar{a}^*,\bar{e}^*)$. The symmetry is obvious as $\bar{a}^*=\bar{e}^*$. ■

The following corollaries state two direct implications for the design of more secure embedding functions.

Corollary 3. *If and only if the given cover source is* homogeneous*, i. e., $t_0=t_1$, Alice's best strategy is* random uniform embedding *(strategy (E.ii) from Section II-D).*

Proof: The 'if' direction follows from the fact that for $t_0=t_1$, it holds that:

$$\bar{a}^*=\frac{(1-t_1)(1+t_0)}{2\cdot(1-t_0t_1)}=\frac{(1-t_0)(1+t_0)}{2\cdot(1-{t_0}^2)}=\frac{1}{2}. \quad (29)$$

Alice changes each of the two symbols with probability $\bar{a}=1/2$. With $k=1$ and $n=2$, this fulfills the definition of random uniform embedding.

If $t_0<t_1$, it holds that:

$$\bar{a}^*=\frac{(1-t_1)(1+t_0)}{2\cdot(1-t_0t_1)}=\frac{1}{2}\cdot\underbrace{\left(\frac{\overbrace{t_0-t_1}^{<0}+(1-t_0t_1)}{1-t_0t_1}\right)}_{>1}>\frac{1}{2}. \quad (30)$$

This proves the 'only-if' direction. ■

Corollary 4. *If and only if one of the cover symbols allows for* perfect *steganography, then Alice's best strategy is* naïve adaptive embedding *(strategy (E.i) from Section II-D).*

Proof: Perfect steganography is only possible if the PMF of at least k symbols ($k=1$ in our model) is invariant to embedding. Inserting the formal condition, $t_0=1$ (from Remark 4), into the equilibrium condition:

$$\bar{a}^*=\frac{(1-t_1)(1+t_0)}{2\cdot(1-t_0t_1)}=\frac{(1-t_1)\cdot 2}{2\cdot(1-t_1)}=1. \quad (31)$$

Alice always changes the better suitable symbol. This fulfills the definition of naïve adaptive embedding. Whenever $t_0>1$ it follows, that

$$t_0(t_1+1)>t_1+1 \quad\Leftrightarrow\quad t_0t_1-1>t_1-t_0. \quad (32)$$

Rewriting Equation (24) yields:

$$\bar{a}^*=\frac{1}{2}+\frac{1}{2}\cdot\underbrace{\left(\frac{t_1-t_0}{t_0t_1-1}\right)}_{<1}<1. \quad (33)$$

This proves the 'only-if' direction. ■

From the uniqueness of the equilibrium and the preceding corollaries follows another property of our model.

Corollary 5. *If $t_0>1$, there are no dominated strategies and thus no dominant strategy equilibria (DSE) in our model.*

Proof: From Corollary 4 it follows that, unless $t_0=1$, the equilibrium given in Theorem 1 defines strategies that put positive probability on every pure strategy. Such an equilibrium is called *completely mixed equilibrium* and only exists if there is no pure or mixed strategy of any player that is strictly or weakly dominated by a convex combination of her other strategies [24]. Therefore, there are no dominant strategies and thus no dominant strategy equilibria. ■

It is easy to see that in the corner case $t_0=1$, the pure strategies $\bar{a}^*=\bar{e}^*=1$ are dominant pure strategies and form a dominant strategy equilibrium.

C. Pay-off in Equilibrium

Now that we determined the equilibrium strategies for Alice, respectively Eve, we can calculate the pay-off in equilibrium.

Corollary 6. *The expected* AER *in equilibrium is*

$$\chi(\bar{a}^*,\bar{e}^*)=\frac{(t_0+1)(t_1+1)-4}{4(t_0t_1-1)}. \quad (34)$$

This corollary follows directly from inserting the equilibrium conditions (Theorem 1) into Lemma 5.

A closer look at the equilibrium strategies reveals that they are *equalizer strategies* [24]. Equalizer strategies yield the same expected payoff for each player, regardless of the (pure or mixed) strategy chosen by the other player.

Corollary 7. *The equilibrium strategies $\bar{a}^*$, respectively $\bar{e}^*$ are equalizer strategies.*

Proof: From the proof of Theorem 1 we know that $\chi(\bar{a}^*, \bar{e}^*) = \chi(\bar{a}^*, \bar{e}) = \chi(\bar{a}, \bar{e}^*)$. Thus, if Alice plays her equilibrium strategy $\bar{a}^*$, Eve's strategy $\bar{e}$ does not influence the pay-off and vice versa. From this property it follows that $\bar{a}^*$ and $\bar{e}^*$ are equalizer strategies. ∎

Corollary 8. *If Alice (Eve) plays her equilibrium strategy, she balances Eve's (Alice's) advantage over choosing a specific position.*

Proof: The corollary follows directly from the fact that equalizer strategies make the other player indifferent to the strategies of the opponent [24]. ∎

This means that the heterogeneity in the cover source is exactly offset by the probabilities of the mixed strategies. Equilibria in high-dimensional spaces may be hard to find [25]. Starting with the solution concept of equalizer strategies might render this problem tractable as it reduces the search space.

Summarizing this section, we have proven that for this instantiation of the framework

- our adaptive steganography game has a unique symmetric Nash equilibrium in equalizer strategies (Thm. 1; Cor. 7);
- random uniform embedding is only optimal for homogeneous covers (Cor. 3); and
- naïve adaptive embedding is only optimal when perfect steganography is possible (Cor. 4).

The optimal strategies depend on the level of heterogeneity of the cover source, albeit in a non-linear manner.

V. Imperfect Recovery

In this section we relax the arguably unrealistic assumption that Eve is able to perfectly recover the order of possible embedding positions. However, both players know the (average) recovery rate r, which is akin a global constant. In our model with two positions, we define r as follows:

Definition 8 (Recovery rate)**.** *The* recovery rate r *is the probability that Eve can correctly recover the order of the symbols, i. e.,* $\hat{\boldsymbol{y}}^{(\mathbf{1})} = \boldsymbol{y}^{(\mathbf{1})}$.

In practice, the recovery rate is an empirical property (hence "rate") of the adaptivity criterion and the embedding function. As the criterion is not explicit in the stylized model, we can use the shortcut of Definition 8.

With the introduction of imperfect recoverability, we need to adjust the pay-off function.

Lemma 6. *The pay-off function with recovery rate r is:*

$$\begin{aligned}\chi_r(\bar{a},\bar{e}) := \frac{1}{2} &+ \left(\frac{1-t_0}{2(t_0+1)}\right)\bar{a} + \left(\frac{1-t_1}{2(t_1+1)}\right)\bar{e} \\ &+ \left(\frac{t_0t_1-1}{(t_0+1)(t_1+1)}\right)\bar{a}\bar{e} + \; r\cdot\Bigg[-\frac{1}{2} + \frac{1}{t_1+1} + \left(\frac{t_1-1}{t_1+1}\right)\bar{e} \\ &+ \left(\frac{t_0t_1-1}{(t_0+1)(t_1+1)}\right)\bar{a} + 2\left(\frac{1-t_0t_1}{(t_0+1)(t_1+1)}\right)\bar{a}\bar{e}\Bigg]. \end{aligned} \tag{35}$$

Proof: Imperfect recovery is modeled by a mixture of correct and incorrect recovery. The pay-off function from Lemma 5 holds with probability r for the case of correct recovery. With probability $(1-r)$, the pay-off function is given by the terms in columns 6–8 of Table I for the case of incorrect recovery. Overall:

$$\begin{aligned}\chi_r(\bar{a},\bar{e}) = r\cdot\chi(\bar{a},\bar{e}) + (1-r)\cdot\Bigg(&\frac{1}{2} + \Big(\frac{1-t_0}{2(t_0+1)}\Big)\bar{a} \\ &+ \Big(\frac{1-t_1}{2(t_1+1)}\Big)\bar{e} + \Big(\frac{t_0t_1-1}{(t_0+1)(t_1+1)}\Big)\bar{a}\bar{e}\Bigg). \end{aligned} \tag{36}$$

Inserting Eq. (22) into Eq. (36) and rearranging yields Eq. (35). ∎

It is sufficient to study the interval $1/2 \leq r \leq 1$ because with $n = 2$, Eve can always invert the output of her recovery function to improve her rate to $r = 1 - r'$, where $r' < 1/2$ is her original rate. Next, we update the equilibrium conditions.

Theorem 2. *There exists a unique (asymmetric) Nash equilibrium in mixed strategies for $r \neq 1/2$. In this equilibrium it holds that:*

$$\bar{a}_r^* = \frac{(1-t_1)(1+t_0)}{2(1-t_0t_1)}, \tag{37}$$

$$\bar{e}_r^* = \frac{1}{2} - \frac{t_0-t_1}{2(2r-1)(t_0t_1-1)}. \tag{38}$$

Proof: The partial derivatives of the pay-off function are:

$$\begin{aligned}\frac{\partial\chi_r(\bar{a},\bar{e})}{\partial\bar{a}} =&\Big((2r-1)\frac{t_0t_1-1}{2(t_0+1)(t_1+1)} + \frac{t_1-t_0}{2(t_0+1)(t_1+1)}\Big) \\ &+ \Big((2r-1)\frac{1-t_0t_1}{(t_0+1)(t_1+1)}\Big)\bar{e}, \end{aligned} \tag{39}$$

$$\begin{aligned}\frac{\partial-\chi_r(\bar{a},\bar{e})}{\partial\bar{e}} =& -\Big((2r-1)\frac{t_1-1}{2(t_1+1)}\Big) \\ &- \Big((2r-1)\frac{1-t_0t_1}{(t_0+1)(t_1+1)}\Big)\bar{a}. \end{aligned} \tag{40}$$

Setting both derivatives to zero yields the strategies.

Inserting $\bar{a}_r^*$ in the partial derivative of the second term of Eq. (36) (factor $(1-r)$), which describes the case where Eve is not able to recover the order of the positions, eliminates all factors containing $\bar{e}$ in this term. The same was already shown for the first term of Eq. (36) (factor r) in the proof of Theorem 1. Some algebra shows that $\chi_r(\bar{a}, \bar{e}_r^*)$ is independent of $\bar{a}$ as well and thus, with the same arguments as in the proof of Theorem 1, $(\bar{a}_r^*, \bar{e}_r^*)$ is a Nash equilibrium. The uniqueness

follows from the fact that no combination of pure strategies is a Nash equilibrium (for $t_0 > 1; r \neq 1/2$). ∎

Note that this equilibrium is no longer symmetric: Alice follows the same strategy as with perfect recoverability, whereas Eve uses a different one.

Equation (38) implicates that Eve's strategy is not well-defined for $r = 1/2$. We handle this special case separately.

Corollary 9. *The pay-off function $\chi_{\frac{1}{2}}$ is linear in $\bar{a}$ and independent of $\bar{e}$. (Eve cannot influence the pay-off.) Alice's best strategy is $\bar{a} = 1$ (naïve adaptive embedding).*

Proof: Inserting $r = 1/2$ into Equation (35), yields:

$$\chi_{\frac{1}{2}}(\bar{a}, \bar{e}) = \frac{t_1 + 3}{4(t_1 + 1)} + \Big(\frac{t_1 - t_0}{2(t_0 + 1)(t_1 + 1)}\Big)\bar{a}, \qquad (41)$$

which is linear in $\bar{a}$ and independent of $\bar{e}$. The slope is positive whenever $t_0 < t_1$. Therefore $\bar{a} = 1$ is the maximum. ∎

The insight here is limited: the special case reminds us that if the stego object does not leak any information about the values of the adaptivity criterion, Eve has no advantage if she tries to recover it.

For $r \neq 1/2$, we find that the equilibrium strategies are still equalizer strategies, and the game outcome is the same as in the case of perfect recovery.

Corollary 10. *With recovery rate r, the equilibrium strategies are equalizer strategies and the pay-off in equilibrium is:*

$$\chi_r(\bar{a}_r^*, \bar{e}_r^*) = \frac{(t_0 + 1)(t_1 + 1) - 4}{4(t_0 t_1 - 1)}. \qquad (42)$$

Proof: From the proof of Theorem 2 follows that the players cannot influence the pay-off when the other player uses her equilibrium strategy. Thus, $\bar{a}_r^*$ and $\bar{e}_r^*$ are equalizer strategies. The pay-off follows from combining Equations (35), (37) and (38). ∎

It is very interesting to find that, excluding the corner case $r = \frac{1}{2}$, the equilibrium pay-off of the game is independent of the recovery rate r. If Alice plays her equilibrium strategy, she does not need to worry about the risk of Eve being able to recover the likely embedding positions via the adaptivity criterion. If a comparable result generalizes to practical scenarios (with gentle assumptions), it could become a cornerstone for the design of secure adaptive steganography.

VI. Numerical Illustration

In this section we numerically illustrate and interpret selected results of Sections III and IV. We plot the variables of interest in the parameter space $t_0, t_1 \in [1, 4]$ and $t_0 \leq t_1$.

Figure 4(a) shows the symmetric optimal adaptive strategy of Alice ($\bar{a}^*$) and Eve ($\bar{e}^*$) as a function of the model parameters t_0 and t_1. Higher values of the strategy variable indicate that the more suitable of both embedding positions is changed, respectively inspected, more often. Values at the diagonal $t_0 = t_1$ illustrate Corollary 3. If the cover source is homogeneous, random uniform embedding is optimal. The boundary line $t_0 = 1$ illustrates Corollary 4. If the more suitable position allows for perfect steganography, it is used with certainty. This is the case where naïve adaptive embedding is optimal.

Regions of perfect steganography can also be identified in Figure 4(b) (mind the rotated base). They are characterized by an error rate at its theoretical maximum of $1/2$: Eve cannot do better than random guessing.

The remaining parameter space is hard to interpret in these graphs because adjusting t_0 or t_1 affects both the heterogeneity and the entropy of the cover source. Figures 5(a) and 5(b) show this interdependence. Entropy is measured in bits and best interpreted as an upper bound for the secure capacity (cf. Sect. III-C). We use $\Delta\,\mathrm{KLD}$, introduced in Section III-E, as a metric for the level of heterogeneity. Higher values indicate more heterogeneous cover sources. Zero indicates homogeneity.

To compare like with like, we select two sets of constant entropy ($H \in \{2.2, 3.6\}$ bit, annotated in the figures) and adjust (t_0, t_1) jointly to vary the level of heterogeneity within these sets. Heterogeneity is the most important prerequisite for adaptive steganography, therefore Figure 6 compares strategies and pay-offs as a function of the level of heterogeneity while keeping everything else constant. In both subfigures, black lines refer to higher entropy, gray lines to lower entropy.

Figure 6(a) reports the optimal adaptive embedding strategies ($\bar{a}^*$) in solid lines. In the equilibrium, Alice uses random uniform embedding ($\bar{a}^* = 1/2$) only if the cover source is homogenous and shifts more and more probability mass to the more suitable position as the level of heterogeneity increases. This increase is steeper for cover sources with higher entropy. Since the equilibrium is symmetric (cf. Theorem 1), the solid lines also display Eve's optimal adaptive detection strategies.

For comparison, the dashed lines in Fig. 6(a) show Alice's choice of $\bar{a}$ in the distortion minimization paradigm. More specifically, we minimize the KLD between the cover and stego distribution. This is tractable in our stylized model, but infeasible in almost all practical scenarios. Observe that the information-theoretic criterion shifts the probability mass to the more suitable position more aggressively than the game-theoretic solution, but it does not coincide with naïve adaptive embedding for the given parameter range. Arguably, game-theoretically optimal adaptive embedding uses less suitable positions more often to prevent Eve from ignoring them and to force her to respond with the game-theoretic strategy.

Figure 6(b) shows Alice's pay-off in terms of Eve's average detection error rate (AER). Higher values indicate more secure steganography. Observe the level shift between high and low entropy. Consistent with the theoretical bound, high-entropy cover sources offer more security for a fixed message length. But the error rate is not constant, unlike the theoretical bound. (Also the low entropy line increases strictly monotonically, which is hardly visible at this scale.) The reasons for this difference is that Eve is constrained to a local detector and therefore cannot use an information-theoretically optimal detector. This is a consequence of our intention to model realistic (and thus bounded) steganalysts. Against this kind of steganalysts, more heterogeneous cover sources offer more security. But can we conclude that (optimal) adaptive embedding is worth pursuing?

To answer this question, note that we do not plot separate error rates for the benchmark where Alice minimizes the KLD, or for any other canonical strategy. This is because in our model, Eve's optimal adaptive detector is an equalizer

strategy (cf. Corollary 7). This implies by definition that the pay-off is independent of the opponent's action. Therefore, the dashed lines in Figure 6(a) lead to exactly the same error rates. And so do naïve adaptive or random uniform embedding. (Both would be horizontal lines in the coordinate system of Figure 6(a).) In this sense, adaptive embedding does not improve the steganographic security if the steganalyst already uses the optimal adaptive detector. But Alice must play her equilibrium strategy to prevent Eve from doing something else that could be more harmful to Alice than the equilibrium payoff.

VII. Related Work

The idea of adaptive embedding is almost as old as research on digital steganography. (See [1, pp. 48] for a survey and [26], [27], [28] for more recent examples.) However, the choice of the adaptivity criterion that directs the selection of embedding positions has not become an exact science. It seems that many authors apply judgment or heuristics inspired by known steganalysis methods. When reporting security gains over non-adaptive random uniform embedding, they often seem to disobey Kerckhoffs' principle by not considering that the steganalyst knows the adaptivity criterion and can estimate its values for the cover from the stego object.

This article extends our conference publication [9], which first motivated to study adaptive steganography with game theory in order to overcome the shortcomings sketched above. The conference paper contrasted optimal adaptive strategies against the information-theoretic benchmark (minimizing KLD) using a cover model with a simple step-function. This article introduces a complete framework with a substantially refined terminology and notation.

Several derived works fit into the proposed framework without mentioning it. In [29], we use binary covers of length n and allow the steganalyst to query the most likely value for one position from an oracle, mimicking cover estimation in practical steganalysis. In [30], the steganographer changes the values of *exactly* k positions in covers of length n. The steganalyst can aggregate information from all positions. Another variant of the model implements independent embedding. It lets the steganographer change k values *on average*, reflecting that some values might already carry the right steganographic semantic in the cover [31]. Denemark and Fridrich [32] independently extend the model of [9] to Gaussian covers with LSB matching as embedding operation. They report equilibria for $n = 2$ and second the qualitative results of our works.

Recent empirical results show that a steganalyst who examines only the most likely embedding positions [10] or weights all positions according to their approximate embedding probability [11] can detect several state-of-the-art embedding schemes better than detectors not using this information. It seems that the loss of detection power due to imprecise knowledge of the selection channel is rather small, as captured by our imperfect recovery scenario (Section V). It always pays off to use imprecise knowledge about likely embedding positions rather than none [33].

We are aware of three other independent publications using game theory in the broader context of steganography. Back in 1998, Ettinger [34] proposed a game between a steganographer who chooses the embedding rate and an *active attacker* who chooses the distortion rate subject to constraints on the utility of the channel. This differs from mainstream steganography research because the protection goal is availability, not undetectability. Ker [35] uses game theory to find strategies in the special case of *batch steganography*, where the hidden message can be spread over many cover objects. The steganalyst anticipates this and tries to detect the existence of any secret message (pooled steganalysis). Orsdemir et al. [36] point out a strategic component in practical steganography and steganalysis. They devise a *meta-game* where the steganographer chooses between two embedding functions and the steganalyst decides against which of the two functions a single classifier should be trained. As the embedding functions are black boxes, the equilibria of this matrix game do not directly inform the design of secure embedding functions or optimal detectors.

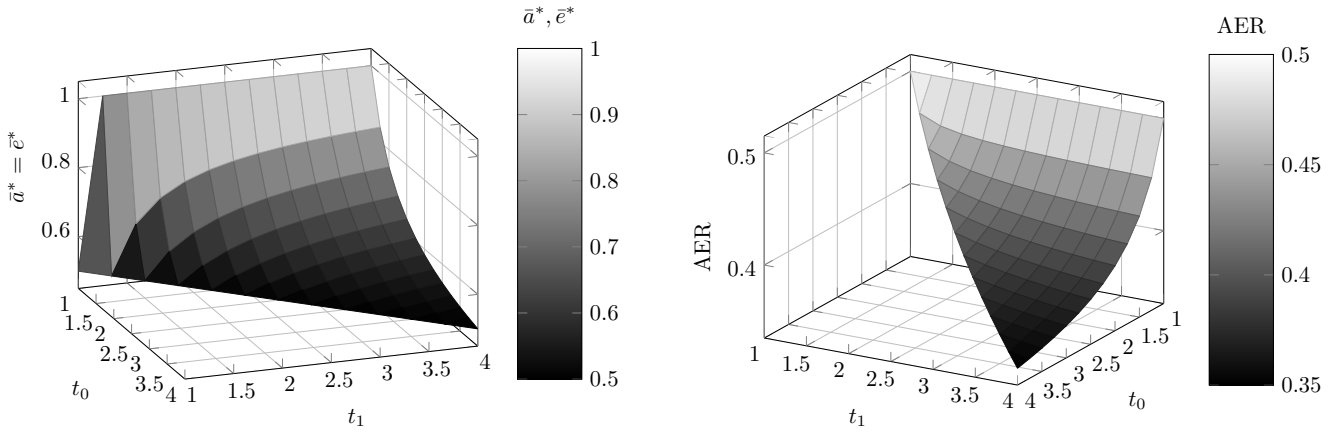
Katzenbeisser and Petitcolas [37] give a challenge-response protocol, called "game" by the conventions in cryptology, to formalize the advantage of computationally bounded steganalysts. We build on this protocol to obtain a pay-off metric under equal priors and augment it by inserting both players' strategies to make it a game in the sense of game theory [8].

Recent themes at the intersection of machine learning and security are adversarial classification [38] and signal processing [39]. Although there is no direct counterpart to our analysis of adaptive steganography, interesting parallels exist and the applicability of the results for steganalysis based on machine-learning and signal detection seems worth exploring.

VIII. Conclusion

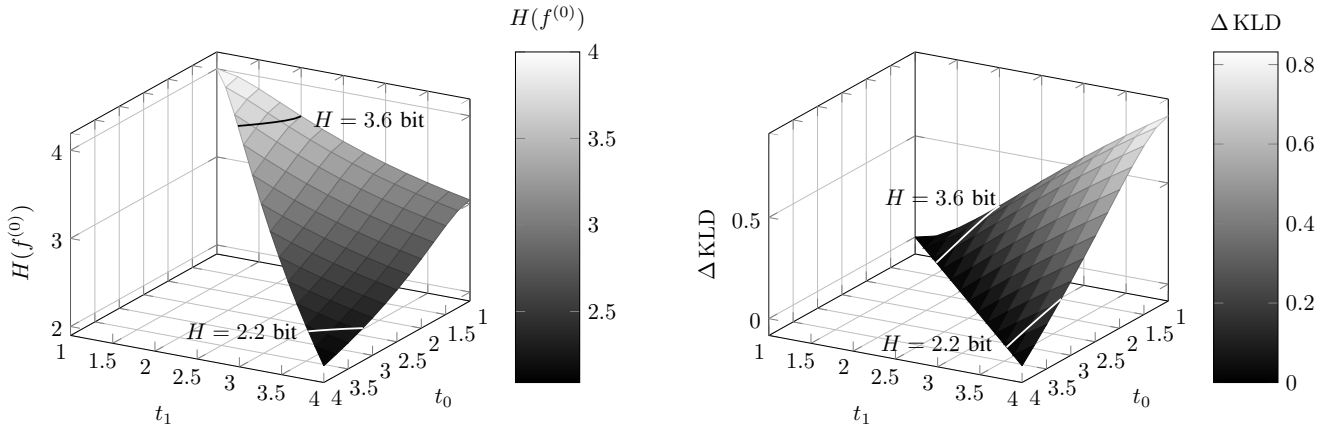
The main contribution of this work is threefold. First, we present a universal game-theoretic framework to model adaptive embedding in the presence of an attacker who anticipates this behavior and can recover the likely embedding positions from the stego object. The framework offers a novel way to analytically study the security of adaptive steganography while fully respecting Kerckhoffs' principle. Second, we instantiate the framework with a stylized two-symbol model and solve the game for equilibrium conditions. We find unique symmetric equilibria in equalizer strategies, making the opponent indifferent to the choice of embedding positions or detector weights. Third, we relax the initial assumption of perfect recovery. We find that in our model the embedding strategy is independent of the recovery rate.

All results depend on a number of assumptions: the players know the marginal cover distribution, covers consist of two a priori independent symbols, the steganographer replaces exactly one bit, the steganalyst inspects only one position. Thus, many limitations apply when transferring our results to practical systems. Nevertheless, a solid theory not only helps to guide the design of future adaptive embedding and detection functions with qualitative insights, but also to identify promising avenues to solve the general problem more rigorously. Among the results of this article, equalizer strategies and the invariance to the recovery rate seem to have the best chances to influence future works.



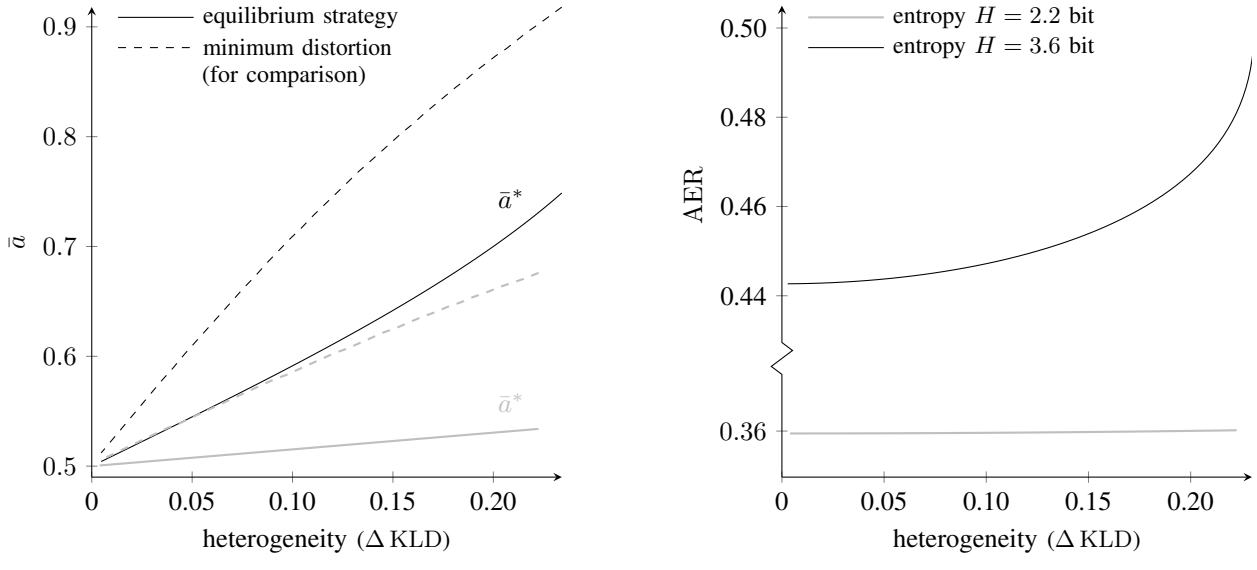
(a) (b)

Fig. 4. Optimal adaptive embedding ($\bar{a}^*$) and detection ($\bar{e}^*$) strategy (a). Alice's equilibrium pay-off measured by the average error rate (AER) (b).



(a) (b)

Fig. 5. Entropy of the cover source (in bits) as a function of the model parameters (a). Level of heterogeneity with annotated sets of constant entropy (b).



(a) (b)

Fig. 6. Embedding strategies (a) and AER of optimal adaptive detection (b) as functions of the level of heterogeneity for two values of constant entropy.

More generally, we regard this stream of work as a step towards adding more theoretical rigor to practical steganography and steganalysis. This might help to narrow the gap between two diverging strands, strong theorems that apply to non-existing cover sources on the one hand, and methods that just work, but little can be said with confidence about their design decisions and security properties on the other. At the time of writing, the biggest research challenges towards this end seem to be the incorporation of non-trivial dependence structures in the cover model as well as adapting and validating the framework for high-dimensional detectors based on machine learning.

The game we introduce is characterized by Alice's objective to minimize the information flow to Eve. As the amount of available information is endogenous in our setup, we do not have discrete information sets like in classical game theory. Our game might constitute a new class of games that could be called *information hiding games*.

ACKNOWLEDGMENT

This research was funded by Deutsche Forschungsgemeinschaft (DFG) under grant "Sichere adaptive Steganographie" and by Archimedes Privatstiftung, Innsbruck, Austria.

APPENDIX A
PROOF OF LEMMA 2

Proof: We carry out the proof for $\mathcal{P}_{(y_0)}$. First, we insert $\bar{a} = 1$ into Eq. (12), simplify, and then expand using Eq. (11):

$$\mathcal{P}_{(y_0)}(u,v) = \frac{{t_0}^{u+(-1)^u} \cdot {t_1}^v}{d_0 \cdot d_1}. \tag{43}$$

We use shorthand $\mathbb{X}_0 \subset \mathbb{X}$ for the set of all even elements in $\mathbb{X}$, and $\mathbb{X}_1 = \mathbb{X} \setminus \mathbb{X}_0$. (The subscript indicates the LSB.) Now, starting from the definition of KLD [2, for example]:

$$\begin{aligned}
&\mathrm{KLD}(\mathcal{P}_0, \mathcal{P}_{(y_0)}) = \\
&= \sum_{u\in\mathbb{X}}\sum_{v\in\mathbb{X}} \mathcal{P}_0(u,v) \cdot \log \frac{\mathcal{P}_0(u,v)}{\mathcal{P}_{(y_0)}(u,v)} && (44)\\
&= \sum_{v\in\mathbb{X}} \left(\sum_{u\in\mathbb{X}_0} \frac{{t_0}^u \cdot {t_1}^v}{d_0 \cdot d_1} \log \left(\frac{{t_0}^u \cdot {t_1}^v}{d_0 \cdot d_1} \cdot \frac{d_0 \cdot d_1}{{t_0}^{u+1} \cdot {t_1}^v} \right) \right. \\
&\qquad \left. + \sum_{u\in\mathbb{X}_1} \frac{{t_0}^u \cdot {t_1}^v}{d_0 \cdot d_1} \log \left(\frac{{t_0}^u \cdot {t_1}^v}{d_0 \cdot d_1} \cdot \frac{d_0 \cdot d_1}{{t_0}^{u-1} \cdot {t_1}^v} \right) \right) && (45)\\
&= \sum_{v\in\mathbb{X}} \left(\sum_{u\in\mathbb{X}_0} \frac{{t_0}^u \cdot {t_1}^v}{d_0 \cdot d_1} \log \frac{1}{t_0} + \sum_{v\in\mathbb{X}_1} \frac{{t_0}^u \cdot {t_1}^v}{d_0 \cdot d_1} \log t_0 \right) && (46)\\
&= \sum_{v\in\mathbb{X}}\sum_{u\in\mathbb{X}} (-1)^{u+1} \cdot \frac{{t_0}^u \cdot {t_1}^v}{d_0 \cdot d_1} \log t_0 && (47)\\
&= \log t_0 \cdot \frac{1}{d_0 \cdot d_1} \cdot \sum_{u\in\mathbb{X}} (-1)^{u+1} \cdot {t_0}^u \cdot \underbrace{\sum_{v\in\mathbb{X}} {t_1}^v}_{=d_1} && (48)\\
&= \log t_0 \cdot \frac{1}{d_0} \cdot (-1) \cdot \sum_{u=0}^{2^\ell - 1} (-t_0)^u. && (49)
\end{aligned}$$

Now, using a closed form for the sum of the geometric series:

$$\begin{aligned}
&= \log t_0 \cdot \frac{1-t_0}{1-{t_0}^{2^\ell}} \cdot (-1) \cdot \frac{1-(-t_0)^{2^\ell}}{1-(-t_0)} && (50)\\
&= \log t_0 \cdot \frac{t_0 - 1}{t_0 + 1}. && (51)
\end{aligned}$$

The proof for $\mathrm{KLD}(\mathcal{P}_0, \mathcal{P}_{(y_1)})$ is analogous. ∎

APPENDIX B
PROOF OF LEMMA 4

Proof: False positives occur if decide classifies a symbol drawn from $f_{t_i}^{(0)}$ as S (for stego).

$$\begin{aligned}
\alpha_i &= \sum_{u=0}^{2^{(\ell-1)}-1} f_{t_i}^{(0)}(2u) \overset{\text{Eq. (1)}}{=} \sum_{u=0}^{2^{(\ell-1)}-1} \frac{(t_i)^{2u}}{d_i} && (52)\\
&= \frac{\frac{{t_i}^{2^\ell}-1}{{t_i}^2-1}}{\frac{{t_i}^{2^\ell}-1}{t_i-1}} = \frac{t_i - 1}{{t_i}^2 - 1} = \frac{1}{t_i + 1}. && (53)
\end{aligned}$$

False negatives occur if decide classifies a symbol drawn from $f_{t_i}^{(1)}$ as C (for cover).

$$\beta_i = \sum_{u=0}^{2^{(\ell-1)}-1} f_{t_i}^{(1)}(2u+1) \tag{54}$$

We rewrite in terms of $f_{t_i}^{(0)}$ (with the help of Lemma 1):

$$= \sum_{u=0}^{2^{(\ell-1)}-1} \frac{f_{t_i}^{(0)}(2u+1)}{t_i} \overset{\text{Eq. (1)}}{=} \sum_{u=0}^{2^{(\ell-1)}-1} \frac{(t_i)^{2u+1}}{d_i \cdot t_i}. \tag{55}$$

After reducing t_i from the right hand side of Eq. (55), the term equals the right hand side of Eq. (52) and it follows that

$$\mathrm{AER} := \frac{\alpha_i + \beta_i}{2} = \frac{1}{t_i + 1}. \tag{56}$$

∎

REFERENCES

[1] R. Böhme, *Advanced Statistical Steganalysis*. Springer, Berlin Heidelberg, 2010.
[2] C. Cachin, "An information-theoretic model for steganography," *Information and Computation*, vol. 192, pp. 41–56, 2004.
[3] R. Böhme, "An epistemological approach to steganography," in *Information Hiding*, ser. Lecture Notes in Computer Science, S. Katzenbeisser and A.-R. Sadeghi, Eds., vol. 5806. Springer, Berlin Heidelberg, 2009, pp. 15–30.
[4] N. Hopper, J. Langford, and L. von Ahn, "Provably secure steganography," in *Advances in Cryptology – CRYPTO 2002*, ser. Lecture Notes in Computer Science, M. Yung, Ed., vol. 2442. Springer, Berlin Heidelberg, 2002, pp. 119–123.
[5] F. Petitcolas, R. Anderson, and M. Kuhn, "Information hiding – a survey," *Proceedings of the IEEE, Special Issue on Protection of Multimedia Content*, vol. 87, no. 7, pp. 1062–1078, 1999.
[6] J. Fridrich, M. Goljan, P. Lisonek, and D. Soukal, "Writing on wet paper," *IEEE Transactions on Signal Processing*, vol. 53, no. 10, pp. 3923–3935, Oct. 2005.
[7] T. Filler, J. Judas, and J. Fridrich, "Minimizing embedding impact in steganography using trellis-coded quantization," in *Media Forensics and Security II*, N. D. Memon, J. Dittmann, A. M. Alattar, and E. J. Delp III, Eds., vol. 7541. SPIE, 2010, p. 754105.

[8] J. von Neumann and O. Morgenstern, *Theory of Games and Economic Behavior*. Princeton University Press, 1944.
[9] P. Schöttle and R. Böhme, "A game-theoretic approach to content-adaptive steganography," in *Information Hiding*, ser. Lecture Notes in Computer Science, M. Kirchner and D. Ghosal, Eds., vol. 7692. Springer, Berlin Heidelberg, 2012, pp. 125–141.
[10] W. Tang, H. Li, W. Luo, and J. Huang, "Adaptive steganalysis against wow embedding algorithm," in *Proceedings of the 2nd ACM Workshop on Information Hiding and Multimedia Security*, 2014, pp. 91–96.
[11] T. Denemark, V. Sedighi, V. Holub, R. COGRANNE, and J. Fridrich, "Selection-channel-aware rich model for steganalysis of digital images," in *IEEE International Workshop on Information Forensics and Security (WIFS)*, 2014, pp. 48–53.
[12] J. Fridrich, *Steganography in Digital Media: Principles, Algorithms, and Applications*. Cambridge University Press, New York, NY, USA, 2009.
[13] T. Filler and J. Fridrich, "Complete characterization of perfectly secure stego-systems with mutually independent embedding operation," in *ICASSP '09: Proceedings of the 2009 IEEE International Conference on Acoustics, Speech and Signal Processing*. Washington, DC, USA: IEEE Computer Society, 2009, pp. 1429–1432.
[14] J. Fridrich, "Minimizing the embedding impact in steganography," in *Proceedings of ACM Multimedia and Security Workshop (MM&SEC)*. New York, NY, USA: ACM, 2006, pp. 2–10.
[15] T. Filler, A. D. Ker, and J. Fridrich, "The square root law of steganographic capacity for markov covers," in *Media Forensics and Security*, E. J. Delp III, J. Dittmann, N. D. Memon, and P. W. Wong, Eds., vol. 7254, no. 1. SPIE, 2009, p. 725408.
[16] Y. Wang and P. Moulin, "Perfectly secure steganography: Capacity, error exponents, and code constructions," *IEEE Transactions on Information Theory*, vol. 54, no. 6, pp. 2706–2722, June 2008.
[17] A. D. Ker, "The square root law in stegosystems with imperfect information," in *Information Hiding*, ser. Lecture Notes in Computer Science, R. Böhme, P. Fong, and R. Safavi-Naini, Eds., vol. 6387. Springer, Berlin Heidelberg, 2010, pp. 145–160.
[18] ——, "A curiosity regarding steganographic capacity of pathologically nonstationary sources," in *Media Watermarking, Security, and Forensics III*, N. D. Memon, J. Dittmann, A. M. Alattar, and E. J. Delp III, Eds., vol. 7880. SPIE, 2011, p. 78800E.
[19] P. Schöttle, S. Korff, and R. Böhme, "Weighted stego-image steganalysis for naive content-adaptive embedding," in *4th IEEE International Workshop on Information Forensics and Security (WIFS 2012)*. IEEE, 2012, pp. 193–198.
[20] E. Lam and J. Goodman, "A mathematical analysis of the DCT coefficient distributions for images," *IEEE Transactions on Image Processing*, vol. 9, no. 10, pp. 1661–1666, Oct. 2000.
[21] S. Inusah and T. J. Kozubowski, "A discrete analogue of the Laplace distribution," *Journal of Statistical Planning and Inference*, vol. 136, no. 3, pp. 1090–1102, 2006.
[22] J. C. Harsanyi, "Games with incomplete information played by "Bayesian" players, I-III Part I. The basic model," *Management Science*, vol. 14, no. 3, pp. 159–182, 1967.
[23] J. Nash, "Non-cooperative games," *The Annals of Mathematics*, vol. 54, no. 2, pp. 286–295, 1951.
[24] V. Pruzhansky, "Some interesting properties of maximin strategies," *International Journal of Game Theory*, vol. 40, no. 2, pp. 351–365, 2011.
[25] N. Nisan, T. Roughgarden, E. Tardos, and V. V. Vazirani, *Algorithmic game theory*. Cambridge University Press Cambridge, 2007.
[26] T. Pevný, T. Filler, and P. Bas, "Using high-dimensional image models to perform highly undetectable steganography," in *Information Hiding*, ser. Lecture Notes in Computer Science, R. Böhme, P. Fong, and R. Safavi-Naini, Eds., vol. 6387. Springer, Berlin Heidelberg, 2010, pp. 161–177.
[27] T. Denemark, J. Fridrich, and V. Holub, "Further study on the security of SUNIWARD," *Proceedings SPIE, Electronic Imaging, Media Watermarking, Security, and Forensics*, vol. 9028, pp. 2–6, 2014.
[28] V. Sedighi, J. Fridrich, and R. Cogranne, "Content-adaptive pentary steganography using the multivariate generalized gaussian cover model," *Proceedings SPIE, Electronic Imaging, Media Watermarking, Security, and Forensics*, vol. 9409, pp. 94 090H–94 090H–13, 2015.
[29] B. Johnson, P. Schöttle, and R. Böhme, "Where to hide the bits?" in *GameSec 2012*, ser. Lecture Notes in Computer Science, J. Grossklags and J. Walrand, Eds., no. 7638. Springer, Berlin Heidelberg, 2012, pp. 1–17.
[30] B. Johnson, P. Schöttle, A. Laszka, J. Grossklags, and R. Böhme, "Bitspotting: Detecting optimal adaptive steganography," in *12th International Workshop on Digital Forensics and Watermarking (IWDW 2013)*, ser. Lecture Notes in Computer Science, Y.-Q. Shi, H.-J. Kim, and F. Pérez-González, Eds., vol. 8389. Springer, Berlin Heidelberg, 2014, pp. 3–18.
[31] P. Schöttle, B. Johnson, A. Laszka, J. Grossklags, and R. Böhme, "A game-theoretic analysis of content-adaptive steganography with independent embedding," in *Proceedings of the 21st European Signal Processing Conference (EUSIPCO)*, 2013.
[32] T. Denemark and J. Fridrich, "Detection of content adaptive LSB matching: a game theory approach," in *Proceedings SPIE, Electronic Imaging, Media Watermarking, Security, and Forensics*, vol. 9028, 2014, pp. 902 804–902 804–12.
[33] V. Sedighi and J. Fridrich, "Effect of imprecise knowledge of the selection channel on steganalysis," in *Proceedings of the 3rd ACM Workshop on Information Hiding and Multimedia Security*, 2015, pp. 33–42.
[34] M. Ettinger, "Steganalysis and game equilibria," in *Information Hiding*, ser. Lecture Notes in Computer Science, D. Aucsmith, Ed., vol. 1525. Springer, Berlin Heidelberg, 1998, pp. 319–328.
[35] A. D. Ker, "Batch steganography and the threshold game," in *Security, Steganography, and Watermarking of Multimedia Contents IX*, E. J. Delp III and P. W. Wong, Eds., vol. 6505, no. 1. SPIE, 2007, p. 650504.
[36] A. Orsdemir, O. Altun, G. Sharma, and M. Bocko, "Steganalysis-aware steganography: Statistical indistinguishability despite high distortion," in *Security, Forensics, Steganography, and Watermarking of Multimedia Contents X*, E. J. Delp III, P. W. Wong, J. Dittmann, and N. D. Memon, Eds., vol. 6819, no. 1. SPIE, 2008, p. 681915.
[37] S. Katzenbeisser and F. A. P. Petitcolas, "Defining security in steganographic systems," in *Security and Watermarking of Multimedia Contents IV*, E. J. Delp III and P. W. Wong, Eds., vol. 4675, no. 1. SPIE, 2002, pp. 50–56.
[38] N. Dalvi, P. Domingos, Mausam, S. Sanghai, and D. Verma, "Adversarial classification," in *Proceedings of the 10th ACM SIGKDD International Conference on Knowledge Discovery and Data Mining*. New York, NY, USA: ACM, 2004, pp. 99–108.
[39] M. Barni and B. Tondi, "The source identification game: An information-theoretic perspective," *IEEE Transactions on Information Forensics and Security*, vol. 8, no. 3, pp. 450–463, Mar. 2013.

Pascal Schöttle is a member of the Security and Privacy Lab at Universität Innsbruck, Austria. He received his MSc degree in IT Security from Ruhr-University Bochum and his Ph.D. degree in computer science from the University of Münster, Germany.

His research interests focus on multimedia security and steganography in particular, and include asymmetric cryptography and network anomaly detection.

Rainer Böhme is Professor for Security and Privacy at the Institute of Computer Science, Universität Innsbruck, Austria. Prior to that he was Assistant Professor of Information Systems and IT Security at the University of Münster in Germany and Postdoctoral Fellow at the International Computer Science Institute in Berkeley, California.

His research interests include multimedia security, digital forensics, privacy-enhancing technologies, as well as economics of information security and privacy and virtual currencies. He holds a Master's degree in Communication Science and Economics and a Doctorate in Computer Science, both from Technische Universität Dresden in Germany.