Adaptive Steganography and Steganalysis with Fixed-Size Embedding

Benjamin Johnson[ae], Pascal Schöttle[b], Aron Laszka[c],
Jens Grossklags[d], and Rainer Böhme[b]

[a]Cylab, Carnegie Mellon University, USA
[b]Department of Information Systems, University of Münster, Germany
[c]Institute for Software Integrated Systems, Vanderbilt University, USA
[d]College of Information Sciences and Technology, Pennsylvania State University, USA
[e]School of Information, University of California, Berkeley, USA

Abstract. We analyze a two-player zero-sum game between a steganographer, Alice, and a steganalyst, Eve. In this game, Alice wants to hide a secret message of length k in a binary sequence, and Eve wants to detect whether a secret message is present. The individual positions of all binary sequences are independently distributed, but have different levels of predictability. Using knowledge of this distribution, Alice randomizes over all possible size-k subsets of embedding positions. Eve uses an optimal (possibly randomized) decision rule that considers all positions, and incorporates knowledge of both the sequence distribution and Alice's embedding strategy.
Our model extends prior work by removing restrictions on Eve's detection power. We give defining formulas for each player's best response strategy and minimax strategy; and we present additional structural constraints on the game's equilibria. For the special case of length-two binary sequences, we compute explicit equilibria and provide numerical illustrations.

Keywords: Game Theory, Content-Adaptive Steganography, Security

1 Introduction

In steganography, the objective of a steganographer is to hide a secret message in a communication channel. The objective of her counterpart, the steganalyst, is to detect whether the channel contains a message [29]. Digital multimedia, such as JPEG images, are the most commonly studied communication channels in this context; but the theory can be applied more generally to any data stream having some irrelevant components and an inherent source of randomness [10].

In contrast to random uniform embedding, where the steganographer chooses her message-hiding positions along a pseudo-random path through the communication channel, content-adaptive steganography leverages the fact that different parts of a communication channel may have different levels of predictability [2, 4]. All content-adaptive embedding schemes have in common that they try to identify less predictable embedding positions. These schemes can be roughly divided into locally calculated criteria and distortion minimizing criteria. An example for the first category is the assumption

that areas with a high local variance are more suitable, e.g., [13]. The second category assumes that embedding positions introducing less distortion are preferable, e.g., [15]. The claimed purpose of all adaptivity criteria is to identify a (partial) ordering of all available embedding positions according to their suitability for embedding.

For example, digital images often have areas of homogeneous color where any slight modification would be noticed, whereas other areas are heterogeneous in color so that subtle changes to a few pixels would still appear natural. It follows that if a steganographer wants to modify image pixels to communicate a message, she should prefer to embed in these heterogeneous areas.

Our model abstracts this concept of content-adaptivity, by considering a communication channel as a random variable over binary sequences, where each position in the sequence has a different level of predictability. The predictability of each position is observable by both Alice, a content-adaptive steganographer, and Eve, a computationally-unbounded steganalyst; and we apply a game-theoretic analysis to determine each player's optimal strategy for embedding and detection, respectively.

We show that if Alice changes exactly k bits of a binary cover sequence, then Eve's best-response strategy can be expressed as a multilinear polynomial inequality of degree k in the sequence position variables. In particular, when $k = 1$, this polynomial inequality is a linear aggregation formula similar to what is typically used in practical steganalysis, e. g., [11]. Conversely, given any strategy by Eve to separate cover and stego objects, Alice has a best-response strategy that minimizes a relatively-simple summation over Eve's strategic choices. We give formulas for both players' minimax strategies, and explain why the straightforward linear programming solution for computing these strategies is not efficiently implementable for realistic problem sizes. We give structural constraints to the players' equilibrium strategies; and in the case where there are only two embedding positions, we classify all equilibria, resolving an open question from [31]. Furthermore, we bridge the two research areas of game-theoretic approaches and information-theoretic optimal steganalysis, and conjecture that the main results of earlier works still hold when the steganalyst is conservatively powerful.

The rest of the paper is organized as follows. In Section 2, we briefly review related work. In Section 3, we describe the details of our game-theoretic model. Section 4 contains our analysis of the general case; and in Section 5, we compute and illustrate the game's equilibria for the special case of sequences of length two. We conclude the paper in Section 6.

2 Related work

Game theory is a mathematical framework to investigate competition between strategic players with contrary goals [34]. Game theory gains more and more importance in practically all areas concerned with security ranging from abstract models of security investment decisions [14, 17] to diverse applied scenarios such as the scheduling of patrols at airports [30], the modeling of Phishing strategies [6], network defense [23], and team building in the face of a possible insider threat [22].

The application of game theory has also found consideration in the various subdisciplines of information hiding including research on covert channels [16], anonymity

[1], watermarking [24] and, of course, steganography.[1] Similarly, game-theoretic approaches can be found in the area of multimedia forensics [3, 33].

In content-adaptive steganography [4], where Alice chooses the positions into which she embeds a message and Eve tries to anticipate these positions to better detect the embedding, the situation is naturally modeled using game theory.

Practical content-adaptive steganography schemes, on the other hand, have typically relied primarily on the notion of unpredictability to enhance the security of embedded messages. In fact, the early content-adaptive schemes not only preferred less predictable areas of images, but restricted all embedding changes to the least predictable areas, e. g., [9]. Prior works examining adaptive embedding have dubbed this strategy *naïve adaptive embedding*, and have shown it to be a non-optimal strategy in progressively more general settings [5, 18, 31]. It was shown in [5] that the steganalyst can leverage her knowledge about the specific adaptive embedding algorithm from [9] to detect it with better accuracy than even random uniform embedding. In [31] it was shown for the first time that, if the steganalyst is strategic, it is never optimal for the steganographer to deterministically embed in the least predictable positions. The game-theoretic analysis in [31] was restricted to a model with two embedding positions, where Eve could only look in one position. A subsequent extension of that model [18] allowed the steganographer to change multiple bits in an arbitrary-sized cover sequence, but maintained limiting restrictions on the power of the steganalyst, by requiring her to make decisions on the basis of only one position. Another extension generalizes the model by introducing a non-uniform cost of steganalysis and models the problem as a quasi-zero-sum game [21].

Another extension of this research stream expanded the power of Eve but required Alice to embed independently in each position [32]. Other authors have studied steganography using game-theoretical models. In 1998, Ettinger [8] proposed a two-player, zero-sum game between a steganographer and an active steganalyst whose purpose it is to interrupt the steganographic communication; Ker [20] uses game theory to find strategies in the special case of batch steganography, where the payload can be spread over many cover objects. The steganalyst anticipates this and tries to detect the existence of any secret message (so-called pooled steganalysis); and Orsdemir et al. [26] frame the competition between steganographer and steganalyst with the help of *set theory*. The steganographer has the possibility to use either a naïve or a sophisticated strategy, where in the sophisticated strategy she incorporates statistical indistinguishability constraints. By this they devise a meta-game. The only other game-theoretical approach that is also concerned with content-adaptive embedding, the most common approach in modern steganography, e. g., [12, 28], is [7]. Here, the authors examine the embedding operation of LSB matching with a content-adaptive embedding strategy and a multivariate Gaussian cover model.

This work directly extends [19], which first introduced the game theoretic model studied in this paper. Compared to that work, we have added several new results constraining the game's equilibrium strategies. First, we give formal constraints determining when the game admits or does not admit trivial equilibria. We use these constraints to show that under the non-trivial conditions, Alice can affect her payoff by changing

[1] See [27] for an introduction to the area of information hiding.

her embedding strategy at key positions. Finally, we use these structural results to prove that under relatively general conditions, it is not optimal against an adaptive classifier to naïvely embed in the least biased positions. As an additional contribution, we give a constructive proof that our simplified representation of Eve's mixed strategy is a surjective reduction.

3 Game-Theoretic Model

To describe our game-theoretic model, we specify the set of players, the set of states that the world can be in, the set of choices available to the players, and the set of consequences as a result of these choices. Because our game is a randomized extension of a deterministic game, we first present the structure of the deterministic game, and follow up afterwards with details of the randomization.

3.1 Players

The players are Alice, a steganographer, and Eve, a steganalyst. Alice wants to send a message through a communication channel, and Eve wants to detect whether the channel contains a message. At times, we find it convenient to also mention Nature, the force causing random variables to take realizations, and Bob, the message recipient; although Nature and Bob are not players in a game-theoretic sense because they are not strategic.

3.2 Events

Our event space Ω is the set $\{0,1\}^N \times \{C,S\}$. An event consists of two parts: a binary sequence $x \in \{0,1\}^N$ and a steganographic state $y \in \{C,S\}$, where C stands for *cover* and S for *stego*. The binary sequence represents what Eve observes on the communication channel. The steganographic state tells whether or not a message is embedded in the sequence. In the randomized game, neither of these two states is known by the players until after they make their choices. To define payoffs for the finite game, we simply assume that some event has been chosen by Nature so that the world is in some fixed state (x,y).

Figure 1 illustrates an event with player interaction as a block diagram. Following the diagram, Alice embeds a secret message of length k into the binary sequence x; Nature determines whether the original cover or the modified stego object appears on the communication channel; Eve observes the sequence appearing on the channel and makes a decision as to whether or not it contains a message; and (not relevant to our analysis but useful for narrative closure) Bob extracts the message, if it happened to be there.

3.3 Choices

Alice's (pure strategy) choice is to select a size-k subset I of $\{0,\dots,N-1\}$, which represents the positions into which she embeds her encoded message, by flipping the value of the given sequence at each of the positions in I.

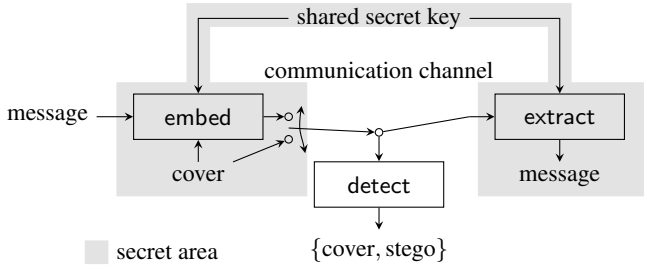


Fig. 1. Block diagram of a steganographic communication system

Eve's (pure strategy) choice is to select a subset E_S of $\{0,1\}^N$, which represents the set of sequences that she classifies as stego objects (i.e., sequences containing a secret message). Objects in $E_C := \{0,1\}^N \setminus E_S$ are classified as cover objects (i.e., sequences not containing a secret message).

3.4 Consequences

Suppose that Alice chooses a pure strategy $I \subseteq \{0,\ldots,N-1\}$, Eve chooses a pure strategy $E_S \subseteq \{0,1\}^N$, and Nature chooses a binary sequence x and a steganographic state y. Then, Eve wins 1 if she classifies x correctly (i.e., either she says stego and Nature chose stego, or she says cover and Nature chose cover), and she loses 1 if her classification is wrong. The game is zero-sum so that Alice's payoff is the negative of Eve's payoff. Table 1 formalizes the possible outcomes as a zero-sum payoff matrix.[2]

Table 1. Payoffs for (Eve, Alice)

	steganographic state	
Eve's decision for x	C	S
$x \in E_C$	$(\ 1,-1)$	$(-1,\ 1)$
$x \in E_S$	$(-1,\ 1)$	$(\ 1,-1)$

3.5 Randomization

In the full randomized game, we have distributions on binary sequences and steganographic states. We also have randomization in the players' strategies. To describe the nature of the randomness, we start by defining two random variables on our event space

[2] The payoff matrix and the zero sum property might be different if false positives and false negatives result in different profits, respectively losses.

Ω. Let $X : \Omega \to \{0,1\}^N$ be the random variable which takes an event to its binary sequence and let $Y : \Omega \to \{C, S\}$ be the random variable which takes an event to its steganographic state. We proceed through the rest of this section by first describing the structure of the distribution on Ω; next describing the two players' mixed strategies; and finally, by giving the players' payoffs as a consequence of their mixed strategies.

Steganographic States The event $Y = S$ happens when Nature chooses the steganographic state to be stego; and this event occurs with probability p_S. We also define $\Pr_\Omega[Y = C] := p_C = 1 - p_S$. From Eve's perspective, p_S is the prior probability that she observes a stego sequence on the communication channel. A common convention in steganography (following a similar convention in cryptography) is to equate the prior probabilities p_C and p_S of the two steganographic states, so that Eve observes a stego sequence with exactly 50% probability. Our results describing equilibria for this model carry through with arbitrary prior probabilities; so we retain the notations p_S and p_C in several subsequent formulas. Note however, that with highly unequal priors, the game may trivialize because the prior probabilities can dominate other incentives. For this reason, we do require equal priors for some structural theorems; and we also use equal priors in our numerical illustrations.

Binary Sequences The distribution on binary sequences depends on the value of the steganographic state. If $Y = C$, then the steganographic state is cover, and X is distributed according to a *cover distribution* $\mathcal{C}$; if $Y = S$, then the steganographic state is stego, and X is distributed according to a *stego distribution* $\mathcal{S}$.

With this notation in hand, we may define, for any event $(X = x, Y = y)$:

$$\begin{aligned} \Pr_\Omega[(x,y)] &= \Pr_\Omega[Y = y] \cdot \Pr_\Omega[X = x | Y = y] \\ &= \begin{cases} p_C \cdot \Pr_{\mathcal{C}}[X = x] & \text{if } y = C \\ p_S \cdot \Pr_{\mathcal{S}}[X = x] & \text{if } y = S\,. \end{cases} \end{aligned} \tag{1}$$

We will define the distributions $\mathcal{C}$ and $\mathcal{S}$ after describing the players' mixed strategies.

Players' Mixed Strategies We next describe the mixed strategy choices for Alice and Eve. Recall that a mixed strategy is a probability distribution over pure strategies.

In a mixed strategy, Alice can probabilistically embed into any given subset of positions, by choosing a probability distribution over size-k subsets of $\{0, \ldots, N-1\}$. To describe a mixed strategy, for each $I \subseteq \{0, \ldots, N-1\}$, we let a_I denote the probability that Alice embeds into each of the positions in I.

A mixed strategy for Eve is a probability distribution over subsets of $\{0,1\}^N$. Suppose that Eve's mixed strategy assigns probability e_S to each subset $S \subseteq \{0,1\}^N$. Overloading notation slightly, we define $e : \{0,1\}^N \to [0,1]$ via

$$e(x) = \sum_{S \subseteq \{0,1\}^N : x \in S} e_S \quad . \tag{2}$$

Each $e(x)$ gives the total probability for the binary sequence x that Eve classifies the sequence x as stego. Note that this "projected" representation of Eve's mixed strategy given in Equation (2) requires specifying 2^N real numbers, whereas the canonical representation of her mixed strategy using the notation e_S would require specifying 2^{2^N} real numbers. For this reason, we prefer to use the projection representation. Fortunately, the projected representation contains enough information to determine both players' payoffs, because it determines the classifier's success rates. In the reverse direction, we may also construct a true mixed strategy from a reduced representation, as evidenced by the subsequent lemma.

Reduced Representation of Eve's Mixed Strategy The following lemma shows that the mapping from the canonical representation of Eve's mixed strategy to the projected representation is surjective, so we may express results using the simpler representation without loss of generality.

Lemma 1. *For every function* $e : \{0,1\}^N \mapsto [0,1]$*, there exists a distribution* e_S*,* $S \subseteq \{0,1\}^N$*, satisfying Equation* (2)*.*

Proof. We prove the above lemma using a constructive proof. More specifically, we provide an algorithm that can compute an appropriate distribution e_S, $S \subseteq \{0,1\}^N$, from an arbitrary function $e : \{0,1\}^N \mapsto [0,1]$. First, order the sequences by their $e(x)$ values in a non-increasing order, and denote them $x^1, x^2, \ldots, x^{2^N}$ (i.e., without loss of generality, assume $e(x^1) \geq e(x^2) \geq \ldots \geq e(x^{2^N})$). Second, assign probabilities to subsets of sequences as follows. Let the first subset of sequences be $S^0 = \{\}$, and let its probability be $e_{S^0} = 1 - e(x^1)$. Next, let the second subset be $S^1 = \{x^1\}$, and let its probability be $e_{S^1} = e(x^1) - e(x^2)$. Then, let the third subset be $S^2 = \{x^1, x^2\}$ and its probability be $e_{S^2} = e(x^2) - e(x^3)$. Similarly, let the $(k+1)$th subset be $S^k = \{x^1, x^2, \ldots, x^k\}$, and let its probability be $e_{S^k} = e(x^k) - e(x^{k+1})$. Finally, let the last subset be $S^{2^N} = \{x^1, x^2, ..., x^{2^N}\}$, and let its probability be $e_{S^{2^N}} = e(x^{2^N})$.

We have to show that the output of the algorithm 1) is a distribution (i.e., the probabilities sum up to one) and 2) satisfies Equation (2). First, the sum of the resulting probabilities is

$$
\begin{aligned}
& e_{S^0} + e_{S^1} + e_{S^2} + \ \ldots \ + e_{S^{2^N}} && (3)\\
=& 1 - e(x^1) + e(x^1) - e(x^2) + e(x^2) - e(x^3) + \ \ldots \ + e(x^{2^N}) && (4)\\
=& 1\ . && (5)
\end{aligned}
$$

Second, for an arbitrary sequence x^k, we have

$$
\begin{aligned}
\sum_{S \subseteq \{0,1\}^N \,:\, x^k \in S} e_S =& \sum_{l=k}^{2^N} e_{S^l} && (6)\\
=& e(x^k) - e(x^{k+1}) + e(x^{k+1}) - e(x^{k+2}) + \ldots + e(x^{2^N}) && (7)\\
=& e(x^k)\ . && (8)
\end{aligned}
$$

Therefore, we have that the resulting distribution satisfies Equation (2), which concludes our proof. □

Note that the resulting distribution is relatively simple, since it assigns a non-zero probability to at most $N^2 + 1$ subsets only (and even less than that if some sequences have zero $e(x)$ values). It is easy to see that we cannot do any better than this generally, in the sense that there exists an infinite number of e functions, for which no distribution with a smaller support can exist.

Cover Distribution In the cover distribution $\mathcal{C}$, the coordinates of X are independently distributed so that

$$\Pr_{\mathcal{C}}[X = x] = \prod_{i=0}^{N-1} \Pr_{\mathcal{C}}[X_i = x_i]. \tag{9}$$

The bits are not identically distributed however. For each i we have

$$\Pr_{\mathcal{C}}[X_i = 1] = f_i, \tag{10}$$

where $\langle f_i \rangle_{i=0}^{N-1}$ is a monotonically-increasing sequence from $\left(\frac{1}{2}, 1\right)$. Note that this assumption is without loss of generality because, in applying the abstraction of a communication channel into sequences, we can always flip 0 s and 1s to make 1s more likely; and we can re-order the positions from least to most predictable.

For notational convenience, we define

$$\tilde{f}_i = 2f_i - 1. \tag{11}$$

We construe $\tilde{f}_i$ as a measure of the bias of the predictability at position i. If the bias at some position is close to zero, then the value of that position is not very predictable, while if the bias is close to 1, the value of the position is very predictable.

Putting it all together, the cover distribution is defined by

$$\begin{aligned} \Pr_{\mathcal{C}}[X = x] &= \prod_{x_i=1} f_i \cdot \prod_{x_i=0} (1 - f_i) \\ &= \prod_{i=0}^{N-1} \left(1 - f_i + x_i \tilde{f}_i\right). \end{aligned} \tag{12}$$

Stego Distribution The stego distribution $\mathcal{S}$ depends on Alice's choice of an embedding strategy. Let $I \subseteq \{0, \ldots, N-1\}$, and for each $x \in \{0,1\}^N$ let x_I denote the binary sequence obtained from x by flipping the bits at all the positions in I. The stego distribution is obtained from the cover distribution by adjusting the likelihood that each x occurs, assuming that for each I, with probability a_I Alice flips the bits of x in all the positions in I.

More formally, suppose that Alice embeds into each subset $I \subseteq \{0, \ldots, N-1\}$ with probability a_I. We then have

$$\begin{aligned}
\Pr_{\mathcal{S}}[X = x] &= \sum_I a_I \cdot \Pr_{\mathcal{C}}[X = x_I] \\
&= \sum_I a_I \cdot \prod_{i \notin I} \Pr_{\mathcal{C}}[X_i = x_i] \cdot \prod_{i \in I} \Pr_{\mathcal{C}}[X_i = 1 - x_i] \\
&= \sum_I a_I \cdot \prod_{i \notin I} \left(1 - f_i + x_i \tilde{f}_i\right) \cdot \prod_{i \in I} \left(f_i - x_i \tilde{f}_i\right). \qquad (13)
\end{aligned}$$

Player Payoffs In the full game, the expected payoff for Eve can be written as:

$$\begin{aligned}
u(Eve) = \; & \Pr_{\Omega}[X \in E_S \text{ and } Y = S] && \text{(true positive)} \\
& + \Pr_{\Omega}[X \in E_C \text{ and } Y = C] && \text{(true negative)} \\
& - \Pr_{\Omega}[X \in E_S \text{ and } Y = C] && \text{(false positive)} \\
& - \Pr_{\Omega}[X \in E_C \text{ and } Y = S] && \text{(false negative)} \qquad (14)
\end{aligned}$$

and this can be further computed as

$$\begin{aligned}
u(Eve) &= p_S \Pr_{\mathcal{S}}[X \in E_S] + p_C \Pr_{\mathcal{C}}[X \in E_C] - p_C \Pr_{\mathcal{C}}[X \in E_S] - p_S \Pr_{\mathcal{S}}[X \in E_C] \\
&= \sum_{x \in \{0,1\}^N} \Big[\; e(x) p_S \Pr_{\mathcal{S}(a)}[X = x] \\
&\qquad + (1 - e(x)) p_C \Pr_{\mathcal{C}}[X = x] \\
&\qquad - (1 - e(x)) p_S \Pr_{\mathcal{S}(a)}[X = x] \\
&\qquad - e(x) p_C \Pr_{\mathcal{C}}[X = x] \Big] \\
&= \sum_{x \in \{0,1\}^N} \big(2e(x) - 1\big) \big(p_S \Pr_{\mathcal{S}(a)}[X = x] - p_C \Pr_{\mathcal{C}}[X = x]\big). \qquad (15)
\end{aligned}$$

The terms $\Pr_{\mathcal{C}}[X = x]$ and $\Pr_{\mathcal{S}(a)}[X = x]$ are defined in Equations (12) and (13), respectively. Note that we write $\mathcal{S} = \mathcal{S}(a)$ to clarify that the distribution $\mathcal{S}$ depends on Alice's mixed strategy a.

In summary, Eve's payoff is the probability that her classifier is correct minus the probability that it is incorrect; and the game is zero-sum so that Alice's payoff is exactly the negative of Eve's payoff.

4 Model Analysis

In this section, we present our analytical results. We begin by describing best response strategies for each player. Next, we describe in formal notation the minimax strategies for each player. Finally, we present several results which give structural constraints on the game's Nash equilibria.

4.1 Best Responses

To compute best responses for Alice and Eve, we assume that the other player is playing a fixed strategy, and determine the strategy for Alice (or Eve) which minimizes (or maximizes) the payoff in Equation (15) as appropriate.

Alice's Best Response Given a fixed strategy e for Eve, Alice's goal is to minimize the payoff in Equation (15). However, since she has no control over the cover distribution $\mathcal{C}$, this goal can be simplified to that of minimizing

$$\begin{aligned}
&\sum_{x\in\{0,1\}^N}(2e(x)-1)\cdot p_S\mathrm{Pr}_{S(a)}[X=x]\\
=&p_S\sum_{x\in\{0,1\}^N}(2e(x)-1))\cdot\sum_{I\subseteq\{0,\dots,N-1\}}a_I\mathrm{Pr}_{\mathcal{C}}[X=x_I]\\
=&p_S\sum_{I\subseteq\{0,\dots,N-1\}}a_I\sum_{x\in\{0,1\}^N}(2e(x)-1))\cdot\mathrm{Pr}_{\mathcal{C}}[X=x_I]\,.
\end{aligned}$$

This formula is linear in Alice's choice variables, so she can minimize its value by putting all her probability on the sum's least element. A best response for Alice is thus to play a pure strategy I that minimizes

$$\sum_{x\in\{0,1\}^N}(2e(x)-1))\cdot\mathrm{Pr}_{\mathcal{C}}[X=x_I]. \tag{16}$$

Of course, several different I might simultaneously minimize this sum. In this case, Alice's best response strategy space may also include a mixed strategy that distributes her embedding probabilities randomly among such I.

Eve's Best Response Given a fixed strategy for Alice, Eve's goal is to maximize her payoff as given in Equation (15). So, for each x, she should choose $e(x)$ to maximize the term of the sum corresponding to x. Specifically, if $p_S\mathrm{Pr}_{S(a)}[X=x]-p_C\mathrm{Pr}_C[X=x]>0$, then the best choice is $e(x)=1$; and if the strict inequality is reversed, then the best choice is $e(x)=0$. If the inequality is an equality, then Eve may choose any value for $e(x)\in[0,1]$ and still be playing a best response.

Formally, her optimal decision rule is

$$e(x)=\begin{cases}1 & \text{if } \frac{\mathrm{Pr}_{\Omega}[Y=S|X=x]}{\mathrm{Pr}_{\Omega}[Y=C|X=x]}>1\,,\\ 0 & \text{if } \frac{\mathrm{Pr}_{\Omega}[Y=S|X=x]}{\mathrm{Pr}_{\Omega}[Y=C|X=x]}<1\,,\\ \text{any } p\in[0,1] & \text{if } \frac{\mathrm{Pr}_{\Omega}[Y=S|X=x]}{\mathrm{Pr}_{\Omega}[Y=C|X=x]}=1\,.\end{cases} \tag{17}$$

For a fixed sequence x, the condition for classifying x as stego can be rewritten as:

$$\begin{aligned}
1 <& \frac{\mathrm{Pr}_{\Omega}[Y=S|X=x]}{\mathrm{Pr}_{\Omega}[Y=C|X=x]} \\
=& \frac{\mathrm{Pr}_{\Omega}[X=x]}{\mathrm{Pr}_{\Omega}[X=x]} \cdot \frac{\mathrm{Pr}_{\Omega}[Y=S|X=x]}{\mathrm{Pr}_{\Omega}[Y=C|X=x]} \\
=& \frac{\mathrm{Pr}_{\Omega}[Y=S]}{\mathrm{Pr}_{\Omega}[Y=C]} \cdot \frac{\mathrm{Pr}_{\Omega}[X=x|Y=S]}{\mathrm{Pr}_{\Omega}[X=x|Y=C]} \\
=& \frac{p_S}{p_C} \frac{\mathrm{Pr}_{\mathcal{S}}[X=x]}{\mathrm{Pr}_{\mathcal{C}}[X=x]} \\
=& \frac{p_S}{p_C} \frac{\sum_I a_I \cdot \prod_{i \notin I} \left(1 - f_i + x_i \tilde{f}_i\right) \cdot \prod_{i \in I} \left(f_i - x_i \tilde{f}_i\right)}{\prod_{i=0}^{N-1} \left(1 - f_i + x_i \tilde{f}_i\right)} \\
=& \frac{p_S}{p_C} \sum_I a_I \prod_{i \in I} \left(\frac{f_i - x_i \tilde{f}_i}{1 - f_i + x_i \tilde{f}_i} \right) \\
=& \frac{p_S}{p_C} \sum_I a_I \prod_{i \in I} \left(\frac{f_i}{1 - f_i} - x_i \frac{\tilde{f}_i}{f_i(1 - f_i)} \right).
\end{aligned} \tag{18}$$

Note that Eve's decision rule is written as a multilinear polynomial inequality of degree at most k in the binary sequence x, and that the number of terms in the formula is $\binom{N}{k}$. When k is a constant relative to N (as it typically is in practical applications), then $\binom{N}{k}$ is polynomial in N, and Eve's optimal decision rule can be applied for each binary sequence in time that is polynomial in the length of the sequence.

4.2 Minimax Strategies

A minimax strategy in a two-player game is a mixed strategy of one player that maximizes her payoff assuming that the other player is going to respond with an optimal pure strategy [34].

Eve's minimax strategy is given by

$$\underset{e}{\mathrm{argmax}} \left(\min_I \Big(\sum_{x \in \{0,1\}^N} (2e(x) - 1)(p_S \mathrm{Pr}_{\mathcal{C}}[X = x_I] - p_C \mathrm{Pr}_{\mathcal{C}}[X = x]) \Big) \right); \tag{19}$$

while Alice's minimax strategy is given by

$$\begin{aligned}
\underset{a}{\mathrm{argmin}} \Bigg(\max_{E_S} \Big(& \sum_{x \in E_S} \big(p_S \mathrm{Pr}_{\mathcal{S}(a)}[X = x] - p_C \mathrm{Pr}_{\mathcal{C}}[X = x]\big) \\
& + \sum_{x \in E_C} \big(p_C \mathrm{Pr}_{\mathcal{C}}[X = x] - p_S \mathrm{Pr}_{\mathcal{S}(a)}[X = x]\big) \Big) \Bigg).
\end{aligned} \tag{20}$$

Each minimax strategy can be determined (recursively) as the solution to a linear program involving the payoff matrix for Alice's and Eve's pure strategies. Unfortunately, Eve's pure strategy space has size 2^{2^N} so it is computationally intractable to find the minimax strategies using this method even for $N = 5$.

4.3 Nash Equilibria

In this subsection, we present structural constraints for Nash equilibria [25]. We begin with a lemma giving natural conditions under which Eve's classifier must respect the canonical partial ordering on binary sequences. It shows that the classifier must essentially divide the set of all binary sequences into low and high, with high sequences classified as cover and low sequences classified as stego. Then, we give specific constraints on the distribution priors relative to the position biases that determine whether or not the game admits trivial equilibria – in which Eve's classifier is constant for all binary sequences. If either the priors are too imbalanced, or the position biases are too small, then the game will admit such trivial equilibria. In more prototypical parameter regions, however, the game does not admit trivial equilibria. Next, we show that when Eve's classifier is non-trivial, Alice can affect the outcome of Eve's detector, and hence her own payoff by changing her embedding probability for one position in the sequence. Finally, we show that in the non-trivial equilibrium setting, it is not optimal for Alice to embed naïvely in only the least biased positions.

Sequence Ordering in Eve's Equilibrium Strategy

Lemma 2. *Define a partial ordering on* $\{0,1\}^N$ *by* $x < z$ *iff* $x_i \leq z_i$ *for* $i = 0, \ldots, N-1$ *and* $x_i < z_i$ *for at least one* i*. Then whenever Alice's embedding strategy satisfies the constraint* $\frac{p_S}{p_C} \sum_I a_I \prod_{i \in I} \left(\frac{f_i}{1-f_i} - x_i \frac{\tilde{f}_i}{f_i(1-f_i)} \right) \neq 1$ *for the sequence* x*, the following condition holds:*

- *If Eve classifies* x *as stego and* $z < x$*, then Eve classifies* z *as stego too.*
- *If Eve classifies* x *as cover and* $x < z$*, then Eve classifies* z *as cover too.*

Proof. Suppose Eve classifies x as stego. Then from the conditions on Eve's best response (Equations (17) and (18)), we have that $\frac{p_S}{p_C} \sum_I a_I \prod_{i \in I} \left(\frac{f_i}{1-f_i} - x_i \frac{\tilde{f}_i}{f_i(1-f_i)} \right) \geq 1$; and by the hypothesis of the lemma, the inequality is strict. Suppose $z < x$. Then the value of $\frac{p_S}{p_C} \sum_I a_I \prod_{i \in I} \left(\frac{f_i}{1-f_i} - z_i \frac{\tilde{f}_i}{f_i(1-f_i)} \right)$ is at least the value of the same expression with x replacing z. So this value is also greater than 1, and so Eve also classifies z as stego. The proof of the reverse direction is analogous. □

This lemma implies that in any Nash equilibrium, the set of all binary sequences can be divided into three disjoint sets, low sequences which Eve's likelihood test proscribes a clear value of stego, high sequences which Eve's test proscribes as clearly cover, and a small set of mid-level boundary sequences on which Eve's behavior is not obviously constrained. Furthermore, changing 0s to 1s in a clearly-cover sequence keeps it cover, and changing 1s to 0s in a clearly-stego sequence keeps it stego.

Constraints on Parameters to Guarantee Nontrivial Equilibria Next we give a key parameter constraint on the prior probabilities of cover and stego that determines the complexity of equilibrium strategies for both players. Essentially if the priors are too far apart relative to the sequence position biases, then the game admits trivial equilibria – in which Eve's classifier is constant; while if they are sufficiently close together then it does not.

Lemma 3. *Suppose that*

$$\prod_{i=0}^{k-1} \frac{1-f_i}{f_i} < \frac{p_C}{p_S} < \prod_{i=0}^{k-1} \frac{f_i}{1-f_i}. \tag{21}$$

Then in any equilibrium, Eve classifies 0^N *as stego and* 1^N *as cover.*

Moreover, if either inequality is reversed strictly, then there exists an equilibrium in which Alice plays a pure strategy of the form $I = \{0, \ldots, k-1\}$ *(naïve adaptive embedding), and Eve's classifier is constant.*

Proof. Since $\langle f_i \rangle_{i=0}^{N-1}$ is monotonically increasing, we have that for any size-k subset $I \subseteq \{0, \ldots N-1\}$,

$$\prod_{i\in I} \frac{1-f_i}{f_i} \leq \prod_{i=0}^{k-1} \frac{1-f_i}{f_i} \text{ and } \prod_{i=0}^{k-1} \frac{f_i}{1-f_i} \leq \prod_{i\in I} \frac{f_i}{1-f_i}. \tag{22}$$

Consequently, for any mixed strategy $\langle a_I \rangle_{I\subseteq\{0,\ldots,N-1\}}$ of Alice, we have

$$\sum_I a_I \prod_{i\in I} \frac{1-f_i}{f_i} \leq \prod_{i=0}^{k-1} \frac{1-f_i}{f_i} \text{ and } \prod_{i=0}^{k-1} \frac{f_i}{1-f_i} \leq \sum_I a_I \prod_{i\in I} \frac{f_i}{1-f_i}. \tag{23}$$

The above, together with Equation (21) now implies that for any $\langle a_I \rangle_{I\subseteq\{0,\ldots,N-1\}}$,

$$\frac{p_S}{p_C} \sum_I a_I \prod_{i\in I} \frac{1-f_i}{f_i} < 1 < \frac{p_S}{p_C} \sum_I a_I \prod_{i\in I} \frac{f_i}{1-f_i}.$$

Using Eve's decision rule from Equation (18), the left inequality above implies that Eve's best response strategy for the sequence 1^N is to classify it as cover. The right inequality implies that Eve's best response to the sequence 0^N is to classify it as stego.

If the first inequality is reversed strictly, then

$$\frac{p_S}{p_C} \prod_{i=0}^{k-1} \frac{1-f_i}{f_i} > 1.$$

So if Alice embeds in exactly the positions $0, \ldots, k-1$, then Eve's best response (see Equation (18)) will classify 1^N as stego; and since her decision inequality is strict, by Lemma 2, she will classify all sequences as stego. In this circumstance, the payoff for Alice is independent of her strategy. Thus both players are playing a best response, and the strategy configuration is an equilibrium.

Similarly, if the second inequality is reversed strictly, then

$$\frac{p_S}{p_C}\prod_{i=0}^{k-1}\frac{f_i}{1-f_i} < 1.$$

So if Alice embeds in exactly the positions $0,\ldots,k-1$, then Eve's best response will classify 0^N as cover; and again by Lemma 2 she will classify all sequences as cover. Again Alice has no incentive to change her strategy, and this configuration is an equilibrium. □

Impact of Alice's Strategy on a Nontrivial Classifier The next result shows explicitly that if Eve's classifier is non-trivial, then there is a sequence x and a position i that witnesses a change from stego to cover depending only on that position. We use this lemma as a tool for allowing Alice to change her payoff by adjusting her strategy in response to a fixed classifier.

Lemma 4. *Suppose that Eve classifies 0^N as stego and 1^N as cover. Then there exists at least one position i and a sequence x such that $x_i = 0$ and Eve classifies x as stego (with some positive probability), but when the value of x at position i is flipped to 1, then Eve classifies the modified sequence as cover.*

Proof. Starting with 0^N, flip the bit in each position sequentially from position 0 to $N-1$ until after N steps, the sequence becomes 1^N. Since Eve says stego at the beginning, and cover by the end, there must be a step at which she changes from (probably) stego to cover. The sequence x at this step, and position i at this step serve as witnesses to the lemma's claim. □

Exclusion of Naïve Adaptive Embedding Strategies Our last equilibrium result combines the previous lemmas to show that under relatively mild constraints on the game's parameters, there is no equilibrium in which Alice embeds in exactly the k least biased positions. This result compares well with a result from [19] which showed the same property for a steganography game in which Eve's observational power was more restricted.

The first constraint for the theorem says only that the priors for the stego and cover distributions are not too imbalanced, in comparison to the position biases. The second constraint says that the parameters do not naturally make Eve indifferent on sequence classification against pure strategies. This constraint is satisfied if, for example, the position biases are drawn randomly from a continuous distribution; and it is used only to avoid navigating the logic of pathological cases in which Eve's classifier acts arbitrarily when her likelihood test is inconclusive.

Theorem 1. *Suppose that $k < N$ and the following conditions hold:*

1.

$$\prod_{i=0}^{k-1}\frac{1-f_i}{f_i} < \frac{p_C}{p_S} < \prod_{i=0}^{k-1}\frac{f_i}{1-f_i}, \tag{24}$$

2.

$$\forall x \in \{0,1\}^N, \frac{p_S}{p_C} \prod_{i=0}^{k-1} \left(x_i \frac{1-f_i}{f_i} + (1-x_i)\frac{f_i}{1-f_i} \right) \neq 1. \tag{25}$$

Then there does not exist an equilibrium in which Alice embeds in exactly the k least biased positions.

Proof. Suppose by way of contradiction, that Alice plays a pure strategy by embedding in positions $0, \ldots, k-1$, and that the strategy configuration is an equilibrium. Since Eve is playing a best response to Alice, she classifies an input x as stego whenever

$$\frac{p_S}{p_C} \prod_{i=0}^{k-1} \left(x_i \frac{1-f_i}{f_i} + (1-x_i)\frac{f_i}{1-f_i} \right) > 1;$$

and classifies x as cover when the inequality (for x) is reversed. The inequality is never an equality by assumption, so that Eve's decision is necessarily determined by the binary values of x at positions $0, \ldots, k-1$.

Note that since $k < N$, Alice is not embedding in position $N-1$; and Eve's classifier does not depend on position $N-1$.

By Lemma 3, Eve classifies 0^N as stego and 1^N as cover; so by Lemma 4, there is a position $i \in \{0, \ldots, k-1\}$ and sequence $x \in \{0,1\}^N$ such that $x_i = 0$ and Eve classifies x as stego, but when x_i is flipped to 1, Eve classifies the resulting sequence as cover.

Let J be the set $I \setminus \{i\} \cup \{N-1\}$; and suppose Alice changes her pure strategy from I to J. Let $Pr_{S(I)}[X = x]$ denote the probability of the sequence x appearing on the communication channel in the stego distribution under the original strategy I, and let $Pr_{S(J)}[X = x]$ denote the same probability under Alice's new strategy J. Our goal is now to show that

$$\sum_{x:e(x)=1} Pr_{S(J)}[X = x] < \sum_{x:e(x)=1} Pr_{S(I)}[X = x].$$

Let us group all sequences according to their values on positions other than i and $N-1$. For a binary sequence $x \in \{0,1\}^N$, we write x as zw where $z \in \{0,1\}^{N-2}$ records the $N-2$ binary values of x for positions other than i or $N-1$; and w records the binary values of x at positions i and $N-1$. Let X_z and X_w denote the random variables associated with the respective parts of the sequence x, and let S_z and S_w denote the stego distributions restricted to the parts of the sequence z and w respectively.

Now let x be any sequence that Eve classifies as stego, so that $e(x) = 1$. We assume for now that z (the components of x at positions other than i and $N-1$) is fixed. Since the conditions of Lemma 2 are satisfied, increasing x at position i can only move the classifier from stego to cover, or leave it the same. Moreover, changing x in position $N-1$ does not affect Eve's classifier at all. Given that Eve classifies x as stego, there are only two possible cases. Either

1. Eve classifies all four sequences zw with $w \in \{00, 10, 01, 11\}$, as stego, or
2. Eve classifies exactly the two sequences zw with $w \in \{00, 01\}$ as stego.

In the first case, the change in strategy from I to J does not change the value of

$$\sum_{\{w:e(zw)=1\}} Pr_S[X = zw],$$

since for fixed z,

$$\sum_{w\in\{00,10,01,11\}} Pr_{S(J)}[X = zw] = \sum_{w\in\{00,10,01,11\}} Pr_{S(I)}[X = zw].$$

In the second case, however, the probabilities of stego sequences differ.
In the case of the original distribution I, we have

$$\begin{aligned}
&\sum_{w\in\{00,10\}} Pr_{S(I)}[X = zw] \\
&= Pr_{S(I)}[X = z00] + Pr_{S(I)}[X = z10] \\
&= Pr_{S_z(I)}[X_z = z] \cdot (Pr_{S_w(I)}[X_w = 00] + Pr_{S_w(I)}[X_w = 01]) \\
&= Pr_{S_z(I)}[X_z = z] \cdot \left(\frac{f_i}{1-f_i} \cdot \frac{1-f_{N-1}}{f_{N-1}} + \frac{f_i}{1-f_i} \cdot \frac{f_{N-1}}{1-f_{N-1}} \right) \\
&= Pr_{S_z(J)}[X_z = z] \cdot \frac{f_i}{1-f_i} \cdot \frac{f_{N-1}^2 + (1-f_{N-1})^2}{f_{N-1}(1-f_{N-1})};
\end{aligned}$$

while in the case of the modified distribution J, we have

$$\begin{aligned}
&\sum_{w\in\{00,10\}} Pr_{S(J)}[X = zw] \\
&= Pr_{S_z(J)}[X_z = z] \cdot (Pr_{S_w(J)}[X_w = 00] + Pr_{S_w(J)}[X_w = 01]) \\
&= Pr_{S_z(J)}[X_z = z] \cdot \left(\frac{1-f_i}{f_i} \cdot \frac{f_{N-1}}{1-f_{N-1}} + \frac{1-f_i}{f_i} \cdot \frac{1-f_{N-1}}{f_{N-1}} \right) \\
&= Pr_{S_z(J)}[X_z = z] \cdot \frac{1-f_i}{f_i} \cdot \frac{f_{N-1}^2 + (1-f_{N-1})^2}{f_{N-1}(1-f_{N-1})} \\
&= Pr_{S_z(I)}[X_z = z] \cdot \frac{1-f_i}{f_i} \cdot \frac{f_{N-1}^2 + (1-f_{N-1})^2}{f_{N-1}(1-f_{N-1})} \\
&= \left(\frac{1-f_i}{f_i} \right)^2 \sum_{w\in\{00,10\}} Pr_{S(I)}[X = zw] \\
&< \sum_{w\in\{00,10\}} Pr_{S(I)}[X = zw].
\end{aligned}$$

By Lemma 4, this second case must occur for at least one stego sequence x; therefore, summing over all x with $e(x) = 1$ and grouping these x according to their z components, we see that the total probability of stego sequences

$$\sum_{x:e(x)=1} Pr_S[X = x]$$

is smaller under the distribution $S(J)$ than under the distribution $S(I)$. Thus Alice can strictly increase her payoff in the game by changing her strategy; and so the configuration is not an equilibrium. □

The theorem shows that if the game has non-trivializing parameter conditions, then it is not optimal for Alice to use only the least biased positions. Rather, she should also use additional positions that may not be taken into consideration by Eve. We conjecture an even stronger result holds – namely that Alice must actually use all N of the positions – under additional reasonable and precise parameter constraints. Two avenues for pursuing this conjecture include formulating more restrictive constraints that avoid navigating Eve's indeterminate actions on boundary sequences, or examining Eve's allowable equilibrium actions on boundary sequences more directly. We leave the precise statement and proof of this conjecture for future work.

In the following section, we explicitly compute all equilibria in the case of length-two sequences and an embedding size of $k = 1$.

5 Numerical Illustration

In this section, we instantiate our model with the special case of flipping a single bit ($k = 1$) in sequences of length two ($N = 2$). In this setting, Alice's pure strategy space is $\{\{0\}, \{1\}\}$; and since $a_{\{1\}} = 1 - a_{\{0\}}$, her mixed strategy space can be represented by a single value $a_0 = a_{\{0\}} \in [0, 1]$. Eve's pure strategy space is represented by the set of all $[0, 1]$-valued functions on $\{\binom{0}{0}, \binom{0}{1}, \binom{1}{0}, \binom{1}{1}\}$. Throughout this section we assume that cover and stego objects are equally likely, i.e., $p_C = p_S = \frac{1}{2}$. Notice that the assumption of equal priors implies the conditions from Equation (21) which guarantee only non-trivial equilibria.

5.1 Alice's Minimax Strategy

To compute Alice's minimax strategy, we first divide Alice's strategy space into three regions based on Eve's best response:

Lemma 5. *The following table gives Eve's best response for each sequence x as a function of a_0.*

Alice's strategy	*Eve's best response* $x =$			
	$\binom{0}{0}$	$\binom{0}{1}$	$\binom{1}{0}$	$\binom{1}{1}$
$a_0 < \theta_1$	*S*	*C*	*S*	*C*
$\theta_1 < a_0 < \theta_2$	*S*	*S*	*S*	*C*
$\theta_2 < a_0$	*S*	*S*	*C*	*C*

where $\theta_1 = \frac{(1-f_0)\tilde{f}_1}{f_0+f_1-1}$ *and* $\theta_2 = \frac{f_0\tilde{f}_1}{f_0+f_1-1}$.

Proof. We prove Eve's optimal decision for the four realizations separately.

$\binom{0}{0}$: Eve always classifies $\binom{0}{0}$ as stego.

$$\begin{aligned}
&\Pr_{\mathcal{C}}\left[X = \binom{0}{0}\right] = \\
&(1-f_0)(1-f_1) < a_0 f_0 (1-f_1) + (1-a_0)(1-f_0) f_1 \\
&\qquad = \Pr_{\mathcal{S}(a_0)}\left[X = \binom{0}{0}\right],
\end{aligned}$$

since $(1-f_0)(1-f_1) < f_0(1-f_1)$ and $(1-f_0)(1-f_1) < (1-f_0)f_1$.

$\binom{0}{1}$: Eve classifies $\binom{0}{1}$ as cover when $a_0 < \frac{(1-f_0)\tilde{f}_1}{f_0+f_1-1} := \theta_1$.

$$\begin{aligned}
&\Pr_{\mathcal{C}}\left[X = \binom{0}{1}\right] = \\
&\qquad (1-f_0) f_1 \overset{!}{>} a_0 f_0 f_1 + (1-a_0)(1-f_0)(1-f_1) \\
&\qquad\qquad = \Pr_{\mathcal{S}(a_0)}\left[X = \binom{0}{1}\right] && \Leftrightarrow \\
&(1-f_0)(f_1 - 1 + f_1) > a_0 (f_0 f_1 - 1 + f_0 + f_1 - f_0 f_1) && \Leftrightarrow \\
&\frac{(1-f_0)\tilde{f}_1}{f_0 + f_1 - 1} > a_0
\end{aligned}$$

$\binom{1}{0}$: Eve classifies $\binom{1}{0}$ as cover when $a_0 > \frac{f_0\tilde{f}_1}{f_0+f_1-1} := \theta_2$.

$$\begin{aligned}
&\Pr_{\mathcal{C}}\left[X = \binom{1}{0}\right] = \\
&\qquad f_0(1-f_1) \overset{!}{>} a_0 (1-f_0)(1-f_1) + (1-a_0) f_0 f_1 \\
&\qquad\qquad = \Pr_{\mathcal{S}(a_0)}\left[X = \binom{1}{0}\right] && \Leftrightarrow \\
&f_0(1-f_1) - f_0 f_1 > a_0 (1 - f_0 - f_1 + f_0 f_1 - f_0 f_1) && \Leftrightarrow \\
&\frac{-f_0 \tilde{f}_1}{1 - f_0 - f_1} < a_0
\end{aligned}$$

$\binom{1}{1}$: Eve always classifies $\binom{1}{1}$ as cover.

$$\begin{aligned}
&\Pr_{\mathcal{C}}\left[X = \binom{0}{0}\right] = \\
&\qquad f_0 f_1 > a_0 (1-f_0) f_1 + (1-a_0) f_0 (1-f_1) \\
&\qquad\qquad = \Pr_{\mathcal{S}(a_0)}\left[X = \binom{0}{0}\right],
\end{aligned}$$

since $f_0 f_1 > (1-f_0) f_1$ and $f_0 f_1 > f_0(1-f_1)$.

Finally, $\theta_1 < \theta_2$ always holds, since $(1 - f_0) < f_0$. □

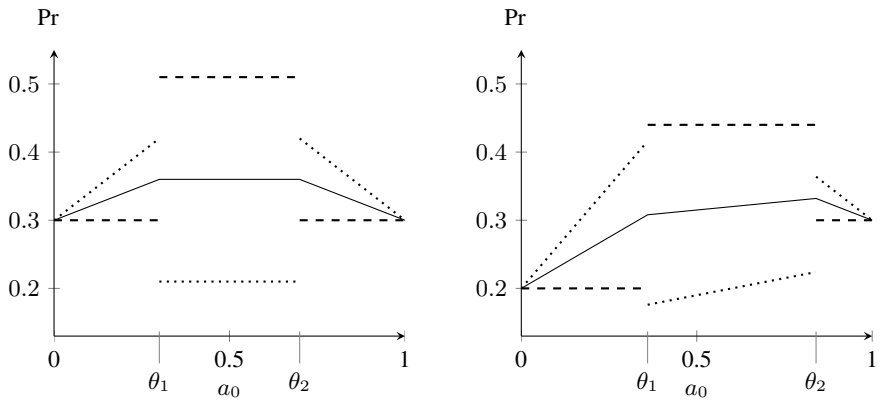
Theorem 2. *The strategy* $(\theta_2, 1 - \theta_2)$ *is a minimax strategy for Alice.*

Proof. First, for each region, we compute the derivative of Alice's payoff as a function of a_0 given that Eve always uses her best response. Then, we have that Alice's payoff is

- strictly increasing when $a_0 < \theta_1$,
- strictly decreasing when $a_0 > \theta_2$,
- and, when $\theta_1 \leq a_0 \leq \theta_2$, it is strictly increasing if $f_0 \neq f_1$, and it is constant if $f_0 = f_1$.

Thus, we have that $a_0 = \theta_2$ always attains the maximum. □

Note that embedding uniformly into both positions ($a_0 = \frac{1}{2}$) is optimal only if the biases are uniform ($f_0 = f_1$); and embedding only in the first position would be optimal only if the bias of the first position were zero ($\tilde{f}_0 = 0$) or if the bias of the second position were one ($\tilde{f}_1 = 1$). This confirms the results from [31], which also considers a two position game but allows Eve to look at only one position.



(a) $f_0 = 0.7, f_1 = 0.7$ (b) $f_0 = 0.7, f_1 = 0.8$

Fig. 2. Eve's false positive rate (dashed line), false negative rate (dotted line) and her overall misclassification rate (solid line) as a function of a_1, assuming that Eve plays a best response to Alice.

Figure 2 depicts Eve's error rates and the resulting overall misclassification rate as a function of Alice's strategy $(a_0, 1 - a_0)$. Figure 2(a) shows a homogeneous f, while Figure 2(b) shows a heterogeneous f. It can be seen that neither the false positive rate (dashed line) nor the false negative rate (dotted line) is continuous and that the discontinuities occur at the points θ_1 and θ_2, the points where Eve changes her optimal decision rule. Nonetheless, the overall misclassification rate (solid line) is continuous, which leads to the conclusion that this rate leverages out the discontinuities and thus is a good measure of the overall accuracy of Eve's detector.

5.2 Eve's Minimax Strategy

Theorem 3. *Eve's minimax strategy* $e_{minimax}$ *is* $e_{minimax}\binom{0}{0} = e_{minimax}\binom{0}{1} = 1$, $e_{minimax}\binom{1}{1} = 0$, *and*

$$e_{minimax}\binom{1}{0} = p = \frac{\tilde{f}_0}{f_0 + f_1 - 1} \,. \tag{26}$$

Proof. Since the game is zero sum, Eve's strategy is a minimax strategy if Alice's minimax strategy is a best response to it [34]. Therefore, it suffices to show that Alice has no incentives for deviating from her own minimax strategy when Eve uses $e_{minimax}$. Alice's best response to $e_{minimax}$ is

$$\begin{aligned}
&\operatorname*{argmax}_{a_0 \in [0,1]} \Bigg\{ - \Pr_{\mathcal{S}(a_0)}\left[X = \binom{0}{0}\right] - \Pr_{\mathcal{S}(a_0)}\left[X = \binom{0}{1}\right] \\
&\qquad\qquad + (1-2p)\Pr_{\mathcal{S}(a_0)}\left[X = \binom{1}{0}\right] + \Pr_{\mathcal{S}(a_0)}\left[X = \binom{1}{1}\right] \Bigg\} \\
&= \operatorname*{argmax}_{a_0 \in [0,1]} \Big\{ - a_0 f_0 (1-f_1) - (1-a_0)(1-f_0) f_1 \\
&\qquad\qquad - a_0 f_0 f_1 - (1-a_0)(1-f_0)(1-f_1) \\
&\qquad\qquad + (1-2p)\big[a_0(1-f_0)(1-f_1) + (1-a_0) f_0 f_1\big] \\
&\qquad\qquad + a_0(1-f_0) f_1 + (1-a_0) f_0 (1-f_1) \Big\} \\
&= \operatorname*{argmax}_{a_0 \in [0,1]} \Big\{ a_0 \left[2 - 4f_0 - 2p\left(1 - f_0 - f_1\right)\right] + \text{ const}(f,p) \Big\} \,.
\end{aligned}$$

If $p = \frac{\tilde{f}_0}{f_0 + f_1 - 1}$, then the value of the above optimization problem does not depend on a_0. Consequently, Alice has no incentives for deviating from her minimax strategy. □

It follows immediately from the theorem that Eve's minimax decision function is deterministic if and only if the cover is homogeneous ($f_0 = f_1$). This is interesting from the perspective of practical steganography, as all practical detectors are deterministic although embedding functions are pseudo-random and covers are heterogeneous.

6 Conclusion

We analyzed a two-player game between Alice, a content-adaptive steganographer, and Eve, an unbounded steganalyst. In keeping with a strict application of Kerckhoffs' principle to steganography, we allowed Eve access to Alice's embedding strategy, the cover source distribution, and unbounded computational power. Under these assumptions, we formalized processes both for constructing an optimal content-adaptive embedding strategy under the assumption of an optimal classifier, and for constructing an optimal detector under the assumption of an optimal embedding strategy.

Our formalism applies to arbitrary-sized cover sequences, although implementing the formalism for large covers remains a computational challenge. For the special case of a two-bit cover sequence, we exemplified an optimal classifier/embedding pair, and illustrated its structure in terms of the classification error rates.

For the practical steganalyst, our results give direction to the optimal detection of strategic embedding, and for optimal embedding against a strategic detector. In particular, Eve's optimal classifier should be monotone in the cover's predictability metric; and Alice's optimal adaptive embedding strategy should not naïvely use only the least biased positions. We also showed that a deterministic classifier can be sub-optimal for covers with heterogeneous predictability.

In our detailed analysis of length-two cover sequences, Alice's optimal randomized embedding strategy changed each part of the cover with some positive probability, and with more sophisticated structural constraints on the game's parameters, we expect that an analogous result can be proven for larger covers. It remains for future work to prove this conjecture and more directly address the computational tractability of implementing optimal strategies.

Acknowledgments: We thank the reviewers for their comments on an earlier version of this paper. We gratefully acknowledge support by the Penn State Institute for Cyber-Science. The second author's research visit at Penn State was supported under Visiting Scientists Grant N62909-13-1-V029 by the Office of Naval Research (ONR), and the third author's research visit at Penn State was supported by the Campus Hungary Program.

References

1. Acquisti, A., Dingledine, R., Syverson, P.: On the economics of anonymity. In: Wright, R. (ed.) Financial Cryptography, LNCS, vol. 2742, pp. 84–102. Springer-Verlag, Berlin, Germany (2003)
2. Anderson, R.: Stretching the limits of steganography. In: Anderson, R. (ed.) Information Hiding (1st International Workshop), LNCS, vol. 1174, pp. 39–48. Springer-Verlag, Berlin, Germany (1996)
3. Barni, M., Tondi, B.: The source identification game: An information-theoretic perspective. IEEE Transactions on Information Forensics and Security 8(3), 450–463 (Mar 2013)
4. Böhme, R.: Advanced Statistical Steganalysis. Springer-Verlag, Berlin, Germany (2010)
5. Böhme, R., Westfeld, A.: Exploiting preserved statistics for steganalysis. In: Fridrich, J. (ed.) Information Hiding. LNCS, vol. 3200, pp. 82–96. Springer-Verlag, Berlin, Germany (2004)
6. Chia, P., Chuang, J.: Colonel Blotto in the Phishing war. In: Baras, J., Katz, J., Altman, E. (eds.) Decision and Game Theory for Security, vol. 7037, pp. 201–218. Springer-Verlag, Berlin, Germany (2011)
7. Denemark, T., Fridrich, J.: Detection of content adaptive LSB matching: A game theory approach. In: Alattar, A., Memon, N., Heitzenrater, C. (eds.) Proceedings SPIE, Media Watermarking, Security, and Forensics. vol. 9028, p. 902804. SPIE and IS&T (2014)
8. Ettinger, M.: Steganalysis and game equilibria. In: Aucsmith, D. (ed.) Information Hiding (2nd International Workshop). LNCS, vol. 1525, pp. 319–328. Springer-Verlag, Berlin, Germany (1998)
9. Franz, E.: Steganography preserving statistical properties. In: Petitcolas, F. (ed.) Information Hiding (5th International Workshop). LNCS, vol. 2578, pp. 278–294. Springer-Verlag, Berlin, Germany (2003)

10. Fridrich, J.: Steganography in Digital Media: Principles, Algorithms, and Applications. Cambridge University Press, New York, NY (2009)
11. Fridrich, J., Goljan, M.: On estimation of secret message length in LSB steganography in spatial domain. In: Delp, E., Wong, P. (eds.) Proceedings SPIE, Security, Steganography, and Watermarking of Multimedia Contents VI. vol. 5306, pp. 23–34. SPIE (2004)
12. Fridrich, J., Kodovskỳ, J.: Multivariate Gaussian model for designing additive distortion for steganography. In: Proceedings of the IEEE International Conference on Acoustics, Speech and Signal Processing (ICASSP). pp. 2949–2953. Vancouver, BC (May 2013)
13. Fridrich, J., Du, R.: Secure steganographic methods for palette images. In: Pfitzmann, A. (ed.) Information Hiding (3rd International Workshop). LNCS, vol. 1768, pp. 47–60. Springer-Verlag, Berlin, Germany (2000)
14. Grossklags, J., Christin, N., Chuang, J.: Secure or insure?: A game-theoretic analysis of information security games. In: Proceedings of the 17th International World Wide Web Conference (WWW). pp. 209–218. Beijing, China (Apr 2008)
15. Guo, L., Ni, J., Shi, Y.: Uniform embedding for efficient JPEG steganography. IEEE Transactions on Information Forensics and Security 9(5), 814–825 (May 2014)
16. Hélouët, L., Zeitoun, M., Degorre, A.: Scenarios and covert channels: Another game... Electronic Notes in Theoretical Computer Science 119(1), 93–116 (Feb 2005)
17. Johnson, B., Böhme, R., Grossklags, J.: Security games with market insurance. In: Baras, J., Katz, J., Altman, E. (eds.) Decision and Game Theory for Security, vol. 7037, pp. 117–130. Springer-Verlag, Berlin, Germany (2011)
18. Johnson, B., Schöttle, P., Böhme, R.: Where to hide the bits? In: Grossklags, J., Walrand, J. (eds.) Decision and Game Theory for Security. LNCS, vol. 7638, pp. 1–17. Springer-Verlag, Berlin, Germany (2012)
19. Johnson, B., Schöttle, P., Laszka, A., Grossklags, J., Böhme, R.: Bitspotting: Detecting optimal adaptive steganography. In: Shi, Y., Kim, H., Pérez-González, F. (eds.) Digital-Forensics and Watermaking, pp. 3–18. LNCS, Springer, Berlin, Germany (2014)
20. Ker, A.: Batch steganography and the threshold game. In: Delp, E., Wong, P. (eds.) Proceedings SPIE, Security, Steganography, and Watermarking of Multimedia Contents IX. vol. 6505, pp. 0401–0413. SPIE (2007)
21. Laszka, A., Foldes, A.: Modeling content-adaptive steganography with detection costs as a quasi-zero-sum game. Infocommunications Journal 5(4), 33–43 (Dec 2013)
22. Laszka, A., Johnson, B., Schöttle, P., Grossklags, J., Böhme, R.: Managing the weakest link: A game-theoretic approach for the mitigation of insider threats. In: Proceedings of the 18th European Symposium on Research in Computer Security (ESORICS). pp. 273–290. Egham, UK (September 2013)
23. Maillé, P., Reichl, P., Tuffin, B.: Interplay between security providers, consumers, and attackers: A weighted congestion game approach. In: Baras, J., Katz, J., Altman, E. (eds.) Decision and Game Theory for Security, vol. 7037, pp. 67–86. Springer-Verlag, Berlin, Germany (2011)
24. Moulin, P., Ivanovic, A.: The zero-rate spread-spectrum watermarking game. IEEE Transactions on Signal Processing 51(4), 1098–1117 (Apr 2003)
25. Nash, J.: Non-cooperative games. The Annals of Mathematics 54(2), 286–295 (Sep 1951)
26. Orsdemir, A., Altun, O., Sharma, G., Bocko, M.: Steganalysis-aware steganography: Statistical indistinguishability despite high distortion. In: Delp, E., Wong, P., Dittmann, J., Memon, N. (eds.) Proceedings SPI, Security, Forensics, Steganography, and Watermarking of Multimedia Contents X. vol. 6819, p. 681915. SPIE (2008)
27. Petitcolas, F.: Introduction to information hiding. In: Katzenbeisser, S., Petitcolas, F. (eds.) Information Hiding Techniques for Steganography and Digital Watermarking, pp. 1–14. Recent Titles in the Artech House Computer Security Series, Artech House, Boston, MA (2000)

28. Pevný, T., Filler, T., Bas, P.: Using high-dimensional image models to perform highly undetectable steganography. In: Böhme, R., Fong, P., Safavi-Naini, R. (eds.) Information Hiding (12th International Conference). LNCS, vol. 6387, pp. 161–177. Springer-Verlag, Berlin, Germany (2010)
29. Pfitzmann, A., Hansen, M.: Anonymity, unlinkability, undetectability, unobservability, pseudonymity, and identity management-a consolidated proposal for terminology
30. Pita, J., Jain, M., Ordónez, F., Portway, C., Tambe, M., Western, C., Paruchuri, P., Kraus, S.: Using game theory for Los Angeles airport security. AI Magazine 30(1), 43–57 (Spring 2009)
31. Schöttle, P., Böhme, R.: A game-theoretic approach to content-adaptive steganography. In: Kirchner, M., Ghosal, D. (eds.) Information Hiding (14th International Conference). LNCS, vol. 7692, pp. 125 – 141. Springer-Verlag, Berlin, Germany (2012)
32. Schöttle, P., Laszka, A., Johnson, B., Grossklags, J., Böhme, R.: A game-theoretic analysis of content-adaptive steganography with independent embedding. In: Proceedings of the 21st European Signal Processing Conference (EUSIPCO). Marrakech, Morocco (Sep 2013)
33. Stamm, M., Lin, W., Liu, K.: Forensics vs. anti-forensics: A decision and game theoretic framework. In: Proceedings of the 2012 IEEE International Conference on Acoustics, Speech and Signal Processing (ICASSP). pp. 1749–1752. Kyoto, Japan (Mar 2012)
34. von Neumann, J., Morgenstern, O.: Theory of Games and Economic Behavior. Princeton University Press, Princeton, NJ (1944)