Steganalysis Features for Content-Adaptive JPEG Steganography

Tomáš Denemark, *Student Member, IEEE*, Mehdi Boroumand, *Student Member, IEEE*, and Jessica Fridrich, *Fellow, IEEE*

Abstract—**All modern steganographic algorithms for digital images are content adaptive in the sense that they restrict the embedding modifications to complex regions of the cover which are difficult to model for the steganalyst. The probabilities with which the individual cover elements are modified (the selection channel) are determined jointly by the size of the embedded payload and content complexity. The most accurate detection of content-adaptive steganography is currently achieved with detectors built as classifiers trained on cover and stego features that incorporate the knowledge of the selection channel. While selection-channel-aware features have been proposed for detection of spatial domain steganography, an equivalent for the JPEG domain does not exist. Since modern steganographic algorithms for JPEG images are currently best detected with features formed by histograms of noise residuals split by their JPEG phase, we use such feature sets as a starting point in this paper and extend their design to incorporate the knowledge of the selection channel. This is achieved by accumulating in the histograms a quantity that bounds the expected absolute distortion of the residual. The proposed features can be computed efficiently and provide a substantial detection gain across all tested algorithms especially for small payloads.**

Index Terms—**Steganalysis, adaptive steganography, selection channel, JPEG, detection, security.**

I. Introduction

Today, the most secure steganographic schemes for digital images represented either in the spatial or JPEG domain are content adaptive in the sense that they execute embedding changes primarily in complex regions of the cover image capitalizing on the inability of the steganalyst to detect the traces of embedding in content that is hard to model [1]–[7]. Today's detectors of such schemes are built using machine learning, such as binary classifiers

The following authors are with the Department of Electrical and Computer Engineering, Binghamton University, NY, 13902, USA. Email: {tdenema1,mboroum1,fridrich}@binghamton.edu

trained on examples of cover and stego images represented with higher-order statistics of noise residuals, the so-called rich media models. This also applies to modern steganographic schemes that hide messages in quantized DCT coefficients from a JPEG file, UED (Uniform Embedding Distortion) [6], [7], and J-UNIWARD [5]. The most accurate detection of such JPEG steganography is currently achieved with features that are computed in the spatial domain [8]–[11] rather than from quantized DCT coefficients [12].

A potential weakness of content-adaptive schemes is that the rule that drives the distribution of the embedding change probabilities among individual elements of the cover is, by the Kerckhoffs' principle, also available to the steganalyst, who can use it to improve the detection. Recently, the spatial rich model (SRM) [13] has been modified to incorporate content adaptivity within the feature design [2], [14]. This was achieved in a heuristic manner by accumulating some function of the embedding change probabilities in co-occurrences of noise residuals. Such selection-channel-aware features improve the detection of adaptive algorithms to a varying degree depending on how strong the content adaptivity is. The embedding algorithm WOW [1] suffered the most from such attacks while the security of S-UNIWARD [5], HILL [3], and MVG [15] decreased only marginally.

The way the selection channel is incorporated in SRM cannot be used for detection of JPEG steganography because the embedding and the steganalysis domains are different. In particular, the embedding changes applied to an 8×8 block of quantized DCT coefficients affect all 64 pixels and the modifications are no longer limited to ± 1 changes but can have a much larger amplitude depending also on the JPEG quality factor. Pixel change rate thus no longer properly characterizes the distortion at a pixel. On the other hand, knowing the embedding change probabilities of quantized DCT coefficients it is possible to compute the expected value of the distortion at each pixel. In this paper, we show that by accumulating such a quantity in histograms of JPEG-phase-aware noise residuals [8]–[10], it is possible to construct spatial rich features that provide more accurate detection of current content-adaptive JPEG algorithms. The improvement appears to be the largest for small payloads and diminishes for large payloads when the embedding algorithm loses most of its content adaptivity.

This paper starts in the next section with a summary of basic concepts and notational conventions. In Section III,

we review steganalysis features that utilize JPEG phase. These features are subsequently made aware of the selection channel in Section IV. Guided by the requirement of a reasonable computational complexity, we select the most promising and computationally efficient variant of the quantity that will be accumulated in the histograms of residuals forming the feature vector. At the end of this section, we describe the final form of the selection-channel-aware features in a pseudo-code. In Section V, we subject the newly proposed features to tests on three content-adaptive JPEG algorithms on a wide range of payloads and two JPEG quality factors. We also investigate the robustness of the features to imprecisely determined payload and selection channel. The paper is summarized in Section VI where we also discuss possible extensions of this work.

II. NOTATION AND BASIC CONCEPTS

Boldface font is reserved for vectors and matrices, the calligraphic font for sets, and capital letters will be used to denote random variables with the corresponding lower-case letter as their realizations. The elements of a matrix will be denoted with the corresponding italic font with subscript indices. The cardinality of a finite set $\mathcal{B}$ is denoted $|\mathcal{B}|$. We use the notation $[P]$ for the Iverson bracket $[P] = 1$ when P is true and $[P] = 0$ when P is false.

For easier technical description, we only consider $n_1 \times n_2$ 8-bit grayscale images with n_1 and n_2 multiples of 8. A JPEG image will be represented with an array of quantized DCT (discrete cosine transform) coefficients of the same dimensions as the pixel representation of the image, $\mathbf{c} \in \{-1023, \ldots, 1024\}^{n_1 \times n_2}$. Often, it will be useful to consider a block representation of $\mathbf{c}$. The (a,b)th 8×8 block of DCT coefficients, $1 \le a \le n_1/8$, $1 \le b \le n_2/8$, is formed by c_{kl} with k,l restricted to $1 + 8(a-1) \le k \le 8a$, $1 + 8(b-1) \le l \le 8b$, and will be denoted $\mathbf{c}^{(a,b)}$. The individual elements of $\mathbf{c}^{(a,b)}$ are $c_{kl}^{(a,b)}$, this time with $0 \le k \le 7$, $0 \le l \le 7$, hoping that no confusion will be created by using the indices k,l for two different purposes – when used in c_{kl}, their range is $1 \le k \le n_1, 1 \le l \le n_2$ while in a block, as in $c_{kl}^{(a,b)}$, their range is $0, \ldots, 7$.

The (k,l)th DCT basis, $0 \le k,l \le 7$, is an 8×8 matrix $\mathbf{f}^{(k,l)} = (f_{ij}^{(k,l)})$, $0 \le i,j \le 7$:

$$f_{ij}^{(k,l)} = \frac{w_k w_l}{4} \cos \frac{\pi k(2i+1)}{16} \cos \frac{\pi l(2j+1)}{16}, \quad (1)$$

where $w_0 = 1/\sqrt{2}$ and $w_k = 1$ for $k > 0$. By decompressing the (a,b)th block of DCT coefficients, we obtain a corresponding block of 8×8 pixels $x_{ij}^{(a,b)}$, $0 \le i,j \le 7$:

$$x_{ij}^{(a,b)} = \sum_{k,l=0}^{7} f_{kl}^{(i,j)} q_{kl} c_{kl}^{(a,b)}, \quad (2)$$

where q_{kl} are the elements of the JPEG luminance quantization matrix. Note that in (2), the pixel values are *not rounded*. Putting all blocks into one $n_1 \times n_2$ matrix,

the decompressed (non-rounded) image is represented with a matrix $\mathbf{x} \in \mathbb{R}^{n_1 \times n_2}$.

Finally, we note that a,b will be strictly used to index blocks, k,l for DCT coefficients, and i,j for pixels with the same range and conventions applied to both i,j and k,l.

III. FEATURES BASED ON JPEG PHASE

Today, there exist numerous steganalysis features that are suitable for detection of JPEG steganography. Early embedding schemes, such as F5 [16], model-based steganography [17], Jsteg [18], OutGuess [19], and Steghide [20], are best detected using statistics formed from quantized DCT coefficients, such as the JPEG Rich Model (JRM) [12]. Unfortunately, JRM is far less effective for detecting modern JPEG steganography, examples of which are UED [6], [7] and J-UNIWARD [5], which adapt their embedding changes to cover content. Such schemes are currently best detected with features assembled as histograms of noise residuals split by their JPEG phase defined as the location w.r.t. the 8×8 pixel grid: DCT Residuals (DCTR) [8], PHase Aware Rich Model (PHARM) [9], and Gabor Filter Residuals (GFR) [10]. The splitting by phase is effective because the impact of the stego signal on pixels in a decompressed JPEG image depends on the JPEG phase.

In this section, we provide enough detail about steganalysis features based on JPEG phase to be able to explain in the next section their new proposed variant that incorporates the knowledge of the selection channel.

The DCTR [8], PHARM [9], and GFR [10] features are formed from noise residuals computed by convolving the decompressed (non-rounded) JPEG image $\mathbf{x}$ (2) with kernel $\mathbf{g} \in \mathbb{R}^{k_1 \times k_2}$,

$$\mathbf{r}(\mathbf{x}, \mathbf{g}) = \mathbf{x} \star \mathbf{g}. \quad (3)$$

We note that because the convolution uses no padding (implemented with 'valid' in Matlab), $\mathbf{r} \in \mathbb{R}^{n_1' \times n_2'}$ with $n_1' = n_1 - k_1 + 1$ and $n_2' = n_2 - k_2 + 1$. Next, the residual is quantized,

$$\mathbf{r}(\mathbf{x}, \mathbf{g}, Q) = Q_{\mathcal{Q}}(\mathbf{r}(\mathbf{x}, \mathbf{g})/q), \quad (4)$$

where $Q_{\mathcal{Q}}$ is a quantizer with centroids $\mathcal{Q} = \{0, 1, 2, \ldots, T\}$, q is a fixed quantization step and T a truncation threshold. Each residual is used to compute the following 64 histograms, $0 \le m \le T$, $0 \le i,j \le 7$:

$$h_m^{(i,j)}(\mathbf{x}, \mathbf{g}, Q) = \sum_{a=1}^{\lfloor n_1'/8 \rfloor} \sum_{b=1}^{\lfloor n_2'/8 \rfloor} [|r_{ij}^{(a,b)}(\mathbf{x}, \mathbf{g}, Q)| = m]. \quad (5)$$

All $T+1$ values, $h_0^{(i,j)}, \ldots, h_T^{(i,j)}$, from each histogram are concatenated into a vector of $64 \times (T+1)$ values and these vectors are then concatenated for kernels $\mathbf{b}$ from some filter bank $\mathcal{B}$. To reduce the feature dimensionality, $64 \times (T+1) \times |\mathcal{B}|$, and make the bins better populated, certain bins in the concatenated histograms are merged based on symmetries of $\mathbf{g}$ and DCT bases. The DCTR

feature set uses a filter bank with $|\mathcal{B}_{DCTR}| = 64$ kernels $\mathbf{g}$ corresponding to 64 DCT bases $\mathbf{f}^{(k,l)}$, $0 \leq k, l \leq 7$. In PHARM, the kernels are obtained by convolving nine small-support pixel predictors with 100 random projection kernels (a total of $|\mathcal{B}_{PHARM}| = 900$ kernels), while in GFR $|\mathcal{B}_{GFR}| = 256$ Gabor filters (four support sizes, two Gabor phases, and 32 orientations) are employed. We note that the size, $k_1 \times k_2$, of the kernels in all three feature sets satisfies $1 \leq k_1, k_2 \leq 15$, which means that no kernel ever intersects more than four 8×8 pixel blocks. Finally, we wish to point out that the PHARM feature vector as described in [9] uses only $T = 2$ bins as the $T + 1$st bin is redundant (the sum of all three bins is equal to the number of pixels). In our case, however, we will be accumulating some other quantity in the histogram bins and it will no longer be true that the sum of the bins is constant. This increases the dimensionality of the the proposed version of PHARM (see the next section) from the original 12,600 to 18,900.

IV. RESIDUAL DISTORTION MEASURE

To incorporate the selection channel into the feature design, we inspired ourselves with the selection-channel-aware version of the SRM [13] called maxSRM [14], where the co-occurrences of noise residuals accumulated the embedding change probabilities. Porting this concept directly to the features from the previous section for steganalysis of JPEG images is, however, not possible because the embedding changes are executed in the DCT domain and the embedding modifies the pixel values in the decompressed JPEG image $\mathbf{x}$ by a wide range of values rather than by ± 1.

Our approach is inspired by the following observation. In the pixel domain, when the embedding changes by ± 1 are equiprobable with probability β, the change rate is one half of the expected absolute value (or a square) of the pixel embedding distortion: $\frac{1}{2}\left(\beta|1| + \beta|-1|\right) = \frac{1}{2}\left(\beta(1)^2 + \beta(-1)^2\right) = \beta$. Thus, an equivalent quantity for JPEG-domain steganography would be the expected distortion of the noise residual (3). Because the embedding changes in the DCT domain by ± 1 are again equiprobable, the expected distortion in the pixel domain is zero due to the linearity of inverse DCT and the linearity of the convolution. To measure the distortion, it is thus natural to use some measure of the statistical spread, such as the expected value of the square of the residual distortion or its absolute value. To this end, we first derive the properties of the random variable representing the embedding distortion in the residual domain and then investigate several different quantities of statistical spread as a distortion measure. The criterion we use to select the final measure is driven by computational complexity.

We denote the quantized DCT coefficients in the (a, b)th block of the cover and stego image by $c_{kl}^{(a,b)}$ and $s_{kl}^{(a,b)} = c_{kl}^{(a,b)} + w_{kl}^{(a,b)}$, respectively, where $w_{kl}^{(a,b)}$ are the embedding changes, which are independent realizations of random variables $W_{kl}^{(a,b)}$ attaining the values in $\{-1, 0, 1\}$

with probabilities $\{\beta_{kl}^{(a,b)}, 1 - 2\beta_{kl}^{(a,b)}, \beta_{kl}^{(a,b)}\}$ determined by the steganographic scheme and the payload size. We stress that this model of embedding fits all modern JPEG steganographic algorithms, including both versions of UED and J-UNIWARD. Recalling (2), the difference between the non-rounded pixel values in the decompressed cover and stego images, $x_{ij}^{(a,b)} = \sum_{k,l=0}^{7} f_{kl}^{(i,j)} q_{kl} c_{kl}^{(a,b)}$ and $y_{ij}^{(a,b)} = \sum_{k,l=0}^{7} f_{kl}^{(i,j)} q_{kl} s_{kl}^{(a,b)}$, respectively, is:

$$z_{ij}^{(a,b)} = y_{ij}^{(a,b)} - x_{ij}^{(a,b)} = \sum_{k,l=0}^{7} f_{kl}^{(i,j)} q_{kl} w_{kl}^{(a,b)}. \qquad (6)$$

Because the embedding changes are mutually independent and because

$$E[W_{kl}^{(a,b)}] = 0, \qquad (7)$$

$$Var[W_{kl}^{(a,b)}] = 2\beta_{kl}^{(a,b)}, \qquad (8)$$

we have

$$E[Z_{ij}^{(a,b)}] = 0, \qquad (9)$$

$$Var[Z_{ij}^{(a,b)}] = 2\sum_{k,l=0}^{7} (f_{kl}^{(i,j)})^2 q_{kl}^2 \beta_{kl}^{(a,b)}, \qquad (10)$$

where we remind that, by our convention, $Z_{ij}^{(a,b)} = \sum_{k,l=0}^{7} f_{kl}^{(i,j)} q_{kl} W_{kl}^{(a,b)}$ is the random variable whose realization is $z_{ij}^{(a,b)}$.

From (3) and the linearity of convolution, the residual distortion, the difference between the residuals of stego and cover images, $\boldsymbol{\rho} \in \mathbb{R}^{n_1' \times n_2'}$, can thus be expressed as

$$\boldsymbol{\rho}(\mathbf{w}) = \mathbf{r}(\mathbf{y}, \mathbf{g}) - \mathbf{r}(\mathbf{x}, \mathbf{g}) = \mathbf{z}(\mathbf{w}) \star \mathbf{g}. \qquad (11)$$

Technically, $\boldsymbol{\rho}$ also depends on the kernel $\mathbf{g}$ but, in order to declutter the notation, we only explicitly write the dependence on the embedding changes $\mathbf{w}$ as these are the most important. Since the kernels $\mathbf{g}$ for the features discussed in the previous section never intersect more than four different 8×8 pixel blocks, when computing a specific value of $\rho_{ij}^{(a,b)}(\mathbf{w})$ (11), it will generally depend on either one 8×8 block, when the kernel is positioned within one JPEG block, two blocks when the kernel straddles two adjacent blocks, or four blocks. Because the inverse DCT is linear and the residual also depends linearly on the non-rounded pixel values, each value of $\rho_{ij}^{(a,b)}(\mathbf{w})$ is thus a linear combination of 64, 128, or 256 values of $\mathbf{w}$ coming from four 8×8 blocks. In order to formalize this linear relationship, we will associate a given residual value (and thus a value of $\rho_{ij}^{(a,b)}$) with the position of the upper left corner of the kernel $\mathbf{g}$ when performing the convolution. Due to this convention, the value of $\rho_{ij}^{(a,b)}(\mathbf{w})$ is thus a linear combination of w_{kl} from four blocks with block indices (a, b), $(a+1, b)$, $(a, b+1)$, and $(a+1, b+1)$. Introducing the pair of indices $(u, v) \in \{0, 1\} \times \{0, 1\}$, we can write

$$\rho_{ij}^{(a,b)}(\mathbf{w}) = \sum_{k,l=0}^{7} \sum_{u,v=0}^{1} \alpha_{kl}^{(u,v)}(i, j, \mathbf{g}) w_{kl}^{(a+u,b+v)}, \qquad (12)$$

where $0 \le u, v \le 1$ and $0 \le i, j, k, l \le 7$. Following (6) and (11), the coefficients $\alpha_{kl}^{(u,v)}(i, j, \mathbf{g})$ depend on the kernel $\mathbf{g}$, and the phase (i, j). From (6), we see that they also depend on the coefficients of the inverse DCT and the quantization steps q_{kl}. They are, however, independent of the content or the embedding scheme and can be in principle computed. In this paper, we will not need their explicit form and will only do with the fact that the distortion $\rho_{ij}^{(a,b)}$ can be expressed this way (12).

To better explain the coefficients $\alpha_{kl}^{(u,v)}$, we note that, for example, since the DCTR and GFR features use 8×8 kernels $\mathbf{g}$, for phase $(0,0)$ only 64 values of $\alpha_{kl}^{(u,v)}(i, j, \mathbf{g})$ will generally be non-zero. For phases $(0, k)$, $(k, 0)$, $k > 0$, there will be 128 non-zero values, and for the remaining 49 phases there will be 256 non-zero $\alpha_{kl}^{(u,v)}(i, j, \mathbf{g})$.

Because $E[W_{kl}^{(a,b)}] = 0$ for all (a, b) and k, l, we have $E[\rho_{ij}^{(a,b)}(\mathbf{W})] = 0$ as well. Thus, we will take some measure, δ, of the statistical spread of the distortion $\rho_{ij}^{(a,b)}(\mathbf{W})$ as a quantity that should be accumulated in the histograms of residuals (5) in a similar fashion as the embedding change probabilities are accumulated in the selection-channel-aware maxSRM [14], $0 \le m \le T$, $0 \le i, j \le 7$:

$$\overline{h}_m^{(i,j)}(\mathbf{x}, \mathbf{g}, Q, \boldsymbol{\beta}) =$$
$$\sum_{a=1}^{\lfloor n_1'/8 \rfloor} \sum_{b=1}^{\lfloor n_2'/8 \rfloor} [|r_{ij}^{(a,b)}(\mathbf{x}, \mathbf{g}, Q)| = m] \cdot \delta(\rho_{ij}^{(a,b)}(\mathbf{W})). \quad (13)$$

In (13), $\overline{h}_m^{(i,j)}$ stands for the selection-channel-aware version of the histograms (5) and $(i, j) \in \{0, \ldots, 7\}^2$ is the JPEG phase. Note that since the distribution of $\mathbf{W}$ depends on $\boldsymbol{\beta}$ (8), so does $\overline{h}_m^{(i,j)}$.

The two most frequently used measures of statistical spread of a random variable are the standard deviation and the expectation of the absolute value, the latter being considered as a more robust measure. We thus study the following measures of statistical spread of $\rho_{ij}^{(a,b)}(\mathbf{W})$ (c.f., (12)) :

$$\delta_{std}(\boldsymbol{\beta})_{ij}^{(a,b)} = \sqrt{Var[\rho_{ij}^{(a,b)}(\mathbf{W})]}$$
$$= \sqrt{2 \sum_{k,l=0}^{7} \sum_{u,v=0}^{1} (\alpha_{kl}^{(u,v)}(i,j,\mathbf{g}))^2 \beta_{kl}^{(a+u,b+v)}} \quad (14)$$

$$\delta_{EA}(\boldsymbol{\beta})_{ij}^{(a,b)} = E[|\rho_{ij}^{(a,b)}(\mathbf{W})|]. \quad (15)$$

Note that the distribution of the random variable $\mathbf{W}$ is fully described using $\boldsymbol{\beta}$. This is why in (14)–(15), we point out the dependency of δ_{std} and δ_{EA} on the change rates $\boldsymbol{\beta}$. To clarify the above expressions, $\delta_{std}(\boldsymbol{\beta})_{ij}^{(a,b)}$ stands for the ijth element in the (a,b)th block in matrix $\delta_{std}(\boldsymbol{\beta}) \in \mathbb{R}^{n_1' \times n_2'}$ and the same applies to $\delta_{EA}(\boldsymbol{\beta})$. Note that both $\delta_{std}(\boldsymbol{\beta})$ and $\delta_{EA}(\boldsymbol{\beta})$ depend on the change rates $\boldsymbol{\beta}$ (the selection channel), which is an $n_1 \times n_2$ array of embedding change probabilities arranged in the same fashion as the DCT coefficients, and on the kernel $\mathbf{g}$.

Neither (14) or (15) are, unfortunately, suitable for practical usage. The standard deviation $\delta_{std}(\boldsymbol{\beta})$ can be computed for all (a, b) and a given kernel $\mathbf{g}$ using one convolution $\mathbf{A} \star \boldsymbol{\beta}$, where $\mathbf{A}$ is a 16×16 matrix with four 8×8 blocks $\mathbf{A}^{(u,v)} = ((\alpha_{kl}^{(u,v)}(i, j, \mathbf{g}))^2)_{k,l=0}^7$

$$\mathbf{A}(i, j, \mathbf{g}) = \begin{pmatrix} \mathbf{A}^{(0,0)} & \mathbf{A}^{(0,1)} \\ \mathbf{A}^{(1,0)} & \mathbf{A}^{(1,1)} \end{pmatrix}. \quad (16)$$

However, because there are 64 phases and $|\mathcal{B}|$ kernels, one thus needs to compute $64 \times |\mathcal{B}|$ convolutions, which is rather expensive even for the smallest filter bank of DCTR and completely prohibitive for PHARM with 900 filters.

The problem with $\delta_{EA}(\boldsymbol{\beta})$ is that it cannot be computed analytically and Monte Carlo estimation requires at least 200 simulated embeddings to obtain a value accurate within 10% (determined experimentally for DCTR and J-UNIWARD at 0.4 bpnzac, bits per non-zero AC DCT coefficient). This increases the number of required convolutions by a factor of 200. One possibility is to approximate the sum (over all indices) in $\rho_{ij}^{(a,b)}(\mathbf{w})$ (12) with a Gaussian random variable $\mathcal{N}(0, \sigma^2)$ for which one can easily verify that $E[|\mathcal{N}(0, \sigma^2)|] = 2\sigma/\sqrt{2\pi}$. Unfortunately, this brings us back to the prohibitive complexity of evaluating the variance. Also, note that with this approximation $\delta_{EA}(\boldsymbol{\beta})$ and $\delta_{std}(\boldsymbol{\beta})$ coincide.

To resolve the complexity issues, we turned our attention to how the JPEG-phase-aware features are formed [8]–[10]. They are computed in two steps by first decompressing the JPEG image to the spatial domain and then evaluating merely $|\mathcal{B}|$ convolutions. To substantially decrease the complexity, we will strive to keep a similar two-stage process. To achieve this goal, we switch to an upper bound of $|\rho_{ij}^{(a,b)}(\mathbf{w})|$:

$$|\boldsymbol{\rho}(\mathbf{w})| \le |\mathbf{z}| \star |\mathbf{g}|, \quad (17)$$

and then further bound

$$|z_{ij}^{(a,b)}| \le \sum_{k,l=0}^{7} |f_{kl}^{(i,j)}| \cdot q_{kl} \cdot |w_{kl}^{(a,b)}|. \quad (18)$$

Because $E[|W_{kl}^{(a,b)}|] = 2\beta_{kl}^{(a,b)}$, we have for the expectation

$$E[|Z_{ij}^{(a,b)}|] \le 2 \sum_{k,l=0}^{7} |f_{kl}^{(i,j)}| q_{kl} \beta_{kl}^{(a,b)} \triangleq t_{ij}^{(a,b)}(\boldsymbol{\beta}). \quad (19)$$

Finally, using (17)–(19) $\delta_{EA}(\boldsymbol{\beta})$ can be bounded by

$$\delta_{EA}(\boldsymbol{\beta}) = E[|\boldsymbol{\rho}(\mathbf{W})|] \le \mathbf{t}(\boldsymbol{\beta}) \star |\mathbf{g}| \triangleq \delta_{uSA}(\boldsymbol{\beta}), \quad (20)$$

which can be efficiently evaluated by first computing $t_{ij}^{(a,b)} = 2 \sum_{k,l=0}^{7} |f_{kl}^{(i,j)}| q_{kl} \beta_{kl}^{(a,b)}$ by blocks (this is as computationally demanding as decompressing a JPEG image) and then convolving $\mathbf{t}(\boldsymbol{\beta})$ with the absolute value of the kernel. We used the subscript 'uSA' (Upper bounded Sum of Absolute values) for the bounding quantity.

We observed an approximately quadratic dependence between δ_{uSA} and δ_{EA}, $\delta_{uSA} \propto \delta_{EA}^2$, when used within the DCTR, PHARM, and GFR features. Thus, to obtain
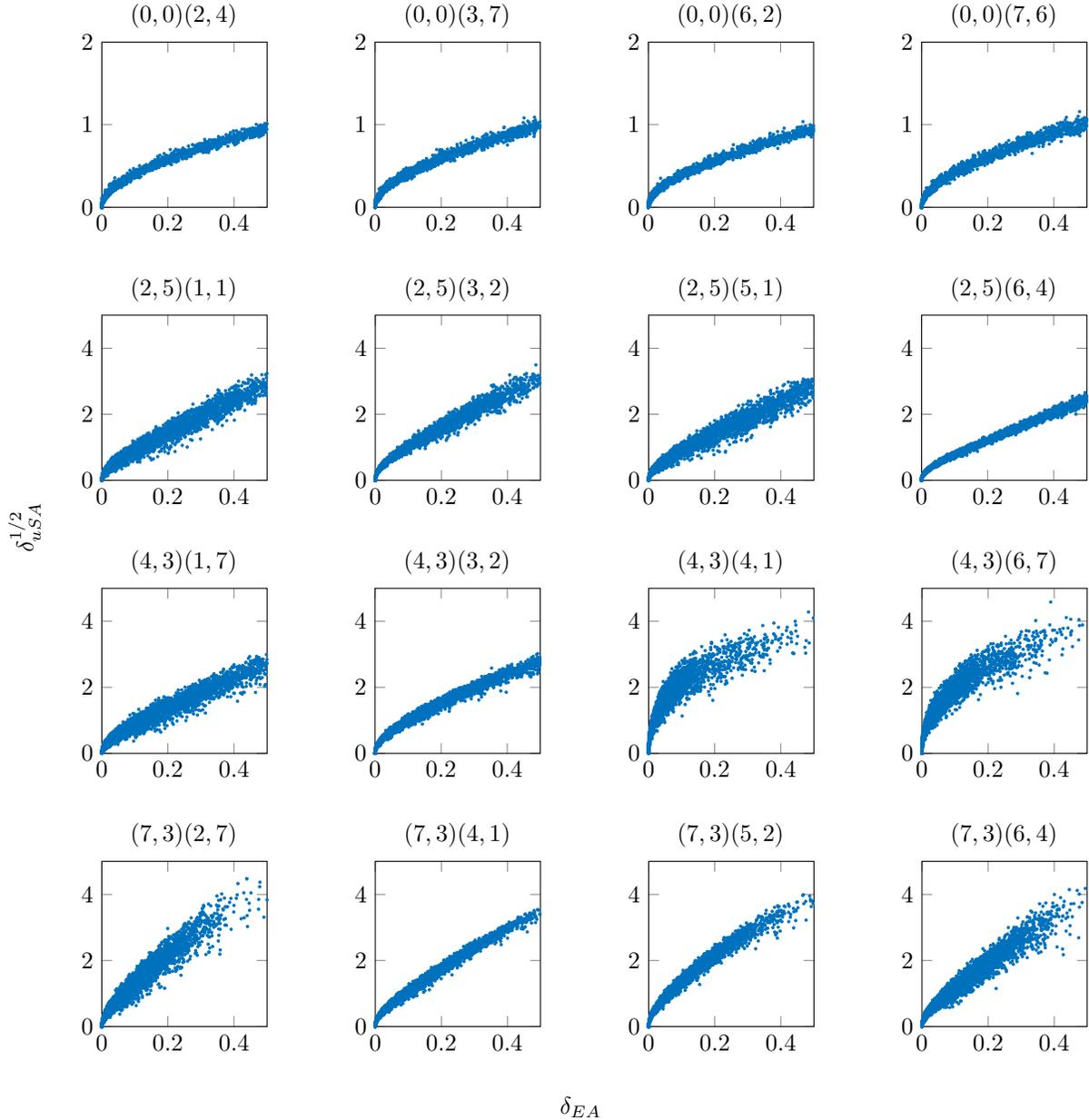
Figure 1. Plot of $\delta_{uSA}^{1/2}$ versus δ_{EA} for one BOSSbase image for the DCTR filter bank. The first number pair above each scatter plot indicates the DCTR kernel (the spatial frequency of the DCT mode) while the second pair is the JPEG phase. Note that the square root forces an approximate linear relationship between both quantities.

a quantity that is more closely related to the expectation of the residual distortion, we use the square root $\delta_{uSA}^{1/2}(\boldsymbol{\beta})$ meaning that it is applied to the $n_1' \times n_2'$ matrix $\delta_{uSA}(\boldsymbol{\beta})$ elementwise. The above claims are supported by Figure 1, which shows $\delta_{uSA}^{1/2}(\boldsymbol{\beta})_{ij}^{(a,b)}$ versus $\delta_{EA}(\boldsymbol{\beta})_{ij}^{(a,b)}$ across all blocks (a,b) for sixteen different combinations of DCTR kernels $\mathbf{g}$ and JPEG phases. The values of δ_{EA} were obtained using Monte Carlo simulations by embedding the image '1013.pgm' from BOSSBase 1000-times. The first ordered pair above each plot shows the spatial frequency of the DCT kernel while the second ordered pair is the JPEG phase i,j. Note that with the exception of kernel-phase combinations $(4,3),(4,1)$ and $(4,3),(6,7)$, there appears

to be an approximate linear relationship between $\delta_{uSA}^{1/2}$ and δ_{EA}. Qualitatively similar results were observed for the PHARM and GFR filter banks. The square root thus indeed makes $\delta_{uSA}^{1/2}(\boldsymbol{\beta})$ a rather good (and much more computationally efficient!) approximation of $\delta_{EA}(\boldsymbol{\beta})$.

A. Final feature design

We now summarize in pseudo-code and in Figure 2 the final design of the features that will be subjected to experimental tests in the next section. In the pseudo-code below, $\boldsymbol{\beta} \in \mathbb{R}^{n_1 \times n_2}$ is the selection channel in the form of a matrix of embedding change probabilities of DCT coefficients arranged in the same fashion as unquantized

DCT coefficients, $f_{kl}^{(i,j)}$ are the DCT bases (1), and q_{kl} is the 8×8 JPEG luminance quantization matrix of the investigated image.

1) Select a JPEG-phase-aware feature set, which is equivalent to selecting the filter bank $\mathcal{B} \in \{\mathcal{B}_{DCTR}, \mathcal{B}_{PHARM}, \mathcal{B}_{GFR}\}$.

2) Decompress the JPEG image under investigation to the spatial domain (apply (2) by blocks), denote the non-rounded pixel values with $\mathbf{x}$.

3) For each filter (kernel) $\mathbf{g} \in \mathcal{B}$:

   a) compute the residual $\mathbf{r}(\mathbf{x}, \mathbf{g}) = \mathbf{g} \star \mathbf{x}$ and quantize it $\mathbf{r}(\mathbf{x}, \mathbf{g}, Q) = Q_{\mathcal{Q}}(\mathbf{r}(\mathbf{x}, \mathbf{g})/q)$.

   b) compute $\mathbf{t}(\boldsymbol{\beta}) \in \mathbb{R}^{n_1 \times n_2}$ by blocks, $t_{ij}^{(a,b)}(\boldsymbol{\beta}) = \sum_{k,l=0}^{7} |f_{kl}^{(i,j)}| q_{kl} \beta_{kl}^{(a,b)}$ for all blocks (a,b).

   c) evaluate $\delta_{uSA}^{1/2}(\boldsymbol{\beta}) = \sqrt{\mathbf{t}(\boldsymbol{\beta}) \star |\mathbf{g}|}$ (square root applied in an elementwise fashion to all elements of the $n_1' \times n_2'$ matrix $\mathbf{t}(\boldsymbol{\beta}) \star |\mathbf{g}|$).

   d) Compute the following $64 \times (T+1)$ values $\overline{h}_m^{(i,j)}(\mathbf{x}, \mathbf{g}, Q, \boldsymbol{\beta})$, $0 \le m \le T$, $0 \le i,j \le 7$:

$$\overline{h}_m^{(i,j)}(\mathbf{x}, \mathbf{g}, Q, \boldsymbol{\beta}) = \sum_{a=1}^{\lfloor n_1'/8 \rfloor} \sum_{b=1}^{\lfloor n_2'/8 \rfloor} [|r_{ij}^{(a,b)}(\mathbf{x}, \mathbf{g}, Q)| = m] \cdot \delta_{uSA}^{1/2}(\boldsymbol{\beta})_{ij}^{(a,b)}.$$

$$(21)$$

4) Concatenate $\overline{h}_m^{(i,j)}(\mathbf{x}, \mathbf{g}, Q, \boldsymbol{\beta})$, $\mathbf{g} \in \mathcal{B}$, $0 \le i,j \le 7$, $0 \le m \le T$, and form the final feature vector using the same symmetrization rules as those used for forming the JPEG-phase-aware features from $h_m^{(i,j)}(\mathbf{x}, \mathbf{g}, Q)$ (5).

Note that in the pseudo-code, in contrast to (19) we removed the multiplicative factor "2" from $t_{ij}^{(a,b)}(\boldsymbol{\beta})$ as it does not change the detection performance.

V. Experimental results

In this section, we subject the selection-channel-aware features described in Section IV to tests on real imagery. The experiments are conducted on the standard database BOSSbase 1.01 [21] containing 10,000 grayscale images with 512×512 pixels. We ran the experiments on JPEG images with quality factors 75 and 95.

The steganographic algorithms tested in this section are the original version of UED [7] (UED-SC), its improved version [6] (UED-JC), and J-UNIWARD as described in [5]. We use three JPEG-phase-aware steganalysis feature sets: DCTR [8], PHARM [9], and Gabor Filter Bank (GFR) [10]. We stress that we always compute the quantity $\delta_{uSA}^{1/2}$ from the image under investigation. Thus, if the image is a stego image, we used the change rates $\boldsymbol{\beta}$ computed from costs obtained from the stego image. As such, they will generally be slightly different than the change rates used for embedding because of the effect of embedding changes themselves. It is necessary to carry out the experiments this way because this is exactly what would be happening in practice.

The detection accuracy is evaluated using the minimal total error probability under equal priors, $P_{\mathrm{E}} = \min_{P_{\mathrm{FA}}} \frac{1}{2}(P_{\mathrm{FA}} + P_{\mathrm{MD}})$, achieved on the test set averaged over ten 50/50 splits of the database. The symbols P_{FA} and P_{MD} stand for the false-alarm and missed-detection rates. The classifier is the FLD ensemble [22]. To inform the reader about the statistical significance of the improvements, we state that the mean absolute deviation of P_{E} over the ten ensemble runs ranges between 0.0005 and 0.0046, depending on the feature set, embedding algorithm, payload, and JPEG quality factor (also see Table II).

Figure 3 shows the average detection error $\overline{P}_{\mathrm{E}}$ as a function of payload in bits per non-zero AC DCT coefficient (bpnzac) for three steganographic algorithms and two JPEG quality factors and payloads ranging from 0.05–0.5 bpnzac. The exact numerical values are in Table II. For easy comprehension, color is used to highlight the embedding algorithm. Each combination of the embedding algorithm, payload, and JPEG quality factor has two partially overlapping bars with the solid color fill showing the performance of the selection-channel-aware features computed with $\delta_{uSA}^{1/2}$ while the original features correspond to the patterned column.

The GFR feature set always offers the most accurate detection irrespectively of the feature type, embedding algorithm, payload, and quality factor. Making the features aware of the selection channel generally improves the detection for payload smaller than 0.3 bpnzac. The gain is larger for quality factor 75 than for 95. In some cases, the detection error drops by as much as 8% (UED-JC for 75 quality factor). With increasing payload, the gain decreases, which is natural because the embedding algorithms become less adaptive. For large payloads embedded with UED, the gain may even become negative (the detection slightly worsens). We verified that this loss is not due to the fact that the embedding change probabilities extracted from the stego and cover images differ as the loss remains unchanged when computing the features with embedding probabilities of the cover. This thus points to a small inefficiency associated with the quantity $\delta_{uSA}^{1/2}$ accumulated in the histograms. The significant detection gain for small payloads far outweighs this small loss as it is more difficult to detect smaller payloads.

Even though the goal of this paper is not to benchmark steganography, it is interesting that the order of the three tested steganographic schemes by their empirical security does not change when switching to selection-channel-aware features and does not depend on the feature type either.

A. Robustness study

Making features aware of the selection channel implicitly assumes that the size of the embedded payload is known or known at least approximately. A natural question to ask is whether the lack of knowledge about the payload size has any negative effect on detection accuracy and how big it
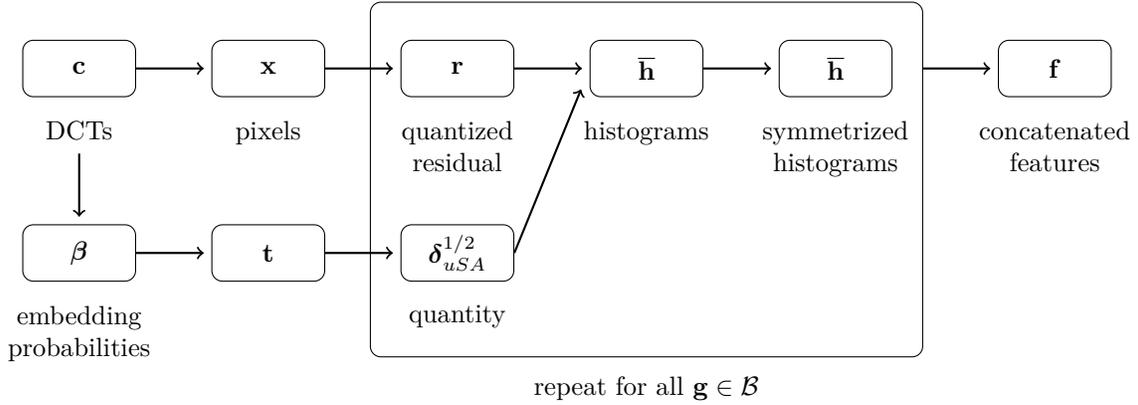
Figure 2. Diagram describing the proposed algorithm for the extracting of the $\delta_{uSA}^{1/2}$ features.

is when compared with the original version of the features that do not incorporate the selection channel. Note that if the payload size is known only approximately, not only the selection channel will be imprecise but also the stego images used for training the classifier will be embedded with a different payload. In other words, the Warden will be faced with a stego-source mismatch. The problem of classifier training with unknown payload has previously been investigated in [23]. Although a detailed study of this topic is clearly outside of the scope of our work, we believe that some limited study has its place in the current paper. In the first experiment in this section, we contrast the difference in detection loss due to stego-source mismatch when the original and selection-channel-aware features are used. We remark that studying only the situation when the selection channel is imprecise (but the stego images are created with the correct payload) does not make sense because, as stated above, the Warden either knows or does not know the payload.

Additionally, it is also worth investigating how much the embedding changes themselves affect the selection channel. When computing the quantity $\delta_{uSA}^{1/2}$ from a stego image, the change rates are obtained from an image that has been modified by embedding and will thus generally be different than the change rates used for embedding (computed from the cover image). The impact of this imprecision on selection-channel-aware features is studied in the second experiment of this section.

Robustness experiment 1. The purpose of this experiment is to assess the loss of detection accuracy when detecting images embedded with relative payload α_{tst} while training a classifier on payload α_{trn}. Since we deal with three features, two quality factors, and three embedding methods, we selected only two cases that we report on in detail (the other cases are qualitatively similar). They correspond to J-UNIWARD with DCTR and $\delta_{uSA}^{1/2}$-DCTR features on 75% quality JPEGs (Figure 4 left) and UED-JC with GFR and $\delta_{uSA}^{1/2}$-GFR features on 75% quality JPEGs (Figure 4 right). Both figures show the detection error $\overline{P}_\mathrm{E}$ as a function of α_{trn}. Each curve corresponds to

one testing payload α_{tst} (differentiated by markers) and one feature set (differentiated by line style). To read the graph, first select a test payload α_{tst} and a feature set (e.g., select one curve). To see the detection error when classifying with a detector trained for α_{trn}, move on the curve left and right. The increase in $\overline{P}_\mathrm{E}$ when moving away from the point on the curve corresponding to $\alpha_{trn} = \alpha_{tst}$ thus informs us about the loss of detection. The figure shows that the loss of detection due to stego-source mismatch for DCTR and $\delta_{uSA}^{1/2}$-DCTR is quite comparable, meaning that the selection-channel-aware DCTR does not suffer from the stego-source mismatch any more than the original DCTR features. Also, the loss is only slowly increasing with the difference $\alpha_{tst} - \alpha_{trn}$, which is comforting to know for practical applications. The figure also shows that while $\delta_{uSA}^{1/2}$-GFR for large α_{tst} does not improve on GFR, it is much more stable to the stego-source mismatch. For example, in the case of UED-JC on 75% quality with $\alpha_{tst} = 0.5$ bpnzac and $\alpha_{trn} = 0.05$ bpnzac the detection error of $\delta_{uSA}^{1/2}$-GFR is 14% lower than that of GFR, even though there is no gain when $\alpha_{trn} = \alpha_{tst} = 0.05$ bpnzac.

Robustness experiment 2. In Table I, we provide a limited scale experiment with J-UNIWARD and UED-JC on JPEG quality factor 75, reporting $\overline{P}_\mathrm{E}$ when the change rates are always computed from covers (even when we are extracting a feature from a stego image) and when they are extracted from the corresponding image. The differences in detection accuracy are well within the statistical spread and thus statistically insignificant. This is in agreement with what was reported in [24], namely that the effect of the embedding changes on selection-channel estimation, and subsequent steganalysis, is negligible.

VI. Conclusions

Steganalysis of content-adaptive steganography needs to take into account the probabilities with which the embedding modifies individual cover elements. However, incorporating this prior probabilistic knowledge (the selection channel) within detectors built as classifiers trained on examples of cover and stego features is quite challenging. The main complication stems from the fact that the
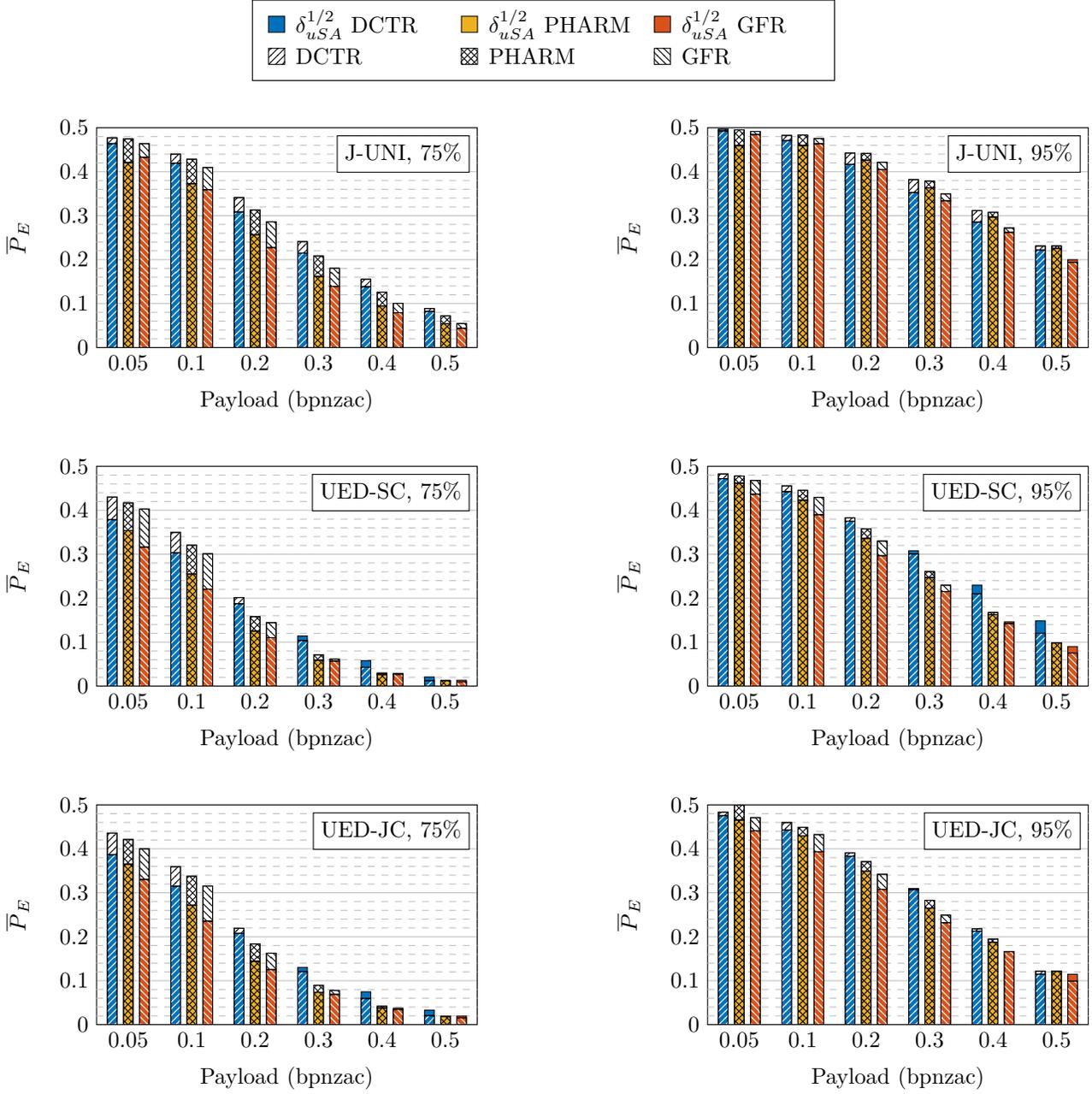
Figure 3. Detection error $\overline{P}_\text{E}$ for three steganographic algorithms for DCTR, GFR, and PHARM features (patterns) and their selection-aware $\delta_{uSA}^{1/2}$ version (solid fill) versus payload, JPEG quality factors 75 and 95.

quantity from which steganalysis features are formed are quantized noise residuals extracted in the pixel domain. When the embedding modifies JPEG DCT coefficients, the impact of embedding on residuals becomes even more complicated. Fortunately, if the residuals are obtained in a linear fashion from pixels, e.g., by convolving the image with a kernel, because the embedding changes are independent, it is possible to derive the impact of embedding on residuals analytically.

In this paper, we investigate several quantities that measure the expected embedding distortion in the residual domain when embedding in the JPEG domain. In order to obtain a distortion measure that can be evaluated

with acceptable computational complexity, we consider an upper bound on the mean absolute residual distortion and transform it in a non-linear manner to make it strongly correlate with the true mean value. The resulting quantity can be efficiently computed using convolutions and is accumulated in residual histograms of three feature sets that are aware of the JPEG phase: DCTR, PHARM, and Gabor Filter Residuals (GFR). These feature sets were selected because they currently provide the most accurate detection of modern steganography in JPEG domain. The selection-channel-aware versions of these features provide further significant detection gain of content-adaptive JPEG steganography, especially for small payloads.
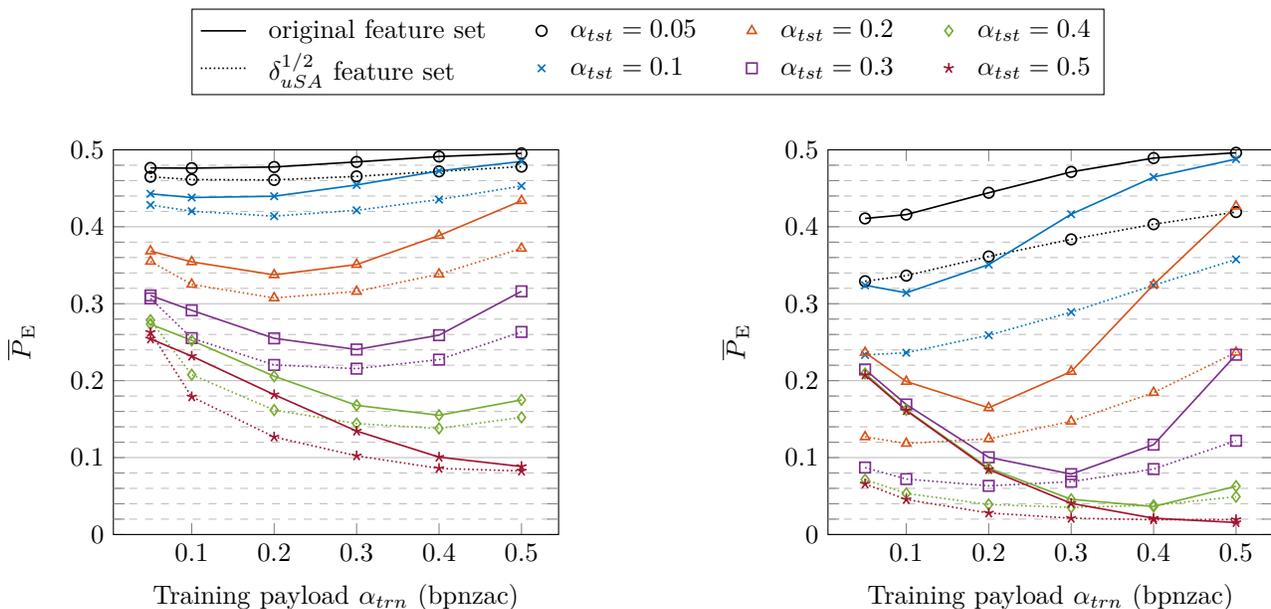
Figure 4. Detection error $\overline{P}_\mathrm{E}$ when training the detector on stego images embedded with payload α_{trn} and testing on α_{tst}. One testing payload corresponds to one curve. Left: J-UNIWARD with DCTR and $\delta_{uSA}^{1/2}$-DCTR, Right: UED-JC with GFR and $\delta_{uSA}^{1/2}$-GFR, both for JPEG quality factor 75.

Table I
DETECTION ERROR $\overline{P}_\mathrm{E}$ FOR J-UNIWARD AND UED-JC WITH $\delta_{uSA}^{1/2}$-DCTR FEATURES WITH CHANGE RATES COMPUTED ALWAYS FROM THE COVER (COVER) AND WITH CHANGE RATES COMPUTED FROM THE IMAGE UNDER INVESTIGATION (TRUE); JPEG QUALITY FACTOR 75. THE STANDARD DEVIATION OF THE RESULTS RANGES BETWEEN 0.0013 AND 0.0030.

J-UNIWARD	0.1	0.2	0.3	0.4	0.5
$\delta_{uSA}^{1/2}$ (cover)	0.4159	0.3083	0.2166	0.1389	0.0836
$\delta_{uSA}^{1/2}$ (true)	0.4189	0.3084	0.2160	0.1401	0.0833
UED-JC					
$\delta_{uSA}^{1/2}$ (cover)	0.3193	0.2101	0.1343	0.0789	0.0335
$\delta_{uSA}^{1/2}$ (true)	0.3179	0.2102	0.1336	0.0785	0.0337

We also investigated the loss of detection accuracy due to imprecise selection channel either due to unknown payload or due to the stego changes themselves. The selection-channel-aware version of the features does not appear any more sensitive to stego-source mismatch than the original feature sets, and for some combinations of mismatched testing and training payloads, they even appear more robust to the stego-source mismach. In agreement with previous art, while the embedding changes themselves introduce a slight imprecision into the selection channel (and thus into the quantity accumulated in histograms) they have a negligible effect on detection accuracy.

The main innovative concept coined in this paper goes beyond building features for JPEG steganography and should be adopted for detection of spatial-domain embedding as well. Indeed, the value of a noise residual is always affected by more than one pixel – by the entire support of the residual kernel. Thus, considering only the embedding change probability of one pixel to which the residual is attributed, as is done in the current state of the art, the maxSRM, is only an approximation of the probabilistic impact of embedding on the *residual*. Since virtually all embedding schemes modify pixel values by ± 1 with equal probabilities, the embedding change probability is proportional to the expected value of the mean absolute distortion of the *pixel*. Thus, in principle the approach described in this paper can and should be applied to steganalysis of spatial-domain steganography as well. With feature sets like the SRM or the projection SRM (PSRM), however, there is one significant complication due to the fact that these feature sets utilize non-linear (min-max) residuals. Since neighboring min-max residuals are dependent, computing the expected absolute distortion of the residual can be quite involved and becomes intractable for residuals with a large support. This topic is elaborated in more detail in [25].

The code for all algorithms (steganographic methods, feature extractors, and classifiers) is available for download from http://dde.binghamton.edu/download/.

REFERENCES

[1] V. Holub and J. Fridrich, "Designing steganographic distortion using directional filters," in *Fourth IEEE International Workshop on Information Forensics and Security*, (Tenerife, Spain), December 2–5, 2012.

[2] W. Tang, H. Li, W. Luo, and J. Huang, "Adaptive steganalysis against WOW embedding algorithm," in *2nd ACM IH&MMSec. Workshop* (A. Uhl, S. Katzenbeisser, R. Kwitt, and A. Piva, eds.), (Salzburg, Austria), pp. 91–96, June 11–13, 2014.

Table II

DETECTION ERROR $\overline{P}_E$ FOR THREE STEGANOGRAPHIC SCHEMES FOR DCTR, GFR, AND PHARM FEATURES AND THEIR SELECTION-AWARE $\delta_{uSA}^{1/2}$ VERSION FOR SELECTED PAYLOADS, JPEG QUALITY FACTORS 75 AND 95.

J-UNI, QF 75%	0.05 bpp	0.1 bpp	0.2 bpp	0.3 bpp	0.4 bpp	0.5 bpp
DCTR	0.4769±0.0013	0.4400±0.0017	0.3412±0.0017	0.2410±0.0030	0.1553±0.0023	0.0887±0.0025
$\delta_{uSA}^{1/2}$ DCTR	0.4635±0.0028	0.4192±0.0015	0.3081±0.0021	0.2148±0.0026	0.1380±0.0019	0.0818±0.0015
GFR	0.4638±0.0019	0.4095±0.0013	0.2861±0.0037	0.1804±0.0029	0.1005±0.0021	0.0546±0.0018
$\delta_{uSA}^{1/2}$ GFR	0.4325±0.0016	0.3589±0.0025	0.2272±0.0026	0.1389±0.0019	0.0792±0.0016	0.0437±0.0008
PHARM	0.4746±0.0024	0.4284±0.0024	0.3131±0.0039	0.2096±0.0029	0.1259±0.0027	0.0720±0.0017
$\delta_{uSA}^{1/2}$ PHARM	0.4204±0.0157	0.3727±0.0069	0.2567±0.0041	0.1626±0.0026	0.0945±0.0019	0.0542±0.0008

UED-SC, QF 75%	0.05 bpp	0.1 bpp	0.2 bpp	0.3 bpp	0.4 bpp	0.5 bpp
DCTR	0.4301±0.0020	0.3497±0.0026	0.2013±0.0011	0.1034±0.0019	0.0427±0.0011	0.0120±0.0005
$\delta_{uSA}^{1/2}$ DCTR	0.3786±0.0018	0.3030±0.0024	0.1871±0.0028	0.1140±0.0027	0.0578±0.0020	0.0206±0.0008
GFR	0.4029±0.0028	0.3010±0.0036	0.1443±0.0030	0.0617±0.0011	0.0266±0.0009	0.0096±0.0006
$\delta_{uSA}^{1/2}$ GFR	0.3159±0.0023	0.2204±0.0018	0.1103±0.0018	0.0567±0.0008	0.0290±0.0011	0.0131±0.0008
PHARM	0.4168±0.0027	0.3207±0.0041	0.1580±0.0016	0.0711±0.0022	0.0296±0.0009	0.0121±0.0006
$\delta_{uSA}^{1/2}$ PHARM	0.3541±0.0023	0.2548±0.0028	0.1251±0.0043	0.0589±0.0019	0.0263±0.0015	0.0120±0.0007

UED-JC, QF 75%	0.05 bpp	0.1 bpp	0.2 bpp	0.3 bpp	0.4 bpp	0.5 bpp
DCTR	0.4358±0.0024	0.3593±0.0020	0.2192±0.0030	0.1204±0.0020	0.0593±0.0014	0.0203±0.0009
$\delta_{uSA}^{1/2}$ DCTR	0.3865±0.0020	0.3146±0.0025	0.2075±0.0032	0.1303±0.0023	0.0747±0.0026	0.0328±0.0014
GFR	0.4100±0.0023	0.3153±0.0024	0.1627±0.0022	0.0779±0.0016	0.0346±0.0018	0.0153±0.0007
$\delta_{uSA}^{1/2}$ GFR	0.3301±0.0040	0.2352±0.0028	0.1252±0.0008	0.0685±0.0017	0.0377±0.0008	0.0193±0.0010
PHARM	0.4213±0.0029	0.3376±0.0043	0.1837±0.0023	0.0895±0.0025	0.0418±0.0005	0.0187±0.0011
$\delta_{uSA}^{1/2}$ PHARM	0.3646±0.0023	0.2714±0.0031	0.1440±0.0029	0.0736±0.0017	0.0374±0.0018	0.0186±0.0009

J-UNI, QF 95%	0.05 bpp	0.1 bpp	0.2 bpp	0.3 bpp	0.4 bpp	0.5 bpp
DCTR	0.4965±0.0013	0.4826±0.0019	0.4424±0.0032	0.3820±0.0029	0.3081±0.0025	0.2310±0.0020
$\delta_{uSA}^{1/2}$ DCTR	0.4924±0.0042	0.4705±0.0026	0.4163±0.0034	0.3524±0.0036	0.2851±0.0030	0.2217±0.0027
GFR	0.4915±0.0019	0.4756±0.0013	0.4215±0.0016	0.3496±0.0018	0.2721±0.0028	0.1936±0.0023
$\delta_{uSA}^{1/2}$ GFR	0.4843±0.0015	0.4634±0.0027	0.4046±0.0024	0.3338±0.0029	0.2617±0.0038	0.1998±0.0026
PHARM	0.4953±0.0017	0.4835±0.0018	0.4416±0.0032	0.3787±0.0023	0.3079±0.0026	0.2311±0.0038
$\delta_{uSA}^{1/2}$ PHARM	0.4592±0.0108	0.4594±0.0026	0.4260±0.0024	0.3634±0.0016	0.2968±0.0025	0.2256±0.0026

UED-SC, QF 95%	0.05 bpp	0.1 bpp	0.2 bpp	0.3 bpp	0.4 bpp	0.5 bpp
DCTR	0.4826±0.0019	0.4555±0.0017	0.3829±0.0015	0.3004±0.0025	0.2095±0.0015	0.1201±0.0019
$\delta_{uSA}^{1/2}$ DCTR	0.4719±0.0026	0.4420±0.0025	0.3752±0.0026	0.3079±0.0020	0.2296±0.0017	0.1484±0.0022
GFR	0.4679±0.0018	0.4294±0.0016	0.3299±0.0022	0.2295±0.0035	0.1420±0.0027	0.0753±0.0012
$\delta_{uSA}^{1/2}$ GFR	0.4366±0.0031	0.3896±0.0019	0.2964±0.0028	0.2149±0.0031	0.1462±0.0029	0.0899±0.0031
PHARM	0.4779±0.0025	0.4455±0.0029	0.3577±0.0035	0.2605±0.0032	0.1674±0.0024	0.0982±0.0021
$\delta_{uSA}^{1/2}$ PHARM	0.4616±0.0021	0.4229±0.0030	0.3360±0.0042	0.2467±0.0044	0.1624±0.0025	0.0983±0.0028

UED-JC, QF 95%	0.05 bpp	0.1 bpp	0.2 bpp	0.3 bpp	0.4 bpp	0.5 bpp
DCTR	0.4835±0.0021	0.4598±0.0025	0.3908±0.0025	0.3095±0.0029	0.2180±0.0026	0.1216±0.0017
$\delta_{uSA}^{1/2}$ DCTR	0.4753±0.0026	0.4426±0.0039	0.3830±0.0040	0.3069±0.0025	0.2128±0.0027	0.1146±0.0030
GFR	0.4709±0.0014	0.4323±0.0016	0.3420±0.0024	0.2492±0.0028	0.1663±0.0028	0.0990±0.0027
$\delta_{uSA}^{1/2}$ GFR	0.4411±0.0020	0.3931±0.0021	0.3077±0.0025	0.2316±0.0034	0.1662±0.0035	0.1107±0.0024
PHARM	0.4994±0.0016	0.4490±0.0019	0.3708±0.0039	0.2828±0.0018	0.1947±0.0022	0.1220±0.0022
$\delta_{uSA}^{1/2}$ PHARM	0.4653±0.0026	0.4297±0.0030	0.3487±0.0030	0.2651±0.0033	0.1875±0.0026	0.1217±0.0032

[3] B. Li, M. Wang, J. Huang, and X. Li, "A new cost function for spatial image steganography," in *Proceedings IEEE, International Conference on Image Processing, ICIP*, (Paris, France), October 27–30, 2014.

[4] V. Sedighi, J. Fridrich, and R. Cogranne, "Content-adaptive pentary steganography using the multivariate generalized Gaussian cover model," in *Proceedings IS&T, Electronic Imaging, Media Watermarking, Security, and Forensics 2016* (A. Alattar and N. D. Memon, eds.), vol. 9409, (San Francisco, CA), February 8–12, 2015.

[5] 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*, vol. 2014:1, 2014.

[6] L. Guo, J. Ni, and Y. Q. Shi, "Uniform embedding for efficient JPEG steganography," *IEEE Transactions on Information Forensics and Security*, vol. 9, no. 5, pp. 814–825, 2014.

[7] L. Guo, J. Ni, and Y. Q. Shi, "An efficient JPEG steganographic scheme using uniform embedding," in *Fourth IEEE International Workshop on Information Forensics and Security*, (Tenerife, Spain), pp. 169–174, December 2–5, 2012.

[8] V. Holub and J. Fridrich, "Low complexity features for JPEG steganalysis using undecimated DCT," *IEEE Transactions on Information Forensics and Security*, vol. 10, no. 2, pp. 219–228, 2015.

[9] V. Holub and J. Fridrich, "Phase-aware projection model for steganalysis of JPEG images," in *Proceedings IS&T, Electronic Imaging, Media Watermarking, Security, and Forensics 2016* (A. Alattar and N. D. Memon, eds.), vol. 9409, (San Francisco, CA), February 8–12, 2015.

[10] X. Song, F. Liu, C. Yang, X. Luo, and Y. Zhang, "Steganalysis of adaptive JPEG steganography using 2D Gabor filters," in *3rd ACM IH&MMSec. Workshop* (P. Comesana, J. Fridrich, and A. Alattar, eds.), (Portland, Oregon), June 17–19, 2015.

[11] V. Holub and J. Fridrich, "Challenging the doctrines of JPEG steganography," in *Proceedings SPIE, Electronic Imaging, Media Watermarking, Security, and Forensics 2014* (A. Alattar, N. D. Memon, and C. Heitzenrater, eds.), vol. 9028, (San Francisco, CA), pp. 02 1–8, February 3–5, 2014.

[12] J. Kodovský and J. Fridrich, "Steganalysis of JPEG images using rich models," in *Proceedings SPIE, Electronic Imaging, Media Watermarking, Security, and Forensics 2012* (A. Alattar, N. D. Memon, and E. J. Delp, eds.), vol. 8303, (San Francisco, CA), pp. 0A 1–13, January 23–26, 2012.

[13] J. Fridrich and J. Kodovský, "Rich models for steganalysis of digital images," *IEEE Transactions on Information Forensics and Security*, vol. 7, pp. 868–882, June 2011.

[14] T. Denemark, V. Sedighi, V. Holub, 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, GA), December 3–5, 2014.

[15] V. Sedighi, R. Cogranne, and J. Fridrich, "Content-adaptive steganography by minimizing statistical detectability," *IEEE Transactions on Information Forensics and Security*, no. 2, pp. 221–234, 2016.

[16] A. Westfeld, "High capacity despite better steganalysis (F5 – a steganographic algorithm)," in *Information Hiding, 4th International Workshop* (I. S. Moskowitz, ed.), vol. 2137 of Lecture Notes in Computer Science, (Pittsburgh, PA), pp. 289–302, Springer-Verlag, New York, April 25–27, 2001.

[17] P. Sallee, "Model-based methods for steganography and steganalysis," *International Journal of Image Graphics*, vol. 5, no. 1, pp. 167–190, 2005.

[18] D. Upham, "Steganographic algorithm JSteg." Software available at http://zooid.org/ paul/crypto/jsteg.

[19] N. Provos, "Defending against statistical steganalysis," in *10th USENIX Security Symposium*, (Washington, DC), pp. 323–335, August 13–17, 2001.

[20] S. Hetzl and P. Mutzel, "A graph-theoretic approach to steganography," in *Communications and Multimedia Security, 9th IFIP TC-6 TC-11 International Conference, CMS 2005* (J. Dittmann, S. Katzenbeisser, and A. Uhl, eds.), vol. 3677 of Lecture Notes in Computer Science, (Salzburg, Austria), pp. 119–128, September 19–21, 2005.

[21] P. Bas, T. Filler, and T. Pevný, "Break our steganographic system – the ins and outs of organizing BOSS," in *Information Hiding, 13th International Conference* (T. Filler, T. Pevný, A. Ker, and S. Craver, eds.), vol. 6958 of Lecture Notes in Computer Science, (Prague, Czech Republic), pp. 59–70, May 18–20, 2011.

[22] J. Kodovský, J. Fridrich, and V. Holub, "Ensemble classifiers for steganalysis of digital media," *IEEE Transactions on Information Forensics and Security*, vol. 7, no. 2, pp. 432–444, 2012.

[23] T. Pevný, "Detecting messages of unknown length," in *Proceedings SPIE, Electronic Imaging, Media Watermarking, Security and Forensics III* (A. Alattar, N. D. Memon, E. J. Delp, and J. Dittmann, eds.), vol. 7880, (San Francisco, CA), pp. OT 1–12, January 23–26, 2011.

[24] V. Sedighi and J. Fridrich, "Effect of imprecise knowledge of the selection channel on steganalysis," in *3rd ACM IH&MMSec. Workshop* (J. Fridrich, P. Comesana, and A. Alattar, eds.), (Portland, Oregon), June 17–19, 2015.

[25] T. Denemark and J. Fridrich, "Improving selection-channel-aware steganalysis features," in *Proceedings IS&T International Symposium on Electronic Imaging 2016* (A. Alattar and N. D. Memon, eds.), (San Francisco, CA), February 14–18 2016.

Tomas Denemark received his MS in mathematics from the Czech Technical University in Prague in 2012 and now pursues the Ph.D degree in the Department of Electrical and Computer Engineering at Binghamton University (SUNY). His research focuses on steganography, steganalysis and machine learning.

Mehdi Boroumand received the B.S. degree from K.N.Toosi University of Technology, Iran and the M.S. degree from Sahand University of Technology, Iran both in Electrical Engineering. He is currently pursuing his Ph.D. degree in Electrical Engineering at Binghamton University (SUNY). His areas of research include steganography, steganalysis and machine learning.

Jessica Fridrich holds the position of Professor of Electrical and Computer Engineering at Binghamton University (SUNY). She has 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, digital watermarking, and digital image forensic. Dr. Fridrich's research work has been generously supported by the US Air Force and AFOSR. Since 1995, she received 19 research grants totaling over $9 mil for projects on data embedding and steganalysis that lead to more than 160 papers and 7 US patents. Dr. Fridrich is an IEEE Fellow and a ACM member.