Improving Selection-Channel-Aware Steganalysis Features

***Tomáš Denemark and Jessica Fridrich**, Department of ECE, SUNY Binghamton, NY, USA, {tdenema1,fridrich}@binghamton.edu,*
***Pedro Comesaña-Alfaro**, Department of Signal Theory and Communications, University of Vigo, Spain, pcomesan@gts.uvigo.es*

Abstract

Currently, the best detectors of content-adaptive steganography are built as classifiers trained on examples of cover and stego images represented with rich media models (features) formed by histograms (or co-occurrences) of quantized noise residuals. Recently, it has been shown that adaptive steganography can be more accurately detected by incorporating content adaptivity within the features by accumulating the embedding change probabilities (change rates) in the histograms. However, because each noise residual depends on an entire pixel neighborhood, one should accumulate the embedding impact on the residual rather than the pixel to which the residual is formally attributed. Following this observation, in this paper we propose the expected value of the residual L_1 distortion as the quantity that should be accumulated in the selection-channel-aware version of rich models to improve the detection accuracy. This claim is substantiated experimentally on four modern content-adaptive steganographic algorithms that embed in the spatial domain.

Motivation

Modern content-adaptive steganography dates back to 2010 when HUGO (Highly Undetectable steGO) was introduced [22]. It incorporated syndrome-trellis codes [6] as the most innovative element that is currently used in all modern steganographic schemes operating in any domain. Such advanced coding techniques gave the steganographer control over where the embedding changes are to be executed by specifying the costs of modifying each pixel. The costs, together with the payload size, determine the probability with which a given pixel is to be modified during embedding. These probabilities, also called change rates, are recognized as the so-called selection channel.

Since the costs of virtually all content-adaptive embedding techniques are not very sensitive to the embedding changes themselves [25], they are also available to the steganalyst. For simpler embedding paradigms, such as the Least Significant Bit (LSB) replacement combined with naive adaptive embedding, researchers have shown how a publicly known selection channel can be used to improve the WS detector [23]. Modern adaptive steganographic schemes for digital images [13, 19, 26, 16, 24], however, do not use LSB replacement or naive adaptive embedding, and their detection requires detectors built with machine learning.

The prevailing trend is to represent images using rich media models, such as the Spatial Rich Model (SRM) [7], Projection Rich Model (PSRM) [14], and their numerous variants designed for the spatial domain [2], JPEG domain [17, 12, 15, 27], and for color images [8, 9]. Such rich models are concatenations of histograms (for projection type rich models [14] and phase-aware models [15, 12, 27]) or co-occurrences of quantized noise residuals obtained with a variety of linear and non-linear pixel predictors. In [28], the authors proposed to compute the co-occurrences in the SRM only from a fraction of pixels with the highest embedding change probability. Even though this decreased the amount of data available for steganalysis, the authors showed that the embedding algorithm WOW could be detected with a markedly better accuracy. A generalization of this approach was later proposed that utilized the statistics of all pixels by accumulating the maximum of the four pixel change rates in the co-occurrences of four neighboring residuals. This version of the SRM called maxSRM [5] improved the detection of all content-adaptive algorithms to a varying degree. The idea was, however, not extensible to spatial-domain rich features for detection of JPEG steganography [15, 12, 27] or to projection type features because the residuals depend on numerous pixels and one can no longer associate a pixel (or a DCT coefficient) change rate with a given residual sample. This paper resolves this issue by replacing the change rate with the expected value of the residual distortion as the quantity that should be accumulated in the histograms (for JPEG phase-aware features and projection type features) and in co-occurrences (for SRM).

This extension is relatively straightforward for linear residuals since the relationship tying the embedding domain and the residual domain is linear. If the embedding changes are executed independently,[1] one can easily compute the expected value of the embedding distortion in the residual domain analytically. A major complication, however, occurs for non-linear residuals due to the necessity to compute marginals of high-dimensional probability mass functions. This is why the emphasis of this paper is on rich representations formed from linear residuals. An extension of the idea presented in this paper to phase-aware JPEG features appears in [3].

In the next section, we include a brief overview of the SRM, PSRM, and maxSRM to prepare the ground for the third section, where we describe the quantity that

[1] This is true for all current steganographic schemes with the notable exception of steganography that synchronizes the selection channel [4, 20].

will be accumulated in the histograms (PSRM) and co-occurrences of quantized noise residuals (SRM) in the selection-channel-aware version of such features. Since the PSRM is extremely computationally demanding, we only work with a subset of its features that come from linear ('spam' type) residuals of dimension 1,980. In the fourth section, we show that making this relatively compact feature space properly aware of the selection channel achieves state-of-the-art performance with the ensemble classifier. The paper is concluded in the fifth section, where we summarize the contribution and outline how the proposed idea can be executed for phase-aware JPEG features.

Preliminaries: SRM, PSRM, and maxSRM

In this section, we review the basics of the SRM, its projection version, the PSRM, and the selection-channel-aware maxSRM. This is done in order to make the paper self-contained and easier to read.

The symbols $\mathbf{X},\mathbf{Y}\in\{0,\ldots,255\}^{n_1\times n_2}$ will be used exclusively for two-dimensional arrays of pixel values in an $n_1\times n_2$ grayscale cover and stego image, respectively. Elements of a matrix will be denoted with the corresponding lower case letter. The pair of subscripts i,j will always be used to index elements in an $n_1\times n_2$ matrix. The cardinality of a finite set $\mathcal{S}$ will be denoted $|\mathcal{S}|$.

SRM

Both the SRM and the PSRM extract the same set of noise residuals from the image under investigation. They differ in how they represent their statistical properties. The SRM uses four dimensional co-occurrences while the PSRM uses histograms of residual projections.

A noise residual is an estimate of the image noise component obtained by subtracting from each pixel its estimate (expectation) obtained using a pixel predictor from the pixel's immediate neighborhood. Both rich models use 45 different pixel predictors of two different types – linear and non-linear. Each linear predictor is a shift-invariant finite-impulse response filter described by a kernel matrix $\mathbf{K}^{(\mathrm{pred})}$. The noise residual $\mathbf{Z}=(z_{kl})$ is a matrix of the same dimension as $\mathbf{X}$:

$$\mathbf{Z}=\mathbf{K}^{(\mathrm{pred})}\star\mathbf{X}-\mathbf{X}\triangleq\mathbf{K}\star\mathbf{X}. \tag{1}$$

In (1), the symbol $'\star'$ denotes the convolution with $\mathbf{X}$ mirror-padded so that $\mathbf{K}\star\mathbf{X}$ has the same dimension as $\mathbf{X}$. This corresponds to the 'conv2' Matlab command with the parameter 'same'.

An example of a simple linear residual is $z_{ij}=x_{i,j+1}-x_{ij}$, which is the difference between a pair of horizontally neighboring pixels. In this case, the residual kernel is $\mathbf{K}=\begin{pmatrix}-1 & 1\end{pmatrix}$, which means that the predictor estimates the pixel value as its horizontally adjacent pixel. This predictor is used in submodel 'spam14h' in the SRM.

All non-linear predictors in the SRM are obtained by taking the minimum or maximum of up to five residuals obtained using linear predictors. For example, one can predict pixel x_{ij} from its horizontal or vertical neighbors, obtaining thus one horizontal and one vertical residual $\mathbf{Z}^{(\mathrm{h})}=(z^{(\mathrm{h})}_{ij})$, $\mathbf{Z}^{(\mathrm{v})}=(z^{(\mathrm{v})}_{ij})$:

$$z^{(\mathrm{h})}_{ij}=x_{i,j+1}-x_{ij}, \tag{2}$$

$$z^{(\mathrm{v})}_{ij}=x_{i+1,j}-x_{ij}. \tag{3}$$

Using these two residuals, one can compute two non-linear 'minmax' residuals as:

$$z^{(\min)}_{ij}=\min\{z^{(\mathrm{h})}_{ij},z^{(\mathrm{v})}_{ij}\}, \tag{4}$$

$$z^{(\max)}_{ij}=\max\{z^{(\mathrm{h})}_{ij},z^{(\mathrm{v})}_{ij}\}. \tag{5}$$

The next step in forming the SRM involves quantizing $\mathbf{Z}$ with a quantizer $Q_{-T,T}$ with centroids $\mathcal{Q}_{-T,T}=\{-Tq,(-T+1)q,\ldots,Tq\}$, where $T>0$ is an integer threshold and $q>0$ is a quantization step:

$$r_{ij}\triangleq Q_{-T,T}(z_{ij}),\ \forall i,j. \tag{6}$$

The next step in forming the SRM feature vector involves computing a co-occurrence matrix of fourth order, $\mathbf{C}^{(\mathrm{SRM})}\in\mathcal{Q}^4_{-T,T}$, from four (horizontally and vertically) neighboring values of the quantized residual r_{ij} (6) from the entire image:[2]

$$c^{(\mathrm{SRM})}_{d_0d_1d_2d_3}=\sum_{i,j=1}^{n_1,n_2-3}[r_{i,j+k}=d_k,\forall k=0,\ldots,3], \tag{7}$$

$$d_k\in\mathcal{Q}_{-T,T}, \tag{8}$$

where $[P]$ is the Iverson bracket, which is equal to 1 when the statement P is true and to 0 when it is false. Note that the dimensionality of the co-occurrence is $|\mathcal{Q}_{-T,T}|^4=5^4=625$. To keep the co-occurrence bins well-populated and thus statistically significant, the authors of the SRM used $T=2$ and $q\in\{1,1.5,2\}$. Finally, symmetries of natural images are leveraged to further marginalize the co-occurrence matrix to decrease the feature dimension and better populate the SRM feature vector (see Section II.C of [7]). For example, the 625 bins get reduced to 169 bins after symmetrization, while two 625-dimensional co-occurrences of min and max residuals can be symmetrized to 330.

In [5], the authors proposed to use a slightly different 'd2' scan for the co-occurrence that gives slightly better overall detection results. Formally, the co-occurrence obtained using the 'd2' scan can be written as

$$\begin{aligned}c^{(\mathrm{SRMd2})}_{d_0d_1d_2d_3}=&\sum_{i,j=1}^{n_1,n_2-3}[r_{i,j}=d_0,r_{i,j+1}=d_1,\\&r_{i+1,j+2}=d_2,r_{i+1,j+3}=d_3]\\&+\sum_{i,j=1}^{n_1,n_2-3}[r_{i-1,j}=d_0,r_{i-1,j+1}=d_1,\\&r_{i,j+2}=d_2,r_{i,j+3}=d_3].\end{aligned} \tag{9}$$

[2] This is an example of a horizontal co-occurrence.

The vertical version of this co-occurrence is defined similarly and also involves two terms.

The total dimension of the SRM with three quantization steps is 34,671. A smaller version of the SRM with a single quantization step $q = x \in \{1, 1.5, 2\}$ will be denoted as SRMqx, and it consists of 12,753 features.

PSRM

The predictors and residuals used in the PSRM are the same as those used in the SRM. Unlike the SRM, which captures the statistical properties of residuals using four-dimensional co-occurrences, the PSRM uses the first-order statistics (histograms) of projections of residuals onto multiple random directions. Given a noise residual $\mathbf{Z}$, a slightly simplified algorithm for computing the PSRM is:

1. Generate ν random matrices $\mathbf{\Pi}^{(k)} \in \mathbb{R}^{r\times s}, k \in \{1,\ldots,\nu\}$.
   - r, s are uniformly randomly selected from $\{1,\ldots,s_{max}\}$, where $s_{max} > 0$ is an integer parameter,
   - the elements of $\mathbf{\Pi}^{(k)}$ are independent realizations of a standard normal random variable $\mathcal{N}(0,1)$,
   - the elements are normalized so that the Frobenius norm[3] $\left\|\mathbf{\Pi}^{(k)}\right\|_F = 1$.
2. For each $k \in \{1,\ldots,\nu\}$, compute the residual projections $\mathbf{P}^{(k)} \triangleq \mathbf{Z} * \mathbf{\Pi}^{(k)}$.
3. For linear residuals, quantize $|p_{ij}^{(k)}|/q$ with a quantizer Q_T with $T+1$ centroids $\mathcal{Q}_T = \{1/2, 3/2, \ldots, T+1/2\}$:

   $$\tilde{p}_{ij}^{(k)} = Q_T(|p_{ij}^{(k)}|/q). \tag{10}$$

   For non-linear residuals, quantize $p_{ij}^{(k)}/q$ with a quantizer $Q'_{-T,T}$ with $2T+2$ centroids $\mathcal{Q}'_{-T,T} = \{-T - 1/2, -T+1/2, \ldots, T+1/2\}$:

   $$\tilde{p}_{ij}^{(k)} = Q'_{-T,T}(p_{ij}^{(k)}/q). \tag{11}$$

4. Compute ν separate histograms of the quantized values:

$$\begin{aligned} h_m^{(k)} &= \left|\left\{(i,j)\middle|\tilde{p}_{ij}^{(k)} = m+1/2\right\}\right|, \\ &\quad m \in \{0,1,\ldots,T-1\}, \\ &\quad k \in \{1,\ldots,\nu\} \text{ for linear residuals,} \qquad (12) \\ h_m^{(k)} &= \left|\left\{(i,j)\middle|\tilde{p}_{ij}^{(k)} = m+1/2\right\}\right|, \\ &\quad m \in \{-T,\ldots,T-1\}, \\ &\quad k \in \{1,\ldots,\nu\} \text{ for non-linear residuals.} \\ &\qquad (13) \end{aligned}$$

Symmetries of natural images are also used to make the histograms better populated. Depending on the residual and the projection matrix $\mathbf{\Pi}^{(k)}$, the PSRM utilizes up to eight symmetries (rotation by multiples of 90 degrees, mirroring, etc.) for each random random matrix $\mathbf{\Pi}^{(k)}$.

The standard parameter setup for the PSRM is as follows. The number of projections per residual is $\nu = 55$, the maximum projection matrix size $s_{max} = 8$, the quantization step $q = 1$, and the histogram threshold $T = 3$. This setup gives the PSRM the dimensionality of $12,870$, which is similar to that of SRMqx.

maxSRM

The selection-channel-aware SRM called maxSRM [5] is built in the same manner as the SRM [7] but the process of forming the co-occurrence matrices is modified to consider the embedding change probabilities $\hat{\beta}_{ij}$ estimated from the analyzed image:[4]

$$\begin{aligned} c_{d_0d_1d_2d_3}^{(\text{maxSRM})} = &\sum_{i,j=1}^{n_1,n_2-3} \max_{k=0,\ldots,3} \hat{\beta}_{i,j+k} \\ &\times [r_{i,j+k} = d_k, \forall k = 0,\ldots,3]. \end{aligned} \tag{14}$$

Above, $\mathbf{C}^{(\text{maxSRM})}$ denotes the selection-channel-aware version of the co-occurrence $\mathbf{C}^{(\text{SRM})}$. In other words, instead of increasing the corresponding co-occurrence bin by 1, the maximum of the embedding change probabilities taken across the four residuals is added to the bin. Thus, groups of four pixels with small probability of being changed affect the co-occurrence values to a smaller degree than groups where at least one pixel is likely to change. The rest of the process of forming the SRM stays exactly the same, including the symmetrization by sign and direction and merging co-occurrences into submodels (see [7, 5] for details). The maxSRM feature set has the same dimensionality as the SRM, which is 34,671 or 12,753 for maxSRMqx.

Finally, we note that the 'd2' scan of the maxSRM co-occurrence is obtained similarly to (9):

$$\begin{aligned} c_{d_0d_1d_2d_3}^{(\text{maxSRMd2})} = &\sum_{i,j=1}^{n_1,n_2-3} \bar{b}_{ij} \times [r_{i,j} = d_0, r_{i,j+1} = d_1, \\ &r_{i+1,j+2} = d_2, r_{i+1,j+3} = d_3] \\ &+ \sum_{i,j=1}^{n_1,n_2-3} \underline{b}_{ij} \times [r_{i-1,j} = d_0, \\ &r_{i-1,j+1} = d_1, r_{i,j+2} = d_2, \\ &r_{i,j+3} = d_3], \end{aligned} \tag{15}$$

[3] The Frobenius norm of matrix $\mathbf{A}$ is defined as $\|\mathbf{A}\|_F = \sqrt{\text{trace}(\mathbf{A}^{\mathrm{T}}\mathbf{A})}$.

[4] This means that we assume that the payload size is known to the steganalyst. Fortunately, as [25] shows, not knowing the payload size exactly leads to a rather gradual loss of detection accuracy – it is still better to use an imprecise payload size than none.

where

$$\bar{b}_{ij} = \max\{\hat{\beta}_{i,j}, \hat{\beta}_{i,j+1}, \hat{\beta}_{i+1,j+2}, \hat{\beta}_{i+1,j+3}\}, \tag{16}$$

$$\underline{b}_{ij} = \max\{\hat{\beta}_{i-1,j}, \hat{\beta}_{i-1,j+1}, \hat{\beta}_{i,j+2}, \hat{\beta}_{i,j+3}\}. \tag{17}$$

Replacing change rates with L_1 distortion of residuals

As pointed out in the introduction, there is a discrepancy in maxSRM in the sense that we accumulate the embedding change probabilities of *pixels* in the co-occurrence bins of *residuals*. Thus, we need to move away from pixel change rates to some measure of the residual distortion. After all, if the features were formed from pixel values rather than residuals, the change rates are proportional to the expected value of the L_1 (and L_2) distortion. This is because in most modern steganographic schemes the cover pixel x_{ij} is modified to $y_{ij} = x_{ij} + 1$ and $y_{ij} = x_{ij} - 1$ with the same probability β_{ij} and thus $E[|x_{ij} - y_{ij}|] = E[|x_{ij} - y_{ij}|^2] = 2\beta_{ij}$.

We explain the approach only for linear residuals and then discuss the issues with non-linear residuals.

Linear residuals

We recall that a linear residual $\mathbf{Z}$ in SRM is obtained by convolving the image with a kernel, this time we make the dependence of $\mathbf{Z}$ on the image explicit:

$$\mathbf{Z}^{(\mathrm{SRM})}(\mathbf{X}) = \mathbf{K} \star \mathbf{X}, \tag{18}$$

and in coordinates:

$$z_{ij}^{(\mathrm{SRM})}(\mathbf{X}) = \sum_{k,l} K_{kl} x_{i-k,j-l}. \tag{19}$$

The specific range for the indices k and l depends on the kernel support. Note that in PSRM the residual is additionally convolved with a projection matrix $\mathbf{\Pi}$:

$$\mathbf{Z}^{(\mathrm{PSRM})}(\mathbf{X}) = \mathbf{\Pi} \star (\mathbf{K} \star \mathbf{X}) = (\mathbf{\Pi} \star \mathbf{K}) \star \mathbf{X}, \tag{20}$$

due to the associativity of convolution. Thus, irrespectively of whether we deal with a linear residual from SRM or PSRM, the quantity whose sample statistic is collected (either a fourth-order co-occurrence or a histogram) is obtained by convolving the image with a kernel.

For steganographic schemes minimizing an additive distortion, message embedding is equivalent to adding noise whose distribution depends on the pixel location:

$$y_{ij} = x_{ij} + \xi_{ij},$$

where ξ_{ij} are independent random variables attaining their values in $\{-1, 0, 1\}$ with probabilities $\beta_{ij}, 1 - \beta_{ij}, \beta_{ij}$. Thus, each element of the difference $\mathbf{Z}(\mathbf{Y}) - \mathbf{Z}(\mathbf{X})$ is a random variable with

$$E[z_{ij}(\mathbf{Y}) - z_{ij}(\mathbf{X})] = E[\sum_{k,l} K_{kl} \xi_{i-k,j-l}] = 0, \tag{21}$$

$$Var[z_{ij}(\mathbf{Y}) - z_{ij}(\mathbf{X})] = 2\sum_{k,l} K_{kl}^2 \beta_{i-k,j-l}. \tag{22}$$

While it is straightforward to evaluate the expectation of the L_2 norm

$$\begin{aligned} E\left[\left(z_{ij}(\mathbf{Y}) - z_{ij}(\mathbf{X})\right)^2\right] &= 2\sum_{k,l} K_{kl}^2 \beta_{i-k,j-l} \\ &= Var[z_{ij}(\mathbf{Y}) - z_{ij}(\mathbf{X})] \\ &\triangleq \sigma_{ij}^2 \end{aligned} \tag{23}$$

due to the independence of embedding changes, it is much more difficult to compute the expectation of the absolute value, $E[|z_{ij}(\mathbf{Y}) - z_{ij}(\mathbf{X})|]$. We will thus consider a simplification and assume that $z_{ij}(\mathbf{Y}) - z_{ij}(\mathbf{X})$ is a zero-mean Gaussian random variable with variance (22). In this case, it is easy to evaluate

$$E[|z_{ij}(\mathbf{Y}) - z_{ij}(\mathbf{X})|] = \frac{2}{\sqrt{\pi}} \sqrt{\sum_{k,l} K_{kl}^2 \beta_{i-k,j-l}} \propto \sigma_{ij}. \tag{24}$$

In our experiments, the Gaussian approximation of the expectation of the L_1 distortion (24) worked much better than the L_2 distortion. This is why in the rest of this paper, we only use σ_{ij} as the quantity that will be accumulated in co-occurrences in SRM and in histograms in PSRM of linear residuals.

For SRM, the selection-channel-aware features built from a linear residual will be formed by replacing the change rates $\hat{\beta}_{ij}$ in Eqs. (7) (for the horizontal and vertical scans) and (9) (for the 'd2' scan) with σ_{ij}:

$$\begin{aligned} c_{d_0 d_1 d_2 d_3}^{\sigma\mathrm{SRM}} = \sum_{i,j=1}^{n_1, n_2 - 3} & \max_{k=0,\ldots,3} \sigma_{i,j+k} \\ & \times [r_{i,j+k} = d_k, \forall k = 0, \ldots, 3]. \end{aligned} \tag{25}$$

For the PSRM, the histograms of linear residuals (12) are replaced with their σ version:

$$h_m^{(k)\sigma} = \sum_{i,j=1}^{n_1, n_2} \sigma_{ij} \times [\tilde{p}_{ij}^{(k)} = m + 1/2], \tag{26}$$

$$m \in \{0, 1, \ldots, T-1\}, k \in \{1, \ldots, \nu\}. \tag{27}$$

At this point, we remark on the dimensionality of the σ-version of the PSRM. In the original PSRM, linear residuals are represented using only T bins because the last, $T+1$-st bin with centroid at $T + 1/2$ is uniquely determined by the other bins (the sum $\sum_{m \in \mathcal{Q}} h_m^{(k)} = n_1 n_2$). This is not true for $h_m^{(k)\sigma}$ as the sum of all $T+1$ bins is no longer equal to the number of residual values $n_1 n_2$. In our experiments, we did not see any statistically significant benefit in using all $T+1$ bins in $h_m^{(k)\sigma}$, which is why in the σ-version of the PSRM, we also skip the last $T+1$-st bin to keep the same feature dimensionality. Similarly, histograms of non-linear residuals in PSRM are represented using only $2T$ bins (both the first and the last values corresponding to centroids $-T - 1/2$ and $T + 1/2$ are skipped) and we keep the same arrangement in the σ-version.

Table 1. Detection of three steganographic algorithms for two payloads on BOSSbase 1.01 using the original maxSRM features and their proposed σmaxSRM form.

	0.2 bpp			0.4 bpp		
	$\overline{P}_{\mathrm{E}}$	min $\overline{P}_{\mathrm{E}}$	max $\overline{P}_{\mathrm{E}}$	$\overline{P}_{\mathrm{E}}$	min $\overline{P}_{\mathrm{E}}$	max $\overline{P}_{\mathrm{E}}$
HILL						
maxSRMq2d2	0.3181	0.3149	0.3228	0.2238	0.2174	0.2278
σmaxSRMq2d2	0.3075	0.3015	0.3109	0.2132	0.2104	0.2146
WOW						
maxSRMq2d2	0.2472	0.2400	0.2530	0.1658	0.1601	0.1732
σmaxSRMq2d2	0.2449	0.2397	0.2509	0.1620	0.1569	0.1694
MVG						
maxSRMq2d2	0.3291	0.3228	0.3336	0.2309	0.2287	0.2347
σmaxSRMq2d2	0.3205	0.3160	0.3239	0.2202	0.2138	0.2303

Non-linear residuals

The situation for non-linear minmax residuals is significantly more complicated because the residuals whose minimum (maximum) is computed are generally dependent random variables. In the most extreme case, which corresponds to the 'minmax41' submodel in EDGE5x5 residual of SRM, the minimum (maximum) is taken over four values that each depend on 15 neighboring pixel values out of the local 5×5 neighborhood. Computing the expectation of the L_1 or L_2 norm thus requires marginalization of a 25-dimensional probability mass function with 3^{25} values. We were unable to find an algorithm whose computational complexity would be sufficiently low to make the feature extractor run in reasonable time. Out of the ideas that have been explored, we list the following.

One could work with simplifying assumptions, such as the Gaussianity of the underlying residuals and repeatedly leverage the analytic expression for the distribution of the minimum / maximum of two Gaussian variables [21]. The Gaussianity assumption will, however, be invalid for residuals with a small support, such as the 'minmax54' first-order residual, and this deviation will lead to suboptimality.

Another possibility is to estimate the expectation $E[|z_{ij}^{\min}(\mathbf{Y}) - z_{ij}^{\min}(\mathbf{X})|]$ using Monte-Carlo simulation, which is a rather expensive alternative. Nevertheless, this approach will tell us how much can be theoretically gained.

Experiments

This section contains the results of all experiments. They were conducted on the standard BOSSbase 1.01 [1] database containing 10,000 grayscale images with 512×512 pixels. The detection accuracy is evaluated using the minimal total error probability on the testing set under equal priors, $P_{\mathrm{E}} = \min_{P_{\mathrm{FA}}} \frac{1}{2}(P_{\mathrm{FA}} + P_{\mathrm{MD}})$, returned by the FLD ensemble [18] averaged over ten 50/50 splits of the database into a pair of training and testing sets.

Before moving to the actual experiments, we summarize the terminology. The version of maxSRM with the quantity $E[|z_{ij}(\mathbf{Y}) - z_{ij}(\mathbf{X})|]$ accumulated in co-occurrences will be denoted σmaxSRM. We note that for linear residuals σ_{ij} is computed using (24) while for minmax residuals, it is obtained using Monte Carlo simulations by embedding the image under investigation 500 times. For the PSRM, we only use the 'spam' type submodels corresponding to linear residuals. This gives our feature set dimensionality of 1,980. This feature set will be abbreviated σspamPSRM.

Our first experiment demonstrates the potential of the proposed idea. We work with σmaxSRMq2d2 (d2 standing for the d2 scan of co-occurrences) with 12,753 features. Table 1 shows the results for three steganographic schemes, WOW [13], HILL [19], and MVG [26] with ternary embedding and Gaussian pixel residual model, and two payloads contrasting the detection error for the original maxSRMq2d2 and the proposed σmaxSRMq2d2. The improvement in the detection error $\overline{P}_{\mathrm{E}}$ ranges from 0.3% to almost 1.5%, depending on the embedding algorithm and payload.

The second experiment was executed with the spam part of the PSRM (with linear residuals only). We compare the spamPSRM subset of PSRM with σspamPSRM (both dimensionality 1,980) because no other selection-channel-aware version of PSRM currently exists. The results appear in Figures 1–4. The improvement in the detection error is significant across all embedding algorithms and payloads, especially for HILL and WOW where the improvement in $\overline{P}_{\mathrm{E}}$ ranges between 2.5% and 6%. In fact, this relatively small σspamPSRM with 1,980 features for HILL and MVG achieves comparable or better detection error than the computationally much more expensive maxSRMd2 (34,671 features). For HILL with payload 0.4 bpp the σspamPSRM improves on maxSRMd2 by 0.5% and even on PSRM by 2.3%. Only in the case of S-UNIWARD the σspamPSRM does not significantly improve on spamPSRM.

Conclusions

Detection of modern content-adaptive steganography requires detectors built using machine learning fed with examples of cover and stego objects represented in a feature space. Currently, it is an open problem how to choose a suitable feature representation that would incorporate the knowledge of the embedding change probabilities of individual image elements, the selection channel. These probabilities are approximately available to the steganalyst because the pixel costs that were used for embedding can

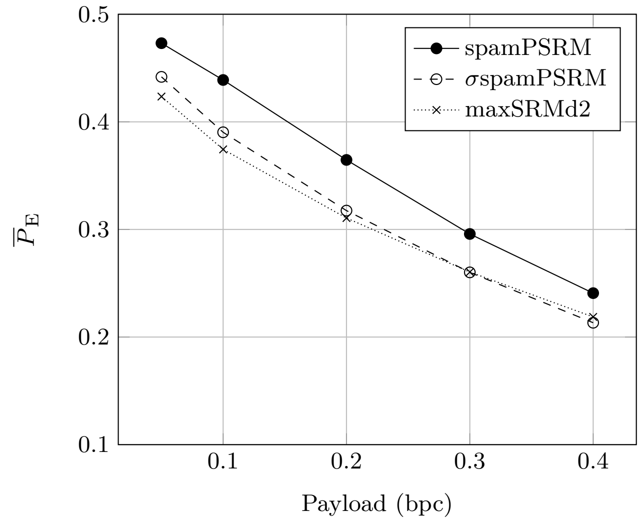

Figure 1. *Detection error $\overline{P}_{\mathrm{E}}$ for HILL with spamPSRM, σspamPSRM, and maxSRMd2*

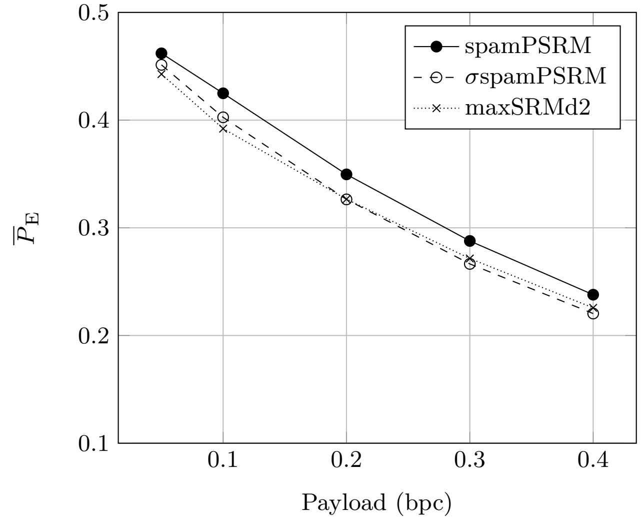

Figure 2. *Detection error $\overline{P}_{\mathrm{E}}$ for MVG with spamPSRM, σspamPSRM, and maxSRMd2*

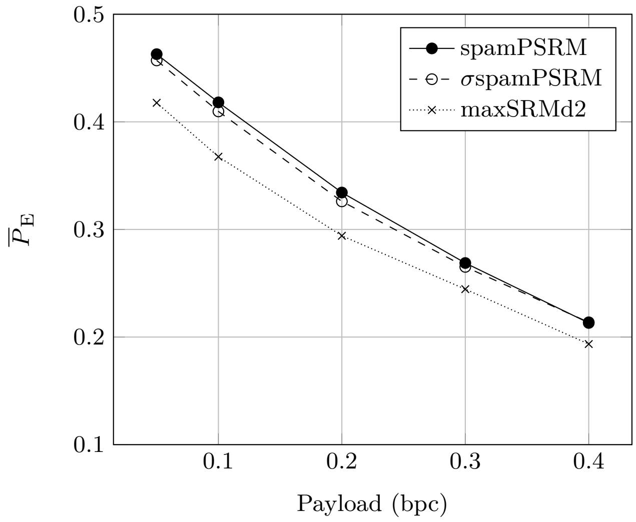

Figure 3. *Detection error $\overline{P}_{\mathrm{E}}$ for S-UNIWARD with spamPSRM, σspamPSRM, and maxSRMd2*

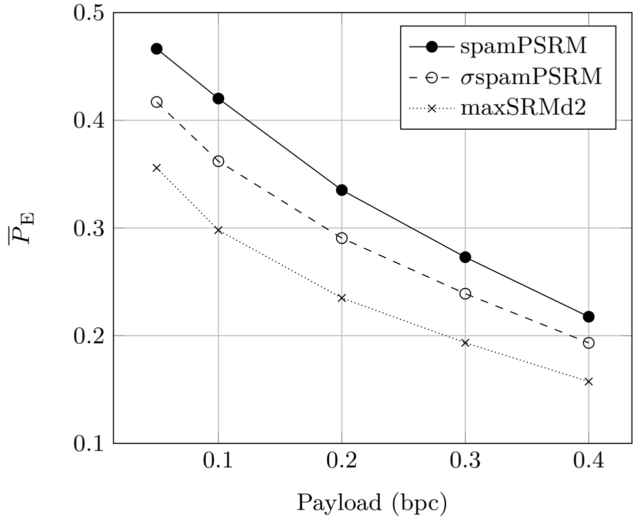

Figure 4. *Detection error $\overline{P}_{\mathrm{E}}$ for WOW with spamPSRM, σspamPSRM, and maxSRMd2*

be relatively accurately estimated from the stego image and because the imprecise knowledge of the payload size does not affect steganalysis accuracy much.

The current approach (the so-called maxSRM feature set) calls for accumulating the pixel change probabilities in co-occurrences of noise residuals. However, because potentially many pixels contribute to one residual sample, one should compute the statistical impact of the embedding changes on the residual and accumulate this quantity instead. To this end, in this paper we propose the expected value of the L_1 residual distortion due to embedding. For linear pixel predictors, the impact of embedding on the residual is easily obtained from the independence of embedding changes and the assumed Gaussianity of the distortion. For non-linear (min-max) predictors, however, the expectation of the L_1 distortion is difficult to obtain analytically due to the necessity to compute the expectation of a minimum (maximum) of up to five dependent random variables that themselves depend on up to 25 pixels. In this paper, we compute such expectations using Monte Carlo simulations.

The proposed idea is applied to the SRM feature set and a subset of the PSRM that is built only from linear residuals (dimensionality 1,980). This reduction of the PSRM feature vector was needed to keep the computational complexity low. Experiments with three embedding schemes and the SRMq2d2 feature set showed that the proposed quantity indeed improves the detection by 0.5–1.5% depending on the embedding algorithm and payload. In the case of the PSRM, the improvement was quite substantial. Compared with the same subset of the original PSRM, the detection error dropped by up to 6% and was comparable and sometimes even slightly lower (for HILL and MVG) than using the entire (and much more computationally demanding) maxSRMd2 model.

We wish to stress that the proposed modification of the rich models does not increase their dimensionality. When the models are restricted only to the subset obtained from linear residuals, the increase in the computational complexity is negligible since the expectation of the L_1 distortion for one residual can be obtained using three convolutions.

Finally, this framework opens up the possibility to extend selection awareness to features computed in the spatial domain for steganalysis of JPEG steganographic algorithms. This applies only to algorithms that are adaptive to content, such as J-UNIWARD [16] and UED [10, 11]. The proposed approach, suitably modified to keep a low computational complexity indeed provides a significant detection boost [3].

The feature extractor code for σspamSRM is available from `http://dde.binghamton.edu/download/feature_extractors/`.

Acknowledgments

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.

References

[1] P. Bas and T. Filler and T. Pevný, "Break Our Steganographic System – the Ins and Outs of Organizing BOSS," in Information Hiding, 13th International Conference, vol. 6958 of Lecture Notes in Computer Science, pp. 59-70 , Prague, Czech Republic, 2011.

[2] L. Chen and Y.Q. Shi and P. Sutthiwan and X. Niu, "A Novel Mapping Scheme for Steganalysis," in International Workshop on Digital Forensics and Watermaking, vol. 7809 of Lecture Notes in Computer Science, pp. 19–33, Springer Berlin Heidelberg, 2013.

[3] T. Denemark and M. Boroumand and J. Fridrich, "Steganalysis Features for Content-Adaptive JPEG Steganography," IEEE Transactions on Information Forensics and Security, under review, 2015.

[4] T. Denemark and J. Fridrich, "Improving Steganographic Security by Synchronizing the Selection Channel," in 3rd ACM IH&MMSec. Workshop, pp. 5–14, Portland, Oregon, 2015.

[5] T. Denemark and V. Sedighi and V. Holub and R. Cogranne and J. Fridrich, "Selection-Channel-Aware Rich Model for Steganalysis of Digital Images," in IEEE International Workshop on Information Forensics and Security, Atlanta, Georgia, 2014.

[6] T. Filler and J. Judas and J. Fridrich, "Minimizing Additive Distortion in Steganography Using Syndrome-Trellis Codes," IEEE Transactions on Information Forensics and Security, vol. 6(3), pp. 920–935, 2011.

[7] J. Fridrich and J. Kodovský, "Rich Models for Steganalysis of Digital Images," IEEE Transactions on Information Forensics and Security, vol. 7(3), pp. 868–882, 2011.

[8] M. Goljan and R. Cogranne and J. Fridrich, "Rich Model for Steganalysis of Color Images," in IEEE International Workshop on Information Forensics and Security, Atlanta, Georgia, 2014.

[9] M. Goljan and J. Fridrich, "CFA-Aware Features for Steganalysis of Color Images," in Proceedings SPIE, Electronic Imaging, Media Watermarking, Security, and Forensics 2015, vol. 9409, pp. 94090V, San Francisco, California, 2015.

[10] L. Guo and J. Ni and Y.Q. Shi, "An Efficient JPEG Steganographic Scheme Using Uniform Embedding," in IEEE International Workshop on Information Forensics and Security, pp. 169–174, Tenerife, Spain, 2012.

[11] L. Guo and J. Ni and Y.Q. Shi, "Uniform Embedding for Efficient JPEG Steganography," IEEE Transactions on Information Forensics and Security, vol. 9(5), pp. 814–825, 2014.

[12] V. Holub and J. Fridrich, "Low Complexity Features for JPEG Steganalysis Using Undecimated DCT," IEEE Transactions on Information Forensics and Security, vol. 10(2), pp. 219–228, 2015.

[13] V. Holub and J. Fridrich, "Designing Steganographic Distortion Using Directional Filters," in IEEE International Workshop on Information Forensics and Security, pp. 234–239, Tenerife, Spain, 2012.

[14] V. Holub and J. Fridrich, "Random Projections of Residuals for Digital Image Steganalysis," IEEE Transactions on Information Forensics and Security, vol. 8(12), pp. 1996–2006, 2013.

[15] V. Holub and J. Fridrich, "Phase-Aware Projection Model for Steganalysis of JPEG Images," in Proceedings SPIE, Electronic Imaging, Media Watermarking, Security, and Forensics 2015, vol. 9409, pp. 94090T, San Francisco, California, 2015.

[16] V. Holub and 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, vol. 2014(1), pp. 1–13, 2014.

[17] J. Kodovský and J. Fridrich, "Steganalysis of JPEG Images Using Rich Models," in Proceedings SPIE, Electronic Imaging, Media Watermarking, Security, and Forensics 2012, vol. 8303, pp. 83030A, San Francisco, California, 2012.

[18] J. Kodovský and J. Fridrich and V. Holub, "Ensemble Classifiers for Steganalysis of Digital Media," IEEE Transactions on Information Forensics and Security, vol. 7(2), pp. 432–444, 2012.

[19] B. Li and M. Wang and J. Huang, "A new cost function for spatial image steganography," in Proceedings IEEE, International Conference on Image Processing, ICIP, pp. 4206-4210, Paris, France, 2014.

[20] B. Li and M. Wang and X. Li and S. Tan and J. Huang, "A Strategy of Clustering Modification Directions in Spatial Image Steganography," IEEE Transactions on Information Forensics and Security, vol. 10(9), pp. 1905–1917, 2015.

[21] S. Nadarajah and S. Kotz, "Exact Distribution of the Max-Min of Two Gaussian Random Variables," IEEE Transactions on VLSI Systems, vol. 16(2), pp. 210–212, 2008.

[22] T. Pevný and T. Filler and P. Bas, "Using High-Dimensional Image Models to Perform Highly Undetectable Steganography," in Information Hiding, 12th International Conference, vol. 6387 of Lecture Notes in Computer Science, pp. 161–177, Calgary, Canada, 2010.

[23] P. Schöttle and S. Korff and R. Böhme, "Weighted Stego-Image Steganalysis for Naive Content-Adaptive Embedding," in IEEE International Workshop on Information Forensics and Security, pp. 193–198, Tenerife, Spain, 2012.

[24] V. Sedighi and R. Cogranne and J. Fridrich, "Content-adaptive steganography by minimizing statistical detectability" IEEE Transactions on Information Forensics and Security, vol. 11(2), pp. 221–234, 2016.

[25] V. Sedighi and J. Fridrich, "Effect of Imprecise Knowledge of the Selection Channel on Steganalysis," in 3rd ACM IH&MMSec. Workshop, pp. 33–42, Portland, Oregon, 2015.

[26] V. Sedighi and 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, vol. 9409, pp. 94090H, San Francisco, California, 2015.

[27] X. Song and F. Liu and X. Luo and Y. Zhang, "Steganalysis of Adaptive JPEG Steganography Using 2D Gabor Filters," in 3rd ACM IH&MMSec. Workshop, pp. 15–23, Portland, Oregon, 2015.

[28] W. Tang and H. Li and W. Luo and J. Huang, "Adaptive Steganalysis Against WOW Embedding Algorithm," in 2nd ACM IH&MMSec. Workshop, pp. 91–96, Salzburg, Austria, 2014.

Author Biography

Tomáš Denemark received his MS in mathematics from the Czech Technical University in Prague in 2012 and now pursues his PhD at Binghamton University. He focuses on steganography and steganalysis.

Jessica Fridrich is Professor of Electrical and Computer Engineering at Binghamton University. She received her PhD in Systems Science from Binghamton University in 1995 and MS in Applied Mathematics from Czech Technical University in Prague in 1987. Her main interests are in steganography, steganalysis, and digital image forensic. Since 1995, she has received 20 research grants totaling over $9 mil that lead to more than 160 papers and 7 US patents.

Pedro Comesaña-Alfaro received both the Telecommunications Engineering (specialized in both Computer Science and Communications) and Ph. D degrees from the University of Vigo, Spain, in 2002 and 2006, respectively. Since 2012 he is Associate Professor in the School of Telecommunications Engineering, University of Vigo. His research interests lie in the areas of digital watermarking, information security, multimedia forensics and digital communications. He received the Best Paper Award of IEEE-WIFS 2014 and IWDW 2011.