On the Role of Side Information in Steganography in Empirical Covers

Jessica Fridrich

Department of ECE, SUNY Binghamton, NY, USA

ABSTRACT

In an attempt to alleviate the negative impact of unavailable cover model, some steganographic schemes utilize the knowledge of the so-called "precover" when embedding secret data. The precover is typically a higher-resolution (unquantized) representation of the cover, such as the raw sensor output before it is converted to an 8-bit per channel color image. The precover object is only available to the sender but not to the Warden, which seems to give a fundamental advantage to the sender. In this paper, we provide theoretical insight for why side-informed embedding schemes for empirical covers might provide high level of security. By adopting a piece-wise polynomial model corrupted by AWGN for the content, we prove that when the cover is sufficiently non-stationary, embedding by minimizing distortion w.r.t. the precover is more secure than by preserving a model estimated from the cover (the so-called model-based steganography). Moreover, the side-informed embedding enjoys four times lower steganographic Fisher information than LSB matching.

1. MOTIVATION

The problem of steganography is to devise a scheme using which secret messages can be passed to another party by hiding them in cover objects so that a traffic-monitoring entity called Warden cannot distinguish between genuine cover objects and objects carrying secret data.[2] In steganography by cover modification, the secret is embedded by making changes to the cover. If the cover-source distribution is known and available to the communicating parties as well as the Warden, the rate of perfectly secure steganographic communication is positive[22,23] even when the actions of both the sender and the (possibly active) Warden are power limited. When the cover source is empirical in nature[1] (e.g., digital media files acquired by a sensor), the individual cover elements, such as pixels or JPEG DCT coefficients follow a highly non-stationary distribution that reflects the content as well as numerous sources of noise. This enormous complexity makes perfect security essentially unachievable in practice – the Warden always seems able to find a representation of the covers within which the actions of the sender can be detected, forcing thus the sender to embed with a vanishing rate as the cover size increases (the so-called square root law of imperfect steganography[5,7,14–17]).

To alleviate the negative impact of unavailable cover model, some steganographic schemes make use of the knowledge of the so-called "precover" when embedding secret data. The precover is usually a higher-resolution (unquantized) representation of the cover, such as the raw image before it is JPEG compressed or raw sensor output before it is converted to an 8-bit per channel color image, such as TIFF or JPEG. Such side-informed schemes provide a very high level of security in practice[10,18,19,21] at least when the security is measured empirically using feature-based steganalyzers implemented using machine learning. The failure of current steganalysis to reliably detect side-informed schemes should, however, be taken with a grain of salt because it could simply mean that current steganalysis lacks the right models (feature spaces). In short, to the best knowledge of the author the role of side-information in steganography in empirical covers is largely unclear with a surprising lack of rigorous arguments.

This paper is an attempt to shed more light on this intriguing topic while focusing on finding as simple formalization as possible that already provides valuable insight. In Section 4, we formalize the concepts of precover and cover sources. In particular, we model images as sequences of segments on which the content follows a linearly parametrizable polynomial model corrupted by additive white Gaussian noise (AWGN) similar in nature to the model investigated in.[3] Here, the content is captured by segments' boundaries and the model

E-mail: fridrich@binghamton.edu; http://www.ws.binghamton.edu/fridrich

parameters. Following a commonly adopted conservative viewpoint, we grant the Warden with a complete knowledge of the model (the so-called fully informed Warden) while the sender needs to estimate the model parameters. This allows us in Section 3 to quantify security in the limit of small payloads by the steganographic Fisher information in the leading term of the Taylor expansion of KL divergence between the cover and stego distributions.

In Section 4, we describe three different embedding methods – embedding while preserving an estimated model, Quantization-Index Modulation (QIM), and the simple Least Significant Bit matching (LSBM), and analyze them by computing their steganographic Fisher information. In Section 4.4, we prove that for a highly non-stationary cover (a statement that can be precisely quantified), the QIM has a lower Fisher information than when the sender embeds by preserving an estimated cover model. The QIM is also always more secure than the uninformed LSBM whose steganographic Fisher information is four times larger than that of QIM. Conversely, for less complex covers, the sender is better off to embed while preserving the estimated cover distribution instead of applying the QIM. The paper is concluded in Section 5.

Calligraphic font is reserved for sets, while capital letters with their corresponding lower-case letters are used for random variables and their realizations. Matrices will be upright boldface symbols (e.g., $\mathbf{A}_{rs}$ is the rsth element of matrix $\mathbf{A}$) with $\mathbf{A}^{\mathrm{T}}$ standing for the transpose. For a statement P, the Iverson bracket $[P] = 1$ when P is true, and it is zero when P is false. We reserve the symbols $\mathbb{R}$ and $\mathbb{Z}$ for the set of real numbers and integers and $\mathbf{I}_n$ for an $n \times n$ unity matrix. Gaussian distribution with mean vector $\mu \in \mathbb{R}^n$ and covariance matrix $\mathbf{C} \in \mathbf{R}^{n\times n}$ will be denoted $N(\mu, \mathbf{C})$. Finally, we use $h_2(\beta) = -\beta \log_2 \beta - (1-\beta)\log_2(1-\beta)$ for the binary entropy function.

Given a countable set of scalar bin centroids, $\mathcal{M} = \{m_j\}$, $m_j < m_{j+1}$, a scalar quantizer is a mapping $Q_{\mathcal{M}} : \mathbb{R} \to \mathcal{M}$, defined as $Q_{\mathcal{M}}(x) = \arg\min_{m_j \in \mathcal{M}} |x - m_j|$. In this paper, we will assume that $Q_{\mathcal{M}}$ is uniform, $m_j = j\triangle$, $j \in \mathbb{Z}$, where $\triangle > 0$ is the bin width. A uniform quantizer with bin width $\triangle$ will be denoted $Q_\triangle$.

2. PRECOVER AND COVER SOURCE

An n-element precover source will be represented using a random variable $Z \triangleq (Z_1, \ldots, Z_n)$ divided into $S \geq 1$ segments containing pixels with indices from $\mathcal{N}_i \triangleq \{n_{i-1}+1, \ldots, n_i\}$, $i = 1, \ldots, S$, where $n_0 = 0$, $n_i < n_{i+1}$, $n_S = n$ are segments' boundaries. On each segment, the content is modeled using a polynomial of degree d corrupted by a AWGN:

$$Z^{(i)} = \mathbf{H}\theta^{(i)} + \Xi^{(i)}, \tag{1}$$

where $Z^{(i)} \triangleq (Z_{n_{i-1}+1}, \ldots, Z_{n_i})^{\mathrm{T}}$, $\theta^{(i)} \triangleq (\theta_0^{(i)}, \ldots, \theta_d^{(i)})^{\mathrm{T}}$, $\Xi^{(i)} \sim N(0, \sigma^2 \mathbf{I}_{|\mathcal{N}_i|})$, and

$$\mathbf{H} = \begin{pmatrix} 1 & 1 & \cdots & 1 \\ 1 & 2 & \cdots & 2^d \\ \cdots & \cdots & \cdots & \cdots \\ 1 & |\mathcal{N}_i| & \cdots & |\mathcal{N}_i|^d \end{pmatrix}. \tag{2}$$

This is a non-stationary model with content that is smooth, spatially-correlated on each segment with discontinuities (edges) at segment boundaries. Equivalently, one can say that Z_k are mutually independent continuous-valued random variables $Z_k \sim N(\mu_k, \sigma^2)$ with

$$\mu_k \triangleq \sum_{l=0}^{d} \theta_l^{(i)} (k - n_{i-1})^l, \quad k \in \mathcal{N}_i. \tag{3}$$

The log-likelihood of observing $\mathbf{z}^{(i)} = (z_{n_{i-1}+1}, \ldots, z_{n_i})$ given $\theta^{(i)}$ is

$$\log L(\mathbf{z}^{(i)}|\theta^{(i)}) = -\frac{|\mathcal{N}_i|}{2}\log(2\pi\sigma^2) - \frac{1}{2\sigma^2}\sum_{k\in\mathcal{N}_i}(z_k - \mu_k)^2. \tag{4}$$

The varying means represent the content, while Ξ stands for various acquisition noise sources. The assumption that the variance σ^2 is constant across the cover is also reasonable as sensor noise is often modeled in this

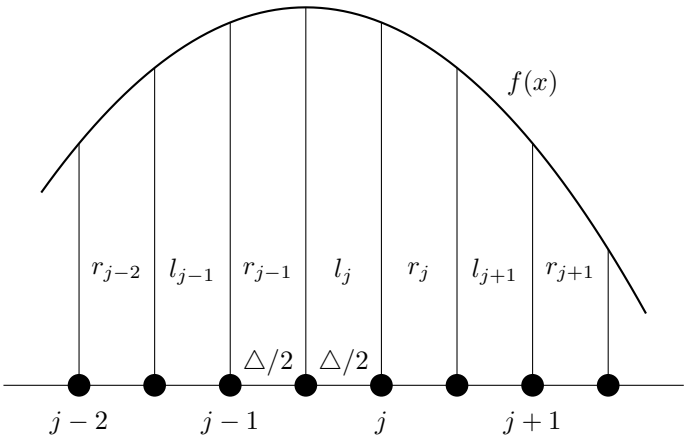

Figure 1. Precover distribution $f(x)$ and the quantization intervals. Note that the p.m.f. of the quantized cover is $p_j = l_j + r_j$.

manner.[11–13] Having said this, some components of this noise depend on the light intensity (shot noise, photo-response non-uniformity, and fixed-pattern noise), but these are typically either suppressed in the final image or small w.r.t. the dominant noise component formed by an approximately spatially uniform readout and electronic noise, which are Gaussian in nature.

For better readability, from now on the symbols i, j, k, and l will be exclusively reserved to index segments, quantization bins, individual pixels, and parameters, respectively. Additionally, for any vector $\mathbf{v} \in \mathbb{R}^n$, $\mathbf{v}^{(i)} \triangleq (v_{n_{i-1}+1}, \ldots, v_{n_i})^{\mathrm{T}}$.

A cover source X corresponding to a precover source Z is obtained by applying $Q_\triangle$ to each element of Z, $X = (X_1, \ldots, X_n) = (Q_\triangle(Z_1), \ldots, Q_\triangle(Z_n))$ with $X_k \sim p^{(k)}$ a probability mass function (p.m.f.) on $\mathcal{M}$:

$$p_j^{(k)} \triangleq \Pr(X_k = m_j) = \int_{(m_j-1/2)\triangle}^{(m_j+1/2)\triangle} f_k(x)\mathrm{d}x, \tag{5}$$

where $f_k(x) = (2\pi\sigma^2)^{-1/2} \exp\left(-(x-\mu_k)^2/(2\sigma^2)\right)$ is the Gaussian density for $N(\mu_k, \sigma^2)$. From Figure 1, $p_j^{(k)} = l_j^{(k)} + r_j^{(k)}$, where

$$r_j^{(k)} \triangleq \int_{j\triangle}^{(j+1/2)\triangle} f_k(x)\mathrm{d}x, \quad l_j^{(k)} \triangleq \int_{(j-1/2)\triangle}^{j\triangle} f_k(x)\mathrm{d}x. \tag{6}$$

3. ASSUMPTIONS

In this section, we lay out the basic assumptions of our study, including the sender's ignorance, properties of embedding schemes, and a measure of security.

3.1 Sender's ignorance

Assuming that the noise properties of images do not change over time and across images, the sender can estimate the noise variance, σ^2, with an arbitrarily high accuracy simply by taking many content-less pictures, such as blue sky shots.*

*Carrying this task in practice may be quite elaborate, but not impossible, as, e.g., the power of the dark current noise depends on the exposure and ambient temperature.

We also assume that the sender can segment the image and thus has access to n_i. The means, μ_k, on each segment, however, are scene dependent and must be estimated as they are corrupted by noise. The best the sender can do is to estimate μ_k from the corresponding segment of the precover.

Due to the cover-model linearity w.r.t. θ, the maximum-likelihood (ML) estimator of θ is minimum-variance unbiased (MVU):

$$\hat{\theta}^{(i)} = (\mathbf{H}^{\mathrm{T}}\mathbf{H})^{-1}\mathbf{H}^{\mathrm{T}}Z^{(i)} \sim N\left(\theta^{(i)}, \frac{1}{|\mathcal{N}_i|}\mathbf{I}^{-1}(\theta^{(i)})\right), \tag{7}$$

where $\mathbf{I}(\theta^{(i)}) \in \mathbb{R}^{(d+1)\times(d+1)}$ is the Fisher information matrix, for which from (3) and (4):

$$\mathbf{I}_{rs}(\theta^{(i)}) = -E\left[\frac{\partial^2 \log L(\mathbf{y}^{(i)}|\theta^{(i)})}{\partial\theta_r^{(i)}\partial\theta_s^{(i)}}\right] = \frac{1}{\sigma^2}\sum_{k\in\mathcal{N}_i}(k-n_{i-1})^{r+s} = \frac{1}{\sigma^2}(\mathbf{H}^{\mathrm{T}}\mathbf{H})_{rs}. \tag{8}$$

With sender's estimate $\hat{\theta}^{(i)} = (\hat{\theta}_0^{(i)}, \ldots, \hat{\theta}_d^{(i)})^{\mathrm{T}}$, $k \in \mathcal{N}_i$,

$$\hat{\mu}_k = \sum_{l=0}^{d}\hat{\theta}_l^{(i)}(k-n_{i-1})^l. \tag{9}$$

3.2 Embedding schemes

A fair comparison of different embedding schemes requires fixing the payload. However, the relationship between the payload and the number, character, and placement of embedding changes will generally depend on the details of the embedding, including the coding scheme, assignment of pixel costs, content-adaptivity criteria, etc.[†]

To obtain a more general result, in this paper we compare three different ways the sender can *execute embedding changes* once the pixels that are to be modified have already been determined by the embedding algorithm for a *given payload*. To further simplify the analysis, we will assume that for a given number of modified pixels (or, for a given change rate β), the modified pixels are selected pseudo-randomly and independently of the cover element values. Thus, given a change rate β, on average $|\mathcal{N}_i|\beta$ randomly selected pixels will be modified in the ith segment. Apparently, this assumption is not valid for content-adaptive embedding schemes and for JPEG steganography that avoids modifying DCT coefficients equal to zero. Section 4.4.1 contains more discussion of this issue.

The above assumptions essentially allow us to compare security by evaluating the Fisher information (11) w.r.t. the change rate β. It is true that even for the simplest case when the sender minimizes the number of embedding changes, the same change rate allows embedding different payloads (e.g., LSBM may use ternary codes while some other schemes, such the QIM below, are limited to binary codes). However, in the asymptotic limit of $\beta \to 0$, the payloads are the same for a rather wide class of embedding methods. This is because by replacing the pixel value, x, with $x + W$, where W follows some integer-valued distribution, one can embed $h_W(\beta) = h_2(\beta) + \beta H(W|W \neq 0)$ bits per pixel. However, because $h_W(\beta) = -\beta(1 + o(\beta))\log_2\beta$, for small change rates the payloads are asymptotically identical independently of W.

3.3 Measuring security

It is common in security studies to grant the opponent the absolute knowledge (the worst case scenario). Thus, we will assume that the Warden knows μ_k, n_i, and σ exactly.[‡] The Warden is also assumed to be passive.

[†]For example, in Ref. [19], the cover elements that are to be modified, $x_k \neq y_k$, are determined using BCH codes on small blocks to minimize the total embedding distortion $D(\mathbf{x}, \mathbf{y}; \rho) = \sum_{k=1}^{n}\rho_k(x_k)\cdot[x_k \neq y_k]$, where $\rho_k(x_k)$ is the embedding cost at the kth DCT coefficient determined from the rounding error and the value x_k. In contrast, the authors of Ref. [21] used syndrome-trellis codes[6] with embedding costs that consider both the rounding error and the entropy of the DCT block to which the coefficient belongs.

[‡]A perhaps more realistic model of Warden's ignorance introduced in Ref. [16] and one we do not use in this paper is based on granting the Warden with a limited access to a cover oracle.

Under these assumptions, it makes sense to measure the security using the KL divergence between the quantized cover and stego distributions as in the framework outlined by Cachin.[2] To further simplify the matters, we resort to asymptotic analysis of small change rates β (payloads) and use the steganographic Fisher information,[4,5,15] which appears in the leading term of the Taylor expansion of the KL divergence between the quantized cover $p^{(k)}$ and stego $q^{(k)}(\beta)$ distributions at $\beta = 0$:

$$D_{\mathrm{KL}}(p^{(k)}||q^{(k)}) = \sum_j p_j^{(k)} \log \frac{p_j^{(k)}}{q_j^{(k)}(\beta)} = \frac{1}{2}\beta^2 I_k(0) + O(\beta^3), \tag{10}$$

where,

$$I_k(0) = \sum_j \frac{1}{p_j^{(k)}} \left(\frac{\mathrm{d}q_j^{(k)}(\beta)}{\mathrm{d}\beta}\Big|_{\beta=0} \right)^2 \tag{11}$$

is the steganographic Fisher information for the kth pixel.

4. ANALYZING THREE EMBEDDING STRATEGIES

We investigate three different strategies for modifying the pixels – two utilize the side-information at the sender but each in a different manner, while the third is the LSBM included as an example of the most common and simplest embedding paradigm studied in steganography today. The first side-informed embedding follows the paradigm of model-based steganography originally introduced in Ref. [20]. Here, the sender estimates the precover model from one instance of the cover (segment) and then modifies the pixels to preserve (force) the quantized model. The second strategy is a simple QIM where the sender quantizes the precover to two interleaved sublattices to minimize the embedding distortion. The ± 1 changes in LSBM are made by flipping an unbiased coin independently of the precover value or the cover model.

4.1 Embedding while preserving estimated model

The sender embeds the secret while preserving the estimated cover distribution, which is the Gaussian $N(\hat{\mu}_k, \sigma^2)$ quantized using $Q_\triangle$. Irrespectively of exactly how the embedding is executed, the pixels from the ith segment of the stego object, $k \in \mathcal{N}_i$, will be realizations of the following random variable:

$$Y_k = Q_\triangle(Z'_k), \tag{12}$$

where

$$Z'_k \sim (1-\beta)N(\mu_k, \sigma^2) + \beta N(\hat{\mu}_k, \sigma^2) \tag{13}$$

is the stego-image mixture (this follows from the random changes assumption). The cover image pixels, on the other hand, $X_k = Q_\triangle(Z_k)$.

By (10) and (11), the KL divergence between $X^{(i)} = (X_{n_{i-1}+1}, \ldots, X_{n_i})$ and $Y^{(i)} = (Y_{n_{i-1}+1}, \ldots, Y_{n_i})$ is

$$D_{\mathrm{KL}}(X^{(i)}||Y^{(i)}) = \sum_{k\in\mathcal{N}_i} D_{\mathrm{KL}}(X_k||Y_k) \approx \frac{1}{2}\beta^2 \sum_{k\in\mathcal{N}_i} I_k(0) \overset{(a)}{=} \frac{1}{2}\beta^2 \sum_{k\in\mathcal{N}_i} \frac{(\mu_k - \hat{\mu}_k)^2}{\sigma^2} + O(\triangle). \tag{14}$$

(Equality (a) is established in Appendix A.1.) Thus, we have the following leading term for the expected value of the KL divergence on the ith segment:

$$E[D_{\mathrm{KL}}(X^{(i)}||Y^{(i)})] \approx \frac{\beta^2}{2\sigma^2} \sum_{k\in\mathcal{N}_i} E[(\mu_k - \hat{\mu}_k)^2] = \frac{\beta^2}{2\sigma^2} \sum_{k\in\mathcal{N}_i} E\left[\left(\sum_{l=0}^{d} (\theta_l^{(i)} - \hat{\theta}_l^{(i)})(k - n_{i-1})^l \right)^2 \right] \tag{15}$$

$$= \frac{\beta^2}{2\sigma^2} \sum_{s=0}^{d} \sum_{r=0}^{d} \frac{1}{|\mathcal{N}_i|} \mathbf{I}_{rs}^{-1}(\theta^{(i)}) \mathbf{I}_{rs}(\theta^{(i)}) \sigma^2 = \frac{\beta^2}{2}, \tag{16}$$

which gives

$$E[D_{\mathrm{KL}}(X||Y)] = \sum_{i=1}^{S} E[D_{\mathrm{KL}}(X^{(i)}||Y^{(i)})] \approx \frac{S\beta^2}{2}. \tag{17}$$

Note that this result does not depend on $\triangle$, n_i, μ, or σ. Intuitively, it makes sense – the more pixels are in the segment, the more accurately the sender can estimate the mean. At the same time, the KL divergence increases linearly with the signal length. These two effects cancel each other.

4.2 Quantization-index modulation

For the purpose of embedding using QIM (previously used in steganography in perturbed quantization[9]), the quantized cover values are assigned parities on two interleaved sublattices covering $\mathcal{M} = (m_j)$, for example as the LSBs of j. The sender embeds the message by changing the parity of some pixels by quantizing the precover value either "up" or "down," depending on the desired parity of the stego pixel y_k, to minimize the distortion between the stego image and the precover.

Under the random modifications assumption, $Y_k \sim q^{(k)}$ with (follow Figure 1):

$$q_j^{(k)} = (1-\beta)p_j^{(k)} + \beta(l_{j+1}^{(k)} + r_{j-1}^{(k)}). \tag{18}$$

Substituting the Fisher information (11) for (18) computed in Appendix A.2 into (10)

$$D_{\mathrm{KL}}(p^{(k)}||q^{(k)}) = \frac{\beta^2\triangle^4}{16\sigma^4}, \tag{19}$$

gives us the final result:

$$D_{\mathrm{KL}}(p||q) = \sum_{i=1}^{S}\sum_{k\in\mathcal{N}_i} D_{\mathrm{KL}}(p^{(k)}||q^{(k)}) = \frac{n\beta^2\triangle^4}{16\sigma^4}. \tag{20}$$

4.3 Embedding using LSB matching

LSBM is a popular embedding scheme in which each cover element (e.g., pixel or DCT coefficient) is changed by $\{-1, 0, 1\}$ to embed a message. The random modifications assumption implies that the impact of LSBM on the first-order statistic of quantized signals is

$$q_j^{(k)} = (1-\beta)p_j^{(k)} + \frac{1}{2}\beta(p_{j+1}^{(k)} + p_{j-1}^{(k)}). \tag{21}$$

By evaluating the Fisher information (see Appendix A.2), we obtain:

$$D_{\mathrm{KL}}(p||q) = \frac{n\beta^2\triangle^4}{4\sigma^4}. \tag{22}$$

4.4 Discussion

First, the comparison of (20) with (22) informs us that QIM has a four-times smaller Fisher information than LSB matching, indicating that it provides better security (one can increase the change rate twice at the same statistical detectability). This is to be expected as QIM distorts the first-order statistic of the cover less than LSBM.

By inspecting (16) and (20), we conclude that QIM is more secure than model preservation when

$$\frac{n\beta^2\triangle^4}{16\sigma^4} < \frac{S\beta^2}{2}, \tag{23}$$

which is equivalent to

$$\overline{n} < \frac{8\sigma^4}{\triangle^4}, \tag{24}$$

where $\overline{n} = n/S$ is the average number of pixels in a segment. This condition is essentially an upper bound on the average segment size or a lower bound on the content complexity (the number of segments). When the image content is sufficiently complex, the sender is unable to estimate the content accurately enough and is better off using the heuristic QIM instead of approximate model preservation.

The bound (24) increases with finer quantization and stronger corrupting noise. This makes sense as stronger noise prevents cover estimation that is accurate enough. Also, while finer quantization decreases the Fisher information for the QIM, it plays no role in model-based steganography.

The QIM is generally an attractive choice for embedding in empirical covers because it is "model-free." Preserving an approximate model is potentially dangerous unless the model is a truthful representation of reality. Forcing the distribution of stego images to follow an approximate model further increases the KL divergence and may also create an opportunity for the Warden to detect the embedding. A good example is the attack on model-based steganography described in.[24] Here, a symmetric model was used for the histogram of DCT coefficients. The lack of deviations from perfect symmetry inherently present in histograms of natural images enabled mounting a successful attack.

4.4.1 Extension to content-adaptive embedding

The analysis of this paper can be extended with some effort to the case when the embedding changes are not uniform across the image. We now state a generalization of the result obtained above for content-adaptive binary embedding schemes that minimize an additive distortion $D(\mathbf{x},\mathbf{y}) = \sum_k \rho_k[x_k \neq y_k]$ with pixel costs $\rho_k > 0$.

As each segment shares one model, it is reasonable to assume that the pixel costs are the same on each segment, $\rho_k = \varrho_i$ for all $k \in \mathcal{N}_i$, $i = 1,\dots,S$. To embed relative payload $0 < \alpha \leq 1$, the embedding changes each pixel $k \in \mathcal{N}_i$ with probability $\beta_i = \frac{e^{-\lambda\rho_i}}{1+e^{-\lambda\rho_i}}$, where $\lambda \geq 0$ satisfies the payload constraint (see, e.g., Chapter 7 in Ref. [8]):

$$\alpha(\lambda) = \frac{1}{n}\sum_{i=1}^{S} |\mathcal{N}_i| h_2\left(\frac{e^{-\lambda\rho_i}}{1+e^{-\lambda\rho_i}}\right). \tag{25}$$

The Fisher information now needs to be expressed w.r.t. α as the concept of the change rate is no longer meaningful. Without providing the full details, it can be shown that in the asymptotic limit of $\alpha \to 0$ (or $\lambda \to \infty$), the FI computed w.r.t. α is determined only by the pixels with the smallest cost, ϱ_{i_0}, $i_0 = \arg\min_i \varrho_i$, giving the following result for the KL divergence between the cover object X and the steganographic mixture Y:

$$D_{\mathrm{KL}}(X||Y) = \frac{1}{2}\left(\frac{\alpha}{\ln\alpha}\right)^2 \frac{n^2}{|\mathcal{N}_{i_0}|^2}\sum_{k\in\mathcal{N}_{i_0}} J_k(0) + O\left((\alpha/\ln\alpha)^3\right), \quad J_k(0) = \sum_j \frac{1}{p_j^{(k)}}\left(\frac{\mathrm{d}q_j^{(k)}(\beta_{i_0})}{\mathrm{d}\beta_{i_0}}\bigg|_{\beta_{i_0}=0}\right)^2. \tag{26}$$

This is natural as for infinitesimaly small α, the only embedding changes will occur in the most textured segment with the smallest ϱ. Note that the scaling is quadratic in terms of $\alpha/\ln\alpha$ rather than α^2. The FI can be computed in the same manner as above and gives the following equivalent of Equation (24):

$$|\mathcal{N}_{i_0}| < \frac{8\sigma^4}{\triangle^4}, \tag{27}$$

which is in agreement with the result obtained for the uniform embedding – instead of a condition on the *average* size of a segment, we get a condition on the size of the *most textured* segment, which is the only segment playing a role in the asymptotics.

5. CONCLUSION

Practitioners of steganography have observed long time ago that empirical steganographic security can be markedly improved when the sender makes use of a higher-resolution representation of covers (the so-called precover) when embedding a secret message. Fundamentally, the side information compensates for the lack of knowledge of the cover model. In this paper, we show that if the cover is sufficiently non-stationary the sender is better off embedding by minimizing the distortion to the precover instead of estimating the non-stationary model and preserving it.

This work should be considered as an initial step in the quest to better understand the role the side information plays in practice. The results were derived under the assumption of i.i.d. Gaussian covers, which is not valid for the JPEG domain. The asymptotic analysis of infinitesimal payloads also makes analysis of content-adaptive embedding singular as only the most textured segment plays any role.

More importantly, however, one should challenge the assumptions made about sender's and Warden's ignorance. Granting full knowledge of the cover to the Warden may seem overly pessimistic given that it is the sender who has side information, not the Warden. Moreover, in practice the Warden, just like the sender, must estimate her model. In fact, it is an open question how to use the side information in the best possible manner in these cases and it is not even clear whether the side information in fact gives any advantage to the sender.

6. ACKNOWLEDGEMENTS

The work on this paper was supported by Air Force Office of Scientific Research under the research grant number FA9950-12-1-0124. The U.S. Government is authorized to reproduce and distribute reprints for Governmental purposes notwithstanding any copyright notation there on. The views and conclusions contained herein are those of the authors and should not be interpreted as necessarily representing the official policies, either expressed or implied of AFOSR or the U.S. Government.

APPENDIX A. TECHNICAL ARGUMENTS

Here, we provide outlines of proofs of some of the statements used in the paper.

A.1 Fisher information for MBS

Let $f(x)$ be a probability density and $\triangle > 0$ a quantization step. For $\mu \in \mathbb{R}$, we define the quantized p.m.f.

$$p_j(\mu) = \int_{(j-1/2)\triangle}^{(j+1/2)\triangle} f(x-\mu)\mathrm{d}x \triangleq F_\triangle(j\triangle - \mu). \tag{28}$$

Note that $F_\triangle(j\triangle) = \triangle f(u_j)$ for some $u_j \in \mathcal{I}_j$, $\mathcal{I}_j \triangleq (j\triangle - \frac{\triangle}{2}, j\triangle + \frac{\triangle}{2})$ by the first mean-value theorem of integration. For the mixture $q(\beta) = (1-\beta)p(0) + \beta p(\mu)$, $\beta > 0$, the FI (11) becomes:

$$\begin{aligned} I(0) &= \sum_j \frac{1}{q_j(0)} \left(\frac{\mathrm{d}q_j(\beta)}{\mathrm{d}\beta}\Big|_{\beta=0} \right)^2 = \sum_j \frac{\big(p_j(\mu) - p_j(0)\big)^2}{p_j(0)} \\ &= \sum_j \frac{\big(F_\triangle(j\triangle - \mu) - F_\triangle(j\triangle)\big)^2}{F_\triangle(j\triangle)} = \mu^2 \sum_j \frac{\big(F'_\triangle(j\triangle)\big)^2}{F_\triangle(j\triangle)} + O(\mu^3) \\ &\approx \mu^2 \sum_j \frac{\big(f(j\triangle + \triangle/2) - f(j\triangle - \triangle/2)\big)^2}{\triangle f(u_j)} = \mu^2 \sum_j \frac{\big(f'(\phi_j)\big)^2}{f(u_j)}\triangle \end{aligned}$$

for $\phi_j \in \mathcal{I}_j$ determined again by the first mean-value theorem for integration. The sum $\sum_j \triangle\big(f'(\phi_j)\big)^2/f(u_j)$ can be approximated by $\int_{\mathbb{R}}\big(f'(x)\big)^2/f(x)$ (up to a term proportional to $\triangle$). For a Gaussian density f, this integral is equal to σ^{-2}, which gives us $I_k(0) \approx \mu^2/\sigma^2$.

This is only an outline of the proof as the approximation of the sum by an integral needs a more precise argument since the sum is not technically a Riemann sum as the arguments of the functions in the numerator and denominator are different.

A.2 Fisher information for QIM and LSBM

We first introduce

$$F^-_\triangle(x) \triangleq \int_{x-\triangle/2}^{x} f(t)\mathrm{d}t, \quad F^+_\triangle(x) \triangleq \int_{x}^{x+\triangle/2} f(t)\mathrm{d}t,$$

and write using Taylor expansion of F^- and F^+ at $x = j\triangle$:

$$l_{j+1} = F^-_\triangle(j\triangle+\triangle) = \sum_{m=0}^{\infty} \frac{F^{-(m)}_\triangle(j\triangle)}{m!}\triangle^m, \quad r_{j-1} = F^+_\triangle(j\triangle-\triangle) = \sum_{m=0}^{\infty} \frac{F^{-(m)}_\triangle(j\triangle)}{m!}(-1)^m\triangle^m,$$

$$p_j = F^-_\triangle(j\triangle) + F^+_\triangle(j\triangle).$$

From (18), after straightforward algebra:

$$\left.\frac{\partial q_j}{\partial\beta}\right|_{\beta=0} = -p_j + l_{j+1} + r_{j-1} = \frac{\triangle^3}{4}f''(j\triangle) + O(\triangle^4).$$

Finally, the Fisher information

$$I(0) = \sum_j \frac{1}{p_j}\left(\left.\frac{\partial q_j}{\partial\beta}\right|_{\beta=0}\right)^2 \approx \sum_j \frac{1}{\triangle f(u_j)}\frac{\triangle^6}{16}\big(f''(j\triangle)\big)^2 \approx \frac{\triangle^4}{8\sigma^4}$$

by approximating the above "Riemann sum" with the integral $\int_{\mathbb{R}}\big(f''(x)\big)^2/f(x)\mathrm{d}x = 2/\sigma^4$.

The same analysis can be carried out for LSBM. By substituting

$$p_{j\pm1} = F_\triangle(j\triangle\pm\triangle) = \sum_{m=0}^{\infty}\frac{F^{-(m)}_\triangle(j\triangle)}{m!}(\pm\triangle)^m,$$

into the expression for q_j (21) and simplifying:

$$\left.\frac{\partial q_j}{\partial\beta}\right|_{\beta=0} = -p_j + (p_{j+1}+p_{j-1})/2 = \frac{\triangle^3}{2}f''(j\triangle) + O(\triangle^4),$$

which finally leads to

$$I(0) = \sum_j \frac{1}{p_j}\left(\left.\frac{\partial q_j}{\partial\beta}\right|_{\beta=0}\right)^2 = \sum_j \frac{1}{\triangle f(u_j)}\frac{\triangle^6}{4}\big(f''(j\triangle)\big)^2 \approx \frac{\triangle^4}{2\sigma^4}.$$

REFERENCES

1. R. Böhme. *Advanced Statistical Steganalysis*. Springer-Verlag, Berlin Heidelberg, 2010.
2. C. Cachin. An information-theoretic model for steganography. *Information and Computation*, 192(1):41–56, July 2004.

3. R. Cogranne, C. Zitzmann, L. Fillantre, F. Retraint, I. Nikiforov, and P. Cornu. A cover image model for reliable steganalysis. In T. Filler, T. Pevný, A. Ker, and S. Craver, editors, *Information Hiding, 13th International Conference*, Lecture Notes in Computer Science, pages 178–192, Prague, Czech Republic, May 18–20, 2011.
4. T. Filler and J. Fridrich. Complete characterization of perfectly secure stegosystems with mutually independent embedding. In *Proceedings IEEE, International Conference on Acoustics, Speech, and Signal Processing*, Taipei, Taiwan, April 19–24, 2009.
5. T. Filler and J. Fridrich. Fisher information determines capacity of ϵ-secure steganography. In S. Katzenbeisser and A.-R. Sadeghi, editors, *Information Hiding, 11th International Conference*, volume 5806 of Lecture Notes in Computer Science, pages 31–47, Darmstadt, Germany, June 7–10, 2009. Springer-Verlag, New York.
6. T. Filler, J. Judas, and J. Fridrich. Minimizing additive distortion in steganography using syndrome-trellis codes. *IEEE Transactions on Information Forensics and Security*, 6(3):920–935, September 2011.
7. T. Filler, A. D. Ker, and J. Fridrich. The Square Root Law of steganographic capacity for Markov covers. In N. D. Memon, E. J. Delp, P. W. Wong, and J. Dittmann, editors, *Proceedings SPIE, Electronic Imaging, Security and Forensics of Multimedia XI*, volume 7254, pages 08 1–11, San Jose, CA, January 18–21, 2009.
8. J. Fridrich. *Steganography in Digital Media: Principles, Algorithms, and Applications*. Cambridge University Press, 2009.
9. J. Fridrich, M. Goljan, and D. Soukal. Perturbed quantization steganography. *ACM Multimedia System Journal*, 11(2):98–107, 2005.
10. J. Fridrich, T. Pevný, and J. Kodovský. Statistically undetectable JPEG steganography: Dead ends, challenges, and opportunities. In J. Dittmann and J. Fridrich, editors, *Proceedings of the 9th ACM Multimedia & Security Workshop*, pages 3–14, Dallas, TX, September 20–21, 2007.
11. G. E. Healey and R. Kondepudy. Radiometric CCD camera calibration and noise estimation. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 16(3):267–276, March 1994.
12. G.C. Holst. *CCD Arrays, Cameras, and Displays*. JCD Publishing & SPIE Press, 2nd edition, 1998.
13. J. R. Janesick. *Scientific Charge-Coupled Devices*, volume Monograph PM83. Washington, DC: SPIE Press - The International Society for Optical Engineering, January 2001.
14. A. D. Ker. A capacity result for batch steganography. *IEEE Signal Processing Letters*, 14(8):525–528, 2007.
15. A. D. Ker. Estimating steganographic Fisher information in real images. In S. Katzenbeisser and A.-R. Sadeghi, editors, *Information Hiding, 11th International Conference*, volume 5806 of Lecture Notes in Computer Science, pages 73–88, Darmstadt, Germany, June 7–10, 2009. Springer-Verlag, New York.
16. A. D. Ker. The square root law in stegosystems with imperfect information. In R. Böhme and R. Safavi-Naini, editors, *Information Hiding, 12th International Conference*, volume 6387 of Lecture Notes in Computer Science, pages 145–160, Calgary, Canada, June 28–30, 2010. Springer-Verlag, New York.
17. A. D. Ker, T. Pevný, J. Kodovský, and J. Fridrich. The Square Root Law of steganographic capacity. In A. D. Ker, J. Dittmann, and J. Fridrich, editors, *Proceedings of the 10th ACM Multimedia & Security Workshop*, pages 107–116, Oxford, UK, September 22–23, 2008.
18. J. Kodovský and J. Fridrich. Steganalysis of JPEG images using rich models. In A. Alattar, N. D. Memon, and E. J. Delp, editors, *Proceedings SPIE, Electronic Imaging, Media Watermarking, Security, and Forensics of Multimedia XIV*, volume 8303, pages 0A 1–13, San Francisco, CA, January 23–26, 2012.
19. V. Sachnev, H. J. Kim, and R. Zhang. Less detectable JPEG steganography method based on heuristic optimization and BCH syndrome coding. In J. Dittmann, S. Craver, and J. Fridrich, editors, *Proceedings of the 11th ACM Multimedia & Security Workshop*, pages 131–140, Princeton, NJ, September 7–8, 2009.
20. P. Sallee. Model-based methods for steganography and steganalysis. *International Journal of Image Graphics*, 5(1):167–190, 2005.
21. C. Wang and J. Ni. An efficient JPEG steganographic scheme based on the block–entropy of DCT coefficents. In *Proc. of IEEE ICASSP*, Kyoto, Japan, March 25–30, 2012.
22. Y. Wang and P. Moulin. Steganalysis of block-structured stegotext. In E. J. Delp and P. W. Wong, editors, *Proceedings SPIE, Electronic Imaging, Security, Steganography, and Watermarking of Multimedia Contents VI*, volume 5306, pages 477–488, San Jose, CA, January 19–22, 2004.

23. Y. Wang and P. Moulin. Perfectly secure steganography: Capacity, error exponents, and code constructions. *IEEE Transactions on Information Theory, Special Issue on Security*, 55(6):2706–2722, June 2008.
24. A. Westfeld and R. Böhme. Exploiting preserved statistics for steganalysis. In J. Fridrich, editor, *Information Hiding, 6th International Workshop*, volume 3200 of Lecture Notes in Computer Science, pages 82–96, Toronto, Canada, May 23–25, 2004. Springer-Verlag, Berlin.