Bitspotting: Detecting Optimal Adaptive Steganography

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

[a]Department of Mathematics, University of California, Berkeley, USA
[b]Department of Information Systems, University of Münster, Germany
[c]Department of Networked Systems and Services,
Budapest University of Technology and Economics, Hungary
[d]College of Information Sciences and Technology, Pennsylvania State University, 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. The earlier work determined where Alice should hide the bits when Eve can only look in one position. Here, we expand Eve's capacity to spot these bits by allowing her to consider all positions. 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 [12]. 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 [5].

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 [1]. 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., [6]. 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 [13]. Furthermore, we bridge the two research areas of game-theoretical 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 [15]. In content-adaptive steganography [1], 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., [4]. Recent work on strategic embedding has dubbed this strategy *naïve adaptive embedding*, and has shown it to be a non-optimal strategy in progressively more general settings [2, 8, 13]. It was shown in [2] that the steganalyst can leverage her knowledge about the specific adaptive embedding algorithm from [4] to detect it with better accuracy than even random uniform embedding. In [13] it was shown for the first time that, if the steganalyst is strategic, it is never optimal for the steganographer to determin-

istically embed in the least predictable positions. The game-theoretic analysis in [13] was restricted to a model with two embedding positions, where Eve could only look in one position. A subsequent extension of that model [8] allowed the steganographer to change multiple bits in an arbitrary-sized cover sequence, but maintained limiting restrictions on the power of the steganlyst, by requiring her to make decisions on the basis of only one position. This paper investigates whether results from these earlier works, including non-optimality of naïve adaptive embedding, extend to the regime in which the steganalyst may consider all positions.

Another extension of this research stream expanded the power of Eve but required Alice to embed independently in each position [14]. Other authors have studied steganography using game-theoretical models, e. g., [3, 9], but none of these works addressed content-adaptive embedding, the most common approach in modern steganography, e. g., [7, 11].

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.

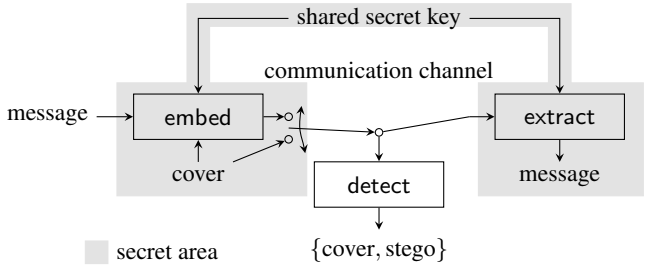


Fig. 1. Block diagram of a steganographic communication system

3.3 Choices

Alice's (pure strategy) choice is to select a size-k subset I of $\{0, \ldots, N-1\}$, 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.

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.[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 Ω. Let $X : \Omega \to \{0,1\}^N$ be the random variable which takes an event to its binary

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

Table 1. Payoffs for (Eve, Alice)

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

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 of an attack occurring or not, 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; and the mapping from the canonical representation to the projected representation is surjective[2] so that we may express results using the simpler representation without loss of generality.

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

$$\Pr_{\mathcal{C}}[X = x] = \prod_{i=0}^{N-1} \Pr_{\mathcal{C}}[X_i = x_i]. \tag{3}$$

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

$$\Pr_{\mathcal{C}}[X_i = 1] = f_i, \tag{4}$$

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 0s and 1s to make 1s more likely; and we can re-order the positions from least to most predictable.

For notational convenience, we define

$$\tilde{f}_i = 2f_i - 1. \tag{5}$$

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

$$\begin{aligned} \Pr_{\mathcal{C}}[X = x] &= \prod_{x_i = 1} f_i \cdot \prod_{x_i = 0} (1 - f_i) \\ &= \prod_{i=0}^{N-1} \left(1 - f_i + x_i \tilde{f}_i\right). \end{aligned} \tag{6}$$

[2] The proof of surjectivity follows directly from using induction on N.

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

$$\begin{aligned}
&\Pr_{\mathcal{S}}[X = x] \\
&= \sum_I a_I \Pr_{\mathcal{C}}[X = x_I] \\
&= \sum_I a_I \cdot \prod_{i \notin I} \Pr_{\mathcal{C}}[X_i = x_i] \cdot \prod_{i \in I} \Pr_{\mathcal{C}}[X_i = 1 - x_i] \\
&= \sum_I a_I \cdot \prod_{i \notin I} \left(1 - f_i + x_i \tilde{f}_i\right) \cdot \prod_{i \in I} \left(f_i - x_i \tilde{f}_i\right). \qquad (7)
\end{aligned}$$

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

$$\begin{aligned}
u(Eve) = &\Pr_{\Omega}[X \in E_S \text{ and } Y = S] && \text{(true positive)} \\
&+ \Pr_{\Omega}[X \in E_C \text{ and } Y = C] && \text{(true negative)} \\
&- \Pr_{\Omega}[X \in E_S \text{ and } Y = C] && \text{(false positive)} \\
&- \Pr_{\Omega}[X \in E_C \text{ and } Y = S] && \text{(false negative)} \qquad (8)
\end{aligned}$$

and this can be further computed as:

$$\begin{aligned}
u(Eve) &= p_S \Pr_{\mathcal{S}}[X \in E_S] + p_C \Pr_{\mathcal{C}}[X \in E_C] - p_C \Pr_{\mathcal{C}}[X \in E_S] - p_S \Pr_{\mathcal{S}}[X \in E_C] \\
&= \sum_{x \in \{0,1\}^N} \Big[e(x) p_S \Pr_{\mathcal{S}(a)}[X = x] \\
&\qquad + (1 - e(x)) p_C \Pr_{\mathcal{C}}[X = x] \\
&\qquad - (1 - e(x)) p_S \Pr_{\mathcal{S}(a)}[X = x] \\
&\qquad - e(x) p_C \Pr_{\mathcal{C}}[X = x] \Big] \\
&= \sum_{x \in \{0,1\}^N} (2e(x) - 1) \\
&\qquad \cdot (p_S \Pr_{\mathcal{S}(a)}[X = x] - p_C \Pr_{\mathcal{C}}[X = x]). \qquad (9)
\end{aligned}$$

The terms $\Pr_{\mathcal{C}}[X = x]$ and $\Pr_{\mathcal{S}(a)}[X = x]$ are defined in Equations (6) and (7), respectively. Note that we write $\mathcal{S} = \mathcal{S}(a)$ to clarify that the distribution $\mathcal{S}$ depends on Alice's mixed strategy. 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 theorems 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 (9) as appropriate.

Alice's Best Response Given a fixed strategy e for Eve, Alice's goal is to minimize the payoff in Equation (9). However, since she has no control over the cover distribution $\mathcal{C}$, this goal can be simplified to that of minimizing

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

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

$$\sum_{x\in\{0,1\}^N}(2e(x)-1))\cdot\text{Pr}_{\mathcal{C}}[X=x_I]. \tag{10}$$

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 (9). So, for each x, she should choose $e(x)$ to maximize the term of the sum corresponding to x. Specifically, if $p_S\text{Pr}_{S(a)}[X=x]-p_C\text{Pr}_C[X=x]>0$, 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

$$e(x)=\begin{cases}1 & \text{if } \frac{\text{Pr}_{\Omega}[Y=S|X=x]}{\text{Pr}_{\Omega}[Y=C|X=x]}>1\,,\\ 0 & \text{if } \frac{\text{Pr}_{\Omega}[Y=S|X=x]}{\text{Pr}_{\Omega}[Y=C|X=x]}<1\,,\\ \text{any } p\in[0,1] & \text{if } \frac{\text{Pr}_{\Omega}[Y=S|X=x]}{\text{Pr}_{\Omega}[Y=C|X=x]}=1\,.\end{cases} \tag{11}$$

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

$$\begin{aligned}
1 <& \frac{\Pr_\Omega[Y = S|X = x]}{\Pr_\Omega[Y = C|X = x]} \\
=& \frac{\Pr_\Omega[X = x]}{\Pr_\Omega[X = x]} \cdot \frac{\Pr_\Omega[Y = S|X = x]}{\Pr_\Omega[Y = C|X = x]} \\
=& \frac{\Pr_\Omega[Y = S]}{\Pr_\Omega[Y = C]} \cdot \frac{\Pr_\Omega[X = x|Y = S]}{\Pr_\Omega[X = x|Y = C]} \\
=& \frac{p_S}{p_C} \frac{\Pr_\mathcal{S}[X = x]}{\Pr_\mathcal{C}[X = x]} \\
=& \frac{p_S}{p_C} \frac{\sum_I a_I \cdot \prod_{i \notin I} \left(1 - f_i + x_i \tilde{f}_i\right) \cdot \prod_{i \in I} \left(f_i - x_i \tilde{f}_i\right)}{\prod_{i=0}^{N-1} \left(1 - f_i + x_i \tilde{f}_i\right)} \\
=& \frac{p_S}{p_C} \sum_I a_I \prod_{i \in I} \left(\frac{f_i - x_i \tilde{f}_i}{1 - f_i + x_i \tilde{f}_i} \right) \\
=& \frac{p_S}{p_C} \sum_I a_I \prod_{i \in I} \left(\frac{f_i}{1 - f_i} - x_i \frac{\tilde{f}_i}{f_i(1 - f_i)} \right). \qquad (12)
\end{aligned}$$

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 [15].

Eve's minimax strategy is given by

$$\operatorname*{argmax}_e \left(\min_I \Big(\sum_{x \in \{0,1\}^N} (2e(x) - 1)(p_S \Pr_\mathcal{C}[X = x_I] - p_C \Pr_\mathcal{C}[X = x]) \Big) \right); \qquad (13)$$

while Alice's minimax strategy is given by

$$\begin{aligned}
\operatorname*{argmin}_a \Bigg(\max_{E_S} \Big(& \sum_{x \in E_S} (p_S \Pr_{\mathcal{S}(a)}[X = x] - p_C \Pr_\mathcal{C}[X = x]) \\
& + \sum_{x \in E_C} (p_C \Pr_\mathcal{C}[X = x] - p_S \Pr_{\mathcal{S}(a)}[X = x]) \Big) \Bigg). \qquad (14)
\end{aligned}$$

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 [10]. We begin with a lemma showing that Eve's classifier in a specific type of equilibrium must respect the canonical partial ordering on binary sequences. We conclude the section with a conjecture about Alice's equilibrium strategy.

Lemma 1. *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 (11) and (12)), 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.

Next, we state a conjecture about Alice's strategy in an equilibrium.

Conjecture 1. Assume equal priors, so that $p_S = p_C = \frac{1}{2}$ and a reasonable $\langle f_i \rangle_{i=0}^{N-1}$. In a Nash equilibrium, Alice uses every $i \in \{0, \ldots, N-1\}$ with non-zero probability.

The proscription that Alice flips each position with some positive probability complements an analogous result proven in [8] for the case in which Eve is restricted to look in only one position. We leave it to future work to characterize the $\langle f_i \rangle_{i=0}^{N-1}$ for which Conjecture 1 holds. For homogeneous $\langle f_i \rangle$ there are simple counter-examples to the conjecture; however, it is important to note that for homogeneous $\langle f_i \rangle$ the definition of adaptive embedding itself is not sensible.

Here, we frame a proof outline for this conjecture. Assume a Nash equilibrium with a and e as the strategies of Alice and Eve, respectively. To obtain a contradiction,

suppose that $i \in \{0, \ldots, N-1\}$ is such that $a_I = 0$ for every I containing i. If x is any sequence that Eve's optimal decision rule classifies as either clearly cover or clearly stego, then Eve's behavior does not depend on the value of x at position i. However, if there are "indifferent" sequences y that Eve's likelihood test proscribes as cover or stego with equal probability, we cannot rule out that Eve may take the position i into account for y. This remains true even though her likelihood test does not proscribe an outcome based on i, and even though she is playing a best response to Alice who is not using position i. Our avenue to proceed is to demonstrate a violation of the equilibrium condition by showing how Alice can increase her payoff by using position i. Toward this end, we can show that, by shifting her embedding probability to sets containing i from sets not containing i, Alice will increase Eve's misclassification probability for sequences that are not on her "indifference boundary". However, it is possible that Eve gains enough advantage from conditioning on i when the special boundary sequences occur to offset this disadvantage. It seems to us that this possibility hinges on structural properties of the sequence $\langle f_i \rangle$.

In the following section, we explicitly compute all equilibria in the case of length-two sequences and a message length of $k = 1$. Note that in this special case, Conjecture 1 is true.

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., $p_C = p_S = \frac{1}{2}$.

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 2. *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 $(1 - f_0) < f_0$. □

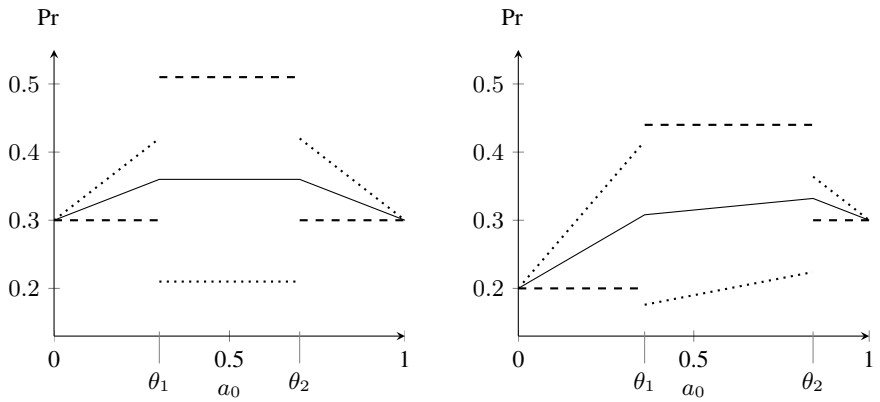
Theorem 1. *The strategy $(\theta_2, 1 - \theta_2)$ is a minimax strategy for Alice.*

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 [13], 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 2. *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*

$$e_{minimax}\binom{1}{0} = p = \frac{\tilde{f}_0}{f_0 + f_1 - 1} \,. \tag{15}$$

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 [15]. 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. In particular, Eve's optimal classifier should be monotone in the cover's predictability metric; and 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 we conjectured an analogous result for larger covers. It remains for future work to prove our conjecture and more directly address the computational tractability of implementing optimal strategies. We seek to either find computable mechanisms for implementing these strategies, or prove hardness results showing that such mechanisms do not exist.

References

1. Rainer Böhme. *Advanced Statistical Steganalysis*. Springer, 2010.
2. Rainer Böhme and Andreas Westfeld. Exploiting preserved statistics for steganalysis. In *Information Hiding*, volume 3200 of *LNCS*, pages 82–96. Springer, 2004.
3. Mark Ettinger. Steganalysis and game equilibria. In *Information Hiding*, volume 1525 of *LNCS*, pages 319–328. Springer, 1998.
4. Elke Franz. Steganography preserving statistical properties. In *Information Hiding*, volume 2578 of *LNCS*, pages 278–294. Springer, 2003.
5. Jessica Fridrich. *Steganography in Digital Media: Principles, Algorithms, and Applications*. Cambridge University Press, New York, NY, USA, 1st edition, 2009.
6. Jessica Fridrich and Miroslav Goljan. On estimation of secret message length in LSB steganography in spatial domain. volume 5306, pages 23–34. SPIE, 2004.
7. Jessica Fridrich and Jan Kodovskỳ. Multivariate Gaussian model for designing additive distortion for steganography. In *ICASSP*. IEEE, 2013.
8. Benjamin Johnson, Pascal Schöttle, and Rainer Böhme. Where to hide the bits? In *GameSec 2012*, volume 7638 of *LNCS*, pages 1–17. Springer.
9. Andrew D. Ker. Batch steganography and the threshold game. volume 6505, page 650504. SPIE, 2007.
10. John Nash. Non-cooperative games. *The Annals of Mathematics*, 54(2):286–295, 1951.
11. Tomáš Pevný, Tomáš Filler, and Patrick Bas. Using high-dimensional image models to perform highly undetectable steganography. In *Information Hiding*, volume 6387 of *LNCS*, pages 161–177. Springer, 2010.
12. Andreas Pfitzmann and Marit Köhntopp. Anonymity, unobservability, and pseudonymity – a proposal for terminology. In *Designing Privacy Enhancing Technologies*, volume 2009 of *LNCS*, pages 1–9. Springer, 2001.
13. Pascal Schöttle and Rainer Böhme. A game-theoretic approach to content-adaptive steganography. In *Information Hiding*, volume 7692 of *LNCS*, pages 125 – 141. Springer, 2012.
14. Pascal Schöttle, Aron Laszka, Benjamin Johnson, Jens Grossklags, and Rainer Böhme. A game-theoretic analysis of content-adaptive steganography with independent embedding. In *EUSIPCO*. IEEE, 2013.
15. John von Neumann and Oskar Morgenstern. *Theory of Games and Economic Behavior*. Princeton University Press, 1944.