Toss that BOSSbase, Alice!

Vahid Sedighi, $^{+}$ ***Jessica Fridrich,*** $^{+}$ ***and Rémi Cogranne,*** $^{\times}$ $^{+}$***Department of Electrical and Computer Engineering, Binghamton University, New York, USA, {vsedigh1,fridrich}@binghamton.edu,*** $^{\times}$***ICD-ROSAS-LM2S, Troyes University of Technology, Troyes, France, remi.cogranne@utt.fr***

Abstract

Steganographic schemes for digital images are routinely designed and benchmarked based on feedback obtained on the standard image set called BOSSbase 1.01. While standardized image sets are important for advancing the field, relying on results from a single source may not provide fair benchmarking and may even lead to designs that are overoptimized and highly suboptimal on other image sources. In this paper, we investigate four modern steganographic schemes for the spatial domain, WOW, S-UNIWARD, HILL, and MiPOD on two more versions of BOSSbase. We observed that with their default settings, the mutual ranking and detectability of all four embedding algorithms can dramatically change across the three image sources. For example, in a version of BOSSbase whose images were cropped instead of resized, all four schemes exhibit almost the same empirical security when steganalyzed with the spatial rich model (SRM). On the other hand, in decompressed JPEG images, WOW is the most secure embedding algorithm out of the four, and this stays true irrespectively of the JPEG quality factor when steganalyzing with both SRM and maxSRM. The empirical security of all four schemes can be increased by optimizing the parameters for each source. This is especially true for decompressed JPEGs. However, the ranking of stego schemes still varies depending on the source. Through this work, we strive to make the community aware of the fact that empirical security of steganographic algorithms is not absolute but needs to be considered within a given environment, which includes the cover source.

Motivation

Currently, steganographic schemes are often developed and benchmarked on standard image sources. By far the most frequently used database is BOSSbase 1.01 [1], which contains 10,000 images taken in the RAW format by seven different cameras, converted to grayscale, downsampled using the Lanczos resampling algorithm with antialiasing turned OFF, and cropped to the final size of 512×512 pixels. Many articles have been published in which this database was the sole source on which steganographers fine-tuned their embedding scheme to obtain the best possible empirical security. However, BOSSbase images are far from what many would consider natural – they are essentially grayscale thumbnails obtained by a script that only a handful of people use.

Because of the rather aggressive downsizing of the original full-resolution RAW files, the content of many BOSSbase images is very complex with apparently rather weak dependencies among neighboring pixel values. The downsizing also effectively suppresses color interpolation artifacts and introduces artifacts of its own. There are images in BOSSbase that are very smooth, e.g., improperly focused images as well as images that are very dark and contain almost no content, such as an image of the Moon. One may thus argue that BOSSbase contains “enough” diversity to be used as a standardized source. On the other hand, virtually all steganographic schemes contain free parameters or design elements, such as an image transform and filter kernels, that are selected based on feedback provided by detectors on BOSSbase. We show that this makes the design overoptimized to a given image source and the embedding suboptimal on different sources. Even after optimizing the parameters of each embedding scheme to the source, universal benchmarking still does not seem possible since the optimized schemes exhibit different empirical security across sources. Additionally, the recently proposed synchronization of embedding changes [4, 12] appears far less effective on images with suppressed noise.

In the next section, we explain the measure of empirical security used in this paper and how it is evaluated. We also describe three versions of BOSSbase that will be investigated, the steganographic algorithms and steganalysis feature sets, as well as the choice of the classifier. In the third section, we start with comparing the empirical security of all algorithms on all three image sources and with two different steganalysis feature sets. Then, in the fourth section we identify the key parameters of each embedding scheme and perform a grid search to find the setting that maximizes the empirical security. The fifth section is devoted to investigating the impact of synchronizing the selection channel in different sources. The paper is concluded in the last section, where we summarize the most important lessons learned.

Setup of experiments

Security of embedding algorithms will be evaluated experimentally by training a binary classifier for the class of cover images and a class of stego images embedded with a fixed relative payload in bits per pixel (bpp), the so-called payload-limited sender. The classifier is the FLD ensemble [10] with two feature representations – the Spatial Rich Model (SRM) [7] and its selection-channel-aware version maxSRMd2 [5]. The security is reported with $\overline{P}_{\mathrm{E}}$, which is the minimal total error probability under equal priors

$$P_{\mathrm{E}} = \frac{1}{2}(P_{\mathrm{FA}} + P_{\mathrm{MD}}) \tag{1}$$

obtained on the testing set averaged over ten 50/50 splits of the image source into training and testing sets. Other measures were proposed in the past, such as the false-alarm-rate for 50% correct detection of stego images [13], FA50, which is more telling about the algorithm security for low false alarms. It has been observed that for the payload-limited sender, the detection statistic that is thresholded in the linearized version of the ensemble classifier [3] when rich models are applied is approximately Gaussian. In this case, both quantities, $\overline{P}_{\mathrm{E}}$ or FA50, would provide the same ranking of stego systems because there is a strictly monotone relationship between them.

For the purpose of this paper, we created the following two new versions of BOSSbase 1.01:

1. BOSSbaseC (C as in Cropped) was obtained using the same script as BOSSbase 1.01 but with the resizing step skipped. The images were centrally cropped to 512×512 pixels right after they were converted from the RAW format to grayscale. Images from this source are less textured but do contain acquisition noise.
2. BOSSbaseJQF (J as in JPEG, QF is the JPEG quality factor) was formed from BOSSbase 1.01 images by JPEG compressing them with quality factor QF$\in \{75, 85, 95\}$ and then decompressing to the spatial domain and representing the resulting image as an 8-bit grayscale. The low-pass character of JPEG compression makes the images less textured and much less noisy.

Figure 1 shows examples of four images from each source. Notice that images from BOSSbaseC appear "zoomed-in" because of the absence of downsizing.

Four embedding algorithms will be investigated in this paper: Wavelet Obtained Weights (WOW) [8], the Spatial version of the UNIversal WAvelet Relative Distortion (S-UNIWARD) [9], High-Low-Low (HILL) [11], and Minimizing the power of the most POwerful Detector (MiPOD) [14], which coincides with the MultiVariate Gaussian (MVG) steganography with a Gaussian residual model [15]. The study is limited to the spatial domain and does not consider JPEG images because the source generally does not play a significant role in JPEG steganography due to the low-pass character of JPEG compression, which tends to even out the differences between various sources.

Empirical security across sources

The purpose of the first experiment is to show that the ranking of steganographic schemes as originally described in the corresponding papers heavily depends on the image source. Figure 2 shows $\overline{P}_{\mathrm{E}}$ as a function of the relative payload in bits per pixel (bpp) for the four embedding algorithms listed in the previous section on BOSSbase 1.01, (first row), BOSSbaseC (second row), and BOSSbaseJ85 (third row) with SRM (left) and maxSRMd2 (right). Note that the ranking as well as the differences between individual embedding algorithms heavily varies depends on the cover source. Most notably, in BOSSbaseJ85, the most secure algorithm is WOW while MiPOD is the least secure, which is the *exact opposite* in comparison with BOSSbase 1.01. Moreover, when detecting with the SRM all four embedding schemes on BOSSbaseC have nearly identical empirical security.

Optimizing steganography for each source

In this section, we investigate how much the empirical security of each algorithm can be improved by adjusting the embedding parameters. This gain is quantified and the optimized embedding algorithms are ranked again for each image source.

We start by describing the parameters with respect to which each embedding scheme was optimized. The description is kept short but, hopefully, detailed enough for a reader familiar with the embedding algorithms to understand the parameters' role. The reader is referred to the corresponding publications for more details.

WOW: This embedding algorithm was designed to prefer making embedding changes at pixels in textured areas defined as regions with an "edge" in the horizontal, vertical, and both diagonal directions. The embedding begins with extracting directional residuals using tensor products of 8-tap Daubechies filters. Three directional filters with 8×8 kernels denoted $\mathbf{K}^{(h)}$, $\mathbf{K}^{(v)}$, and $\mathbf{K}^{(d)}$ are used to extract three directional residuals: $\mathbf{R}^{(h)} = \mathbf{K}^{(h)} \star \mathbf{X}$, $\mathbf{R}^{(v)} = \mathbf{K}^{(v)} \star \mathbf{X}$, and $\mathbf{R}^{(d)} = \mathbf{K}^{(d)} \star \mathbf{X}$, where $'\star'$ denotes a convolution and $\mathbf{X}$ is the matrix of pixel grayscales. In the next step, the so-called embedding suitabilities are computed: $\boldsymbol{\xi}^{(k)} = |\mathbf{R}^{(k)}| \star |\mathbf{K}^{(k)}|$, $k \in \{h, v, d\}$. The embedding cost of changing pixel i, j by $+1$ or -1 is obtained using the reciprocal Hölder norm $\rho_{ij}^{(k)} = \left(|\xi_{ij}^{(h)}|^p + |\xi_{ij}^{(v)}|^p + |\xi_{ij}^{(d)}|^p\right)^{-p}$ with $p = -1$.

To optimize WOW for different image sources, we search for the number of taps in Daubechies filters, $p_1 \in \{2, 4, 8, 16\}$ and the power of the Hölder norm $p_2 = p$.

S-UNIWARD: The pixel embedding costs are obtained from a distortion function defined as the sum of relative absolute differences between wavelet coefficients of cover and stego images. Only the highest frequency band of wavelet coefficients is used in UNIWARD. Formally, we denote the u, vth wavelet coefficient of $\mathbf{X}$ in $k \in \{h, v, d\}$ subband with $W_{uv}^{(k)}(\mathbf{X})$, $\mathbf{W}^{(k)} = \mathbf{K}^{(k)} \star \mathbf{X}$, u, v of the same range as image pixels. S-UNIWARD uses the same kernels formed from 8-tap Daubechies wavelets as WOW. The following non-additive distortion between the cover $\mathbf{X}$ and the stego image $\mathbf{Y}$ is used in UNIWARD:

$$D(\mathbf{X}, \mathbf{Y}) = \sum_{k \in \{h, v, d\}} \sum_{u, v} \frac{|W_{uv}^{(k)}(\mathbf{X}) - W_{uv}^{(k)}(\mathbf{Y})|}{\sigma + |W_{uv}^{(k)}(\mathbf{X})|}, \qquad (2)$$

where $\sigma = 1$ is the stabilizing constant. The embedding cost of the i, jth pixel is defined as $D(\mathbf{X}, \mathbf{Y}_{[ij]})$, where $\mathbf{Y}_{[ij]}$ is a stego image in which only the i, jth pixel was modified by 1.

Similar to WOW, the first parameter p_1 over which we optimize S-UNIWARD is the number of taps of Daubechies filters, $p_1 \in \{2, 4, 8, 16\}$, and the second parameter $p_2 = \sigma$ is the stabilizing constant.

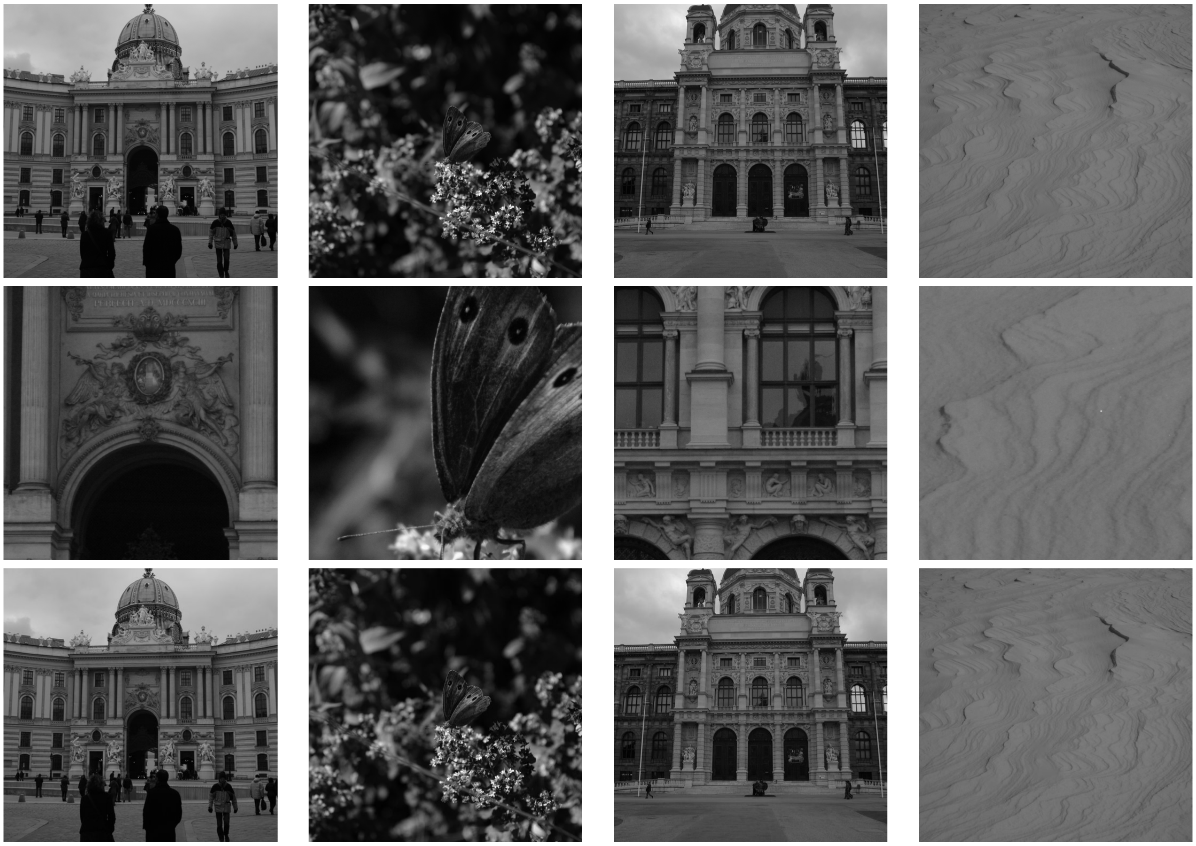

Figure 1. *By rows: Sample images from BOSSbase 1.01, BOSSbaseC, and BOSSbaseJ85.*

HILL: This algorithm originated from WOW. The authors replaced the three directional kernels with one non-directional high-pass 3×3 KB [2] kernel $\mathbf{H}$. HILL thus uses a single residual $\mathbf{R} = \mathbf{X} \star \mathbf{H}$. The pixel costs are then computed using the following formula:

$$\boldsymbol{\rho} = \frac{1}{|\mathbf{R}| \star \mathbf{L}_1} \star \mathbf{L}_2, \qquad (3)$$

where $\mathbf{L}_1$ is an averaging filter of support 3×3 and $\mathbf{L}_2$ is another averaging filter of support 15×15. All operations in (3) are elementwise.

We keep the KB kernel for the high pass filter and search for the support size p_1 of the averaging filter $\mathbf{L}_1$ and the support size p_2 of $\mathbf{L}_2$.

MiPOD: This embedding schemes differs fundamentally from the previous three schemes because it does not start with pixel costs. Instead, based on a residual model, the embedding change probabilities are computed to minimize the power of the most powerful detector. Once the change rates are computed using the method of Lagrange multipliers, the actual embedding proceeds by converting the change rates to costs so that syndrome-trellis codes [6] can be applied. The embedding begins with model estimation. Because MiPOD uses the Gaussian residual model of independent zero-mean residual samples, the model estimation reduces to estimating the variance of the Gaussian distribution at every pixel. When optimizing MiPOD, we searched over the parameters of the variance estimator.

The estimator first extracts a residual using a 2×2 Wiener filter. Then, the residual is locally fitted with a two-dimensional DCT polynomial of degree d in a $k \times k$ sliding window to extract the final residual from which the pixel variance is estimated using sample variance. Finally, before computing the costs the Fisher information is low-pass filtered with an averaging filter of size $l \times l$. MiPOD was optimized w.r.t. the following three parameters: the pair $p_1 = (k, d)$ for the polynomial fit and $p_2 = l$ for Fisher information averaging.

Comments and interpretations

Surprisingly, the least secure embedding scheme on the standard BOSSbase 1.01, WOW, becomes the most secure one on BOSSbaseJQF (Figure 6). And this stays true when steganalyzing with SRM or maxSRM and with the original parameter setting as well as the optimized setting (Figure 5). At the same time, the most secure algorithm on BOSSbase 1.01 when steganalyzing with maxSRM (Figure 3), MiPOD, performs the worst.

Following Table 1, one can say that for each embedding scheme on BOSSbaseJQF it is better to use a smaller support for residual filters. This could be explained by the fact that the decompression removes most of the texture and noise from the image and thus one can estimate the embedding costs (and pixel variances) with better accuracy from a smaller support in an apparent trade off between

Table 1. The original parameters for all four tested embedding schemes and their values optimized on BOSSbaseC and BOSSbaseJ85 for both SRM and maxSRMd2 for payload 0.1 bpp.

		SRM		maxSRMd2	
Source	Scheme	p_1	p_2	p_1	p_2
1.01	WOW	8	-1	4	-0.5
	S-UNI	8	1	4	1
	HILL	3	15	3	11
	MiPOD	(9,8)	7	(5,4)	7
C	WOW	14	-1	14	-1
	S-UNI	12	1	4	5
	HILL	5	11	9	11
	MiPOD	(9,8)	5	(11,10)	11
J85	WOW	4	-3	4	-3
	S-UNI	4	0.2	4	0.2
	HILL	7	3	5	7
	MiPOD	(3,2)	3	(3,2)	3

the estimator variance and bias (estimators with a larger support have a smaller variance but larger bias).

On BOSSbaseJQF, the maximum gain due to the search for optimal parameters was observed for MiPOD and was also achieved with the smallest possible support of the variance estimator. MiPOD might gain more by abandoning the current variance estimator structure and using a simpler variance estimator.

On BOSSbaseC, the most secure scheme with the maxSRM feature set is S-UNIWARD (Figure 4). The search for optimal parameters provides only little gain when steganalyzing with the SRM.

As a curiosity, we point out that on BOSSbaseJQF, maxSRM is performing worse that SRM by about $1-2\%$ for almost all embedding schemes and all three tested quality factors (Figure 6). This is most likely due to the fact that the algorithms' adaptivity is weaker on smooth content, making the utilization of the embedding change probabilities in maxSRM inefficient.

Impact of synchronizing embedding changes across cover sources

Recently, it has been shown that empirical security of embedding schemes built around an additive distortion function can be increased by synchronizing (clustering) the polarity of embedding changes [4, 12]. The synchronization leads to a smaller entropy of the stego noise, which forces a higher change rate but ultimately leads to better security. In [4], this gain was linked to the fact that the selection channel of non-additive embedding schemes is harder to estimate and the fact that steganalysis is most effective with sign-changing kernels.

In this section, we investigate how the gain in empirical security is affected by the image source. Again, the same four embedding algorithms are investigated as in the previous section. They are used as the initial additive scheme from which the non-additive embedding algorithm is built. The reader is referred to the original publications for more details about the embedding algorithms. Finally, before proceeding with the experimental results we note that maxSRM used the change rates of the original embedding schemes for steganalysis of the synchronized (clustered) versions of the embedding algorithms.

Our findings are displayed in Figure 7 and can be summarized as follows. Looking only at the more important maxSRM, synchronizing the embedding changes has the biggest impact in BOSSbaseC (up to 3.6%), a comparatively small impact in BOSSbase 1.01, and virtually no impact in BOSSbaseJ85. Synchronizing the embedding changes also does not change the ranking with respect to the original schemes (in the same source). Overall, CMD provides a larger boost than Synch. Finally, although the selection-channel-aware maxSRM is suboptimal when using the embedding probabilities of the original additive embedding scheme, maxSRM detects much better than SRM in all sources and for all algorithms with the exception of J85 where both are quite comparable.

Conclusions

Standardized image sources are necessary for development of both steganography and steganalysis. Spatial representation of images can, however, be very diverse when it comes to the strength and type of noise as well as the complexity of textures. Statistical properties of pixels can change dramatically after filtering, compression, and resizing. As this paper shows, this diversity makes absolute benchmarking impossible as stego systems may rank very differently in different cover sources even after each stego system has been optimized separately for each source. Among the rather surprising facts revealed in this study is that the algorithm WOW, which is well known to be the least secure among modern content-adaptive schemes in the standard image set BOSSbase 1.01 becomes the most secure in decompressed JPEGs, while the most secure algorithm, MiPOD, becomes the least secure. This remains true when steganalyzing with SRM as well as maxSRM and for both the original versions of the algorithms and their optimized forms. Similarly, the effectiveness of certain boosting measures, such as synchronizing (clustering) the polarity of embedding changes vastly changes across sources. In decompressed JPEGs, this measure is completely ineffective while in sources with noise but suppressed texture, to the contrary, it can have a major positive effect.

In this study, we fixed the steganalysis feature set across the sources, which may skew the results as one could argue that the steganalysis features should, too, be optimized for the source. Besides limiting this paper to a manageable length, another reason why optimization of the features has not been included is the fact that current rich feature representations already contain noise residuals obtained with filters of varying support size and, as such, are expected to be less sensitive to the cover source. Nevertheless, the direction of optimizing the features for cover sources should be investigated.

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, 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. *Advanced Statistical Steganalysis.* Springer-Verlag, Berlin Heidelberg, 2010.

[3] R. Cogranne and J. Fridrich. Modeling and extending the ensemble classifier for steganalysis of digital images using hypothesis testing theory. *IEEE Transactions on Information Forensics and Security*, 10(12):2627–2642, Dec 2015.

[4] T. Denemark and J. Fridrich. Improving steganographic security by synchronizing the selection channel. In J. Fridrich, P. Comesana, and A. Alattar, editors, *3rd ACM IH&MMSec. Workshop*, Portland, Oregon, June 17–19, 2015.

[5] 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.

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

[7] 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.

[8] 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.

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

[10] 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.

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

[12] B. Li, M. Wang, X. Li, S. Tan, and J. Huang. A strategy of clustering modification directions in spatial image steganography. *IEEE Transactions on Information Forensics and Security*, 10(9):1905–1917, September 2015.

[13] T. Pevný and A. D. Ker. Towards dependable steganalysis. In A. Alattar, N. D. Memon, and C. Heitzenrater, editors, *Proceedings SPIE, Electronic Imaging, Media Watermarking, Security, and Forensics 2015*, volume 9409, San Francisco, CA, February 8–12, 2015.

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

[15] V. Sedighi, J. Fridrich, and R. Cogranne. Content-adaptive pentary steganography using the multivariate generalized Gaussian cover model. In A. Alattar, N. D. Memon, and C. Heitzenrater, editors, *Proceedings SPIE, Electronic Imaging, Media Watermarking, Security, and Forensics 2015*, volume 9409, San Francisco, CA, February 8–12, 2015.

Author Biography

Vahid Sedighi received his B.S. degree in Electrical Engineering in 2005 from Shahed University, Tehran, Iran, and his M.S. degree in Electrical Engineering in 2010 from Yazd University, Yazd, Iran. He is currently pursuing the Ph.D degree in the Department of Electrical and Computer Engineering at Binghamton University, State University of New York. His research interests include statistical signal processing, steganography, steganalysis, and machine learning.

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.

Rémi Cogranne holds the position of Associate Professor at Troyes University of Technology (UTT). He has received his PhD in Systems Safety and Optimization in 2011 and his engineering degree in computer science and telecommunication in 2008 both from UTT. He has been a visiting scholar at Binghamton University in 2014-2015. His main research interests are in hypothesis testing, steganalysis, steganography, image forensics and statistical image processing.

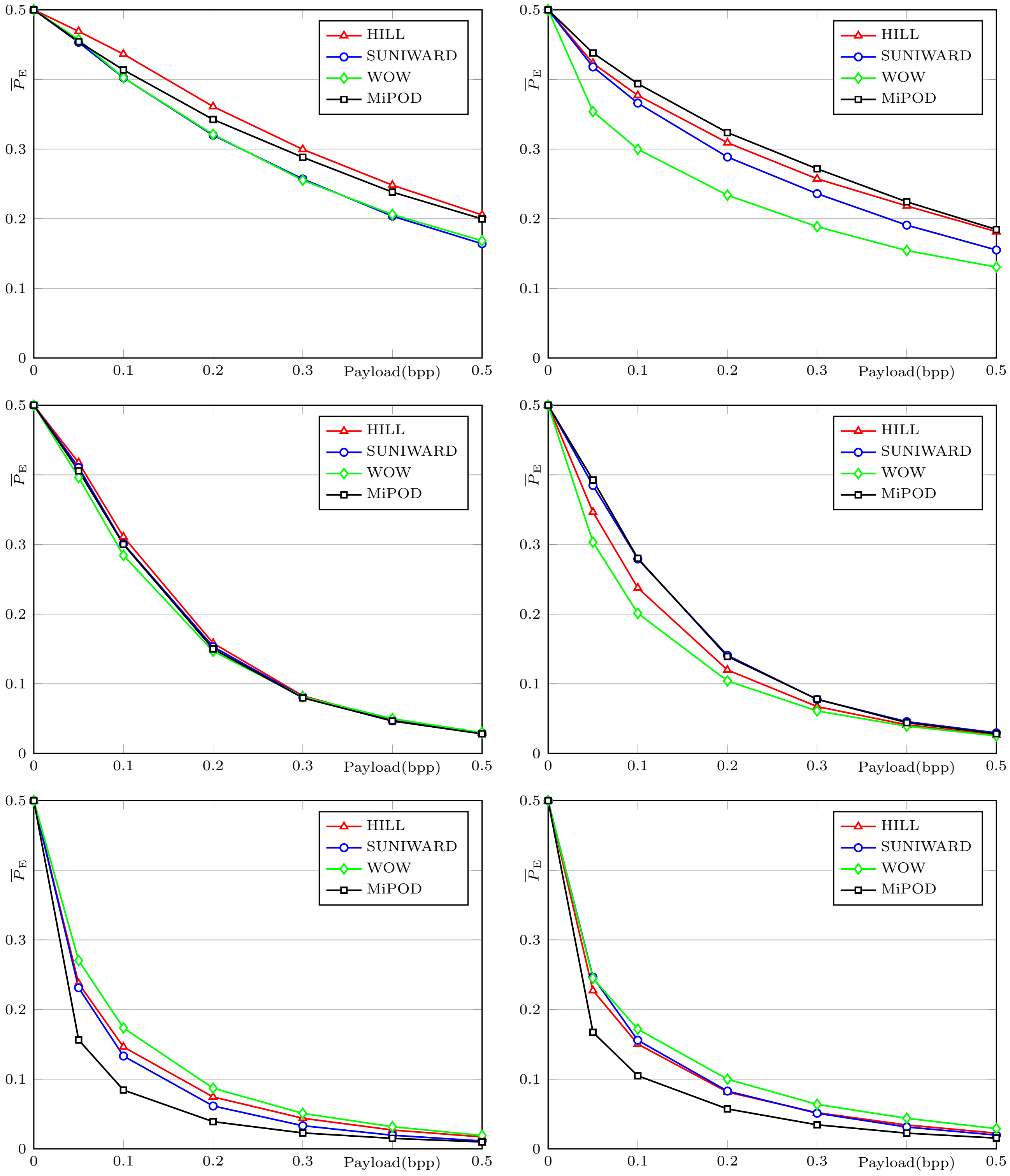

Figure 2. *Steganalysis of four embedding schemes using SRM (left) and maxSRMd2 (right). By rows: BOSSbase 1.01, BOSSbaseC, and BOSSbaseJ85.*

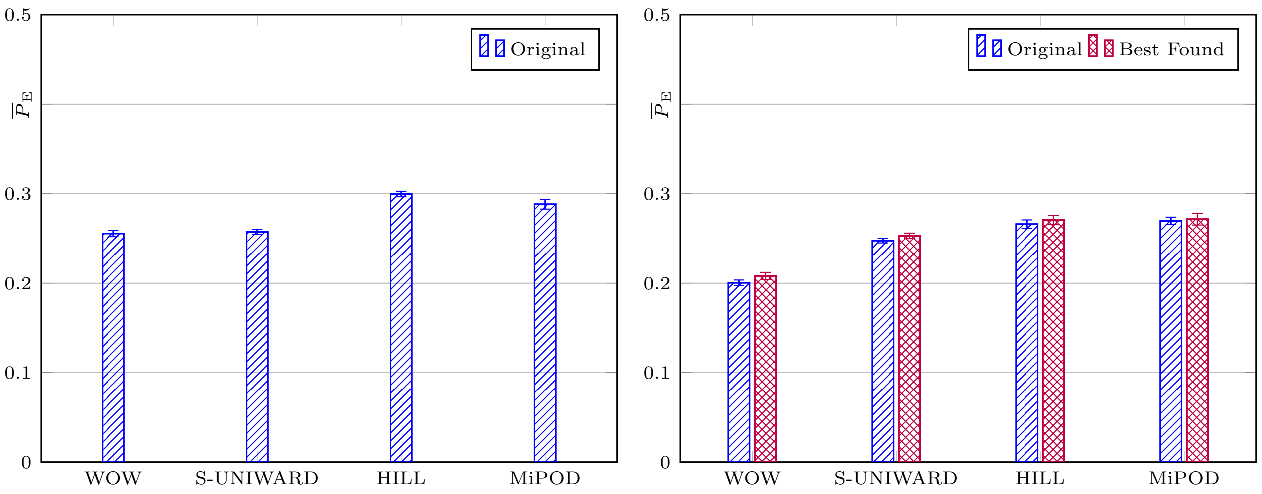

Figure 3. *Gain in empirical security when searching for optimal parameters of each embedding scheme on BOSSbase 1.01 for maxSRMd2 for payload 0.3 bpp. Note that we do not search for optimal parameters for SRM since all embedding schemes have gone through this procedure during their design process.*

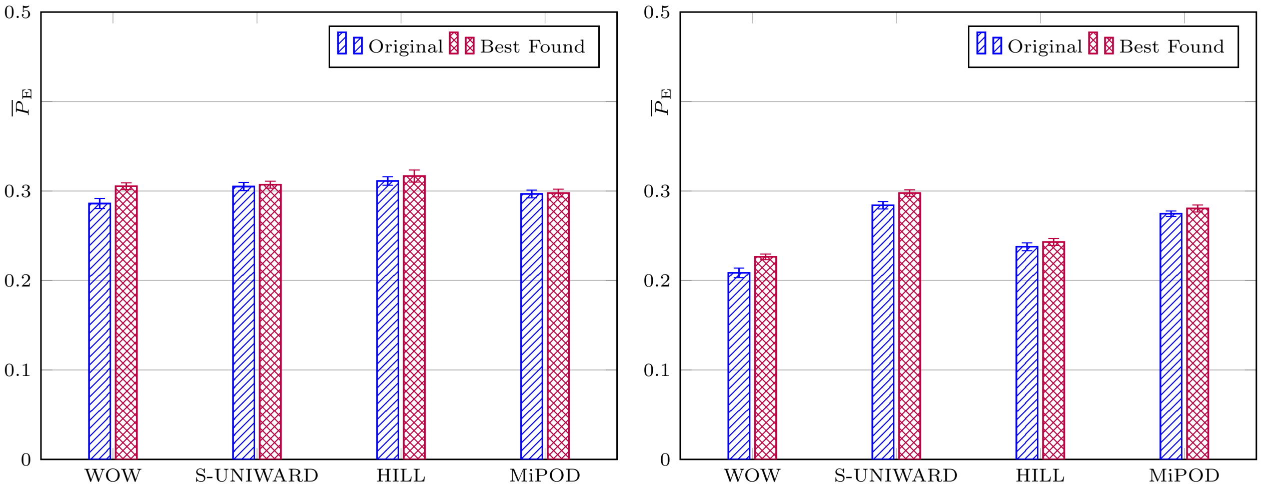

Figure 4. *Gain in empirical security when searching for optimal parameters of each embedding scheme on BOSSbaseC for both SRM and maxSRMd2 for payload 0.1 bpp.*

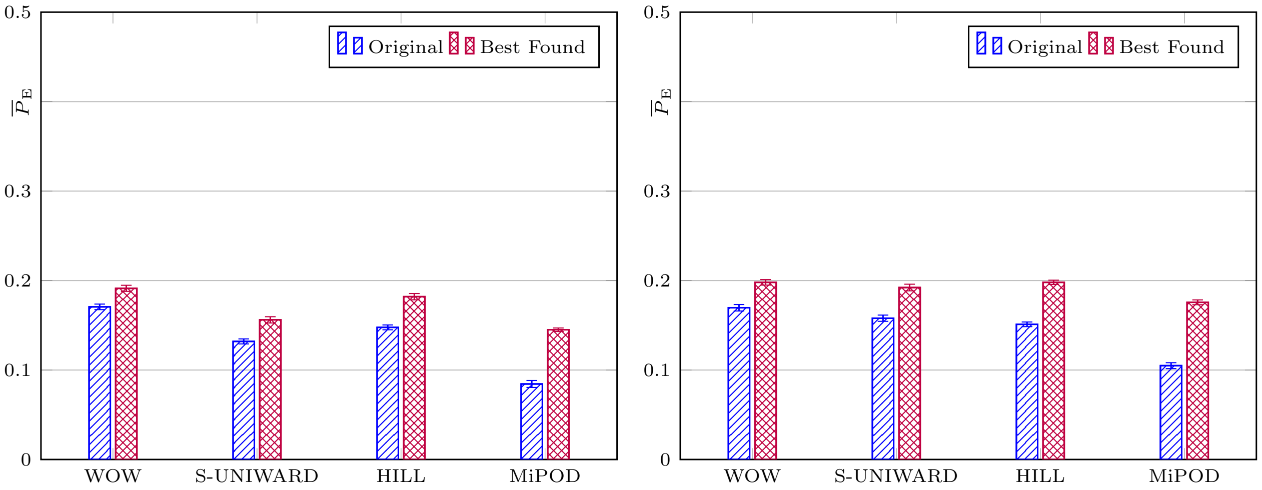

Figure 5. *Gain in empirical security when searching for optimal parameters of each embedding scheme on BOSSbaseJ85 for both SRM and maxSRMd2 for payload 0.1 bpp.*

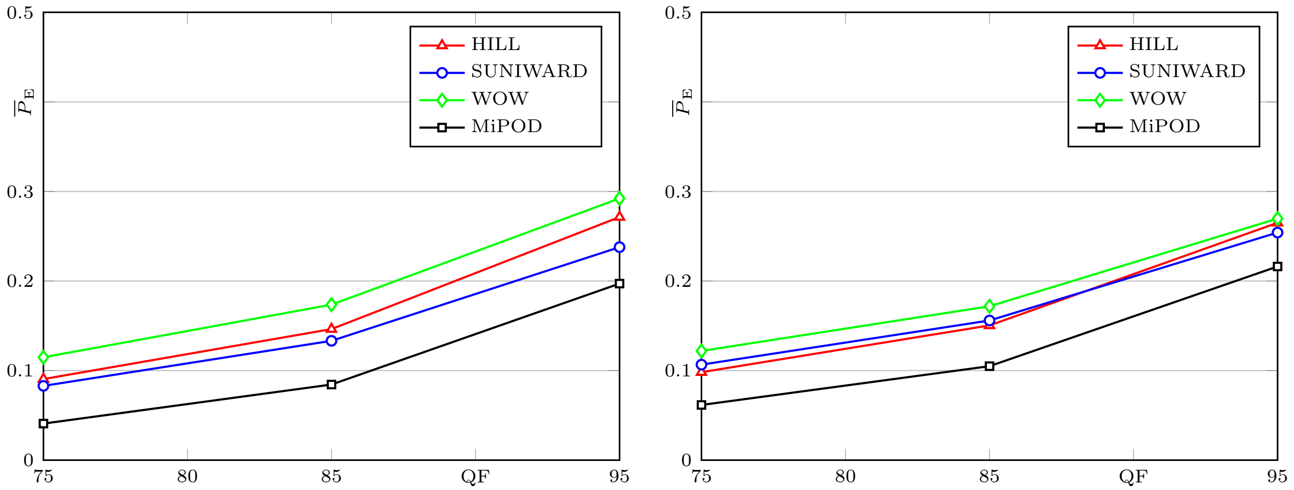

Figure 6. *Steganalysis of four embedding schemes using SRM (left) and maxSRMd2 (right) on BOSSbaseJ75, 85, and 90 for payload 0.1 bpp.*

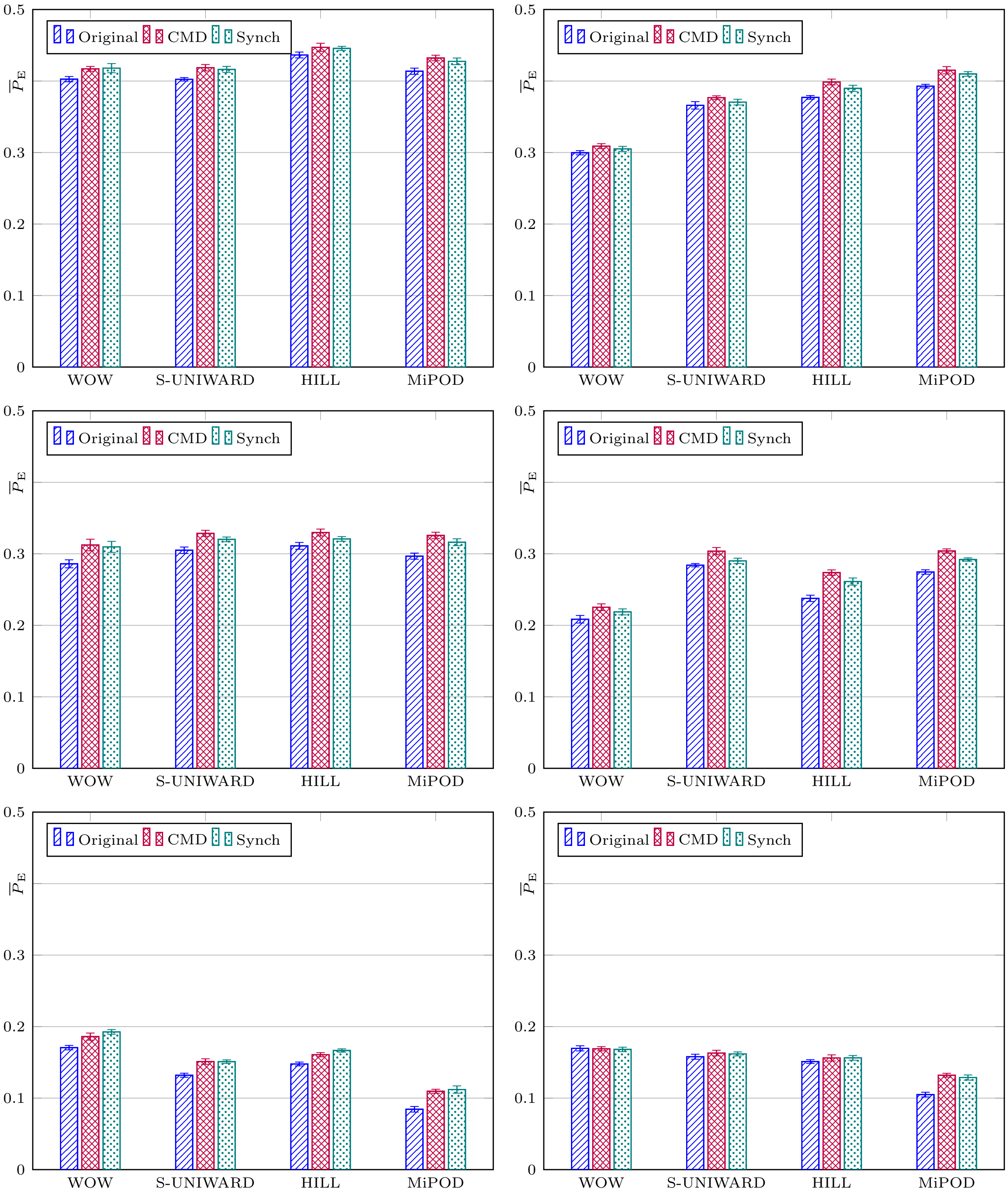

Figure 7. *Gain of synchronizing (clustering) embedding changes versus the original versions of the embedding algorithms on BOSSbase 1.01, BOSSbaseC, and BOSSbaseJQF with QF = 85 (by rows) at payload 0.1 bpp.*