A Game-Theoretic Approach to Content-Adaptive Steganography

Pascal Schöttle and Rainer Böhme

Department of Information Systems, University of Münster, Germany
{pascal.schoettle,rainer.boehme}@uni-muenster.de

Abstract. Content-adaptive embedding is widely believed to improve steganographic security over uniform random embedding. However, such security claims are often based on empirical results using steganalysis methods not designed to detect adaptive embedding. We propose a framework for content-adaptive embedding in the case of imperfect steganography. It formally defines heterogeneity within the cover as a necessary condition for adaptive embedding. We devise a game-theoretic model for the whole process including cover generation, adaptive embedding, and a detector which anticipates the adaptivity. Our solution exhibits a unique equilibrium in mixed strategies. Its location depends on the level of heterogeneity of the cover source, but never coincides with naïve adaptive embedding. The model makes several simplifying assumptions, including independent cover symbols and the steganalyst's ability to recover the adaptivity criterion perfectly.

Keywords: Content-Adaptive Steganography, Game Theory, Security.

1 Introduction and Motivation

In the past couple of years, several so-called *content-adaptive* steganographic schemes have been proposed, e. g., [13,25,26,21,23]. They all have in common that they embed in the locations of the cover medium, which are most suitable for embedding, i. e., where changes are (supposed to be) harder to detect. To find these locations, the schemes specify an *adaptivity criterion*, e. g., the local variance. Most often the superiority of content-adaptive over random uniform embedding is claimed on the grounds of better resistance against selected steganalysis methods, not tailored to detect adaptive embedding. However, such arguments disregard Kerckhoffs' principle [20]: the warden knows the adaptivity criterion as well and may be able to reproduce or estimate its values. In other words, the adaptivity criterion leaks side information to the warden.

Furthermore, most of the adaptive schemes embed the m bits of the secret message M into the m most "secure" locations of the cover medium. From now on, we will call this kind of adaptive embedding *naïve adaptive steganography.* There is initial evidence that this is not optimal. For example, it is shown in [4] that the adaptive embedding function suggested in [9] is less secure than uniform random embedding, if the attacker recalculates the adaptivity criterion.

M. Kirchner and D. Ghosal (Eds.): IH 2012, LNCS 7692, pp. 125–141, 2012.

An implication of this finding is that restricting the steganographer to the most suitable embedding locations can lead to less secure steganography. Therefore, leaving the steganographer with more choice on where to embed may strengthen the resistance to steganalysis methods specifically designed to detect adaptive embedding. As the steganalyst, in turn, can anticipate this behavior, she has to be given choice, too. Game theory is the preferred method to model a situation with two (or more) opponents who can adjust their strategies according to assumptions about the behavior of the other(s). In general, they want to either maximize their gain or minimize their loss in a competitive environment. So-called Nash equilibria [22] are stable situations in this environment, where none of the players would benefit from unilaterally changing her strategy.

Game theory requires that all participants have a parameter of choice. In our case this choice is discrete for both players, steganographer and steganalyst. We model the choice of the steganographer as the decision to embed either in the better location or in the worse. A steganalyst who anticipates adaptive embedding can choose which of the symbols she pays more attention to, depending on their suitability for embedding.

This paper documents a first attempt to develop a rigorous approach to secure content-adaptive steganography. We formulate a game-theoretic model spanning the entire process from cover generation to embedding and detection. For now, we keep the model as simple as possible in order to be able to solve our game, and to calculate theoretical bounds of detectability for arbitrary embedding and detection functions. By this, we are able to prove that naïve adaptive steganography is never optimal and introduce the term of *optimal adaptive steganography* as an adaptive embedding function, which anticipates a steganalysis technique that is aware of content-adaptive embedding and may recover the adaptivity criterion. Depending on the level of heterogeneity, optimal adaptive embedding distributes the embedding changes between more secure and less secure locations.

This paper is organized as follows: Section 2 briefly reviews related work. Section 3 gives a formal definition of heterogeneity and develops our basic model including first conclusions about which strategies are possible at all. Section 4 deals with the game-theoretical payoff function and optimal strategies for both players. The results are discussed in Section 5. Finally, Section 6 draws a conclusion and prioritizes directions for future work.

2 Related Work

The idea of combining game theory with steganographic security was first mentioned by Ettinger in 1998 [7], who proposes zero-sum games to model the contest between a data-hider and a data-attacker. He studies active attackers who not only want to detect, but to suppress hidden communication. Consequently, this approach is less focussed on indistinguishability, but on the maximum capacity which can be hidden robust enough to prevent an attacker, who is bound by a distortion constraint, from suppressing the channel.

Ker [16] uses game theory to find best strategies for a steganographer who can spread her secret message over several homogeneous cover media (batch steganography), and a steganalyst who anticipates this and tries to detect the existence of at least one secret message (pooled steganalysis). He concludes that a (batch) steganographer should either spread her payload as thinly as possible or concentrate it as much as possible. The specific choice of the payoff function precludes to fully explore mixed strategy equilibria. So the author presents min-max and max-min solutions in pure strategies.

To our knowledge there are no other game-theoretic works in the area of steganographic security so far. However, in general, game theory is gaining popularity in the field of information security, e. g. [1,14].

In the context of syndrome coding, Fridrich [10] shows that for sufficiently large covers, it is never optimal to embed only into the symbols which cause the least amount of (additive) distortion. Her result, along with the definition of a *detectability profile*, which mirrors our notion of an adaptivity criterion, is relevant for adaptive steganography. However, her work does not specify a detector. Therefore, it solves an optimization problem and not a game.

3 Our Model

3.1 Definition of Heterogeneity

A precondition for adaptive steganography is heterogeneity within the cover. For example, in images, flat regions are less secure to embed, whereas edges and noisy areas are likely more secure. Until now, there is no formal definition of heterogeneity for the purpose of adaptive embedding. We try to close this gap.

Definition 1 (Cover). *A sequence of n k-bit symbols is called* cover*, if it is a realization of the (cover) distribution P_0. More specifically, every symbol of the cover can take values in $\{0, \ldots, 2^k - 1\}$.*

Cachin [5] defines information-theoretic security of a steganographic system. He assumes that the distribution of the covers P_0 and the distribution of the stego objects P_1 are known. Then he suggests to use the *Kullback–Leibler divergence* (KLD) as a measure of discrepancy between these two distributions. He derives bounds for the detectability of a steganographic embedding function. A lower KLD indicates more similar distributions and thus a more secure embedding function. Therefore, if the embedding function is fixed, it is convenient to base a definition of heterogeneity on KLD.

Definition 2 (Heterogeneity). *A cover is called* heterogeneous*, if it contains (well-defined) areas, where embedding changes result in a lower KLD. I. e., let P_0 be the probability distribution of the cover and $P_{(x_i)}$ be the altered probability distribution after making a specific embedding change at location x_i. Then, the cover is* heterogeneous *iff there exists $i, j \in [1, \ldots, n], i \neq j$ with $KLD(P_0, P_{(x_i)}) \neq KLD(P_0, P_{(x_j)})$. Otherwise the cover is* homogeneous.

So, the simplest model to study adaptive embedding consists of exactly two areas which differ in their detectability of embedding changes.

3.2 Game-Theoretical Setup

Let *Alice* be the steganographer and *Eve* be the steganalyst.

As mentioned in Section 1, Eve has access to the embedding function. This is a realistic assumption and in line with Kerckhoffs' principle. There are discussions on how to interpret this principle for steganography [6,2,8], but Eve's access to the embedding function should be undisputed. Alice does not know the cover distribution P_0, because with that knowledge she could choose her stego objects like realizations of P_0 and could thus perform perfect steganography [24]. Granting Eve access to both distributions P_0 and P_1 (which would be the case for a strict interpretation of Kerckhoffs' principle [8]) would enable her to detect at the information-theoretic bound. This is neither realistic nor interesting to examine. We follow [3,18] where it is argued that a more realistic setup is incomplete information on both sides. With this condition, neither perfect embedding nor best possible detection is practicable and thus, both players have to make choices. In particular, both players have to anticipate the choice of their opponent. By this we are in a classical game-theoretic situation.

As mentioned above, the simplest model to study adaptive embedding consists of exactly two areas. We further specify this to a model with exactly two 2-bit symbols $p_0^{(0)}, p_1^{(0)}$, one better suitable for embedding than the other, i. e., $n = k = 2$. Following the notation in [3], the superscript (0) in $p_i^{(0)}$ denotes a symbol before embedding and the superscript (1) in $p_i^{(1)}$ denotes a symbol after embedding. If symbols are independent (see Sect. 3.6 below), we can think of larger heterogeneous covers as sets of pairs of pixels $(p_0^{(0)}, p_1^{(0)})$ drawn from two equally sized areas of different detectability. The game is repeated for each pair.

Since steganographic security is defined by the indistinguishability between cover and stego objects, we start with the "game" introduced by Katzenbeisser and Petitcolas [15]. Despite the name, their setup is not a game in a game-theoretic sense, but inspired by cryptographic security proofs. We augment it with choice variables in adaptive embedding to make it a veritable game.

Figure 1 shows the extensive form of our game. The different entities in our game are: *Nature*, the steganographer *Alice*, the *Judge*, and the steganalyst *Eve*. Nature generates a cover with exactly two symbols $p_0^{(0)}, p_1^{(0)}$, according to a predefined probability mass function (PMF). Without loss of generality, among the two symbols, $p_0^{(0)}$ is always better or equally suitable for embedding than $p_1^{(0)}$. Upon receiving a heterogeneous cover from Nature, Alice embeds with probability $\bar{a}$ into $p_0^{(0)}$ and with probability $1-\bar{a}$ into $p_1^{(0)}$. The Judge is fair and forwards with constant probability $\mu = 1/2$ either the cover or the stego object to Eve. In a game-theoretic sense, the Judge is a part of Nature. When Eve gets either the cover or the stego, she examines $p_0^{(1)}$ with probability $\bar{e}$ and $p_1^{(1)}$ with probability $1 - \bar{e}$. Then she has to make a decision about the type of object she received.

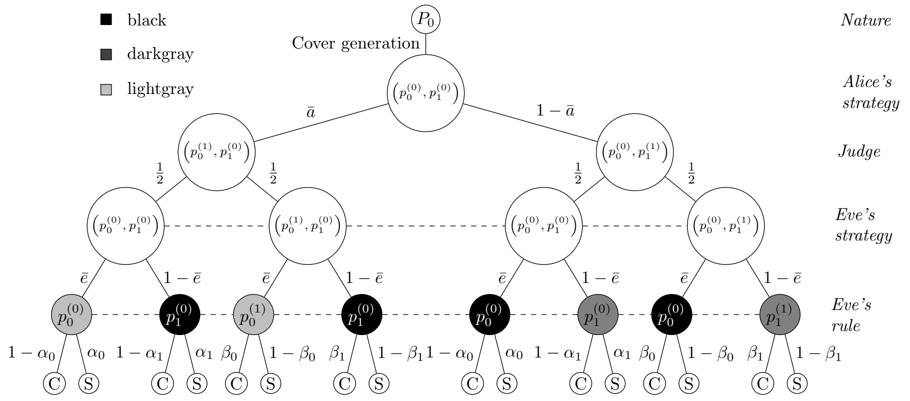


Fig. 1. Content-adaptive game in extensive form. The dashed line indicates Eve's information set, i. e., Eve does not know which of the connected nodes has been reached. α_i, β_i are the false positive, respectively false negative, rates for f_{m_i}, the PMF of $p_i^{(0)}$.

3.3 Embedding Function

We model LSB replacement as embedding function because it is best studied and well tractable. We will introduce one modification in that Alice always has to flip one bit instead of one on average. We justify this by the fact that in practice it is very unlikely that not a single bit has to be flipped. The corresponding probability is 2^{-m} for an m-bit message and thus negligible in m.

Note that changing exactly one symbol is incompatible with the popular simplifying assumption of independent embedding. It makes the symbols dependent in P_1 even if they were independent in P_0. Therefore P_1 cannot be decomposed into a product of the PMFs of its symbols. Other models are certainly conceivable, but not considered in this work.

3.4 Strategies

Alice's parameter of choice is $a \in \{0, 1\}$. A value of $a = 1$ means she embeds in $p_0^{(0)}$, i. e., the better suitable symbol, and $a = 0$ means she embeds in $p_1^{(0)}$. We assume that the order of suitability is perfectly preserved through embedding (not an unrealistic assumption for several so far proposed adaptivity criteria) and thus Eve can recover it. In future models we may relax the assumption of perfect recovery and replace it by a partial recovery.

Eve's parameter of choice is $e \in \{0, 1\}$. A value of $e = 1$ means she examines $p_0^{(1)}$, i. e., the better suitable symbol, and $e = 0$ means she examines $p_1^{(1)}$. We model Eve's decision in the way that she can either examine p_0 or p_1, but not both at the same time. We justify this by the fact that for real-world covers, it is intractable to use all information from the joint distribution of all symbols in the sequence. Although specific steganalysis methods can take all symbols into account, there has to be a weighting decision [19,17] and we model this decision in our model by an exclusive either p_0 or p_1. Note that whenever restricting the adversary, security claims may break down if she is more powerful than assumed.

Game theory differentiates between stable situations in *pure* and in *mixed* strategies. A pure strategy is a strategy where a player deterministically decides what to do, whereas a mixed strategy is a probability distribution over pure strategies. To be able to research the mixed strategies as well, we introduce the random variable A, of which Alice's choice a is a realization and the random variable E, of which Eve's choice e is a realization. Furthermore, let $\bar{a} = \text{prob}(A = 1)$ and $\bar{e} = \text{prob}(E = 1)$ be Alice's, respectively Eve's, parameter in mixed strategies. Now, a value of $\bar{a} = 1/2$ means that Alice embeds randomly without bias and a value of $\bar{e} = 1/2$ means that Eve examines both symbols with the same probability.

3.5 Exclusion of Pure Strategies

Lemma 1. *Under the assumption that $P_0 \neq P_1$ for LSB replacement, i. e., LSB replacement does not preserve the cover distribution perfectly, there is no equilibrium in pure strategies.*

Proof. There are exactly four pure strategies in the above described game.

1. Alice embeds always in $p_0^{(0)}$.
2. Alice embeds always in $p_1^{(0)}$.
3. Eve examines always p_0.
4. Eve examines always p_1.

If Alice follows strategy (1) (i. e., naïve adaptive embedding), Eve's best response would be strategy (3), because she would not gain from examining the other location. Hence, Alice would change her strategy to (2) so that Eve would not get any information from examining $p_0^{(0)}$. Now, Eve would switch to (4) because all information would be in $p_1^{(1)}$. Now, Alice's best response would be strategy (1) again, because Eve will not detect changes there. By this they are in an infinite loop.

So, in every situation in pure strategies, one of the players would benefit from changing her strategy. Therefore no equilibrium exists in pure strategies. □

3.6 Cover Generation Model

We need a model to represent some (simplified) conditions of heterogeneous cover sources. For this, our model should have one parameter m_i to adjust the level of heterogeneity. Now, the distribution P_0 according to which the two ordered symbols $p_0^{(0)}$ and $p_1^{(0)}$ are realised, is a discrete bivariate distribution of $f_{m_0}^{(0)}$ (the PMF of $p_0^{(0)}$) and $f_{m_1}^{(0)}$ (the PMF of $p_1^{(0)}$) with $m_0 \neq m_1$ (if $m_0 = m_1$, we model a homogeneous cover). Here, m_i measures the suitability for embedding. A value of $m_i = 0$ indicates a uniform distribution (i. e., maximal entropy) and allows perfect steganography. With increasing m_i, the entropy and the suitability for embedding decrease. As we assume that $p_0^{(0)}$ is more suitable for embedding, we define $m_0 \leq m_1$. In practice, the order of the symbols is established by the adaptivity criterion. Reordering the cover according to this criterion removes Markov-properties of the cover [8], but maintains some higher-order dependencies not regarded here, because most of them are incognizable or intractable in practice. Therefore, we may assume that the two ordered symbols are independent before embedding.

So the joint PMF of the cover generation $f^{(0)}(x_0, x_1)$ is given by

$$f^{(0)}(x_0, x_1) = f_{m_0}^{(0)}(x_0) \cdot f_{m_1}^{(0)}(x_1). \tag{1}$$

To fulfil the requirements from above, we model the family of probability mass functions depending on m_i as

$$f_{m_i}^{(0)}(x) = (2^k - x)m_i + \frac{1 - \left(\sum_{j=1}^{2^k} j\right) m_i}{2^k}, \quad x \in \{0, \ldots, 2^k - 1\}, \text{ with} \tag{2}$$

$$m_i \in \left[0; \left(\sum_{j=1}^{2^k - 1} j\right)^{-1}\right), \text{ and therefore: } m_i \in \left[0; \frac{1}{6}\right) \text{ for } k = 2. \tag{3}$$

Equation (2) ensures that the sum of masses equals 1 and the masses for the different symbol values are strictly decreasing. The constraints in Equation (3) ensure that the PMF is never negative. Note that the interval has to be open. Otherwise the value $x = 2^k - 1$ would have zero mass. This would allow detection with certainty whenever this value occurs in a stego object after LSB flipping.

Figure 2 visualizes our cover generation model. For two fixed values of m_0, it shows the corresponding PMFs depending on m_1. A lower value of m_0 in the homogeneous case means a higher entropy. A bigger difference between m_0 and m_1 indicates a higher level of heterogeneity within the cover. As can be seen, by changing m_0 and m_1, the entropy as well as the level of heterogeneity change.

3.7 Embedding Impact

Let $f_{m_i}^{(1)}$ be the PMF resulting from always embedding in $p_i^{(0)}$. Then, for single symbol values x_j it holds, that:

$$f_{m_i}^{(0)}(x_j) = \text{prob}(x_j|\text{Cover}) \quad \text{and} \quad f_{m_i}^{(1)}(x_j) = \text{prob}(x_j|\text{Stego}). \tag{4}$$

As we are interested in the distribution after embedding P_1, we now proceed by examining the distribution after embedding in $p_0^{(0)}$ with probability $\bar{a}$ and embedding in $p_1^{(0)}$ with probability $1 - \bar{a}$.

The LSB replacement embedding operation emb(x) simply swaps the values $2j$ by $2j + 1$, and vice versa, for $j \in \{0, \ldots, 2^{k-1}\}$. This can be expressed by

$$\text{emb}(x) := x + (-1)^x \quad \Rightarrow \quad \text{emb}^{-1}(x) = \text{emb}(x). \tag{5}$$

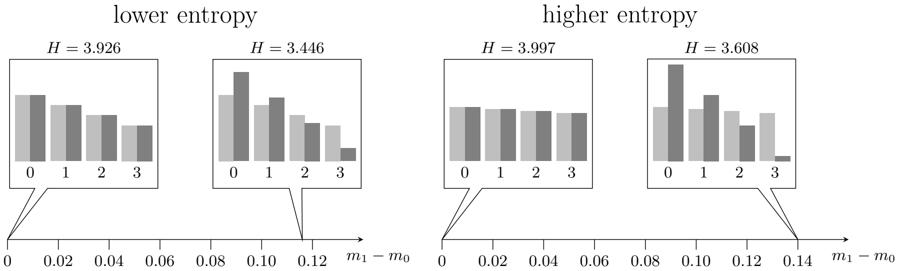


Fig. 2. Cover generation model with increasing levels of heterogeneity from left to right. $f^{(0)}_{m_0}$ is light gray, $f^{(0)}_{m_1}$ is dark gray. Left: $m_0 = 0.05, m_1 \in \{0.05, 0.165\}$. Right: $m_0 = 0.01, m_1 \in \{0.01, 0.15\}$.

Now in our model, where we always embed, it holds that

$$f^{(1)}_{m_i}(x_j) = f^{(0)}_{m_i}(emb^{-1}(x_j)), \quad j \in \{0, 1, \ldots, n\}. \tag{6}$$

This yields the following lemma about $f^{(1)}_{m_i}(x_j)$.

Lemma 2. *In our model, the PMF* $f^{(1)}_{m_i}(x_j)$ *is*

$$f^{(1)}_{m_i}(x_j) = \begin{cases} f^{(0)}_{m_i}(x_j + 1), & : x_j \equiv 0 \pmod 2 \\ f^{(0)}_{m_i}(x_j - 1), & : x_j \equiv 1 \pmod 2 \end{cases} \tag{7}$$

$$= \begin{cases} f^{(0)}_{m_i}(x_j) - m_i, & : x_j \equiv 0 \pmod 2 \\ f^{(0)}_{m_i}(x_j) + m_i, & : x_j \equiv 1 \pmod 2). \end{cases} \tag{8}$$

Proof. From Equation (6) we know that:

$$\begin{aligned} f^{(1)}_{m_i}(x_j) &= f^{(0)}_{m_i}(emb^{-1}(x_j)) \\ &= f^{(0)}_{m_i}(x_j + (-1)^{x_j}) \\ &= \begin{cases} f^{(0)}_{m_i}(x_j + 1), & : x_j \equiv 0 \pmod 2 \\ f^{(0)}_{m_i}(x_j - 1), & : x_j \equiv 1 \pmod 2). \end{cases} \end{aligned} \tag{9}$$

And with Equation (2):

$$\begin{aligned} f^{(1)}_{m_i}(x_j) &= \begin{cases} (2^k - (x_j + 1))m_i + \frac{1-\left(\sum_{j=1}^{2^k} j\right)m_i}{2^k}, & : x_j \equiv 0 \pmod 2 \\ (2^k - (x_j - 1))m_i + \frac{1-\left(\sum_{j=1}^{2^k} j\right)m_i}{2^k}, & : x_j \equiv 1 \pmod 2 \end{cases} \\ &= \begin{cases} f^{(0)}_{m_i}(x_j) - m_i, & : x_j \equiv 0 \pmod 2 \\ f^{(0)}_{m_i}(x_j) + m_i, & : x_j \equiv 1 \pmod 2). \end{cases} \end{aligned} \tag{10}$$

□

As Lemma 1 excludes both pure strategies, we get a mixed strategy and thus a mixture distribution of the kind,

$$f^{(1)}(x_0, x_1) = \bar{a}\left(f^{(1)}_{m_0}(x_0) \cdot f^{(0)}_{m_1}(x_1)\right) + (1 - \bar{a})\left(f^{(0)}_{m_0}(x_0) \cdot f^{(1)}_{m_1}(x_1)\right). \qquad (11)$$

To quantify the overall information Eve can potentially gain from the embedding function, we can numerically calculate the KLD between $f^{(0)}$ and $f^{(1)}$ as benchmark for a numerical analysis. This is certainly precluded for real covers.

3.8 Eve's Decision

The parameter on which Eve's choice relies is $\bar{e}$. This indicates to which symbol she assigns a higher weight. This symbol will influence her decision and thus her false positive and false negative rates more. Conveniently, as will be shown in this paragraph, the false positive rate equals the false negative rate in our model. So we have only one variable of interest, the *equal error rate* (*EER*).

Recall that we have a strictly decreasing PMF and thus for P_0 it holds that,

$$f^{(0)}_{m_i}(0) > f^{(0)}_{m_i}(1) > f^{(0)}_{m_i}(2) > f^{(0)}_{m_i}(3). \qquad (12)$$

Therefore, we know from Lemma 2 that in pure strategies it holds that,

$$f^{(1)}_{m_i}(1) > f^{(1)}_{m_i}(0) > f^{(1)}_{m_i}(3) > f^{(1)}_{m_i}(2). \qquad (13)$$

This is sufficient to derive Eve's optimal decision rule $DR(x_j)$ between "Cover" and "Stego" for individual symbols.

Lemma 3. *Eve's best decision rule for individual symbol values x_j is:*

$$DR(x_j) = \begin{cases} \text{Cover}, & : x_j \equiv 0 \pmod 2 \\ \text{Stego}, & : x_j \equiv 1 \pmod 2. \end{cases} \qquad (14)$$

Proof. The decision rule implements the *maximum a posteriori* (MAP) estimation, which can be found, for example, in [11]. Here it is important to notice that the a priori probability of "Cover" prob(Cover) $= \mu = 1/2$ equals the probability of "Stego" prob(Stego) $= \mu = 1/2$ because the Judge is fair.

The MAP estimation minimizes the decision errors by calculating:

$$\hat{\theta} = \arg\max_{\theta} \text{prob}(\theta|x) = \arg\max_{\theta} \text{prob}(x|\theta) \cdot \text{prob}(\theta). \qquad (15)$$

With $\theta \in \{\text{Cover}, \text{Stego}\}$ and $x = x_j$, this results in

$$\begin{aligned} \hat{\theta} &= \arg\max_{\theta} \text{prob}(x_j|\theta) \cdot \mu \\ &\overset{Eq.\,(4)}{=} \max\left\{f^{(0)}_{m_i}(x_j), f^{(1)}_{m_i}(x_j)\right\} \\ &= \begin{cases} \text{Cover}, & : x_j \equiv 0 \pmod 2 \\ \text{Stego}, & : x_j \equiv 1 \pmod 2, \end{cases} \end{aligned} \qquad (16)$$

because of Equations (12) and (13). □

Thus, in our case with $n = k = 2$, Eve's decides for "Cover" whenever she sees a symbol with value 0 or 2, and "Stego" for values 1 and 3.

Let α_i and β_i be Eve's false positive, respectively, false negative rate for $f_{m_i}^{(0)}$ and $f_{m_i}^{(1)}$. By Lemma 3, her true positive rate $(1-\alpha_i)$ (and consequently the false positive rate as well) is aggregated between the cases where her decision yields "Cover" and the same for the true negative rate $(1 - \beta_i)$ in all other cases.

Lemma 4. *In our model, Eve's false positive rate α_i equals her false negative rate β_i and thus is called* equal error rate EER_i.

$$EER_i = \alpha_i = \beta_i = \frac{1}{2} - m_i, \tag{17}$$

for $i \in \{0, 1\}$.

The proof can be found in Appendix A.1.

Equation (17) is intuitive, as values of $m_i = 0$ indicate an uniform distribution. In this case P_1 would equal P_0, i. e., the same distribution before and after embedding. Therefore the false positive and false negative rate would be 50%, i. e., random guessing. Furthermore, it follows our initial thoughts that a higher value of m_i implies a better detectability, which materializes in a lower EER.

Corollary 1. *The worst case for Eve would be Alice choosing $a \in \{0, 1\}$ and she herself choosing $e = 1 - a$ because by this, her decision would be merely guessing, i. e., $EER = 0.5$.*

The proof can be found in Appendix A.2.

This confirms Lemma 1 that there is no equilibrium in pure strategies, as with every pure strategy, one of the players would benefit from changing her strategy to the opposite. Now we are in the position to solve the game and to identify equilibria in mixed strategies.

4 Solving the Game

The EER described in Section 3.8 can be seen as the payoff function in our zero-sum game. As it is Alice's intention to perform least detectable steganography, her goal is to maximize the EER, whereas it is Eve's goal to maximize her detection rate and thus, to minimize the EER.

4.1 Payoff Function

From Figure 1 and the EER described in Section 3.8, the payoff function $\chi(\bar{a}, \bar{e})$ for mixed strategies can be derived and equals the overall EER. It is stated in the following corollary.

Corollary 2. *In our model, the payoff function in mixed strategies is*

$$\chi(\bar{a}, \bar{e}) = \frac{1}{2} - (\bar{a} \cdot \bar{e} \cdot m_0 + (1 - \bar{a}) \cdot (1 - \bar{e}) \cdot m_1) \tag{18}$$

Proof. From Figure 1 it can be seen, that the nodes of Eve's decision (shaded nodes) can be partitioned into three different situations.

The first situation is that Alice embeds in $p_0^{(0)}$ and Eve anticipates this (lightgray nodes in Figure 1). This situation occurs with probability $\bar{a} \cdot \bar{e}$. When faced with a situation like this, we know from Equation (17) that Eve's EER equals α_0 $(= \beta_0)$. The second possible situation is that Alice embeds in $p_1^{(0)}$ and Eve, again, anticipates this (darkgray nodes in Figure 1). The occurrence probability of this situation is $(1 - \bar{a}) \cdot (1 - \bar{e})$. Again, we know the payoff from Equation (17), which is α_1 $(= \beta_1)$. The third and last situation is that Alice embeds in $p_i^{(0)}$, but Eve inspects the wrong location (black nodes in Figure 1). This situation occurs with probability $(1 - \bar{a}) \cdot \bar{e}$ (for Alice embedding in $p_0^{(0)}$, but Eve examining $p_1^{(1)}$) and $\bar{a} \cdot (1 - \bar{e})$ (for Alice embedding in $p_1^{(0)}$, but Eve examining $p_0^{(1)}$). Here, we know from Corollary 1 that Eve's decision rule is no better than random guessing and thus has an EER of 1/2.

This leads to the following expression for $\chi(\bar{a}, \bar{e})$,

$$\begin{aligned}\chi(\bar{a}, \bar{e}) &= (\bar{a} \cdot \bar{e}) \cdot \alpha_0 + ((1 - \bar{a}) \cdot \bar{e} + \bar{a} \cdot (1 - \bar{e})) \cdot \frac{1}{2} + (1 - \bar{a}) \cdot (1 - \bar{e}) \cdot \alpha_1 \\ &= (\bar{a} \cdot \bar{e}) \cdot \alpha_0 + \frac{\bar{a} + \bar{e} - 2\bar{a}\bar{e}}{2} + (1 - \bar{a}) \cdot (1 - \bar{e}) \cdot \alpha_1. \qquad (19)\end{aligned}$$

From Lemma 4 we know that $\alpha_i = 1/2 - m_i$ and thus:

$$\begin{aligned}\chi(\bar{a}, \bar{e}) &= (\bar{a} \cdot \bar{e}) \cdot (\frac{1}{2} - m_0) + \frac{\bar{a} + \bar{e} - 2\bar{a}\bar{e}}{2} + (1 - \bar{a}) \cdot (1 - \bar{e}) \cdot (\frac{1}{2} - m_1) \\ &= \frac{\bar{a} \cdot \bar{e}}{2} - \bar{a} \cdot \bar{e} \cdot m_0 + \frac{\bar{a} + \bar{e} - 2\bar{a} \cdot \bar{e}}{2} \\ &+ \frac{1}{2} - \frac{\bar{a}}{2} - \frac{\bar{e}}{2} + \frac{\bar{a} \cdot \bar{e}}{2} - (1 - \bar{a}) \cdot (1 - \bar{e}) \cdot m_1 \qquad (20) \\ &= \frac{1}{2} - (\bar{a} \cdot \bar{e} \cdot m_0 + (1 - \bar{a}) \cdot (1 - \bar{e}) \cdot m_1) \qquad (21)\end{aligned}$$

□

4.2 Best Strategies

As the payoff function is the same for both players but with contrary goals, i. e., Alice wants to maximize it, while Eve wants to minimize it, an equilibrium in mixed strategies can be found by looking at the partial derivatives of the payoff function and setting them to zero. With this method we find a unique equilibrium of our model, which happens to be symmetric.

Lemma 5. *In our model, there exists a unique symmetric Nash equilibrium in mixed strategies. In this equilibrium it holds that:*

$$\bar{a}^* = \bar{e}^* = \frac{m_1}{m_0 + m_1} \qquad (22)$$

Proof. The partial derivatives of the payoff function are,

$$\frac{\partial \chi(\bar{a},\bar{e})}{\partial \bar{a}} = -\left(m_0+m_1\right)\cdot \bar{e} + m_1 \tag{23}$$

$$\frac{\partial \chi(\bar{a},\bar{e})}{\partial \bar{e}} = -\left(m_0+m_1\right)\cdot \bar{a} + m_1 . \tag{24}$$

Setting both derivatives to zero yields the values for the equilibrium,

$$-(m_0+m_1)\cdot \bar{e} + m_1 \overset{!}{=} 0 \Leftrightarrow \bar{e}^* = \frac{m_1}{m_0+m_1} \tag{25}$$

$$-(m_0+m_1)\cdot \bar{a} + m_1 \overset{!}{=} 0 \Leftrightarrow \bar{a}^* = \frac{m_1}{m_0+m_1} . \tag{26}$$

□

Inserting these optimal values into $\chi(\bar{a}^*,\bar{e}^*)$ yields the equilibrium EER .

Corollary 3. *In the equilibrium it holds that the EER is,*

$$EER^* = \chi\left(\frac{m_1}{m_0+m_1},\frac{m_1}{m_0+m_1}\right) = \frac{1}{2} - \frac{m_0\cdot m_1}{m_0+m_1} . \tag{27}$$

Proof. Equation (21) can be rearranged to

$$\chi(\bar{a},\bar{e}) = \frac{1}{2} - \left((m_0+m_1)\cdot(\bar{a}\cdot\bar{e}) - m_1\cdot\bar{a} - \bar{e}\cdot m_1 + m_1\right), \tag{28}$$

and using $\bar{e} = \bar{a} = \bar{a}^* = \frac{m_1}{m_0+m_1}$ from Lemma 5 we obtain,

$$\chi(\bar{a}^*,\bar{a}^*) = \frac{1}{2} - \big((m_0+m_1)\cdot(\bar{a}^*)^2 - 2\cdot m_1\cdot\bar{a}^* + m_1\big) \tag{29}$$

$$= \frac{1}{2} - \left((m_0+m_1)\cdot\left(\frac{m_1}{m_0+m_1}\right)^2 - \frac{2\cdot {m_1}^2}{m_0+m_1} + m_1\right) \tag{30}$$

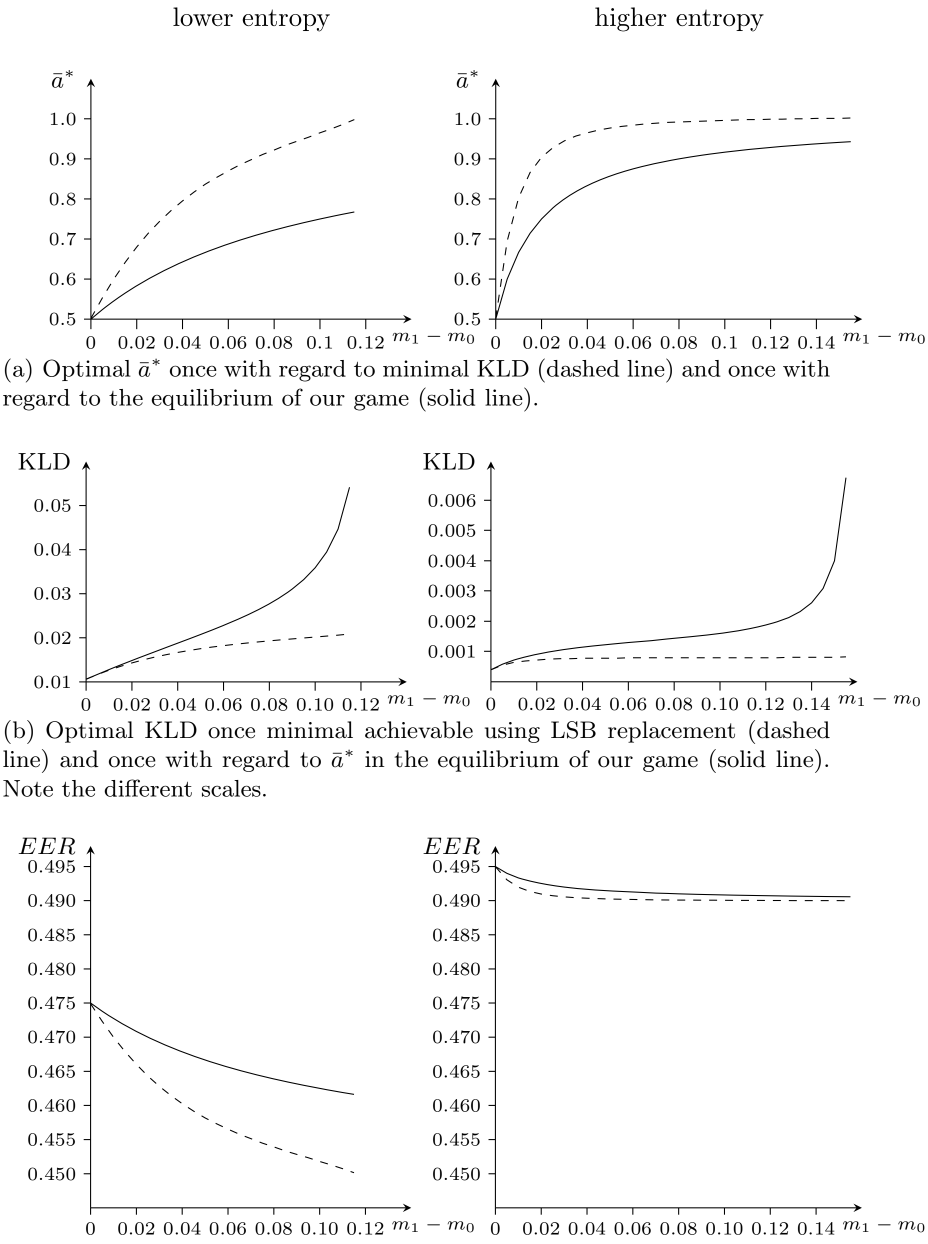
$$= \frac{1}{2} - \left(m_1 - \frac{{m_1}^2}{m_0+m_1}\right) = \frac{1}{2} - \frac{m_0\cdot m_1}{m_0+m_1} . \tag{31}$$

□

With this unique value for $\bar{a}^*$, we say a steganographer performs *optimal adaptive steganography.* It is always less detectable than a steganographer who performs naïve adaptive steganography.

5 Discussion

One implication of our analysis is that the optimal distribution of embedding changes depends on the level of heterogeneity in the cover source. So, steganographer and steganalyst both have to adjust their strategy to the cover source.



(a) Optimal $\bar{a}^*$ once with regard to minimal KLD (dashed line) and once with regard to the equilibrium of our game (solid line).

(b) Optimal KLD once minimal achievable using LSB replacement (dashed line) and once with regard to $\bar{a}^*$ in the equilibrium of our game (solid line). Note the different scales.

(c) Optimal EER with optimal KLD and fixed detector (dashed line) and once in the equilibrium of our game (solid line).

Fig. 3. Comparsion of equilibrium parameters with numerical benchmarks based on KLD, as a function of the level of heterogeneity. Left figures: $m_0 = 0.05, m_1 \in [0.05, 0.165]$. Right figures: $m_0 = 0.01, m_1 \in [0.01, 0.165]$.

The discussion of our results is facilitated by looking at numerical examples in Figure 3. As one requirement for our model was simplicity, we are able to calculate the KLD as benchmark, which is infeasible for real-world cover sources.

Figure 3(a) shows the optimal value of $\bar{a}^*$, once by numerically minimizing KLD (dashed line) and once the value found in the equilibrium (solid line). Figure 3(b) shows the KLD created by the values for $\bar{a}^*$ from the figure above and Figure 3(c) shows the resulting EER. To recall how the corresponding PMFs look like, they are displayed in Figure 2. Figure 3(a) reveals that if Alice's goal was to minimize KLD, she would choose higher values for $\bar{a}^*$, i. e., embed with higher probability in the better suitable location. Furthermore, it can be seen in Figure 3(b) that the KLD generated by Alice's strategy in the equilibrium increases rapidly with an increasing level of heterogeneity. Nonetheless, Figure 3(c) shows that Alice's strategy in the equilibrium implicates a higher EER than in the situation with minimal KLD, and thus more secure steganography against the specific detector defined in our model. By this, both players could perform better, if the other would not follow the strategy in the equilibrium. So, it follows that if Alice tries to minimize the KLD and Eve anticipates this (still being bound to her specific detector), Eve's detection rate would increase and thus Alice would perform less secure steganography.

6 Conclusion and Outlook

The literature is full of content-adaptive embedding schemes, but most of them seem to be designed ad-hoc. Their security relies solely on the opinion of the respective developer that the adaptivity criterion of her choice is good at selecting secure embedding locations. To overcome such design methods in the medium term, we give a first definition of heterogeneity for content-adaptive steganography and specify a model of the entire process, covering the choices of Alice and Eve, and being simple enough to be tractable, both in terms of game-theoretic equilibria and information-theoretic benchmarks.

We show that *naïve adaptive steganography*, the strategy to embed only in the most suitable locations of a heterogeneous cover, is never optimal. We solve our model and find a unique equilibrium of our game, where none of the players would gain from changing her strategy. As a result, we define a new kind of adaptive embedding, the so-called *optimal adaptive steganography*, which takes into account the knowledge of an attacker who can recover (or estimate) the values of the adaptivity criterion as side information.

The way we model the level of adaptivity certainly needs further refinement and, in future works, we may be able to relax some of the restrictions we impose on our model. Furthermore, as mentioned in Section 3.6, changing m_0 and m_1 of the cover generation model influences the entropy and the level of heterogeneity. It would be more convenient if both quantities of interest could be adjusted independently. This is a goal for future models.

It is obvious that a cover model with exactly two locations is not realistic, so there is space for future work. Special attention in these future models has to be be paid to what happens if the parameters $\bar{a}, \bar{e}$, our players' parameters of choice, become divided into n instead of 2 parts. By this, we have to think about how to model the different weights and will most likely come to an optimization problem over the function field, similar to the batch steganography problem stated in [16].

Another open question is the relation between adaptive embedding and steganalysis based on machine learning. As a first remark on the combination of these two areas, [12] states that "[...] it does not appear that giving [Eve] probabilistic information about the selection channel is a weakness".

Another field for future research is the advantage the attacker gains from cover estimation in the case of heterogeneity *within* and *between* covers. By adding this to our model, we end up with a double-stochastic cover generation process. As can be seen by these examples, a rigorous understanding of content-adaptive steganography in theory and practice remains a relevant target for future investigations.

Acknowledgement. This research was funded by the German science foundation (DFG) under grant "Sichere adaptive Steganographie". The second author thanks Andrew Ker for useful discussions on game theory and steganography in general.

References

1. Bensoussan, A., Kantarcioglu, M., Hoe, S.: A Game-Theoretical Approach for Finding Optimal Strategies in a Botnet Defense Model. In: Alpcan, T., Buttyán, L., Baras, J. (eds.) GameSec 2010. LNCS, vol. 6442, pp. 135–148. Springer, Heidelberg (2010)
2. Böhme, R.: An Epistemological Approach to Steganography. In: Katzenbeisser, S., Sadeghi, A.-R. (eds.) IH 2009. LNCS, vol. 5806, pp. 15–30. Springer, Heidelberg (2009)
3. Böhme, R.: Advanced Statistical Steganalysis. Springer, Berlin (2010)
4. Böhme, R., Westfeld, A.: Exploiting Preserved Statistics for Steganalysis. In: Fridrich, J. (ed.) IH 2004. LNCS, vol. 3200, pp. 82–96. Springer, Heidelberg (2004)
5. Cachin, C.: An Information-Theoretic Model for Steganography. In: Aucsmith, D. (ed.) IH 1998. LNCS, vol. 1525, pp. 306–318. Springer, Heidelberg (1998)
6. Cayre, F., Bas, P.: Kerckhoffs–based embedding security classes for WOA data hiding. IEEE Transactions on Information Forensics and Security 3(1), 1–15 (2008)
7. Ettinger, J.M.: Steganalysis and Game Equilibria. In: Aucsmith, D. (ed.) IH 1998. LNCS, vol. 1525, pp. 319–328. Springer, Heidelberg (1998)
8. Filler, T., Ker, A.D., Fridrich, J.: The square root law of steganographic capacity for Markov covers. SPIE, vol. 7254, p. 725408 (2009)
9. Franz, E.: Steganography Preserving Statistical Properties. In: Petitcolas, F.A.P. (ed.) IH 2002. LNCS, vol. 2578, pp. 278–294. Springer, Heidelberg (2003)
10. Fridrich, J.: Minimizing the embedding impact in steganography. In: Voloshynovskiy, S. (ed.) MM&Sec 2006: Proceedings of the 8th Workshop on Multimedia and Security, pp. 2–10. ACM (2006)

11. Fridrich, J.: Steganography in Digital Media: Principles, Algorithms, and Applications, 1st edn. Cambridge University Press, New York (2009)
12. Fridrich, J., Kodovský, J., Holub, V., Goljan, M.: Steganalysis of Content-Adaptive Steganography in Spatial Domain. In: Filler, T., Pevný, T., Craver, S., Ker, A. (eds.) IH 2011. LNCS, vol. 6958, pp. 102–117. Springer, Heidelberg (2011)
13. Fridrich, J., Du, R.: Secure Steganographic Methods for Palette Images. In: Pfitzmann, A. (ed.) IH 1999. LNCS, vol. 1768, pp. 47–60. Springer, Heidelberg (2000)
14. Johnson, B., Grossklags, J., Christin, N., Chuang, J.: Are Security Experts Useful? Bayesian Nash Equilibria for Network Security Games with Limited Information. In: Gritzalis, D., Preneel, B., Theoharidou, M. (eds.) ESORICS 2010. LNCS, vol. 6345, pp. 588–606. Springer, Heidelberg (2010)
15. Katzenbeisser, S., Petitcolas, F.: Defining security in steganographic systems. SPIE, vol. 4675, pp. 50–56 (2002)
16. Ker, A.: Batch steganography and the threshold game. SPIE, vol. 6505, p. 650504 (2007)
17. Ker, A.: Estimating Steganographic Fisher Information in Real Images. In: Katzenbeisser, S., Sadeghi, A.-R. (eds.) IH 2009. LNCS, vol. 5806, pp. 73–88. Springer, Heidelberg (2009)
18. Ker, A.: The Square Root Law in Stegosystems with Imperfect Information. In: Böhme, R., Fong, P.W.L., Safavi-Naini, R. (eds.) IH 2010. LNCS, vol. 6387, pp. 145–160. Springer, Heidelberg (2010)
19. Ker, A., Böhme, R.: Revisiting weighted stego-image steganalysis. SPIE, vol. 6819, p. 681905 (2008)
20. Kerckhoffs, A.: La cryptographie militaire. Journal des Sciences Militaires IX(1), 5–38 (1883)
21. Luo, W., Huang, F., Huang, J.: Edge adaptive image steganography based on LSB matching revisited. IEEE Transactions on Information Forensics and Security 5(2), 201–214 (2010)
22. Nash, J.: Non-cooperative games. The Annals of Mathematics 54(2), 286–295 (1951)
23. Pevný, T., Filler, T., Bas, P.: Using High-Dimensional Image Models to Perform Highly Undetectable Steganography. In: Böhme, R., Fong, P.W.L., Safavi-Naini, R. (eds.) IH 2010. LNCS, vol. 6387, pp. 161–177. Springer, Heidelberg (2010)
24. Wang, Y., Moulin, P.: Perfectly secure steganography: Capacity, error exponents, and code constructions. IEEE Transactions on Information Theory 54(6), 2706–2722 (2008)
25. Wu, H.C., Wu, N.I., Tsai, C.S., Hwang, M.S.: Image steganographic scheme based on pixel-value differencing and LSB replacement methods. In: IEE Proceedings – Vision, Image and Signal Processing, vol. 152(5), pp. 611–615 (2005)
26. Yang, C.H., Weng, C.Y., Wang, S.J., Sun, H.M.: Adaptive data hiding in edge areas of images with spatial LSB domain systems. IEEE Transactions on Information Forensics and Security 3(3), 488–497 (2008)

A Appendix

A.1 Proof of Lemma 4

As mentioned in Section 3.8, Eve's true positive and true negative rate can be calculated as follows:

True Positives TP(x_j):

$$x_j = 0 : \mathrm{TP}(0) = \frac{f^{(0)}_{m_i}(0)}{f^{(0)}_{m_i}(0) + f^{(1)}_{m_i}(0)} = \frac{f^{(0)}_{m_i}(0)}{f^{(0)}_{m_i}(0) + f^{(0)}_{m_i}(1)} \tag{32}$$

$$x_j = 2 : TP(2) = \frac{f^{(0)}_{m_i}(2)}{f^{(0)}_{m_i}(2) + f^{(1)}_{m_i}(2)} = \frac{f^{(0)}_{m_i}(2)}{f^{(0)}_{m_i}(2) + f^{(0)}_{m_i}(3)} \tag{33}$$

$$\Rightarrow (1-\alpha_i) = (f^{(0)}_{m_i}(0) + f^{(0)}_{m_i}(1)) \cdot TP(0) + (f^{(0)}_{m_i}(2) + f^{(0)}_{m_i}(3)) \cdot TP(2) \tag{34}$$

$$= f^{(0)}_{m_i}(0) + f^{(0)}_{m_i}(2) \tag{35}$$

True Negatives TN(x_j):

$$x_j = 1 : \mathrm{TN}(1) = \frac{f^{(1)}_{m_i}(1)}{f^{(0)}_{m_i}(1) + f^{(1)}_{m_i}(1)} = \frac{f^{(0)}_{m_i}(0)}{f^{(0)}_{m_i}(0) + f^{(0)}_{m_i}(1)} = TP(0) \tag{36}$$

$$x_j = 3 : \mathrm{TN}(3) = \frac{f^{(1)}_{m_i}(3)}{f^{(0)}_{m_i}(3) + f^{(1)}_{m_i}(3)} = \frac{f^{(0)}_{m_i}(2)}{f^{(0)}_{m_i}(2) + f^{(0)}_{m_i}(3)} = TP(2) \tag{37}$$

$$\Rightarrow (1-\beta_i) = (f^{(0)}_{m_i}(0) + f^{(0)}_{m_i}(1)) \cdot TN(1) + (f^{(0)}_{m_i}(2) + f^{(0)}_{m_i}(3)) \cdot TN(3) \tag{38}$$

$$= f^{(1)}_{m_i}(1) + f^{(1)}_{m_i}(3) = f^{(0)}_{m_i}(0) + f^{(0)}_{m_i}(2) = (1-\alpha_i) \tag{39}$$

$$\overset{Eq.(2)}{\Leftrightarrow} (1-\alpha_i) = (1-\beta_i) = 4 \cdot m_i + \frac{1-10m_i}{4} + 2 \cdot m_i + \frac{1-10m_i}{4} \tag{40}$$

$$= 6 \cdot m_i + 2 \cdot \frac{1-10m_i}{4} = \frac{2 \cdot m_i + 1}{2} = m_i + \frac{1}{2} \tag{41}$$

$$\Rightarrow EER_i = \alpha_i = \beta_i = \frac{1}{2} - m_i. \tag{42}$$

for $i \in \{0,1\}$. □

A.2 Proof of Corollary 1

If Eve chooses $e = 1 - a$ and $a \in \{0,1\}$, it holds that Alice always embeds in $p^{(0)}_a$ and by this never into $p^{(0)}_e$. From Eq. (6) it follows that $f^{(1)}_{m_a}(x_j) = f^{(0)}_{m_a}(emb^{-1}(x_j))$, but $f^{(1)}_{m_e}(x_j) = f^{(0)}_{m_e}(x_j)$, as there is no embedding in $p^{(0)}_e$. Therefore, it holds that:

$$\left.\begin{array}{r} x_j \in \{0,2\} : \mathrm{TP}(x_j) \\ x_j \in \{1,3\} : \mathrm{TN}(x_j) \end{array}\right\} = \frac{f^{(0)}_{m_e}(x_j)}{f^{(0)}_{m_e}(x_j) + f^{(1)}_{m_e}(x_j)}$$

$$= \frac{f^{(0)}_{m_e}(x_j)}{f^{(0)}_{m_e}(x_j) + f^{(0)}_{m_e}(x_j)} = \frac{f^{(0)}_{m_e}(x_j)}{2 \cdot f^{(0)}_{m_e}(x_j)} = \frac{1}{2}. \tag{43}$$

□