# Effect of Imprecise Knowledge of the Selection Channel on Steganalysis

Vahid Sedighi
Binghamton University
Department of ECE
Binghamton, NY 13902-6000
vsedighi1@binghamton.edu

Jessica Fridrich
Binghamton University
Department of ECE
Binghamton, NY 13902-6000
fridrich@binghamton.edu

## ABSTRACT

It has recently been shown that steganalysis of content-adaptive steganography can be improved when the Warden incorporates in her detector the knowledge of the selection channel – the probabilities with which the individual cover elements were modified during embedding. Such attacks implicitly assume that the Warden knows at least approximately the payload size. In this paper, we study the loss of detection accuracy when the Warden uses a selection channel that was imprecisely determined either due to lack of information or the stego changes themselves. The loss is investigated for two types of qualitatively different detectors – binary classifiers equipped with selection-channel-aware rich models and optimal detectors derived using the theory of hypothesis testing from a cover model. Two different embedding paradigms are addressed – steganography based on minimizing distortion and embedding that minimizes the detectability of an optimal detector within a chosen cover model. Remarkably, the experimental and theoretical evidence are qualitatively in agreement across different embedding methods, and both point out that inaccuracies in the selection channel do not have a strong effect on steganalysis detection errors. It pays off to use imprecise selection channel rather than none. Our findings validate the use of selection-channel-aware detectors in practice.

## Categories and Subject Descriptors

I.4.9 [**Computing Methodologies**]: Image Processing and Computer Vision—*Applications*

## General Terms

Security, Algorithms, Theory

## Keywords

Steganography, steganalysis, adaptive, selection channel

## 1. INTRODUCTION

Steganography in digital images has seen great advances in the recent years. The main paradigm shift occurred in 2010 with the introduction of near-optimal codes [12] that allowed the sender to assign "costs" of changing individual image elements (e.g., pixels) and then embed the secret message while minimizing the sum of costs of all modified pixels. By assigning large costs to pixels in smooth regions and low costs in highly textured content, the embedding is forced to execute the modifications where they would be presumably harder to detect. The first method based on this framework was HUGO [36]. Soon, many other content-adaptive schemes with increasingly improved security operating in both the spatial [22, 25, 33, 32] and JPEG domain [25, 19] appeared. Recently, steganalysts began investigating the possibility of using an approximate knowledge of the embedding change probabilities to better detect adaptive embedding. Indeed, since the pixel costs are driven by content, they can be usually accurately estimated from the stego image because the embedding changes themselves are rather subtle.

The first fundamental insight was given by Schöttle et al. [39] who showed on a simple example that it is advantageous for the sender to deviate from her optimal embedding strategy in exchange for a mismatched detector of the Warden. Framing the interaction between the sender and the Warden within the game theory, the authors showed that the Nash equilibrium, attained in mixed strategies, was an overall better choice for the sender than minimizing the KL divergence between cover and stego objects. The same authors showed in [40] that it is possible to use the knowledge of the embedding change probabilities in naive LSB replacement to improve the weighted-stego detector [27]. Experimental evidence was also presented that embedding schemes whose selection channel is more sensitive to the embedding changes themselves are harder to attack than schemes with a more robust selection channel.

In [8], it was shown that the security of S-UNIWARD [23] was compromised due to a faulty selection channel in which pixels with high and low embedding change probabilities were tightly interleaved. The problem was tied to an improperly selected parameter whose role was to merely stabilize the numerical computation, and it disappeared after adjusting this parameter to produce a selection channel free of artifacts [25].

The first general-purpose attack on content-adaptive steganography appeared in [43] but was, for some reason, presented as an attack specific to WOW [22]. The authors

proposed to form the 4D co-occurrences of quantized noise residuals in the Spatial Rich Model [15] (SRM) only from a fraction of pixels with the lowest costs. This attack, which was later nicknamed "thresholded SRM" (tSRM) in [9] was further improved by forming the co-occurrences from all pixels but letting each pixel contribute to a specific co-occurrence bin only with the maximum of the four probabilities corresponding to pixels whose residuals point to the specific co-occurrence bin (the so-called maxSRM [9]).

All steganalysis attacks that use the selection channel inherently assume that the Warden is able to estimate the embedding change probabilities, which usually strongly depend on the payload size. Estimating the payload size, however, is rather difficult especially for modern embedding schemes whose detection requires high-dimensional rich media models, which substantially complicates the payload estimator construction in practice (e.g., the search for support vector regressor hyperparameters [37]). In fact, it seems hard to substantially improve upon the trivial estimator that always outputs the mean payload [28].

It is thus important to study the effect of payload mismatch on the accuracy of detection when selection-channel-aware detectors are used. The above cited prior art [43, 9] already contains a limited experimental investigation of this issue. In particular, classifiers equipped with the maxSRM feature set appear to lose less detection power due to payload mismatch than the tSRM features.

In this paper, we work with two qualitatively different steganography detectors – classifiers built using machine learning and tests designed in an optimal manner from a cover model. We do so for the widely popular embedding by minimizing distortion and the newly emerged model-based steganography minimizing the power of the most powerful detector within a chosen cover model [41]. Interestingly, both detectors seem to point to the same evidence. First, the loss of detection power due to imprecise knowledge of the selection channel is rather small – it still pays off to use selection-channel-aware detector even with an incorrect payload. Second, the misjudged payload is much less of an issue for detection of the model-based steganography than for minimal-distortion steganography. Finally, the inaccuracy due to estimating the embedding change probabilities from the stego image rather than the cover has a negligible effect for the tested stego schemes.

In the next section, two different steganography detectors that utilize the selection channel are reviewed. Then, in Section 3 we introduce four types of Warden to better show the effect of imprecision in the selection channel on detection accuracy. The paper continues in Section 4, where we describe four steganographic schemes used in our experiments. In Section 5, we explain the setup of our experiments and how the loss (gain) of the detection power will be measured and the results reported. All experiments and their interpretation appear in Section 6. A summary of the paper is given in Section 7.

# 2. STEGANALYSIS WITH THE KNOWLEDGE OF SELECTION CHANNEL

Currently, there exist two major trends in steganalysis of digital images – detectors derived in some sense as optimal using the theory of statistical hypothesis testing based on a cover model [11, 6, 7, 46, 4, 3, 44] and detectors constructed

as classifiers trained on examples of cover and stego images represented in a feature space [10, 34, 42, 48, 1, 18, 15, 2, 29, 24, 21, 17]. The former approach is usually rather successful when the embedding disturbs some statistics of images that can be well described using a model. This is the case of Jsteg, OutGuess [38], and F5 [45], as these algorithms introduce strong artifacts into the first-order statistics of DCT coefficients, and basically all algorithms that use LSB replacement [6, 47, 11, 44, 5]. The latter approach to detection based on trained classifiers is more successful in detecting modern steganographic algorithms that seem to better preserve the cover statistics. To detect such algorithms, one currently needs to form features as higher-order statistics of noise residuals extracted using a diverse bank of pixel predictors, the so-called rich media models [18, 15, 29, 2, 24, 21, 17].

Both types of detectors have been previously adapted for detection of content-adaptive embedding to utilize the prior knowledge of the embedding change probabilities. Since these detection paradigms will be used in this paper, below we explain how this is done. The exposition is kept short while referring the reader to the corresponding publications for more details.

## 2.1 Empirical detectors

While incorporating priors in the form of the probabilistic selection channel within statistical hypothesis testing is usually straightforward (see Section 2.2), it is much less clear how to utilize this information for empirical detectors. The first general-purpose feature set proposed to improve the detection of content-adaptive steganography was the thresholded SRM [43]. The authors discovered that the embedding algorithm WOW is so "overly adaptive" that it paid off to compute the co-occurrence matrices of the SRM only from a fraction $t \leq 1$ of pixels with the smallest embedding cost. This way, the authors avoided using the embedding change probabilities but, in return, had to adjust the threshold $t$ based on the payload size for each embedding method separately. A mismatch between the estimated (assumed) and the true payload size caused a non-negligible detection loss.

In [9], the authors described an idea similar to the tSRM called the maxSRM. Due to space limitations, we only provide a very brief outline of the main idea. Let us assume that $z_{ij}$, $i = 1, \ldots, N_1$, $j = 1, \ldots, N_2$, is a noise residual computed from a grayscale image $\mathbf{x} = (x_{ij})$, $x_{ij} \in \{0, \ldots, 255\}^{N_1 \times N_2}$, using one of the pixel predictors employed in the SRM, $\mathbf{z} = \mathbf{x} - \text{Pred}(\mathbf{x})$. For example, $\text{Pred}(x_{ij}) = (x_{i,j-1} + x_{i,j+1})/2$ estimates the pixel value from its two closest horizontal neighbors. The residual is subsequently quantized with a quantizer $Q_{T,q} : \mathbb{R} \to \mathcal{Q}_{T,q}$ with centroids $\mathcal{Q}_{T,q} = \{-Tq, -(T-1)q, \ldots, (T-1)q, Tq\}$, $r_{ij} = Q_{T,q}(z_{ij})$. In SRM, the features are formed as 4D co-occurrences

$$C_{d_0 d_1 d_2 d_3} = \sum_{i=1}^{N_1} \sum_{j=1}^{N_2} [r_{i,j+k} = d_k, \forall k \in \{0, 1, 2, 3\}], \quad (1)$$

where $[P]$ is the Iverson bracket equal to 1 when the statement $P$ is true and 0 when $P$ is false. In maxSRM, the Iverson bracket in (1) is simply replaced with $\max\{\beta_{ij}, \beta_{i,j+1}, \beta_{i,j+2}, \beta_{i,j+3}\}$, where $\beta_{ij}$ is the probability of changing pixel $x_{ij}$ during embedding computed from $\mathbf{x}$. Note that in order to compute $\beta_{ij}$, one needs to know the size of the embedded payload. On the other hand, in maxSRM there is no need

to search for any parameters (e.g., the threshold $t$) for each payload and embedding scheme. As long as the payload size is given, one can readily form the feature vector. In contrast to tSRM, the maxSRM's detection accuracy degrades more gracefully with a mismatch between the true embedded payload and the assumed payload. (This is apparent from the studies that appeared in [43, 9, 41].)

## 2.2 Model-based optimal detectors

The second type of detector is derived using the theory of statistical hypothesis testing based on a cover model. The detectors described in this section are derived for content-adaptive LSB matching, which is the prevailing paradigm for spatial-domain steganography today. We start by describing the cover model, the embedding operation, and the ensuing stego image model, and finish with a closed-form expression for the deflection coefficient that describes the performance of the asymptotic likelihood ratio test.

During acquisition using an imaging sensor, pixel values become corrupted by noise, which is well modeled as a field of independent Gaussians with spatially-varying variance [26, 14, 20]. Even though the subsequent processing typically applied to images inside a digital camera, such as demosaicking, filtering, color correction, and anti-aliasing, make the noise component quite complicated by introducing dependencies among adjacent pixels, in order to derive the detector in a closed form, we adopt the following simplified multiparametric statistical model. Since the pixels' expectation can be estimated, e.g., using local pixel predictors or denoising, after subtracting the estimated expectation from the pixel value, the resulting noise residual will be modeled as a sequence of independent quantized realizations of Gaussian random variables with zero mean $X_n \sim \mathcal{N}(0, \sigma_n^2)$, $n = 1, \ldots, N$, where $N = N_1 \times N_2$ is the total number of pixels. We note that, besides the acquisition noise, the variance $\sigma_n^2$ also contains the modeling error and will in general strongly depend on the local image content. Note that here we index the image pixels with a one-dimensional index $n$ instead of $ij$ as in the previous section since one can imagine the two-dimensional array $(x_{ij})$ to be unfolded, e.g., by columns. Due to the independence assumption, the exact ordering is unimportant in our study, and we will be switching between the representations back and forth hopefully without causing any misunderstanding on the reader's side.

Without loss of generality, we will assume that the quantization step is $\triangle = 1$. Assuming the fine quantization limit, $1 \ll \sigma_n$ for all $n$, the probability mass function (pmf) of $X_n$ is given by $\mathcal{P}_{\sigma_n} = (p_{\sigma_n}(k))_{k \in \mathbb{Z}}$ with

$$p_{\sigma_n}(k) = \mathbb{P}(x_n = k) \propto (2\pi\sigma_n^2)^{-1/2} \exp\left(-k^2/(2\sigma_n^2)\right). \quad (2)$$

For simplicity, note that we assume that the pixel levels are unbounded. Also, the fine quantization assumption may cease to hold in (nearly) saturated image regions, such as overexposed light sources.

Virtually all steganographic algorithms in spatial domain use LSB matching to execute the actual embedding. Formally, given the cover $\mathbf{x} = (x_1, \ldots, x_N)$, the stego object $\mathbf{y} = (y_1, \ldots, y_N)$ is obtained from $\mathbf{x}$ using the following random process:

$$\mathbb{P}(y_n = x_n + 1) = \mathbb{P}(y_n = x_n - 1) = \beta_n, \quad (3)$$
$$\mathbb{P}(y_n = x_n) = 1 - 2\beta_n,$$

with $0 \leq \beta_n \leq 1/3$ being the so-called change rates. The stego object is thus a sequence of independent mixtures of quantized Gaussians $(Y_1, \ldots, Y_N)$,

$$Y_n \sim \mathcal{Q}_{\sigma_n, \beta_n} = (q_{\sigma_n, \beta_n}(k))_{k \in \mathbb{Z}} \quad (4)$$

with

$$q_{\sigma_n, \beta_n}(k) = \mathbb{P}(Y_n = k) = (1 - 2\beta_n)p_{\sigma_n}(k)$$
$$+ \beta_n p_{\sigma_n}(k+1) + \beta_n p_{\sigma_n}(k-1), \quad (5)$$

Assuming Alice uses optimal codes for embedding, she can communicate up to $R$ nats per pixel

$$R(\boldsymbol{\beta}) = \frac{1}{N} \sum_{n=1}^{N} H(\beta_n), \quad (6)$$

where $H(x) = -2x \ln x - (1 - 2x) \ln(1 - 2x)$ is the ternary entropy function expressed in nats. To obtain the payload in bits per pixel (bpp), one needs to multiply $R$ by $(\ln 2)^{-1}$.

### 2.2.1 The most powerful detector

Since the Warden will never have a perfect knowledge of the change rates $\beta_n$ used by the sender, when building her detector she will use change rates $\boldsymbol{\gamma} = (\gamma_1, \ldots, \gamma_N)$ that might not coincide with $\boldsymbol{\beta} = (\beta_1, \ldots, \beta_N)$. Assuming that both Alice and the Warden use the same cover model and know the noise variances $\sigma_n^2$, for example by estimating them from the given image (see Section 5.2 for more details), the Warden faces the following simple hypothesis test for all $n$:

$$\begin{array}{ll} \mathcal{H}_0 : & x_n \sim \mathcal{P}_{\sigma_n} \\ \mathcal{H}_1 : & x_n \sim \mathcal{Q}_{\sigma_n, \gamma_n}. \end{array} \quad (7)$$

From the Neyman–Pearson Lemma [31], the most powerful test $\delta : \mathbb{Z}^N \to \{\mathcal{H}_0, \mathcal{H}_1\}$ that maximizes the detection power $\pi = \mathbb{P}(\delta(\mathbf{x}) = \mathcal{H}_1 | \mathcal{H}_1)$ for a prescribed false-alarm probability $\alpha = \mathbb{P}(\delta(\mathbf{x}) = \mathcal{H}_1 | \mathcal{H}_0)$ is the Likelihood Ratio Test (LRT), which can be expressed using the statistical independence of pixels as

$$\Lambda(\mathbf{x}, \boldsymbol{\sigma}) = \sum_{n=1}^{N} \Lambda_n = \sum_{n=1}^{N} \log\left(\frac{q_{\sigma_n, \gamma_n}(x_n)}{p_{\sigma_n}(x_n)}\right) \overset{\mathcal{H}_1}{\underset{\mathcal{H}_0}{\gtrless}} \tau. \quad (8)$$

Under the additional assumptions of a large number of pixels ($N \to \infty$), the Lindeberg's version of the Central Limit Theorem implies that[1]

$$\Lambda^\star(\mathbf{x}, \boldsymbol{\sigma}) = \frac{\sum_{n=1}^{N} \Lambda_n - E_{\mathcal{H}_0}[\Lambda_n]}{\sqrt{\sum_{n=1}^{N} Var_{\mathcal{H}_0}[\Lambda_n]}}$$
$$\rightsquigarrow \begin{cases} \mathcal{N}(0, 1) & \text{under } \mathcal{H}_0 \\ \mathcal{N}(\varrho, 1) & \text{under } \mathcal{H}_1 \end{cases}, \quad (9)$$

where $\rightsquigarrow$ denotes the convergence in distribution and

$$\varrho = \frac{\sum_{n=1}^{N} I_n \beta_n \gamma_n}{\sqrt{\sum_{n=1}^{N} I_n \gamma_n^2}} \quad (10)$$

---

[1]See, e.g., the ternary case in [41] for the derivation, which uses the additional assumption of small payload, $\beta_n \ll 1$, which is not necessary to obtain the result but simplifies the derivation.

is the deflection coefficient, which completely characterizes the statistical detectability. In (10), we used $I_n = 2 \cdot \sigma_n^{-4}$ for the Fisher information of LSBM in $\mathcal{N}(0, \sigma_n^2)$ w.r.t. the change rate $\beta_n$ (see [41] for more details).

Because the distribution of $\Lambda^\star(\mathbf{x}, \boldsymbol{\sigma})$ under $\mathcal{H}_0$ does not depend on any unknown parameters, one can set the threshold $\tau$ in (8) to maximize the detection power for any prescribed false-alarm probability $\alpha$ even when the true values of $\beta_n$ are not known to the Warden.

## 3. FOUR TYPES OF WARDEN

In this work, we consider four different types of Warden to investigate how the detection power decreases with increased ignorance of the Warden regarding the selection channel. Below, the Warden types are ordered by the amount of available information.

Empirical detectors will be constructed as binary classifiers trained on a set of cover-stego image pairs represented with the maxSRM feature vector [9] with the embedding change probabilities $\beta_n$ determined based on the Warden type as described below. Besides explaining how the maxSRM feature vector is computed, the construction of the actual empirical detectors (classifiers) also requires specifying the training database. We postpone discussing the training to Section 6, where we describe the experiments and their results. Note that LR detectors do not need any training phase and the Warden makes a decision on each individual image.

The *omniscient Warden* knows exactly the actions of the sender executed during embedding. The empirical detector will be constructed by computing the features for the pair of training cover (stego) images, $\mathbf{x}$ ($\mathbf{y}$), using the probabilities $\beta_{ij}$ computed from the cover image $\mathbf{x}$ assuming the true embedded payload size $R$. (To prevent any misunderstanding, we note that we need to assume that the payload is $R$ even when computing the cover feature.) For detectors implemented as a LRT, we simply use the change rates $\gamma_n = \beta_n$. Note that the deflection coefficient for the omniscient Warden (10) simplifies to:

$$\varrho^\star = \frac{\sum_{n=1}^N I_n \beta_n^2}{\sqrt{\sum_{n=1}^N I_n \beta_n^2}} = \sqrt{\sum_{n=1}^N I_n \beta_n^2}. \qquad (11)$$

This Warden is unrealistic because one cannot assume that the detector has access to the cover image. We include this Warden in our study because it has the highest detection power and serves as a useful upper bound on detection.

The *payload-informed Warden* knows the size of the embedded payload, $R$ (6), but has no access to the cover image. The Warden thus computes the change rates $\hat{\beta}_{ij}$ from the available image whether it is a cover or stego image. For the cover image, the maxSRM feature vector will be the same as for the omniscient Warden (and $\gamma_n = \beta_n$ for the LRT) while for the stego image, the change rates will be slightly different due to the embedding changes themselves ($\gamma_n$ will generally be different from but close to $\beta_n$ for the LRT).

The *fixed-payload Warden* does not know the embedded payload and computes the change rates ($\beta_{ij}$ or $\gamma_n$) from the available image assuming some fixed value of the embedded payload size $\tilde{R}$, which can generally be different from $R$.

Although the Warden could in principle estimate $R$ using a quantitative detector, as already mentioned in the introduction, for modern steganographic schemes whose detection requires high-dimensional rich models, it is fairly difficult to substantially improve upon the trivial estimator that always guesses the medium payload [28].

The *indifferent Warden* assumes non-adaptive embedding. For the empirical detector, this means that the Warden uses the SRM features while the LRT uses $\gamma_n = \gamma$ for all $n$. Note that, indeed, the maxSRM features match (up to a multiplicative constant) those of the original SRM when we set $\beta_{ij} = \beta > 0$ for all $i, j$.

By comparing the omniscient and payload-informed Wardens, we can study the effect of the embedding changes themselves on the estimated selection channel. The fixed-payload Warden is a realistic detector when the detector does not have any information about the embedded payload. By comparing the first three Wardens with the indifferent Warden, we will be able to assess the gain in detection when the Warden uses a selection-channel-aware detector.

## 4. TESTED STEGO SCHEMES

We selected four content-adaptive steganographic techniques that appear to be the current state of the art – WOW [22], S-UNIWARD implemented with the stabilizing constant $\sigma = 1$ as described in [25], HILL [33], and the ternary Multivariate Gaussian (MVG) method originally described in [16] and further improved by replacing the variance estimator as described in Section 5 of [41]. For HILL, we used the KB high-pass filter and the $3 \times 3$ and $15 \times 15$ averaging filters for the two low-pass filters because this setting provided the best security as reported in [33]. In contrast to [41], for simplicity, in MVG we skipped the smoothing of the Fisher information field.

Notice that WOW, S-UNIWARD, and HILL are cost-based schemes in the sense that the sender first identifies the cost of changing each ($n$th) pixel, $\rho_n$, and embeds the payload while minimizing the distortion

$$D(\mathbf{x}, \mathbf{y}) = \sum_{n=1}^N \rho_n [x_n \neq y_n], \qquad (12)$$

which leads to the following embedding change probabilities:

$$\beta_n = e^{-\lambda \rho_n} / (1 + 2e^{-\lambda \rho_n}), \qquad (13)$$

with $\lambda > 0$ determined to satisfy the payload constraint (6). The costs are typically obtained by changing a single pixel by $\pm 1$ and quantifying the impact of this change on selected noise residuals.

In contrast, the MVG scheme first estimates the cover model, the variances $\sigma_n^2$, and then computes the change rates $\beta_n$ that minimize the deflection coefficient under the omniscient Warden (11) and satisfy the payload constraint (6). This constrained optimization problem is easily solved using the method of Lagrange multipliers [16, 41]. In particular, the change rates $\beta_n$ and the Lagrange multiplier $\lambda > 0$ must satisfy $N + 1$ non-linear equations:

$$\beta_n = \frac{1}{\lambda I_n} \ln \frac{1 - 2\beta_n}{\beta_n}, \ n = 1, \dots, N, \qquad (14)$$

$$R = \frac{1}{N} \sum_{n=1}^N H(\beta_n). \qquad (15)$$

To embed the message, e.g., using syndrome-trellis codes, the sender converts the change rates to costs by inverting (13): $\rho_n = \ln(1/\beta_n - 2)$.

## 5. SETUP OF EXPERIMENTS

Our experiments will be conducted on the BOSSbase database ver. 1.01 [13] containing 10,000 512×512 8-bit grayscale images coming from eight different cameras. We will consider two types of sender – the Payload Limited Sender (PLS) and the Random Payload Sender (RPS). The PLS always embeds a message of a fixed relative length $R$ while the payload size embedded by a RPS is chosen uniformly randomly from $[0.05, 0.5]$ bpp. Three payloads will be used for the true embedded payload $R$ for the PLS and the payload assumed by the fixed-payload Warden, $\tilde{R}$: small (0.05 bpp), medium (0.2 bpp), and large (0.5 bpp).

### 5.1 Executing experiments with empirical detectors

All empirical detectors will be built as FLD ensemble classifiers [30] with maxSRM (or SRM) features computed as described in Section 3. The BOSSbase database embedded with a PLS with payload $R$ or with a random payload uniformly randomly distributed on $[0.05, 0.5]$ will be denoted with $\mathcal{B}_R$ and $\mathcal{B}_U$, respectively. The classifiers will be trained either on $\mathcal{B}_R$ or on $\mathcal{B}_U$ as described in the text. To assess the detection performance, the set of cover–stego image pairs from $\mathcal{B}_R$ ($\mathcal{B}_U$) will be randomly split into two parts of the same size – one used for training the ensemble while the other for its testing. We note that the hyperparameters $d_{\mathrm{sub}}$ and the number of base learners $L$ are determined only once by minimizing the out-of-bag error estimate of the testing error using bootstrapping on the entire database as described in [30]. This is repeated for ten random 5000/5000 database splits to obtain the statistical spread and assess the statistical significance of the results.

### 5.2 Executing experiments with the LRT

The asymptotic LRT does not need a training phase as it is capable of detecting steganography in each individual image once the variances $\sigma_n^2$ are known. In all our experiments, we estimate the variances $\sigma_n^2$ from the given image using the variance estimator described in Section 5 of [41]. We note that the specific choice of the variance estimator seems to play a negligible role. We repeated the experiments reported in Section 6 with six other variance estimators and obtained almost identical results.

For the $j$th image, $j = 1, \ldots, 10,000$, the detection performance is completely described by the deflection coefficient (10) or the ROC curve

$$\pi^{(j)}(\alpha) = Q(Q^{-1}(\alpha) - \varrho^{(j)}), \qquad (16)$$

where $Q(x)$ is the complementary cumulative distribution function (the tail probability) of the standard normal random variable $\mathcal{N}(0, 1)$. The deflection coefficient $\varrho^{(j)}$ is obtained from (10), with the Fisher information $I_n$ computed from the $j$th image and $\gamma_n$ computed based on the type of the Warden. To obtain the overall performance of this detector on the chosen image source and assess the statistical significance of the results, we compute the average power for
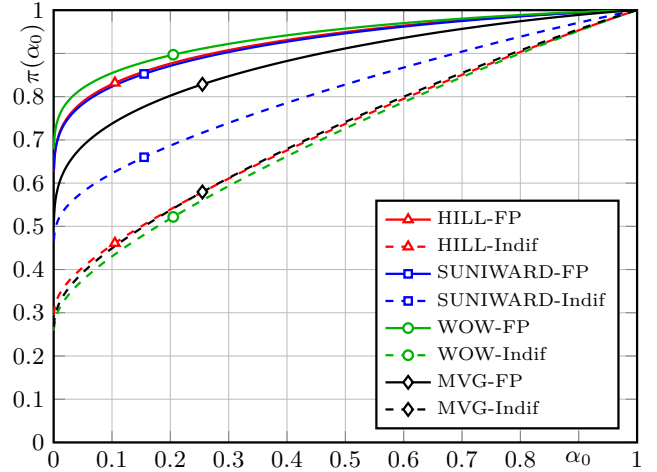


**Figure 1: Examples of ROC curves for the LR detector for the fixed-payload Warden ($\alpha = 0.2$ bpp) and the indifferent Warden for different embedding schemes and the RPS.**

each value[2] of the false alarm $\alpha$ for ten randomly selected subsets of 5000 images:

$$\overline{\pi}(\alpha) = \frac{1}{5000} \sum_{j=1}^{5000} \pi^{(j)}(\alpha). \qquad (17)$$

This quantity indeed correctly describes the expected value of the detector power for a fixed false alarm if the LRT was applied to images from the cover and stego sources. Examples of ROC curves for two types of warden and across four stego schemes for the RPS are shown in Figure 1. The asymmetrical shape is due to the large number of "easy to steganalyze" images in BOSSbase with very large deflection coefficients, such as images that are out of focus or have little content.

### 5.3 Evaluating detection loss/gain

In order to assess the change in the detection accuracy using a single scalar rather than an entire ROC curve, we will use the minimal total detection error under equal priors, which can be expressed as $P_{\mathrm{E}} = \min_\alpha \frac{1}{2}(1 - \overline{\pi}(\alpha) + \alpha)$ for the LR detector. For the empirical detectors, we obtain the value of $P_{\mathrm{E}}$ on the testing set from the ensemble. As explained above, the performance of both the empirical and LR detectors will be reported using the mean value and standard deviation of $P_{\mathrm{E}}$ over ten random database splits.

## 6. RESULTS

In practice, the Warden either knows the size of the embedded payload or she does not. The detection can obviously be more accurate when the payload size is available as the empirical Warden can form the training stego images and extract the maxSRM features with the true payload size. Likewise, the LRT can use the change rates extracted from the cover / stego image. In this case, it only makes sense to

---

[2] The false alarm $\alpha$ was sampled with a fixed step size of $5 \times 10^{-4}$ on $[0, 1]$.

investigate the impact of computing the change rates from the stego image instead of the cover image.

## 6.1 Experiments with the PLS

In our first batch of experiments with the PLS, we thus only inspect the detection loss of the omniscient Warden vs. the payload-informed Warden. We will also study the gain of the payload-informed Warden vs. the indifferent Warden to see the advantage of using the selection-channel-aware detector.

The empirical Warden always trains her classifier on the set of cover/stego images $(\mathcal{B}_0, \mathcal{B}_R)$ with the maxSRM features computed based on the type of the Warden as described in Section 3. In Figure 2, we contrast the detection accuracy loss for the empirical detectors and the LRTs for the PLS with small, medium, and large payloads. The figure clearly shows that the impact of computing the change rates from the stego image rather than the cover image is negligible. For the empirical detector (left), the loss of the detection accuracy between the payload-informed and omniscient wardens is not statistically significant for either tested stego method. Figure 3 depicts the gain of the payload-informed Warden over the indifferent Warden for both types of detector. The gain is quite substantial and almost three times larger for the LR detector than for the empirical one. Given how differently both detectors are built, one can hardly expect even an approximate quantitative match between them. However, notice that the *relative* comparison of embedding schemes w.r.t. each other is approximately preserved in both figures.

## 6.2 Experiments with the RPS

We now turn our attention to the more interesting case of a Warden who does not know the payload size. Here, we will only consider the RPS. The problem of building a detector when the embedded payload is unknown has been investigated in [35], where the author provided experimental evidence that for the best robustness w.r.t. the payload size, the steganalyst should train on a uniform mixture of payloads.

Empirical detectors will be built as binary classifiers trained on $(\mathcal{B}_0, \mathcal{B}_U)$ with maxSRM features computed based on the Warden type. Since we already know that the difference between the omniscient and payload-informed Wardens is negligible, we focus on comparing the payload-informed, fixed-payload, and indifferent Wardens. In the considered case of the RPS, the payload-informed Warden is fictional and can hardly occur in real life. It is included merely as an upper performance bound.

In Figure 4, we show the loss of detection accuracy between the payload-informed Warden and the Warden with payload fixed at the small, medium, and large payload. In accordance with [9], both types of detectors indicate that using the medium fixed payload (0.2 bpp) for estimating the selection channel causes the smallest overall loss of detection performance. Figure 5 shows the gain in detection power between the Warden that uses the knowledge of the selection channel and the indifferent Warden. In both graphs, the bars marked with 0.05, 0.2, and 0.5 correspond to the difference $\overline{P}_{\mathrm{E}}^{(\mathrm{Indif})} - \overline{P}_{\mathrm{E}}^{(\mathrm{FP})}$ for the Warden who estimates the selection channel with payload size fixed to small, medium, and large, while the column marked with TRUE shows $\overline{P}_{\mathrm{E}}^{(\mathrm{Indif})} - \overline{P}_{\mathrm{E}}^{(\mathrm{PI})}$, which is maximal gain one could ob-

tain if the Warden always correctly guessed the true payload size. By comparing the loss in Figure 4 with the gain in Figure 5 for the fixed-payload Warden, it is clear that it is better to use an imprecise selection channel rather than none.

Comparing the detection loss in Figures 2 and 4 across stego methods, we can conclude that the loss of detection power due to mismatched payload is far smaller for the MVG steganography than for the three cost-based schemes, and this is true for both the empirical and LR detectors. We explain this observation for the LRT in the appendix by analyzing the sensitivity of the deflection coefficient w.r.t. the payload size (parameter $\lambda$) for the MVG.

Surprisingly, despite the fact that the empirical and LR detectors are built very differently, the results are qualitatively consistent in terms of relative comparison of losses and gains across the stego methods. Finally, and with a great caution, we note that if it is at all meaningful to relate these two detectors, it seems that the way the knowledge of the selection channel is incorporated in the empirical detector is highly suboptimal as the LR seems to benefit from the awareness of the selection channel much more.

## 7. CONCLUSION

Recently, it has been shown that the detection of content-adaptive steganography can be improved by incorporating in the detector the knowledge of the actions of the sender, which are in turn determined by the content itself. Because steganographic changes themselves are almost always subtle, the Warden can estimate the embedding change probabilities rather accurately as long as the size of the embedded payload is approximately known. Any difference between the assumed and embedded payload size will inevitably lead to a loss of detection power. As discovered in this paper, this loss appears to be rather small and it is advantageous for the Warden to use even imprecisely determined embedding change probabilities than not use them at all.

We establish this for four modern spatial-domain steganographic schemes for classifiers built using machine learning and for likelihood ratio detectors designed in an optimal manner from a cover model. Since both detectors are built from entirely different principles, one cannot expect a quantitative match between them. Nevertheless, both detectors exhibit qualitatively the same behavior and point to the same evidence: 1) the loss of detector power due to imprecise knowledge of the selection channel is rather small, 2) the misjudged payload size is much less of an issue for detection of model-based steganography than for minimal-distortion steganography, 3) the inaccuracy due to estimating the embedding change probabilities from the stego image rather than the cover has a completely negligible effect.
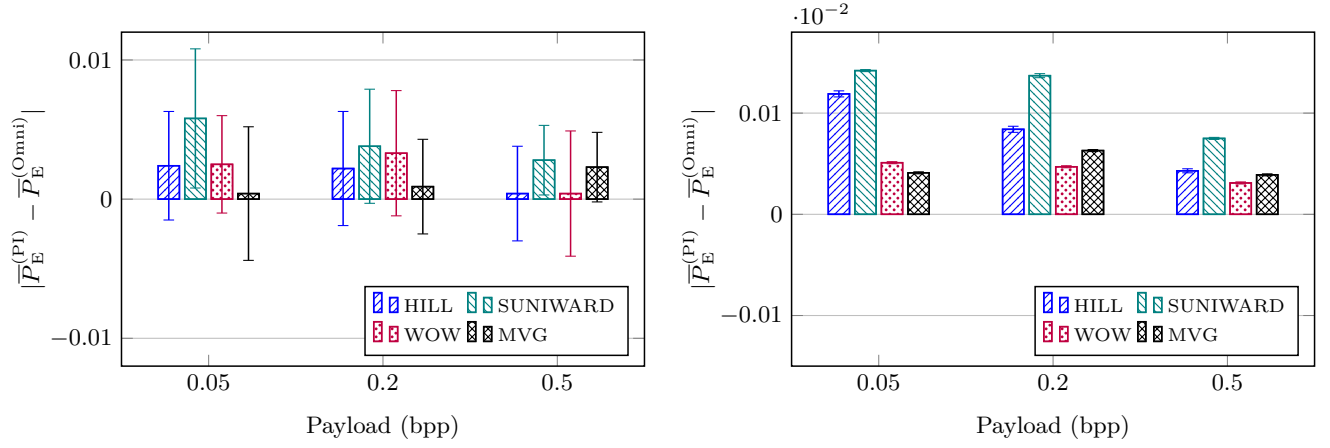
## 8. ACKNOWLEDGMENTS

**Figure 2: Difference in $\overline{P}_E$ when computing the change rates from the stego image (payload-informed Warden, PI) rather than the cover image (omniscient Warden) for four embedding methods and three payloads of the PLS. Left: empirical, Right: LR detector.**
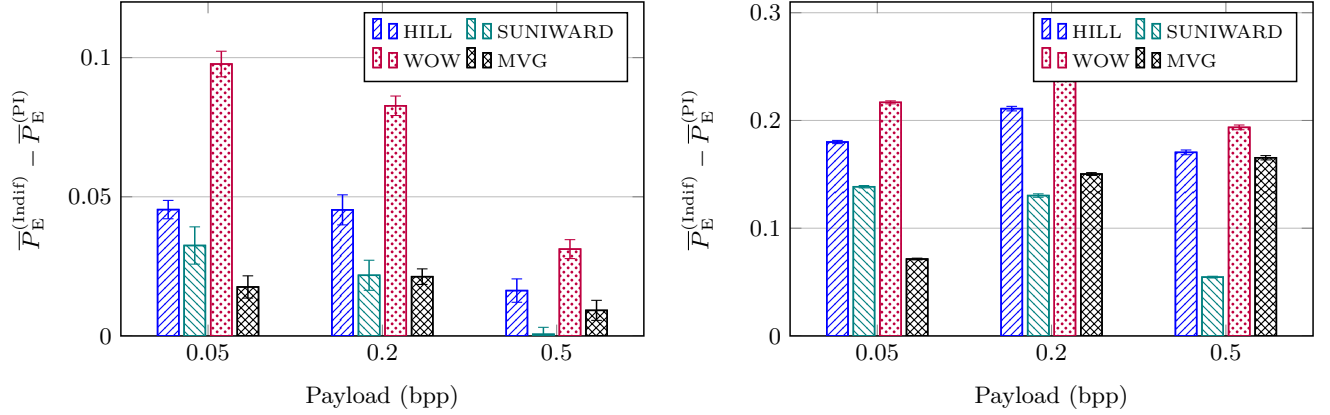


**Figure 3: The gain in detection accuracy when using the knowledge of the selection channel (payload-informed Warden) vs. not using it (indifferent Warden) for three payloads of the PLS.**
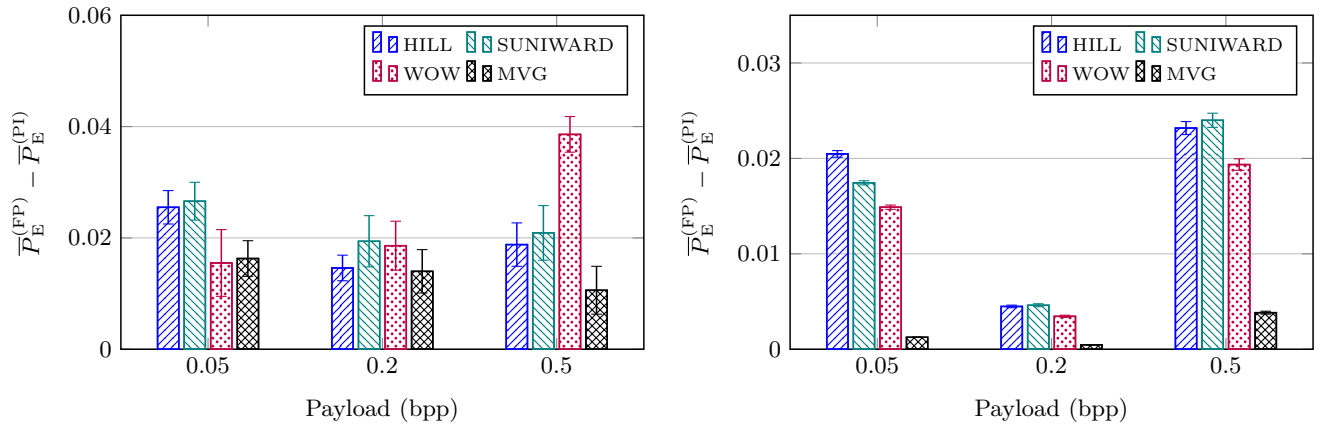


**Figure 4: Loss of detection due to not knowing the payload size. Both graphs show the difference in $\overline{P}_E$ for the fixed-payload and payload-informed Wardens for the RPS. Left: empirical, Right: LR detector.**
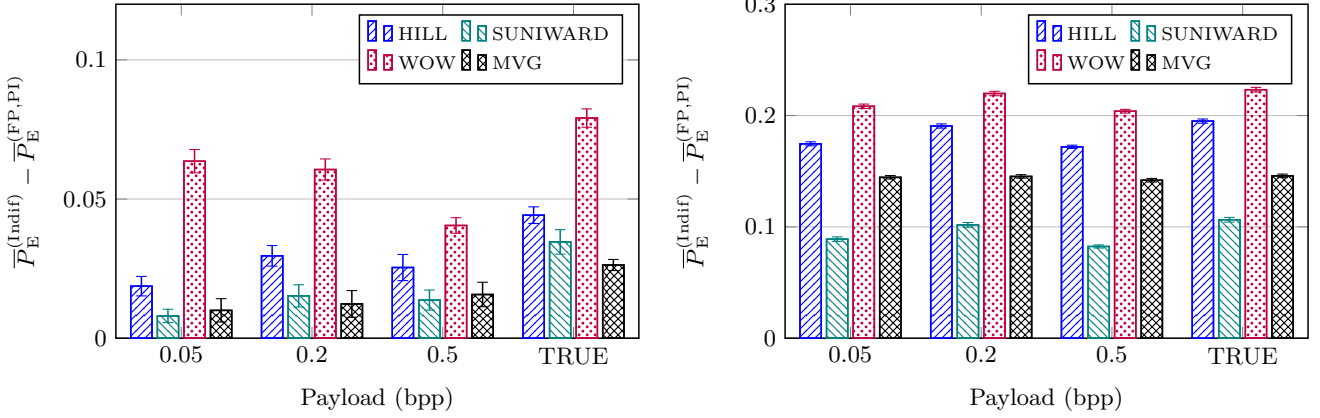
**Figure 5: Gain in detection accuracy when using the knowledge of the selection channel w.r.t. the detector that does not. The three leftmost groups of bars show** $\overline{P}_{\mathrm{E}}^{(\mathrm{Indif})} - \overline{P}_{\mathrm{E}}^{(\mathrm{FP})}$ **when the FP Warden estimates the selection channel with a fixed small, medium, and large payload. The bars marked TRUE corresponds to** $\overline{P}_{\mathrm{E}}^{(\mathrm{Indif})} - \overline{P}_{\mathrm{E}}^{(\mathrm{PI})}$ **when the Warden always estimates the selection channel with the correct payload.**

## APPENDIX

## Deflection coefficient as a function of payload for MVG stego

In MVG, the payload size is controlled by the Lagrange multiplier $\lambda$ (see Eqs. (14)–(15)). In this appendix, we show that the deflection coefficient $\varrho^\star(\lambda)$ (10) for the MVG does not strongly depend on the payload size (on $\lambda$). In particular, we show that the ratio $r(\lambda, \lambda') = \varrho(\lambda')/\varrho^\star(\lambda)$ when the Warden uses change rates $\beta_n(\lambda')$ computed by solving (14) using the Lagrange multiplier $\lambda'$ and for the omniscient Warden, $\varrho^\star$ (11), who uses $\beta_n(\lambda)$, $\lambda \neq \lambda'$:

$$r(\lambda, \lambda') = \frac{\sum_{n=1}^N I_n \beta_n(\lambda) \beta_n(\lambda')}{\sqrt{\sum_{n=1}^N I_n \beta_n^2(\lambda)} \sqrt{\sum_{n=1}^N I_n \beta_n^2(\lambda')}}, \qquad (18)$$

is close to 1 for majority of images from BOSSbase. To this end, we first prove a useful lemma regarding the properties of the solution to the non-linear equation (14) for the change rate $\beta_n$. By $e$ in the lemma we understand the Euler constant.

LEMMA 1. *Let $a > 0$ be such that $a/\ln(a - 2) > 2 + e$, which is equivalent with $a \gtrsim 9.52$. Then, the non-linear equation (14)*

$$\beta = \frac{1}{a} \ln(1/\beta - 2). \qquad (19)$$

*has a unique solution $\beta = \lim_{k \to \infty} \beta^{(k)}$, where $\beta^{(k)}$ is given by the following recursive formula:*

$$\beta^{(0)} = \frac{1}{a},$$
$$\beta^{(k+1)} = \frac{1}{a} \ln(1/\beta^{(k)} - 2), \; k \geq 0. \qquad (20)$$

*Moreover, the subsequences $\beta^{(2l)}$ and $\beta^{(2l+1)}$ are monotone increasing and decreasing, respectively, and*

$$\beta = \frac{\ln a}{a} + \delta, \; |\delta| \leq (4 + c) \frac{\ln a}{a^2}, \qquad (21)$$

*where $c = 2(1 - \frac{1}{5} \ln 10)^{-1} \approx 3.71$.*

PROOF. The convergence and the monotonicity of the subsequences can be established easily using induction. Note that $\beta^{(0)} < \beta^{(1)}$ and $\beta^{(2)} - \beta^{(0)} = \frac{1}{a} \ln \left[ \left( \frac{a}{\ln(a-2)} - 2 \right) / e \right] > 0$ when $a/\ln(a - 2) > 2 + e$. Assuming $\beta^{(k)} \gtrless \beta^{(k-2)}$, we have using (20)

$$\beta^{(k+1)} - \beta^{(k-1)} = \frac{1}{a} \ln \frac{1/\beta^{(k)} - 2}{1/\beta^{(k-2)} - 2} \lessgtr 0, \qquad (22)$$

which establishes the monotonicity of both sequences from the Lemma. Similarly, it can be shown that $\beta^{(k)} \gtrless \beta^{(k+1)} \Rightarrow \beta^{(k+1)} \lessgtr \beta^{(k+2)}$. Due to the monotonicity and boundedness of both sequences, both $\beta^{(2l)}$ and $\beta^{(2l+1)}$, $l = 0, 1, \ldots$, have finite limits, $\beta^{(\mathrm{even})}$ and $\beta^{(\mathrm{odd})}$, which must coincide because

$$0 \leq \beta^{(\mathrm{odd})} - \beta^{(\mathrm{even})} = \frac{1}{a} \ln \frac{1/\beta^{(\mathrm{odd})} - 2}{1/\beta^{(\mathrm{even})} - 2} \leq 0. \qquad (23)$$

Due to space limitations, only an outline of the proof of (21) is given. By repeatedly applying the inequality

$$\frac{-x}{1 - x} \leq \ln(1 - x) \leq -x, \text{ for any } x > 0. \qquad (24)$$

routine manipulations can be used to show that

$$0 < \beta - \beta^{(2)} < \beta^{(1)} - \beta^{(2)} = \ln \ln a + \delta_2, \qquad (25)$$

$$\beta^{(2)} = \frac{\ln a}{a} - \frac{\ln \ln a}{a} + \delta_1, \qquad (26)$$

with $|\delta_2| \leq 4 \ln a/a^2$ and $|\delta_1| \leq c \ln a/a^2$, which establishes (21). $\square$

We now express the ratio $r(\lambda, \lambda')$ (18) using (21). The value of $r$ obviously depends on the profile of the Fisher information $I_n$, $n = 1, \ldots, N$, which strongly depends on content. Without loss on generality, let us assume that $I_n$ are sorted from the smallest to the largest and that $\lambda < \lambda'$. For a given image and $\lambda$, let $\underline{n}(\lambda)$ be the smallest integer such that for all $n \geq \underline{n}$, $\lambda I_n/\ln(\lambda I_n - 2) > 2 + e$, the assumption of the above Lemma. Note that with $\lambda < \lambda'$, we automatically have $\lambda' I_n/\ln(\lambda' I_n - 2) > 2 + e$ as well. Among the images from BOSSbase 1.01, this condition is satisfied

for payload 0.05 bpp, 0.2 bpp, and 0.5 bpp, on average for 99.99%, 99.1%, and 89.2% of all pixels. We can thus split each of the three sums in (18) into two terms – one over pixels for which $\lambda I_n < 9.52$ $(n = 1, \ldots, \underline{n} - 1)$ and for which $\lambda I_n \geq 9.52$ $(n = \underline{n}, \ldots, N)$. Since all change rates are at most $\beta_n(\lambda) \leq 1/3$, we have for any $\lambda, \lambda'$

$$\sum_{n=1}^{\underline{n}-1} I_n \beta_n(\lambda) \beta_n(\lambda') \leq \frac{1}{9} \sum_{n=1}^{\underline{n}-1} I_n = \omega. \qquad (27)$$

Thus, (18) can be written as

$$r(\lambda, \lambda') = \frac{\omega + \mathbf{u} \cdot \mathbf{v}}{\sqrt{\omega + \|\mathbf{u}\|}\sqrt{\omega + \|\mathbf{v}\|}}, \qquad (28)$$

where $\mathbf{u}, \mathbf{v} \in \mathbb{R}^{N-\underline{n}}$ are defined by $(n = 1, \ldots, N - \underline{n})$:

$$u_n = \frac{\ln(\lambda I_{n+\underline{n}}) + \epsilon_{n+\underline{n}}}{\lambda \sqrt{I_{n+\underline{n}}}}, \qquad (29)$$

$$v_n = \frac{\ln(\lambda I_{n+\underline{n}}) + \epsilon'_{n+\underline{n}}}{\lambda' \sqrt{I_{n+\underline{n}}}} \qquad (30)$$

with $\epsilon_n, \epsilon'_n$ bounded from the Lemma

$$|\epsilon_n| \leq (4 + c)\frac{\ln(\lambda I_n)}{\lambda I_n} \qquad (31)$$

$$|\epsilon'_n| \leq (4 + c)\frac{\ln(\lambda I_n)}{\lambda I_n} + \ln(\lambda'/\lambda). \qquad (32)$$

Using Taylor expansion w.r.t. $\omega$ in (28):

$$\frac{\mathbf{u} \cdot \mathbf{v}}{\|\mathbf{u}\| \|\mathbf{v}\|} \left(1 + \mathcal{O}(\omega(\|\mathbf{u}\|^{-2} + \|\mathbf{v}\|^{-2}))\right)$$
$$\leq r(\lambda, \lambda') \leq \frac{\mathbf{u} \cdot \mathbf{v}}{\|\mathbf{u}\| \|\mathbf{v}\|} + \omega/(\|\mathbf{u}\| \|\mathbf{v}\|). \qquad (33)$$

The vectors $\mathbf{u}$ and $\mathbf{v}$ are "almost" collinear because the dominant term in $u_n$ and $v_n$ is $\ln(\lambda I_{n+\underline{n}})$ as both $|\epsilon_n|$ and $|\epsilon'_n|$ are small by (31)-(32). Here, we need to know that for $\lambda, \lambda'$ corresponding to payloads 0.05 bpp, and 0.2 bpp (and for 0.2 bpp and 0.5 bpp), $\lambda'/\lambda \approx 5 - 20$ with a median around 10, while the maximal values of $\lambda I_n$ reach $10^5 - 10^6$ in the vast majority of BOSSbase images.

## A.  REFERENCES

[1] C. Chen and Y. Q. Shi. JPEG image steganalysis utilizing both intrablock and interblock correlations. In *Circuits and Systems, ISCAS 2008. IEEE International Symposium on*, pages 3029–3032, Seattle, WA, May, 18–21, 2008.

[2] L. Chen, Y.Q. Shi, P. Sutthiwan, and X. Niu. A novel mapping scheme for steganalysis. In *Proc. IWDW*, volume 7809 of *LNCS*, pages 19–33. Springer Berlin Heidelberg, 2013.

[3] R. Cogranne and F. Retraint. Application of hypothesis testing theory for optimal detection of LSB matching data hiding. *Signal Processing*, 93(7):1724–1737, July, 2013.

[4] R. Cogranne and F. Retraint. An asymptotically uniformly most powerful test for LSB Matching detection. *IEEE TIFS*, 8(3):464–476, 2013.

[5] R. Cogranne and T. H. Thai. Optimal detection of OutGuess using an accurate model of DCT coefficients. In *Sixth IEEE International Workshop on Information Forensics and Security*, Atlanta, GA, December 3–5, 2014.

[6] R. Cogranne, C. Zitzmann, L. Fillatre, F. Retraint, I. Nikiforov, and P. Cornu. A cover image model for reliable steganalysis. In *Information Hiding, 13th International Conference*, volume 7692 of *LNCS*, pages 178–192, Prague, Czech Republic, May 18–20, 2011.

[7] R. Cogranne, C. Zitzmann, F. Retraint, I. Nikiforov, L. Fillatre, and P. Cornu. Statistical detection of LSB Matching using hypothesis testing theory. In *Information Hiding, 14th International Conference*, volume 7692 of *LNCS*, pages 46–62, Berkeley, California, May 15–18, 2012.

[8] T. Denemark, J. Fridrich, and V. Holub. Further study on the security of S-UNIWARD. In *Proceedings SPIE, Electronic Imaging, Media Watermarking, Security, and Forensics 2014*, volume 9028, pages 05 1–13, San Francisco, CA, February 3–5, 2014.

[9] T. Denemark, V. Sedighi, V. Holub, R. Cogranne, and J. Fridrich. Selection-channel-aware rich model for steganalysis of digital images. In *Proc. IEEE WIFS*, Atlanta, GA, December 3–5, 2014.

[10] H. Farid and L. Siwei. Detecting hidden messages using higher-order statistics and support vector machines. In *Information Hiding, 5th International Workshop*, volume 2578 of *LNCS*, pages 340–354, Noordwijkerhout, The Netherlands, October 7–9, 2002. Springer-Verlag, New York.

[11] L. Fillatre. Adaptive steganalysis of least significant bit replacement in grayscale images. *IEEE Transactions on Signal Processing*, 60(2):556–569, 2011.

[12] T. Filler, J. Judas, and J. Fridrich. Minimizing additive distortion in steganography using syndrome-trellis codes. *IEEE TIFS*, 6(3):920–935, September 2011.

[13] T. Filler, T. Pevný, and P. Bas. BOSS (Break Our Steganography System). http://www.agents.cz/boss, July 2010.

[14] A. Foi, M. Trimeche, V. Katkovnik, and K. Egiazarian. Practical Poissonian-Gaussian noise modeling and fitting for single-image raw-data. *IEEE TIP*, 17(10):1737–1754, Oct. 2008.

[15] J. Fridrich and J. Kodovský. Rich models for steganalysis of digital images. *IEEE TIFS*, 7(3):868–882, June 2011.

[16] J. Fridrich and J. Kodovský. Multivariate Gaussian model for designing additive distortion for steganography. In *Proc. IEEE ICASSP*, Vancouver, BC, May 26–31, 2013.

[17] M. Goljan, R. Cogranne, and J. Fridrich. Rich model for steganalysis of color images. In *Proc. IEEE WIFS*, Atlanta, GA, December 3–5, 2014.

[18] G. Gül and F. Kurugollu. A new methodology in steganalysis : Breaking highly undetactable steganograpy (HUGO). In *Information Hiding, 13th International Conference*, volume 7692 of *LNCS*, pages 71–84, Prague, Czech Republic, May 18–20, 2011.

[19] L. Guo, J. Ni, and Y.-Q. Shi. An efficient JPEG steganographic scheme using uniform embedding. In *Proc. IEEE WIFS*, Tenerife, Spain, December 2–5, 2012.

[20] G. E. Healey and R. Kondepudy. Radiometric CCD camera calibration and noise estimation. *IEEE TPAMI*, 16(3):267–276, March 1994.

[21] V. Holub and J. Fridrich. Low complexity features for JPEG steganalysis using undecimated DCT. *IEEE TIFS*, 10(2):219–228.

[22] V. Holub and J. Fridrich. Designing steganographic distortion using directional filters. In *Proc. IEEE WIFS*, Tenerife, Spain, December 2–5, 2012.

[23] V. Holub and J. Fridrich. Digital image steganography using universal distortion. In *1st ACM IH&MMSec. Workshop*, Montpellier, France, June 17–19, 2013.

[24] V. Holub and J. Fridrich. Random projections of residuals for digital image steganalysis. *IEEE TIFS*, 8(12):1996–2006, December 2013.

[25] V. Holub, J. Fridrich, and T. Denemark. Universal distortion design for steganography in an arbitrary domain. *EURASIP Journal on Information Security, Special Issue on Revised Selected Papers of the 1st ACM IH and MMS Workshop*, 2014:1, 2014.

[26] J. R. Janesick. *Scientific Charge-Coupled Devices*, volume Monograph PM83. Washington, DC: SPIE Press - The International Society for Optical Engineering, January 2001.

[27] A. D. Ker and R. Böhme. Revisiting weighted stego-image steganalysis. In *Proceedings SPIE, Electronic Imaging, Security, Forensics, Steganography, and Watermarking of Multimedia Contents X*, volume 6819, pages 5 1–17, San Jose, CA, January 27–31, 2008.

[28] J. Kodovký and J. Fridrich. Quantitative steganalysis using rich models. In *Proceedings SPIE, Electronic Imaging, Media Watermarking, Security, and Forensics 2013*, volume 8665, San Francisco, CA, February 5–7, 2013.

[29] J. Kodovský and J. Fridrich. Steganalysis of JPEG images using rich models. In *Proceedings SPIE, Electronic Imaging, Media Watermarking, Security, and Forensics 2012*, volume 8303, pages 0A 1–13, San Francisco, CA, January 23–26, 2012.

[30] J. Kodovský, J. Fridrich, and V. Holub. Ensemble classifiers for steganalysis of digital media. *IEEE TIFS*, 7(2):432–444, 2012.

[31] E.L. Lehmann and J.P. Romano. *Testing Statistical Hypotheses, 2nd edition*. Springer, 2005.

[32] B. Li, S. Tan, M. Wang, and J. Huang. Investigation on cost assignment in spatial image steganography. *IEEE TIFS*, 9(8):1264–1277, August 2014.

[33] B. Li, M. Wang, and J. Huang. A new cost function for spatial image steganography. In *Proceedings IEEE ICIP*, Paris, France, October 27–30, 2014.

[34] S. Lyu and H. Farid. Steganalysis using higher-order image statistics. *IEEE TIFS*, 1(1):111–119, 2006.

[35] T. Pevný. Detecting messages of unknown length. In A. Alattar, N. D. Memon, E. J. Delp, and J. Dittmann, editors, *Proceedings SPIE, Electronic Imaging, Media Watermarking, Security and Forensics III*, volume 7880, pages OT 1–12, San Francisco, CA, January 23–26, 2011.

[36] T. Pevný, T. Filler, and P. Bas. Using

high-dimensional image models to perform highly undetectable steganography. In *Information Hiding, 12th International Conference*, volume 6387 of *LNCS*, pages 161–177, Calgary, Canada, June 28–30, 2010. Springer-Verlag, New York.

[37] T. Pevný, J. Fridrich, and A. D. Ker. From blind to quantitative steganalysis. *IEEE TIFS*, 7(2):445–454, 2011.

[38] N. Provos. Defending against statistical steganalysis. In *10th USENIX Security Symposium*, pages 323–335, Washington, DC, August 13–17, 2001.

[39] P. Schöttle and R. Böhme. A game-theoretic approach to content-adaptive steganography. In *Information Hiding, 14th International Conference*, volume 7692 of *LNCS*, pages 125–141, Berkeley, California, May 15–18, 2012.

[40] P. Schöttle, S. Korff, and R. Böhme. Weighted stego-image steganalysis for naive content-adaptive embedding. In *Proc. IEEE WIFS*, Tenerife, Spain, December 2–5, 2012.

[41] V. Sedighi, J. Fridrich, and R. Cogranne. Content-adaptive pentary steganography using the multivariate generalized Gaussian cover model. In *Proceedings SPIE, Electronic Imaging, Media Watermarking, Security, and Forensics 2015*, San Francisco, CA, February 9–11, 2015.

[42] Y. Q. Shi, C. Chen, and W. Chen. A Markov process based approach to effective attacking JPEG steganography. In *Information Hiding, 8th International Workshop*, volume 4437 of *LNCS*, pages 249–264, Alexandria, VA, July 10–12, 2006. Springer-Verlag, New York.

[43] W. Tang, H. Li, W. Luo, and J. Huang. Adaptive steganalysis against WOW embedding algorithm. In *2nd ACM IH&MMSec. Workshop*, Salzburg, Austria, June 11–13, 2014.

[44] T. Thai, R. Cogranne, and F. Retraint. Statistical model of quantized DCT coefficients: Application in the steganalysis of Jsteg algorithm. *IEEE TIP*, 23(5):1–14, May 2014.

[45] A. Westfeld. High capacity despite better steganalysis (F5 – a steganographic algorithm). In *Information Hiding, 4th International Workshop*, volume 2137 of *LNCS*, pages 289–302, Pittsburgh, PA, April 25–27, 2001. Springer-Verlag, New York.

[46] C. Zitzmann, R. Cogranne, L. Fillatre, I. Nikiforov, F. Retraint, and P. Cornu. Hidden information detection based on quantized Laplacian distribution. In *Proc. IEEE ICASSP*, Kyoto, Japan, March 25-30, 2012.

[47] C. Zitzmann, R. Cogranne, F. Retraint, I. Nikiforov, L. Fillatre, and P. Cornu. Statistical decision methods in hidden information detection. In *Information Hiding, 13th International Conference*, volume 7692 of *LNCS*, pages 163–177, Prague, Czech Republic, May 18–20, 2011.

[48] D. Zou, Y. Q. Shi, W. Su, and G. Xuan. Steganalysis based on Markov model of thresholded prediction-error image. In *Proceedings IEEE ICME*, pages 1365–1368, Toronto, Canada, July 9–12, 2006.