Where to Hide the Bits ?

Benjamin Johnson[1], Pascal Schöttle[2], and Rainer Böhme[2]

[1] Department of Mathematics, UC Berkeley, USA
[2] Department of Information Systems, University of Münster, Germany

Abstract. We present a stochastic two-player zero-sum game between defender and attacker related to the security of practical steganography. The defender wants to hide a message in a cover object drawn by nature. The attacker wants to distinguish plain covers from those with a hidden message. We study the case of n-bit covers, independently but not identically distributed to allow for variation in the predictability between parts of the cover. The defender knows the predictability exactly and chooses k embedding positions. The attacker may obtain side information to predict one chosen position of the cover and compare it to the observed object to make a decision. We present a unique mixed strategy Nash equilibrium for this game. It turns out that the attacker's strategy is independent of the number of hidden bits k.

Keywords: Game Theory, Information Hiding, Steganography, Security.

1 Introduction

Steganography is the art and science of hiding the very existence of a secret message by embedding it into inconspicuous cover data, such as image or audio files. A (minimal) steganographic embedding function takes as input a message and a key. It outputs a so-called stego object, which is sent to the recipient, who can extract the hidden message using the shared secret key. Steganalysis, the countermeasure, tries to detect steganography by deciding whether an observed object is cover or stego without knowing the secret key.

Prior work has established the following theory. Given a cover distribution $\mathcal{P}_0$ and a fixed embedding function, the distribution of stego objects $\mathcal{P}_1$ is completely determined. *Perfect steganography* is possible if $\mathcal{P}_0 = \mathcal{P}_1$, and efficient codes exist to embed a message [12]. For perfect steganography, a computationally unbounded steganalyst's chance to make the correct detection decision is no better than random guessing. If $\mathcal{P}_0$ is unknown, but can be sampled efficiently, then the existence of a cryptographic oneway function is sufficient to construct an embedding function that achieves $\mathcal{P}_0 \approx \mathcal{P}_1$ so that computationally bounded steganalysts have only negligible advantage at the detection decision [6]. If $\mathcal{P}_0 \neq \mathcal{P}_1$, we speak of *imperfect steganography* and the degree of imperfection can be quantified by the Kullback–Leibler divergence (KLD) between $\mathcal{P}_0$ and $\mathcal{P}_1$ [2]. All practical steganographic embedding functions proposed for real

J. Grossklags and J. Walrand (Eds.): GameSec 2012, LNCS 7638, pp. 1–17, 2012.

cover media belong to the class of imperfect steganography. This is so because $\mathcal{P}_0$ is unknown, arguably unknowable, and therefore it is virtually impossible to find an embedding function that preserves $\mathcal{P}_0$ exactly [1]. Security bounds for this relevant case of imperfect steganography have been derived mainly under the strong, yet conservative, assumption that the steganalyst knows $\mathcal{P}_0$ [9]. Few extensions exist for the case where both steganographer and steganalyst have incomplete knowledge of $\mathcal{P}_0$ [8]. All these studies predict and experimentally validate asymptotic detectability bounds, but they contain no constructive elements on how to design secure embedding functions.

Due to this deficit of actionable theory, engineering efforts to design practical embedding functions are dominated by heuristics and simulation experiments, often involving machine-learning techniques [10]. One rule of thumb is to minimize the embedding distortion, reflecting the conjecture that $\mathcal{P}_0$ is locally smooth around the realized cover. The actual measures of distortion vary between approaches. Another recurring thread is the idea of *content-adaptive* steganography. It is based on the observation that most cover sources produce *heterogeneous* covers, meaning that if the embedding domain of a cover object is represented by a fixed sequence of symbols, then different positions exhibit different statistical properties. For example, natural images often consist of smooth areas and gradients, but also contain some sharp edges or regions with noisy texture. Because these areas differ in predictability, e. g., in the accuracy of local statistical models, the steganalyst may find it easier to detect subtle embedding changes by deviations from the model in more predictable than in less predictable spots of the cover. To exploit differences in detectability, a content-adaptive embedding function tries to improve steganographic security by concentrating k embedding changes in less predictable areas of a cover, rather than distributing them uniformly over all n possible embedding positions [1].

Our contribution is to narrow the gap between theory and practice with a game-theoretic analysis of the optimal choice of embedding positions in the realistic regime where both steganographer and steganalyst have incomplete knowledge of $\mathcal{P}_0$. As initially pointed out in [11], game theory is the method of choice in this regime, because both players have to decide in which positions they hide or look for evidence of embedding, respectively, in anticipation of the opponent's action. The specific contribution of this work is to extend our model in [11] from the very artificial case of only two positions to covers of size n. Our results here are constructive in the sense that the equilibrium strategy can be efficiently computed for any given vector of predictability.

Here is the structure of our paper. The next Section 2 specifies the game setup and connects it to a specific interpretation of the steganographic communication model. Section 3 presents the solution of the game, starting with message sizes of $k = 1$ bits and generalizing from there. The discussion in Sect. 4 comments on the applicability of our results for the design of secure practical steganography and points to alternative interpretations in the broader field of information security. The final Section 5 concludes with an outlook on open problems.

2 Problem Definition

Let Alice be the defender (steganographer) and Eve be the attacker (steganalyst). Figure 1 visualizes the steganographic communication model. Function embed takes as input the secret message, a fresh cover, and the secret key. It outputs a stego object which is as indistinguishable as possible from covers. The stego object is communicated over the insecure channel. The recipient listening at the other end can apply function extract to retrieve the message. We do not further consider the recipient, but abstract from the necessary coding and cryptography layers which ensure that the recipient can always extract the message correctly. Our game is formulated between Alice, who implements embed, and Eve, who implements function detect. This function outputs a decision of whether the observed object is a plain cover or a stego object. Figure 1 differs from the standard model by allowing Eve to query some side information directly from the cover. Recall that practical steganalysis can often estimate such side information from the observed object. Therefore, we will elaborate below why we require this explicit interaction in our game.

To formalize the role of side information, let the embedding domain of a cover be a random sequence of n symbols with varying predictability from some kind of side information. This side information is fully available to Alice and partly available to Eve. Practical predictors, for instance, exploit spatial correlation in the local neighborhood of a symbol to estimate its most likely value. We assume that both players can exactly quantify the predictability[1] per symbol and order the symbols of the embedding domain by increasing predictability. (Both reordering and potential domain transformations are reversible so that the object on the communication channel always appears in its original order and domain.)

Therefore, we consider a vector $\boldsymbol{X} = (X_0, \ldots, X_{n-1})$ of independent random variables drawn from a binary alphabet $\mathcal{C} = \{0, 1\}$, with realizations typeset in lower case, $\boldsymbol{x} = (x_0, \ldots, x_{n-1})$. Note that real covers may have a larger alphabet, but we settle on bits for a clearer notion of predictability. Moreover, practical embedding functions often work on a vector of binary residuals, such as the sequence of all least significant bits. The monotonically increasing function $f(i) : \{0, \ldots, n-1\} \rightarrow [\frac{1}{2}, 1]$ defines the probability of X_i taking its most likely value. Without loss of generality, let $f(i) = P(X_i = 1)$ for the analysis.

To anchor the two ends of the predictability range, we require $f(0) = \frac{1}{2} + \varepsilon$ and $f(n-1) = 1 - \varepsilon$. We need a strictly positive ε to ensure that we operate in the imperfect steganography regime. If ε were zero, Alice could embed at least one bit into x_0 without risk of detection. Similarly, if $P(X_{n-1} = 1) = 1$, embedding into x_{n-1} would allow detection with certainty.

Alice's action space is to flip k bits of a given cover realization $\boldsymbol{x}$ to embed a hidden message. There exist appropriate key-dependent codes ensuring that

[1] Our notion of *predictability* closely corresponds to the *detectability profile* in [5] or the *adaptivity criterion* in [11]. Both concepts can be interpreted as proxies to estimate the local predictability.

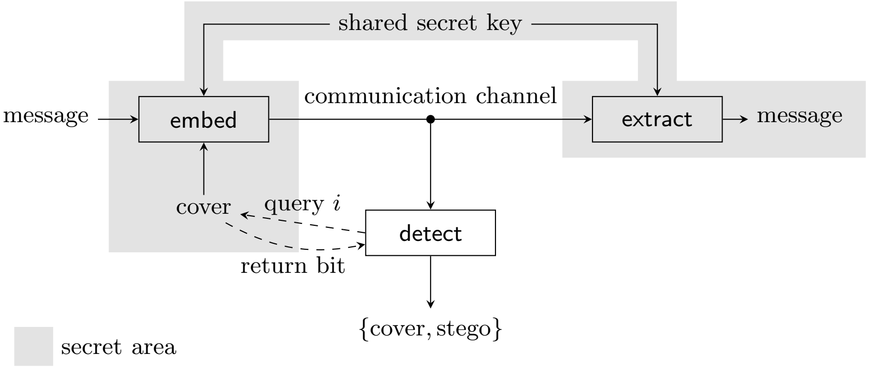

Fig. 1. Block diagram of steganographic communication system with side information

a message of length $k - \mathcal{O}(1)$ can be embedded by changing k cover values irrespective of their position such that the recipient can extract the message with the knowledge of a shared secret key (see for example [4]). Alice chooses a k-sized subset of $\{0, \ldots, n-1\}$ indicating the embedding positions. Her mixed strategy action space is a probability distribution a over all k-sized subsets of $\{0, \ldots, n-1\}$.

Eve tries to decide whether an observed bit vector is a cover or a stego object. We follow the convention in [7] and require that covers and stego objects are equally likely on the communication channel. This can be modeled by assuming that nature flips an unbiased coin to decide if Eve sees Alice's stego object or a cover drawn from the same cover source available to Alice. Eve's optimal decision rule would be a likelihood ratio test using the joint distributions over all cover and stego objects, $\mathcal{P}_0$ and $\mathcal{P}_1$. In practice, however, $\mathcal{P}_0$ and $\mathcal{P}_1$ are unknown and Eve can only make local decisions for individual symbols using a local predictor. We stipulate that Eve can use her knowledge about the marginal distributions of $\mathcal{P}_0$ and $\mathcal{P}_1$ to make optimal local decisions, although this is not always the case for practical steganalysis. While our game might be too optimistic for Eve in this respect, we contrast this by requiring that Eve only looks at one position. To justify this constraint in the basic model, we must assume that the side information necessary for prediction is only available for one position. Therefore, it cannot be estimated from the cover and we must assume an interactive query mechanisms (see Fig. 1). As a result, Eve's mixed strategy action space is a probability distribution e over all n positions for which she can query side information of variable precision, depending on the position's predictability. For all other positions, she cannot tell if $P(X_i = 0) > P(X_i = 1)$ or $P(X_i = 0) < P(X_i = 1)$ in covers. Therefore, she does not gain any information from including the values at these positions in her decision.

It is obvious that Eve's task is very hard in this setup, because if $k = 1$, her advantage over random guessing is not better than ε even if Alice deterministically embeds in the first symbol. If Alice randomizes her strategy, then Eve's advantage shrinks. If Alice embeds more bits, Eve's advantage increases because Alice has to use more predictable (i. e., less secure) positions. Our objective is to quantify by how much, and if (and where) there is an equilibrium.

The following objective function defines a zero-sum game: Alice tries to increase her security by maximizing Eve's decision error, whereas Eve tries to minimize it. We map this to the payoff structure given in Table 1. Note that this payoff matrix induces an objective function based on the equal error rate. For practical applications, the payoff matrix might need adjustment to account for the harm caused by false positive and false negatives, respectively.

Table 1. Payoff for (Eve, Alice)

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

Figure 2 (p. 6) summarizes the game for $k = 1$ in an extensive form graph. From left to right, first, nature draws a cover from $\mathcal{P}_0$, then Alice chooses her single (because $k = 1$) embedding position, creating a stego object (black nodes). A coin flip, invisible to Eve, decides whether she sees the stego or cover object. Then Eve chooses the position she wants to compare with a prediction to make her decision, and outputs the decision result (C for cover or S for stego). Shaded nodes indicate the cases where Eve wins, i. e., she receives positive payoff.

3 Solving the Model

3.1 Preliminaries

We begin by formulating Eve's local decision rule. Eve observes the probability $f(i)$ that bit i is 1. Since $f(i)$ is greater than $\frac{1}{2}$, the object is more likely to be a cover if the observed bit is 1, and more likely to be stego if the observed bit is 0. This constrains Eve's decision rule based on her observation at position i.

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

To simplify the exposition of our equilibrium results, we introduce the notation

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

The function $f(i)$ was introduced as the probability of seeing 1 at position i, and it measures the predictability at position i. The function $\tilde{f}(i)$ can be interpreted as measuring the bias at position i.

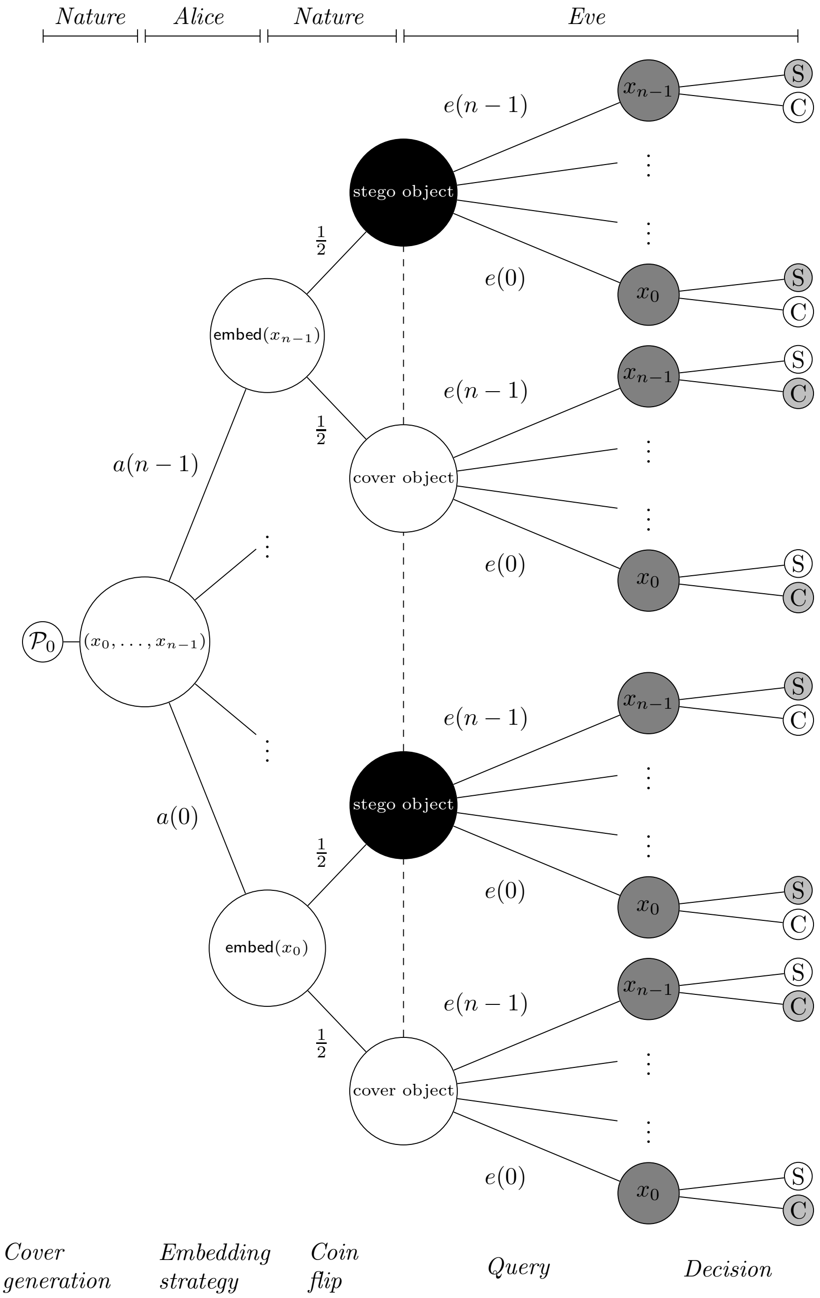

Fig. 2. Extensive form of the game for $k = 1$. The dashed line indicates Eve's information set. The dark gray nodes represent Eve's query strategy and the light gray nodes are the situations in which Eve wins the game.

Recall that Alice's mixed strategy space is a probability distribution over size-k subsets of $\{0, \ldots, n-1\}$. For a subset S of k positions, $a(S)$ is the probability that Alice embeds her bits in these k positions; and we have $\sum_S a(S) = 1$. Overloading notation, let us define the projection of Alice's mixed strategy onto positions to be the total probability that Alice embeds in position i. Formally, we define $a(i)$ for $i \in \{0, \ldots, n-1\}$ as

$$a(i) = \sum_{\{S : i \in S\}} a(S). \tag{3}$$

If Alice embeds in just one position, then $a(i) = a(\{i\})$ and $\sum_{i=0}^{n-1} a(i) = 1$. If Alice embeds k bits, then

$$\sum_{i=0}^{n-1} a(i) = k. \tag{4}$$

Eve's mixed strategy action space is a distribution over positions. Eve queries the bias at position i with probability $e(i)$ and decides stego or cover based only on her observation at position i.

3.2 Game Outcome

We quantify the payoff of Eve and Alice as a function of the bias $\tilde{f}(i)$ at each position, Eve's mixed strategy $e(i)$, and the projection of Alice's mixed strategy onto positions $a(i)$.

Theorem 1 (Game Outcome). *If $\tilde{f}$ is the bias function, e is Eve's mixed strategy, and a is Alice's mixed strategy, then the total expected payoff for (Eve, Alice) is*

$$\left(2 \sum_{i=0}^{n-1} e(i)a(i)\tilde{f}(i), -2 \sum_{i=0}^{n-1} e(i)a(i)\tilde{f}(i) \right). \tag{5}$$

Proof. First assume that Eve looks only at position i. Under this assumption, we may determine the probability she wins the game by enumerating all possible ways the world could be, and adding up the respective probabilities. We may think of the process as an orderly sequence of events. First, nature chooses whether Eve sees a cover object or a stego object by flipping an unbiased coin. The cover object is then instantiated with a realization x_i of position i, with $P(X_i = 1) = f(i)$. If nature chose stego, then Alice flips bit i with probability $a(i)$. Finally, Eve decides whether the object is cover or stego by looking at her observed bit. She decides cover if the bit is 1 and stego if the bit is 0. Table 2 records the events, probabilities, and decision outcomes for each possible case.

Table 2. Game outcome in different states of the world

	Value of x_i				
Reality	Cover	Observed	Probability	Eve's decision	Winner
C	1	1	$\frac{1}{2} \cdot f(i)$	C	Eve
C	0	0	$\frac{1}{2} \cdot (1 - f(i))$	S	Alice
S	1	0	$\frac{1}{2} \cdot f(i) \cdot a(i)$	S	Eve
S	1	1	$\frac{1}{2} \cdot f(i) \cdot (1 - a(i))$	C	Alice
S	0	1	$\frac{1}{2} \cdot (1 - f(i)) \cdot a(i)$	C	Alice
S	0	0	$\frac{1}{2} \cdot (1 - f(i)) \cdot (1 - a(i))$	S	Eve

Legend: C = cover, S = stego

Given that Eve looks only at position i, her probability of winning is

$$\begin{aligned}
&\frac{1}{2}\left(f(i) + f(i)a(i) + (1 - f(i))(1 - a(i))\right) && (6)\\
=&\frac{1}{2}\left(f(i) + f(i)a(i) + 1 - a(i) - f(i) + f(i)a(i)\right) && (7)\\
=&\frac{1}{2}\left(1 + 2f(i)a(i) - a(i)\right) && (8)\\
=&\frac{1}{2} + a(i)\left(f(i) - \frac{1}{2}\right) && (9)\\
=&\frac{1}{2} + a(i)\tilde{f}(i). && (10)
\end{aligned}$$

Hence Eve's total probability of winning is

$$\begin{aligned}
&\sum_{i=0}^{n-1} e(i)\left(\frac{1}{2} + a(i)\tilde{f}(i)\right) && (11)\\
=&\frac{1}{2} + \sum_{i=0}^{n-1} e(i)a(i)\tilde{f}(i), && (12)
\end{aligned}$$

and thus Eve's total expected game payoff is

$$\begin{aligned}
&\mathrm{P}(\text{Eve wins}) \cdot 1 + \mathrm{P}(\text{Eve loses}) \cdot (-1) && (13)\\
=&\left(\frac{1}{2} + \sum_{i=0}^{n-1} e(i)a(i)\tilde{f}(i)\right) \cdot 1 + \left(\frac{1}{2} - \sum_{i=0}^{n-1} e(i)a(i)\tilde{f}(i)\right) \cdot (-1) && (14)\\
=&2\sum_{i=0}^{n-1} e(i)a(i)\tilde{f}(i). && (15)
\end{aligned}$$

The total expected payoff for (Eve, Alice) is thus

$$\left(2\sum_{i=0}^{n-1} e(i)a(i)\tilde{f}(i), -2\sum_{i=0}^{n-1} e(i)a(i)\tilde{f}(i)\right). \tag{16}$$

□

3.3 Nash Equilibria

We now turn our attention to the game's Nash equilibria.

Hiding One Bit. We start with analyzing the case of $k = 1$. This simplifies Alice's mixed strategy action space to a probability distribution over the set $\{0, \ldots, n-1\}$. Recall the convention that $a(i)$ is the probability that Alice embeds into position i.

Lemma 1 (Exclusion of pure strategies). *There is no equilibrium in which either Alice or Eve assigns zero probability to any i.*

Proof. Assume Alice assigns zero probability to position i. Then Eve gains no advantage from assigning positive probability to position i. Hence, Eve's best response would assign zero probability to position i. But then Alice can completely eliminate Eve's advantage by assigning probability 1 to position i. So Alice is not in equilibrium.

Assume Eve assigns zero probability to position i, then Alice can completely eliminate Eve's advantage by assigning probability 1 to position i. But then Eve's best response would be assign probability 1 to position i. So Eve is not in equilibrium. □

It is useful to quantify Eve's advantage from looking at one position and observing the bias. The following two definitions facilitate such quantification.

Definition 1 (Eve's local advantage). *Eve's* local advantage *at position i is $a(i) \cdot \tilde{f}(i)$.*

Definition 2 (Eve's total advantage). *Eve's* total advantage *is the weighted sum over all her local advantages at positions $0, \ldots, n-1$, i. e., $\sum_{i=0}^{n-1}\left(e(i)a(i)\tilde{f}(i)\right)$.*

Observe that from Theorem 1, Eve's expected game payoff is exactly twice her total advantage. Hence we may consider total advantage as a quantity of primary interest. Eve's primary objective is to increase her total advantage, while Alice's primary objective is to reduce it. Our next lemma characterizes the structure of possible equilibria in relation to Eve's local and total advantages.

Lemma 2 (Uniform local advantage condition). *A necessary condition for any equilibrium is that Eve's local advantage is uniform over $i = 0, \ldots, n-1$.*

Proof. Suppose Eve's local advantage is not uniform. Then there is at least one position i where her local advantage is not as high as it is at some other position j. I. e., $a(i) \cdot \tilde{f}(i) < a(j) \cdot \tilde{f}(j)$. Eve can then strictly increase her total advantage by setting $e(j) = e(j) + e(i)$ and then setting $e(i) = 0$. (The resulting difference in her total advantage will be $e(i)(a(j) \cdot \tilde{f}(j) - a(i) \cdot \tilde{f}(i))$, which is positive.) So the situation is not an equilibrium. □

This condition can actually be fulfilled, as shown in the next lemma.

Lemma 3 (Existence of Alice's unique strategy). *In any equilibrium, Alice's strategy to embed one bit is*

$$a(i) = \frac{1}{\tilde{f}(i) \cdot \sum_{j=0}^{n-1} \frac{1}{\tilde{f}(j)}}. \tag{17}$$

Proof. We start with the condition from Lemma 2,

$$a(i) \cdot \tilde{f}(i) = a(j) \cdot \tilde{f}(j) \quad \forall\, i \neq j. \tag{18}$$

This implies that there is a constant C with $a(i) \cdot \tilde{f}(i) = C$ for each i, and hence $a(i) = \frac{C}{\tilde{f}(i)}$ for some C.

Now by the probability axiom,

$$\sum_{i=0}^{n-1} a(i) = 1, \tag{19}$$

so that $\sum_{i=0}^{n} \frac{C}{\tilde{f}(i)} = 1$, and hence $C = \frac{1}{\sum_{i=0}^{n} \tilde{f}(i)}$. It follows that $a(i) = \frac{1}{\tilde{f}(i) \cdot \sum_{j=0}^{n-1} \frac{1}{\tilde{f}(j)}}$. I. e. the two constraints (18) and (19) completely determine $a(i)$. □

Lemma 4 (Game outcome in equilibrium). *The game's outcome for (Eve, Alice) in equilibrium is*

$$\left(\frac{2}{\sum_{j=0}^{n-1} \frac{1}{\tilde{f}(j)}}, \frac{-2}{\sum_{j=0}^{n-1} \frac{1}{\tilde{f}(j)}} \right). \tag{20}$$

Proof. Alice's strategy fixes Eve's total advantage, which in turn fixes Eve's payoff. As Alice has only one candidate strategy in equilibrium, we know Eve's total advantage in equilibrium must be

$$\sum_{i=0}^{n-1} \Big(e(i) a(i) \tilde{f}(i) \Big) = \sum_{i=0}^{n-1} \left(e(i) \frac{\tilde{f}(i)}{\tilde{f}(i) \sum_{j=0}^{n-1} \frac{1}{\tilde{f}(j)}} \right) = \frac{1}{\sum_{j=0}^{n-1} \frac{1}{\tilde{f}(j)}}, \tag{21}$$

hence Eve's payoff in equilibrium is $\frac{2}{\sum_{j=0}^{n-1} \frac{1}{\tilde{f}(j)}}$ and the result follows. □

Turning now to Eve's strategy, we may construe her objective as preserving her total advantage.

Lemma 5 (Uniform weighted bias condition). *A necessary condition for any equilibrium is that $e(i) \cdot \tilde{f}(i)$ is uniform over $i = 0, \ldots, n-1$.*

Proof. Suppose Alice is playing her unique strategy in equilibrium from Lemma 3 and that (for the sake of contradiction) there exist $i \neq j$ with $e(i) \cdot \tilde{f}(i) < e(j) \cdot \tilde{f}(j)$. Then Alice can decrease Eve's total advantage by adopting a new strategy a^* with, $a^*(j) = 0$; $a^*(i) = a(i) + a(j)$; and $a^*(r) = a(r)$ for $r \neq i, j$.

The difference in Eve's total advantage is

$$\sum_{r=0}^{n-1} \Big(e(r)a^*(r)\tilde{f}(r) \Big) - \sum_{r=0}^{n-1} \Big(e(r)a(r)\tilde{f}(r) \Big) \tag{22}$$

$$= e(i)a^*(i)\tilde{f}(i) + w_j a^*(j)\tilde{f}(i) - (e(i)a(i)\tilde{f}(i) + e(j)a(j)\tilde{f}(i)) \tag{23}$$

$$= e(i)(a(i) + a(j))\tilde{f}(i) - (e(i)a(i)\tilde{f}(i) + e(j)a(j)\tilde{f}(j)) \tag{24}$$

$$= e(i)a(j)\tilde{f}(i) - e(j)a(j)\tilde{f}(i) \tag{25}$$

$$= a(j)(e(i)\tilde{f}(i) - e(j)\tilde{f}(j)) \tag{26}$$

$$< 0.$$

So Alice would prefer to change strategies, in violation of the equilibrium condition. □

Lemma 6 (Existence of Eve's unique strategy). *In any equilibrium for the one-bit case, Eve's probability $e(i)$ of looking at position i must be the same as Alice's probability of embedding at position i:*

$$e(i) = \frac{1}{\tilde{f}(i) \cdot \sum_{j=0}^{n-1} \frac{1}{\tilde{f}(j)}}. \tag{27}$$

Proof. The formula follows from the uniform weighted bias condition: $e(i) \cdot \tilde{f}(i) = e(j) \cdot \tilde{f}(j)$ for all $i \neq j$; and the probability constraint on Eve's mixed strategy: $\sum_{j=0}^{n-1} e(i) = 1$. The argument that these conditions uniquely determine a function is given in Lemma 3. □

Theorem 2 (Unique Nash equilibrium). *There is a unique Nash equilibrium for the one-bit game where Alice embeds in position i with probability*

$$a(i) = \frac{1}{\tilde{f}(i) \cdot \sum_{j=0}^{n-1} \frac{1}{\tilde{f}(j)}}, \tag{28}$$

and Eve observes position i with probability

$$e(i) = \frac{1}{\tilde{f}(i) \cdot \sum_{j=0}^{n-1} \frac{1}{\tilde{f}(j)}}, \tag{29}$$

and the expected payoff outcome for (Eve, Alice) is $\left(\frac{2}{\sum_{j=0}^{n-1} \frac{1}{\tilde{f}(j)}}, -\frac{2}{\sum_{j=0}^{n-1} \frac{1}{\tilde{f}(j)}} \right)$.

Proof. See Lemmas 3, 4, and 6. □

Hiding *k* Bits

Lemma 7 (Alice's k-bit strategy). *In any equilibrium, the projection of Alice's mixed strategy distribution onto singleton subsets satisfies*

$$a(i) = \frac{k}{\tilde{f}(i) \cdot \sum_{j=0}^{n-1} \frac{1}{\tilde{f}(j)}}. \tag{30}$$

Proof. First, any equilibrium must satisfy the uniform advantage condition, as the logic from Lemma 2 applies also in the k-bit case. Thus we have

$$a(i) \cdot \tilde{f}(i) = a(j) \cdot \tilde{f}(j) \quad \forall\, i \neq j. \tag{31}$$

Since we also have

$$\sum_{i=0}^{n-1} a(i) = k, \tag{32}$$

the function $a(i)$ is completely determined as $a(i) = \frac{k}{\tilde{f}(i) \cdot \sum_{j=0}^{n-1} \frac{1}{\tilde{f}(j)}}$. □

Lemma 8 (Eve's k-bit strategy). *In any equilibrium, Eve's mixed strategy distribution is*

$$e(i) = \frac{1}{\tilde{f}(i) \cdot \sum_{j=0}^{n-1} \frac{1}{\tilde{f}(j)}}. \tag{33}$$

Proof. Eve's strategy must satisfy the uniform weighted bias condition: $e(i) \cdot \tilde{f}(i)$ is uniform in i; as the logic from Lemma 5 still applies in the k-bit case. Since we also have $\sum_{i=0}^{n-1} e(i) = 1$, these two conditions imply $e(i) = \frac{1}{\tilde{f}(i) \cdot \sum_{j=0}^{n-1} \frac{1}{\tilde{f}(j)}}$. □

Theorem 3 (k-bit Nash equilibria). *There is a Nash equilibrium for the k-bit game where the projection of Alice's distribution onto singleton positions satisfies*

$$a(i) = \frac{k}{\tilde{f}(i) \cdot \sum_{j=0}^{n-1} \frac{1}{\tilde{f}(j)}}, \tag{34}$$

and Eve observes position i with probability

$$e(i) = \frac{1}{\tilde{f}(i) \cdot \sum_{j=0}^{n-1} \frac{1}{\tilde{f}(j)}}, \tag{35}$$

and the expected payoff outcome for (Eve, Alice) is $\left(\frac{2k}{\sum_{j=0}^{n-1} \frac{1}{\tilde{f}(j)}}, -\frac{2k}{\sum_{j=0}^{n-1} \frac{1}{\tilde{f}(j)}} \right)$.

The equilibrium is unique up to the projection of Alice's mixed strategy.

Proof. See Lemmas 7 and 8 for the strategies. For the payoffs, note that Eve's advantage in equilibrium is

$$\sum_{i=0}^{n-1} \left(e(i)a(i)\tilde{f}(i) \right) = \sum_{i=0}^{n-1} \left(e(i) \frac{k\tilde{f}(i)}{\tilde{f}(i) \sum_{j=0}^{n-1} \frac{1}{\tilde{f}(j)}} \right) = \frac{k}{\sum_{j=0}^{n-1} \frac{1}{\tilde{f}(j)}}, \tag{36}$$

so that Eve's payoff in equilibrium is $\frac{2k}{\sum_{j=0}^{n-1} \frac{1}{\tilde{f}(j)}}$.

□

The following two corollaries are easily observable.

Corollary 1. *Eve's mixed strategy in equilibrium is independent of the number of embedded bits.*

Corollary 2. *Eve's expected payoff in equilibrium increases linearly with the number of embedded bits.*

4 Discussion

4.1 Numerical Examples

Figures 3 and 4 display numerical examples of the equilibrium in our game, instantiated with the parameters $k = 1$ and $n = 100$. The red line shows the prediction function $f(i)$; note the right hand scale. The gray bars display Alice's and Eve's identical optimal strategies (left hand scale). In Figure 3, the parameter ε is set relatively high and the prediction function f is linear.

Figure 4 is more realistic. It shows a small ε and a non-linear prediction function f with the majority of positions being relatively well predictable, just like large homogeneous areas in natural images. Both figures show that the value of $a(0)$ is at its maximum. This illustrates again the advantage of content-adaptive embedding over random uniform embedding if the cover source produces heterogeneous covers. Nonetheless, the fact that $a(i) > 0$ for all i suggests that the steganographer should potentially use every available position and not only the least predictable ones, unlike what is seen in many practical schemes.

4.2 Adequacy of Eve's Constraints

We have motivated our game with practical content-adaptive steganography in heterogeneous covers. Its solution can guide the development of more secure embedding functions and detectors implementing the best response against known embedding strategies. Our results recover the conclusions of [11] for the imperfect steganography regime, namely that random uniform embedding is only optimal in homogeneous covers, and naive adaptive embedding (i. e., deterministically choosing the k least predictable symbols) is inferior to optimal adaptive embedding, the equilibrium strategy. The extension to n cover symbols presented here is an important step towards bringing more realism to the model.

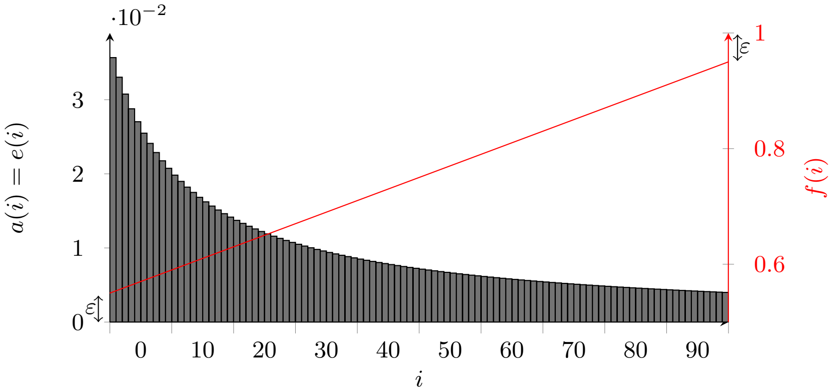

Fig. 3. Equilibrium strategies for $\varepsilon \gg 0$ and a linear function f

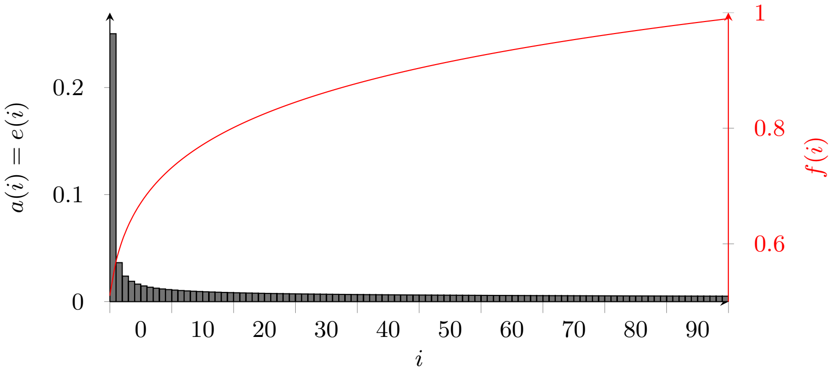

Fig. 4. Equilibrium strategies for $\varepsilon \approx 0$ and a non-linear function f

The biggest remaining obstacle (common to [11] and this paper) is the constraint that Eve can only look at one position at a time. In fact, all practical steganalysis methods are constrained to local – possibly suboptimal – decisions, but their output is an *aggregation* of local decisions for all positions in the observed sequence. Most known steganalysis methods come to a final decision using linear aggregation rules of the form,

$$D(\boldsymbol{x}) = \begin{cases} \text{cover} & \text{for} \quad \sum_0^{n-1} w_i \cdot d(x_i) > \tau \\ \text{stego} & \text{else,} \end{cases} \tag{37}$$

where $d : \mathcal{C} \to \{0, 1\}$ is the local decision function, $\boldsymbol{w}$ is a vector of weights, and τ a decision threshold to adjust the tradeoff between false negatives and false positives. Unfortunately, even a simple game defining Eve's action space as $(\boldsymbol{w}, \tau)$ deprives a straightforward analysis. Our theorems do not generalize to this case because they are based on Eve's advantage, which corresponds to the differences in expected values between cover and stego. This translates to the difference in means of the distribution of the weighted sum of local decisions, but Eve's error rates depend also on higher moments of this distribution. More precisely, they depend on the quantile functions of the distribution for covers and stego objects, respectively. If this problem is solved, we can return to the standard model of steganographic communication as the amount of permissible side information does not need to be artificially constrained.

4.3 Alternative Interpretation

Although we motivated this game with optimal content-adaptive steganography, the underlying information hiding game is general enough to lend itself to alternative interpretations in the broader field of information security. One application is keeping secrets in an organization. Suppose Alice leads a big organization which contains a secret binary state that is extremely sensitive. Think of innovative companies ("Will the new device be a phone or not?"), central banks ("Will interest rates change or not?"), or governments ("Will they really respond to a cyber-attack with conventional warfare?"). In all these cases, Eve, an outside observer, should not be able to distinguish both possible states. However, Alice needs a team of size k to work on projects where knowledge of the state is essential. Her n staff members differ (function f) in their ability to decouple their observable behavior from their knowledge of the state, and Eve has resources to 'probe' (observe, eavesdrop, bribe, ...) exactly one staff member. Disregarding other constraints on team building, the solution to our game tells Alice how to compose her team with minimal risk of leaking the secret state.

4.4 Relation to Adversarial Classification

Our work can be seen as an example for adversarial classification, a term to the best of our knowledge coined by Dalvi et al. [3], who challenge the common assumption in data mining that the data generating process is independent of the data miner's activity. Instead, the authors study instances where the data is actively manipulated by an adversary seeking to increase the error rate in classification problems.

Like our steganographer, an adversary in their model actively manipulates data generated by nature, and a binary classifier tries to distinguish altered from unaltered objects, similar to our steganalyst. Their payoff structure is more complicated, including costs for altering and measuring features, respectively. These costs are offset by utility from successful, respectively erroneous, classifications.

The original work on adversarial classification is presented in a spam detection scenario, where spammers try to "wear out" a Bayes classifier. Nevertheless, the framework is presented in general terms, also suggesting other domains and tasks; but interestingly not steganography.

With theoretical underpinnings in feature-based machine learning theory, adversarial classification may also have the potential to deliver new insights for learning-based universal steganalysis as well as steganographic algorithms leveraging distance metrics in high-dimensional feature space [10].

5 Concluding Remarks

In this paper, we have formulated the problem of hiding multiple bits in a heterogeneous cover sequence as a stochastic two-player zero-sum game between steganographer and steganalyst. The steganographer chooses the embedding positions and the steganalyst applies a local decision rule involving side information to exactly one position. Theorem 3 states the main result: the game has a unique mixed strategy Nash equilibrium. All relevant properties to implement the equilibrium strategies can be efficiently computed from the function describing the heterogeneity in the predictability of cover bits. Corollary 1 stipulates that the steganalyst's equilibrium strategy does not depend on the number of embedded bits. This is a handy property for the construction of detectors, where no knowledge of the hidden message length must be assumed. Corollary 2 states that if the detector follows the equilibrium strategy, its success rate increases linearly with the number of embedded bits. This deviates from the square root law of steganographic capacity, which predicts asymptotically quadratic advantage even for homogeneous covers [9]. The reason for this difference is that our detector is constrained to a locally optimal decision rule.

While local decisions seem to be a good approximation of what is implemented in current steganalysis methods, other simplifications in our model may limit its validity. Most importantly, the constraint on access to side information of one symbol per cover is restrictive. Also giving the steganalyst perfect knowledge of the local predictability appears somewhat unrealistic. Practical content-adaptive embedding functions use different approximations of predictability and, depending on embedding operation and message length, the steganalyst can often recover this proxy pretty well. To account for the remaining uncertainty, future extensions of our game could equip the steganalyst with a noisy version of the true predictability profile. This can be done either by adding an independent random error term or, more realistically, by conditioning the error on the choice of embedding positions. Finally, the impact of the assumption of independent cover symbols needs to be evaluated. It remains to be seen if the useful properties established above can be maintained in a generalized game.

Acknowledgements. This research has received funding from the German national science foundation (DFG) under grant "Sichere adaptive Steganographie". Part of this work has been carried out during Benjamin Johnson's research visit to Germany, which was additionally supported by the University of Münster and the Münster School of Business and Economics, and by the National Science Foundation under ITR award CCF-0424422 (TRUST).

References

1. Böhme, R.: Advanced Statistical Steganalysis. Springer (2010)
2. Cachin, C.: An Information-Theoretic Model for Steganography. In: Aucsmith, D. (ed.) IH 1998. LNCS, vol. 1525, pp. 306–318. Springer, Heidelberg (1998)

3. Dalvi, N., Domingos, P., Mausam, Sanghai, S., Verma, D.: Adversarial classification. In: Proceedings of the 10th ACM SIGKDD International Conference on Knowledge Discovery and Data Mining, pp. 99–108. ACM, New York (2004)
4. Filler, T., Judas, J., Fridrich, J.J.: Minimizing embedding impact in steganography using trellis-coded quantization. In: Memon, N.D., Dittmann, J., Alattar, A.M., Delp, E.J. (eds.) Media Forensics and Security. SPIE Proceedings, vol. 7541, p. 754105. SPIE (2010)
5. Fridrich, J.: Minimizing the embedding impact in steganography. In: Workshop on Multimedia and Security, pp. 2–10. ACM (2006)
6. Hopper, N.J., Langford, J., von Ahn, L.: Provably Secure Steganography. In: Yung, M. (ed.) CRYPTO 2002. LNCS, vol. 2442, pp. 77–92. Springer, Heidelberg (2002)
7. Katzenbeisser, S., Petitcolas, F.A.P.: Defining security in steganographic systems. In: Delp, E.J., Wong, P.W. (eds.) Security, Steganography and Watermarking of Multimedia Contents IV, San Jose, CA, vol. 4675, pp. 50–56 (2002)
8. Ker, A.D.: The Square Root Law in Stegosystems with Imperfect Information. In: Böhme, R., Fong, P.W.L., Safavi-Naini, R. (eds.) IH 2010. LNCS, vol. 6387, pp. 145–160. Springer, Heidelberg (2010)
9. Ker, A.D., Pevný, T., Kodovský, J., Fridrich, J.: The square root law of steganographic capacity. In: MM&Sec 2008: Proceedings of the 10th ACM Workshop on Multimedia and Security, pp. 107–116. ACM, New York (2008)
10. Pevný, T., Filler, T., Bas, P.: Using High-Dimensional Image Models to Perform Highly Undetectable Steganography. In: Böhme, R., Fong, P.W.L., Safavi-Naini, R. (eds.) IH 2010. LNCS, vol. 6387, pp. 161–177. Springer, Heidelberg (2010)
11. Schöttle, P., Böhme, R.: A game-theoretic approach to content-adaptive steganography. In: Ghosal, D., Kirchner, M. (eds.) Information Hiding. LNCS, Springer (to appear, 2012)
12. Wang, Y., Moulin, P.: Perfectly secure steganography: Capacity, error exponents, and code constructions. IEEE Transactions on Information Theory 54(6), 2706–2722 (2008)