Selection-channel-aware rich model for Steganalysis of digital images

Tomas Denemark, Vahid Sedighi, Vojtech Holub, Rémi Cogranne, Jessica Fridrich

Selection-Channel-Aware Rich Model for Steganalysis of Digital Images

Tomáš Denemark, Vahid Sedighi, Vojtěch Holub
Department of ECE
Binghamton University
Binghamton, NY 13902-6000
{tdenema1,vsedigh1}@binghamton.edu
vojtech_holub@yahoo.com

Rémi Cogranne
Member, IEEE
ICD - ROSAS - LM2S
Troyes University of Technology
Troyes, France
remi.cogranne@utt.fr

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

***Abstract*—From the perspective of signal detection theory, it seems obvious that knowing the probabilities with which the individual cover elements are modified during message embedding (the so-called probabilistic selection channel) should improve steganalysis. It is, however, not clear how to incorporate this information into steganalysis features when the detector is built as a classifier. In this paper, we propose a variant of the popular spatial rich model (SRM) that makes use of the selection channel. We demonstrate on three state-of-the-art content-adaptive steganographic schemes that even an imprecise knowledge of the embedding probabilities can substantially increase the detection accuracy in comparison with feature sets that do not consider the selection channel. Overly adaptive embedding schemes seem to be more vulnerable than schemes that spread the embedding changes more evenly throughout the cover.**

I. Introduction

Content-adaptive embedding schemes for digital images change individual pixels with probabilities determined from the local pixel neighborhood in order to execute the embedding changes primarily in regions where they are less detectable, such as textures and noisy areas. The first adaptive methods described in the literature were designed for palette images [11]. One could also argue that schemes that hide message bits in non-zero DCT coefficients are naturally content adaptive since the changes strongly correlate with complex content. In 2010, the Edge Adaptive scheme was designed to hide data in pixel pairs with large differences [16]. The real boom of adaptive schemes started with the advancement of coding schemes [9] capable of embedding messages while nearly optimally minimizing arbitrarily defined additive distortion functions. Examples include HUGO [17], WOW [13], and the UNIWARD family [14].

Detection of steganography that utilizes the knowledge of the selection channel is much less developed. The very first attack of this type was described by Böhme at the rump session at the Information Hiding Workshop in 2005 (and officially published in 2014 [4]). This pertains, however, to a rather special case of public-key steganography implemented using LSB replacement and a specific version of wet paper codes. While attacks derived using the theory of statistical hypothesis testing (e.g., [21], [8], [5]) can incorporate the knowledge of embedding probabilities they are generally not as effective against modern adaptive embedding schemes as machine-learning based methods combined with rich statistical descriptors. Modifying the latter approach to consider the knowledge of the selection channel is, however, not easy as witnessed by the futile effort of the BOSS competition participants [1] attacking HUGO. In 2012, it was shown that an (approximate) knowledge of embedding change probabilities can be used to improve the accuracy of the weighted-stego attack on naive content-adaptive LSB replacement [19]. In [7], the authors managed to utilize rather strong artifacts in the selection channel to mount a very accurate attack on S-UNIWARD with an improperly chosen stabilizing constant (also see [14] for more details). Recently, Tang et al. [20] proposed the first general purpose feature set that utilizes the selection channel and is effective against modern content-adaptive steganography methods. Their attack, which we call in this paper tSRM (thresholded SRM), computes the residual co-occurrences from only t percent of pixels with the highest embedding change probabilities (lowest pixel costs). The value of t that leads to the best detection depends on the embedded payload size and the steganographic scheme. The authors reported the detection only for the WOW algorithm.

Modern content-adaptive embedding schemes mentioned above are all based on the same principle – the sender specifies the costs of changing individual pixels and then embeds the payload with the minimal total expected cost. The costs are determined by the local content, which means the Warden can estimate them from the stego image. If the Warden knows the payload size or if she can estimate it, she can also estimate the actual embedding change probabilities used by the sender (the selection channel) and hopefully mount an even more powerful informed attack. For an ignorant Warden, who does not know the sender's embedding strategy, the interaction between the Warden and the sender can be be formulated as a non-cooperative strategic game with optimal

strategy at the Nash equilibrium, which is generally a different strategy than the one that minimizes the KL divergence between cover and stego objects [18], [6]. Other formulations are certainly possible depending on the information available to the Warden. Ultimately, the problem of content-adaptive steganography and selection-channel-aware steganalysis should be resolved within such a game-theoretic framework with an accurate statistical model for images and optimal Warden's detector. Due to the high complexity of empirical objects [2], such as digital images, and the high complexity of solving the ensuing game, it is however unlikely that optimal practical strategies will ever be identified.

In this paper, we follow the established paradigm of forming joint higher-order statistics of neighboring noise residuals as statistical descriptors. Our approach is reminiscent of the tSRM [20] but incorporates the selection channel in a different way. The four-dimensional co-occurrences are formed from *all* residuals rather than its proper subset, and, instead of populations, each bin holds the sum of maximum values of the four embedding change probabilities at the corresponding residuals. Since this model, which we call maxSRM, uses the statistic from all pixels, we obtain a more accurate detection. Additionally, and in contrast to the tSRM, if the payload size is known or can be estimated no other parameters need to be determined to steganalyze with maxSRM. Furthermore, the detection with maxSRM appears to suffer less when the embedded payload size is unknown.

In the next section, we introduce the common core of all experiments in this paper, including the image source, the classifier used for detection, and three embedding algorithms that will be used in our experiments: WOW [13], S-UNIWARD [14], and its variant called S-UNIGARD. The new maxSRM descriptor is explained in Section III. All experimental results are listed and interpreted in Section IV. Future directions and a summary appear in the last Section V.

II. Experimental setup

All our experiments were carried out on BOSSbase 1.01 [1] containing 10,000 grayscale 512×512 images. The detectors were trained as binary classifiers implemented using the FLD ensemble [15] with default settings. As described in the original publication, the ensemble by default minimizes the total classification error probability under equal priors $P_{\text{E}} = \min_{P_{\text{FA}}}(P_{\text{FA}} + P_{\text{MD}})/2$, where P_{FA} and P_{MD} are the false-alarm of missed-detection probabilities. The random subspace dimensionality and the number of base learners is found by minimizing the out-of-bag (OOB) estimate of the testing error, E_{OOB}, on bootstrap samples of the training set as it is an unbiased estimate of the testing error on unseen data [3]. We evaluate the security using the P_{E} measured on the testing set averaged over ten 5000/5000 database splits denoted as $\overline{P}_{\text{E}}$. The statistical spread is the standard deviation.

We selected three adaptive steganographic techniques that appear to be the state of the art as of writing this paper (May 2014): the Wavelet Obtained Weights (WOW) [13], S-UNIWARD implemented with the stabilizing constant $\sigma = 1$ as described in [14], and its variant that we call S-UNIGARD (described below). All three algorithms follow the paradigm of steganography by minimizing an additive distortion function. Assuming an $n_1 \times n_2$ grayscale cover image $\mathbf{X} = (x_{ij})$, the embedding starts by computing the costs ρ_{ij} of modifying pixel x_{ij} by 1 or by -1 (the costs of both modifications are equal). An optimal embedding scheme hides the secret message while minimizing the total cost of embedding (distortion) $D(\mathbf{X}, \mathbf{Y}) = \sum_{i,j=1}^{n_1,n_2} \rho_{ij}[x_{ij} \neq y_{ij}]$, where $[P]$ is the Iverson bracket $[P] = 1$ when P is true and $[P] = 0$ when P is false, and $\mathbf{Y}$ is the stego image. Such an optimal scheme would modify pixel x_{ij} to $x_{ij} + 1$ with probability β_{ij} (and to $x_{ij} - 1$ with the same probability), where $\beta_{ij} = (1 + e^{\lambda \rho_{ij}})^{-1}$ [9] with $\lambda > 0$ determined from the payload constraint, $\sum_{i,j=1}^{n_1,n_2} H(\beta_{ij}) = Rn$, where $H(x) = -2x \log_2 x - (1 - 2x) \log_2(1 - 2x)$ is the ternary entropy function in bits. In our tests, we used simulators of the embedding that indeed executed the changes with the probabilities β_{ij}. Practical embedding schemes that embed messages with nearly minimal distortion can be built using syndrome–trellis codes [9].

S-UNIGARD is built in the same way as S-UNIWARD with the three Wavelet Daubechies kernels replaced with Gabor filters (hence the letter 'G' replacing 'W' in the embedding scheme name), which are basically a set of differently oriented sinusoidal patterns modulated by a Gaussian kernel. Each kernel is obtained by sampling the following continuous function in $\mathbb{R}^2$ parametrized by the wavelength λ, the orientation angle θ, the phase offset ϕ, and the standard deviation σ of the Gaussian modulation:

$$G_{\lambda,\theta,\phi,\sigma,\gamma}(x, y) = \exp\left(-\frac{u^2 + \gamma^2 v^2}{2\sigma^2}\right) \cos\left(2\pi \frac{u}{\lambda} + \phi\right) \quad (1)$$

$$u = \quad x \cos\theta + y \sin\theta \quad (2)$$

$$v = -x \sin\theta + y \cos\theta \quad (3)$$

In S-UNIGARD, we use $\lambda = 2$, two offsets $\phi \in \{0, \pi/2\}$, 16 directions $\theta \in \{0, \pi/16, \ldots, 15\pi/16\}$, $\gamma = 0.5$, and $\sigma = 1$. The kernels are obtained by sampling $G_{\lambda,\theta,\phi,\sigma,\gamma}(x, y)$ at $x, y \in \{-5, -4, \ldots, 4, 5\}$ giving the filters a support of 11×11 pixels (see all 32 Gabor filters in Figure 1). All kernels are made zero mean (high-pass) by subtracting the kernel mean from all its elements. Assuming the cover is an $n_1 \times n_2$ grayscale image $\mathbf{X} = (x_{ij})$, $1 \leq i \leq n_1$, $1 \leq j \leq n_2$, the cost of changing pixel x_{ij} to $x_{ij} \pm 1$ (obtaining an image $\mathbf{X}_{[i,j]}$) and leaving all other pixels intact is computed in the exact same manner as in S-UNIWARD,

$$\rho_{ij} \triangleq \sum_{k=1}^{32} \sum_{u=1}^{n_1} \sum_{v=1}^{n_2} \frac{|F_{uv}^{(k)}(\mathbf{X}) - F_{uv}^{(k)}(\mathbf{X}_{[i,j]})|}{s + |F_{uv}^{(k)}(\mathbf{X})|}, \quad (4)$$

where $F_{uv}^{(k)}(\mathbf{X}) = (\mathbf{X} \star \mathbf{G}^{(k)})_{uv}$ is the uvth elements of the mirror-padded convolution between $\mathbf{X}$ and the kth Gabor filter $\mathbf{G}^{(k)}$, and s is a positive stabilizing constant. This constant affects the selection channel and needs to be chosen carefully to avoid introducing artifacts into the selection channel [7]. We determined it by a grid search on the grid $\mathcal{G} = \{10^{-15}, 10^{-14}, \dots, 10^{0}\}$ as the value that minimizes the out-of-bag (OOB) detection error on BOSSbase 1.01 [10] when steganalyzing with the 12,753-dimensional SRMQ1 model [12] and the ensemble classifier. The optimum was rather flat around $s \approx 10^{-2}$.

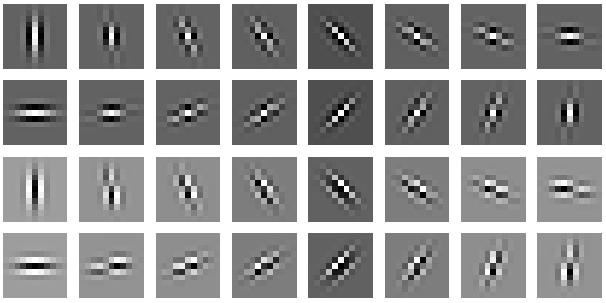

Figure 1. A set of 32 Gabor filters used in S-UNIGARD.

III. maxSRM

The proposed feature set is a variant of the so-called spatial rich model (SRM) described in [12]. The maxSRM is built in the same manner as the SRM 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. The SRM consists of multiple co-occurrence matrices formed by four neighboring quantized noise residual samples. Let us assume that $\mathbf{R} = (r_{ij})$ is one such noise residual, for example, one that was obtained by predicting the pixel value x_{ij} as the average of its horizontal neighbors, $r_{ij} = x_{ij} - (x_{i,j-1} + x_{i,j+1})/2$, quantized to $\mathcal{Q} = \{-2, -1, 0, 1, 2\}$. The SRM uses 4D co-occurrences, which are 4D arrays defined as[1]

$$C_{d_0 d_1 d_2 d_3} = \sum_{i,j=1}^{n_1, n_2 - 3} [r_{i,j} = d_k, \forall k = 0, \dots, 3]. \quad (5)$$

In maxSRM, we modify this definition to

$$\tilde{C}_{d_0 d_1 d_2 d_3} = \sum_{i,j=1}^{n_1, n_2 - 3} \max_{k=0,\dots,3} \hat{\beta}_{i,j+k} [r_{i,j} = d_k, \forall k = 0, \dots, 3]. \quad (6)$$

In other words, instead of adding a 1 to the corresponding co-occurrence bin, we add the maximum of the embedding change probabilities taken across the four residuals. This way, those groups of four of pixels with small probability of being changed will not affect the co-occurrence values much, while those where at least one pixel is likely to change will. We note that the rest of the process of forming the SRM stays exactly the same, including the symmetrization by sign and direction and merging into SRM submodels (see [12] for details). The proposed maxSRM has thus the same dimensionality as the SRM, which is 34,671.

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

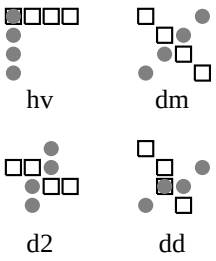

Figure 2. Four types of co-occurrence scan direction.

Table I
Effect of the co-occurrence scans on the detection error $\overline{P}_{\mathrm{E}}$. WOW at 0.4 bpp, the 338-dimensional SQUARE submodel of SRM.

Type	$\overline{P}_{\mathrm{E}}$
hv	22.14±0.26
dm	22.22±0.33
d2	21.71±0.44
dd	21.66±0.29

To further boost the detection, we investigated another design component of the SRM, which is the co-occurrence scan direction. The original SRM uses horizontal and vertical scans (see the case 'hv' in Figure 2). In this paper, we studied three other possibilities shown in the same figure – diagonal and minor-diagonal directions ('dm'), and two 'oblique' directions marked 'd2' and 'dd'. Because the oblique directions do not have a mirror symmetry, they allow collecting twice as much data for the co-occurrences, making them better populated. We observed in our experiments that the oblique directions do provide better detection across all tested algorithms. In Table I, we give a small example of this positive effect with the SQUARE SRM submodel (dimension 338) for the WOW algorithm at 0.4 bpp. The diagonal directions are the worst while the oblique directions are very similar and give (in this case) an improvement of 0.4% in the detection error w.r.t. the 'hv' scan used in SRM. Thus, we decided to include in our tests the version of the maxSRM with all co-occurrence scan directions replaced with the oblique direction 'd2'. We will call this version of the rich model the maxSRMd2.

IV. Experiments

As our first experiment, we provide the detection results for all three embedding algorithms (see Section II) when steganalyzing with SRM, maxSRM, and maxSRMd2 under the ideal case when the steganalyst knows the embedded payload size. The results, shown in Table II,

Table II
AVERAGE DETECTION ERROR $\overline{P}_{\mathrm{E}}$ FOR THREE EMBEDDING ALGORITHMS AND FOUR STEGANALYSIS FEATURE SETS.

Algorithm	Features	0.05	0.1	0.2	0.3	0.4	0.5
WOW	SRM	.4572 ± .0026	.4026 ± .0028	.3210 ± .0038	.2553 ± .0028	.2060 ± .0022	.1683 ± .0023
	maxSRM	.3595 ± .0017	.3025 ± .0033	.2383 ± .0022	.1943 ± .0015	.1623 ± .0038	.1371 ± .0028
	maxSRMd2	.3539 ± .0024	.2997 ± .0023	.2339 ± .0041	.1886 ± .0036	.1543 ± .0036	.1306 ± .0021
	tSRM	.3765 ± .0035	.3160 ± .0032	.2574 ± .0035	.2143 ± .0027	.1815 ± .0026	.1517 ± .0027
S-UNIWARD	SRM	.4533 ± .0026	.4024 ± .0019	.3199 ± .0027	.2571 ± .0016	.2037 ± .0032	.1640 ± .0024
	maxSRM	.4209 ± .0032	.3684 ± .0033	.2981 ± .0032	.2431 ± .0016	.1992 ± .0022	.1633 ± .0028
	maxSRMd2	.4180 ± .0025	.3660 ± .0040	.2886 ± .0025	.2360 ± .0022	.1908 ± .0025	.1551 ± .0019
	tSRM	.4391 ± .0033	.3935 ± .0013	.3199 ± .0027	.2571 ± .0016	.2037 ± .0032	.1640 ± .0024
S-UNIGARD	SRM	.4667 ± .0020	.4214 ± .0035	.3384 ± .0015	.2774 ± .0024	.2278 ± .0033	.1811 ± .0027
	maxSRM	.4195 ± .0030	.3712 ± .0027	.3002 ± .0022	.2466 ± .0022	.2062 ± .0025	.1702 ± .0027
	maxSRMd2	.4170 ± .0024	.3673 ± .0018	.2957 ± .0024	.2409 ± .0035	.1985 ± .0027	.1647 ± .0028
	tSRM	.4335 ± .0022	.3867 ± .0041	.3205 ± .0045	.2660 ± .0029	.2183 ± .0033	.1782 ± .0025

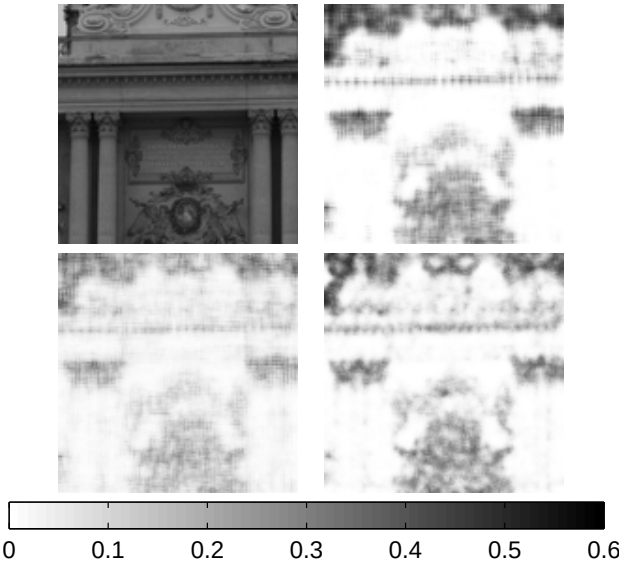

Figure 3. Embedding probability for payload 0.4 bpp using WOW (top right), S-UNIWARD (bottom left), and S-UNIGARD (bottom right) for a 128 × 128 grayscale cover image shown in top left (a 128 × 128 crop of '1013.pgm' from BOSSbase).

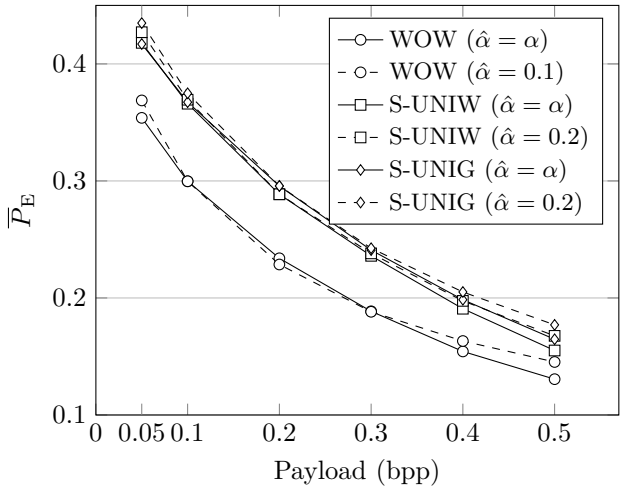

Figure 4. Detection error for all three algorithms when steganalyzing with maxSRMd2 with a fixed test payload ($\hat{\alpha} = 0.1$ for WOW and $\hat{\alpha} = 0.2$ for S-UNIWARD), versus the test payload set to the real payload, $\hat{\alpha} = \alpha$.

point out several rather interesting facts. First, while the security of WOW and S-UNIWARD appears almost the same under the SRM, when the selection channel information is utilized, WOW becomes much more detectable (for small payloads by more than 10%). This is most likely because WOW's adaptivity is stronger in the sense that embedding probabilities of S-UNIWARD are more "spread out" (see Figure 3). Obviously, the difference between SRM and maxSRM will diminish with a decreasing degree of adaptivity of the embedding algorithm. Also notice that while S-UNIGARD appears more secure than S-UNIWARD under SRM, this difference ($\approx 2\%$) becomes negligible when the selection channel is utilized. Finally, the maxSRM is always better than SRM, pointing to the fact that utilizing the selection channel in the proposed manner indeed helps steganalysis. Moreover, the comparison between maxSRM ans maxSRMd2 shows that the 'd2' co-occurrence scan is always better than the default 'hv' making the detection error smaller by as much as $\approx 1\%$. Since the dimensionality of both models is the same, there is no reason not to use maxSRMd2 over maxSRM.

Our next experiment was aimed at finding a fixed testing payload, $\hat{\alpha}$, used for computing the embedding probabilities that would provide an overall good performance when the real payload α is unknown. This will necessarily be a trade off between losing the detection for small versus large payloads. Based on tests with all three algorithms, it appears that a reasonable trade off is achieved when the test payload is fixed to a medium value of $\hat{\alpha} = 0.2$ bpp for S-UNIWARD and S-UNIGARD and to $\hat{\alpha} = 0.1$ bpp for WOW. Figures 4 and 5 show the detection error P_{E} and its increase when steganalyzing with a fixed test payload as opposed to the true payload. The performance drop averaged over payloads is below 1% and exhibits similar values and similar trends across the three tested algorithms.

In Figure 6, we compare the maxSRMd2 with the

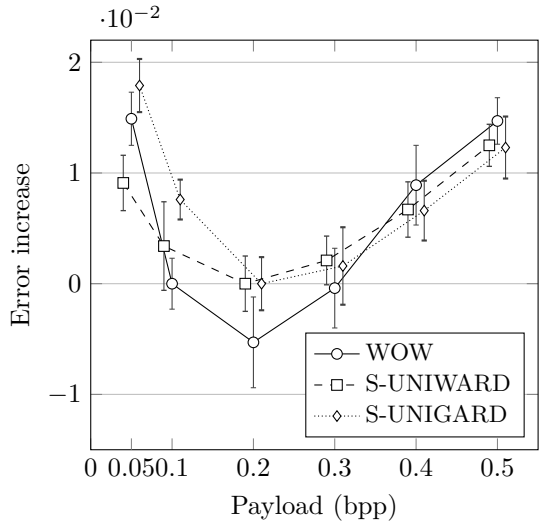

Figure 5. Detection error increase when steganalyzing with the test payload $\hat{\alpha}$ chosen as in the text and the true payload α. The payload (x coordinates) were shifted by a small amount to prevent the markers and error bars from overlapping.

previously proposed tSRM by showing the improvement in detection error over the SRM under the assumption that the real payload is known ($\hat{\alpha} = \alpha$). The threshold t in tSRM was optimized for each tested payload and stego algorithm. While the maxSRMd2 feature set provides better detection for all three algorithms, the tSRM fails to improve detection of S-UNIWARD for all payloads larger than 0.1 bpp and is only marginally effective against S-UNIGARD for large payloads. The maxSRMd2 consistently outperforms tSRM, sometimes by more than 3%. Moreover, when the embedded payload size is known or can be estimated, the maxSRM can be readily used while applying the tSRM requires running potentially expensive experiments to determine the best threshold for each payload. Also, we found out that setting a fixed value of the threshold t in tSRM when the true payload is not known is much trickier. It appears that one needs to either settle for a smaller improvement over the entire range of payloads or sacrifice the improvement (or even take a penalty) for large payloads (see Figure 5 in [20]). Finally, we wish to point out that the maxSRM is a generalization of tSRM because the tSRM feature vector can be computed using maxSRM by preprocessing the embedding change probabilities, β_{ij}, and setting $\beta_{ij} = 1$ when the cost of pixel i, j is within the top t% of costs using and setting it to 0 otherwise.

V. Conclusion

While content-adaptive steganography is nowadays a mature subject, steganalysis that utilizes the probabilistic selection channel is much less developed. Even though detectors built from tractable cover models using the theory of statistical hypothesis testing can incorporate Bayesian priors in a relatively straightforward manner, it is unclear

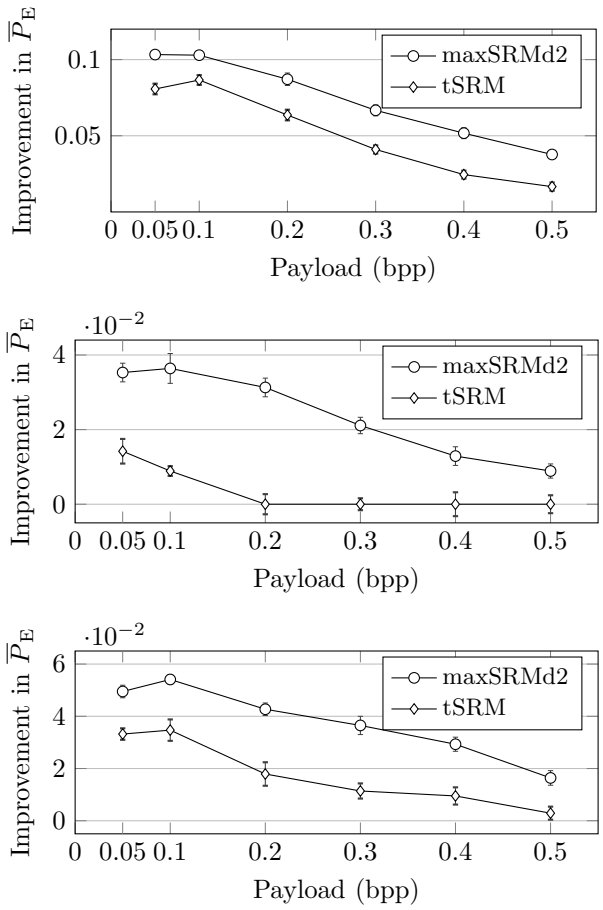

Figure 6. Improvement in detection error when steganalyzing with SRM versus maxSRMd2 or tSRM. The threshold in tSRM was optimized for each payload. From top down: WOW, S-UNIWARD, and S-UNIGARD.

how to adapt the detectors built by training classifiers in heuristically assembled feature spaces. This topic is relevant as such detectors are indispensable for detecting modern content-adaptive steganographic schemes.

In this paper, we propose a variant of the spatial rich model (the so-called maxSRM) modified to incorporate the knowledge of embedding change probabilities. Even though the proposed approach is heuristic, it does bring quite an improvement over features that do not consider the selection channel and it provides an interesting insight into the design of steganographic schemes. While the WOW and S-UNIWARD algorithms exhibit an essentially identical level of statistical detectability when tested with SRM, WOW is much more detectable with the selection-channel-aware maxSRM than S-UNIWARD. This is attributed to the varying degree of adaptivity of both algorithms. Apparently, WOW's selection channel is "overly adaptive," which makes this algorithm more vulnerable to maxSRM than the other algorithms. Moreover, while S-UNIGARD appears more secure than S-UNIWARD under

SRM, this difference ($\approx 2\%$) becomes negligible when the selection channel is utilized. Steganography designers thus need to be aware of how the properties of the selection channel affect statistical detectability when designing future steganographic schemes.

The maxSRM also offers the following three important advantages over the previously proposed thresholded SRM (tSRM): 1) the detection error is always lower, 2) there is no need to determine any parameters when the embedded payload is known or can be estimated, 3) the loss of detection is less severe when the real payload is unknown.

The code for all tested steganographic algorithms as well as maxSRM and maxSRMd2 is available from http://dde.binghamton.edu/download/.

Aknowledgment

This work was supported by Air Force Office of Scientific 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. The work of Rémi Cogranne is also funded by Troyes University of Technology (UTT) strategic program COLUMBO and STEG-DETECT program for scholar mobility. This research has been done while he was a visiting scholar at Binghamton University. Vojtěch Holub is with Digimarc, Inc, Beaverton, OR. This work was done while he was a PhD student at Binghamton University.

References

[1] P. Bas, T. Filler, and T. Pevný. Break our steganographic system – the ins and outs of organizing BOSS. In T. Filler, T. Pevný, A. Ker, and S. Craver, editors, *Information Hiding, 13th International Conference*, volume 6958 of Lecture Notes in Computer Science, pages 59–70, Prague, Czech Republic, May 18–20, 2011.

[2] R. Böhme. Weighted stego-image steganalysis for JPEG covers. In K. Solanki, K. Sullivan, and U. Madhow, editors, *Information Hiding, 10th International Workshop*, volume 5284 of Lecture Notes in Computer Science, pages 178–194, Santa Barbara, CA, June 19–21, 2007. Springer-Verlag, New York.

[3] L. Breiman. Bagging predictors. *Machine Learning*, 24:123–140, August 1996.

[4] M. Carnein, P. Schöttler, and R. Böhme. Predictable rain? Steganalysis of public-key steganography using wet paper codes. In A. Uhl, S. Katzenbeisser, R. Kwitt, and A. Piva, editors, *2nd ACM IH&MMSec. Workshop*, Salzburg, Austria, June 11–13, 2014.

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

[6] T. Denemark and J. Fridrich. Detection of content–adaptive LSB matching (game theory approach). In A. Alattar, N. D. Memon, and C. Heitzenrater, editors, *Proceedings SPIE, Electronic Imaging, Media Watermarking, Security, and Forensics 2014*, volume 9028, pages 04 1–12, San Francisco, CA, February 3–5, 2014.

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

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

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

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

[11] J. Fridrich and R. Du. Secure steganographic methods for palette images. In A. Pfitzmann, editor, *Information Hiding, 3rd International Workshop*, volume 1768 of Lecture Notes in Computer Science, pages 47–60, Dresden, Germany, September 29–October 1, 1999. Springer-Verlag, New York.

[12] J. Fridrich and J. Kodovský. Rich models for steganalysis of digital images. *IEEE Transactions on Information Forensics and Security*, 7(3):868–882, June 2011.

[13] 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.

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

[15] J. Kodovský, J. Fridrich, and V. Holub. Ensemble classifiers for steganalysis of digital media. *IEEE Transactions on Information Forensics and Security*, 7(2):432–444, 2012.

[16] W. Luo, F. Huang, and J. Huang. Edge adaptive image steganography based on LSB matching revisited. *IEEE Transactions on Information Forensics and Security*, 5(2):201–214, June 2010.

[17] T. Pevný, T. Filler, and P. Bas. Using high-dimensional image models to perform highly undetectable steganography. In R. Böhme and R. Safavi-Naini, editors, *Information Hiding, 12th International Conference*, volume 6387 of Lecture Notes in Computer Science, pages 161–177, Calgary, Canada, June 28–30, 2010. Springer-Verlag, New York.

[18] P. Schöttle and R. Böhme. A game-theoretic approach to content-adaptive steganography. In M. Kirchner and D. Ghosal, editors, *Information Hiding, 14th International Conference*, volume 7692 of Lecture Notes in Computer Science, pages 125–141, Berkeley, California, May 15–18, 2012.

[19] P. Schöttle, S. Korff, and R. Böhme. Weighted stego-image steganalysis for naive content-adaptive embedding. In *Fourth IEEE International Workshop on Information Forensics and Security*, Tenerife, Spain, December 2–5, 2012.

[20] W. Tang, H. Li, W. Luo, and J. Huang. Adaptive steganalysis against WOW embedding algorithm. In A. Uhl, S. Katzenbeisser, R. Kwitt, and A. Piva, editors, *2nd ACM IH&MMSec. Workshop*, Salzburg, Austria, June 11–13, 2014.

[21] C. Zitzmann, R. Cogranne, F. Retraint, I. Nikiforov, L. Fillatre, and P. Cornu. Statistical decision methods in hidden information detection. In T. Filler, T. Pevný, A. Ker, and S. Craver, editors, *Information Hiding, 13th International Conference*, Lecture Notes in Computer Science, pages 163–177, Prague, Czech Republic, May 18–20, 2011.