https://doi.org/10.2352/ISSN.2470-1173.2018.07.MWSF-316

Natural Steganography in JPEG Compressed Images

Tomáš Denemark, [+] ***Patrick Bas,*** [×] ***and Jessica Fridrich,*** [+]
[+] ***Department of Electrical and Computer Engineering, Binghamton University, Binghamton, NY, 13902-6000, {tdenema1,fridrich}@binghamton.edu,***
[×] ***CNRS, Ecole Centrale de Lille, CRIStAL Lab, 59651 Villeneuve d'Ascq Cedex, Patrick.Bas@centralelille.fr***

Abstract

In natural steganography, the secret message is hidden by adding to the cover image a noise signal that mimics the heteroscedastic noise introduced naturally during acquisition. The method requires the cover image to be available in its RAW form (the sensor capture). To bring this idea closer to a practical embedding method, in this paper we embed the message in quantized DCT coefficients of a JPEG file by adding independent realizations of the heteroscedastic noise to pixels to make the embedding resemble the same cover image acquired at a larger sensor ISO setting (the so-called cover source switch). To demonstrate the feasibility and practicality of the proposed method and to validate our simplifying assumptions, we work with two digital cameras, one using a monochrome sensor and a second one equipped with a color sensor. We then explore several versions of the embedding algorithm depending on the model of the added noise in the DCT domain and the possible use of demosaicking to convert the raw image values. These experiments indicate that the demosaicking step has a significant impact on statistical detectability for high JPEG quality factors when making independent embedding changes to DCT coefficients. Additionally, for monochrome sensors or low JPEG quality factors very large payload can be embedded with high empirical security.

Introduction

The goal of steganography is to communicate in secrecy by hiding the very presence of the message within a host image called the cover image. The actual embedding involves making small modifications to the cover. The security of such communication is evaluated as the statistical detectability of the introduced changes. In this paper, we assume that the sender has a cover image available in the RAW format, examples of which include Canon's CR2, Nikon's NEF format, and Adobe's Digital Negative (DNG), and desires to communicate secrets in its JPEG compressed form. The RAW file serves as the so-called side-information or precover [25] provided by an acquisition oracle – the digital camera itself.

Side-informed steganography is the most secure form of steganographic communication known today. The first embedding schemes that utilized side-information at the sender were the embedding-while-dithering [16] and perturbed quantization [18] in which the secret was embedded by perturbing the color quantization (and dithering) or the rounding in JPEG compression. The latter direction has been further developed through a series of papers [27, 32, 36, 24, 21] and culminated with SI-UNIWARD [23, 9] as the current state of the art.

The idea to make the embedding modifications resemble noise naturally inserted during acquisition dates back to [2] and the rudimentary stochastic modulation [17], which ignored the important fact that the acquisition noise is independent only in the RAW domain. Franz et al. [13, 15, 14] attempted to estimate the acquisition noise and preserve its dependencies in the developed domain (the true-color domain) by taking multiple scans of the same image on a flat bed scanner. This rather labor intensive method, however, was not practical or secure also because of the inherent difficulty to estimate the acquisition noise properties in the developed domain. A much more practical version of this concept appeared in [11, 10], where the authors showed how multiple JPEG images of the same scene can be used to infer the preferred direction of embedding changes made to quantized DCT coefficients.

Recently, Natural Steganography (NS), that relies on the concept of cover-source switching, has showed a great promise for constructing practical secure steganographic systems [5, 4]. The author showed that a high-capacity steganographic scheme with a rather low empirical detectability can be built when the developing process of a RAW sensor capture is sufficiently simplified, e.g., after gamma correction, bilinear downsampling, and 8-bit quantization of RAW images coming from a monochrome sensor. The impact of embedding is masked as an increased level of photonic (shot) noise due to a larger sensor gain (ISO setting). This is possible because in the raw domain the distribution of the shot noise is well approximated with the heteroscedastic model independently distributed on each photo-site. For a sufficiently simple developer, one can thus arrange the statistical properties of the stego signal to mimic the increased heteroscedastic noise and make the stego image statistically resemble an image taken at a higher ISO setting (a switch in the cover source). The feasibility of this concept was shown in [4] with raw images taken with a Leica M Monochrome Type 230 camera. In a follow-up work [5], the same author extended NS to more complex developers that involved gamma correction and bilinear downsampling as these processes allowed analytic derivation of the acquisition noise properties in the developed domain. In this paper, we make NS more practical by introducing JPEG compression and also by treating the developer as a black box. Similarly to [11, 10], we use multiple instances of developed images in order to design our embedding strategy for each DCT coefficient.

In the next section, we introduce the heteroscedastic model of the acquisition noise and the concept of cover source switching, and study the dependencies of the acquisition noise in the DCT domain. The following section contains the description of the embedding method, which we develop through a cascade of approaches to assess the bounds on its security under various simplifying assumptions. In Section "Database Acquisition and Shot Noise Distribution" we detail the process of acquiring the images

for our experiments and verify our noise modeling assumptions. The proposed variant of NS is put to test in Section "Experiments" which contains all results and their discussion. The paper is closed with a summary and a discussion of potential future directions.

Throughout this paper, we use capital letters for random variables and the corresponding lower-case symbols for their realizations. Matrices are typed in upper-case and vectors in lower-case boldface font.

Natural steganography in JPEG domain

In this section, we introduce the heteroscedastic noise model and study its properties after applying block Discrete Cosine Transform (DCT) as in JPEG compression.

Model in the spatial domain

In natural steganography, the stego signal added to the cover image acquired at ISO_1 is constructed to mimic the additional shot noise to make the stego image look like it was acquired at $ISO_2 > ISO_1$.

The shot noise values in the spatial domain are assumed to be independent realizations of random variables $N_{i,j}$ that follow the heteroscedastic model

$$N_{i,j}^{(1)} \sim \mathcal{N}(0, a_1\mu_{i,j} + b_1) \tag{1}$$

where $\mu_{i,j}$ is the noiseless photo-site value at photo-site i,j, while (a_1, b_1) only depend on the ISO_1 sensitivity and the specific sensor.

The acquired photo-site sample $x_{i,j}^{(1)}$ is thus a realization

$$x_{i,j}^{(1)} = \mu_{i,j} + n_{i,j}^{(1)}, \tag{2}$$

of a Gaussian variable

$$X_{i,j}^{(1)} \sim \mathcal{N}(\mu_{i,j}, a_1\mu_{i,j} + b_1). \tag{3}$$

Because the sum of two independent normally distributed random variables is also normally distributed with the mean and variance the sum of means and variances of both variables, we can write that at ISO_2 the photo-site value is given by $x_{i,j}^{(2)} = x_{i,j}^{(1)} + s_{i,j}$ where $S_{i,j}$ is a random variable representing the stego signal necessary to mimic the image captured at ISO_2:

$$S_{i,j} \sim \mathcal{N}(0, (a_2 - a_1)\mu_{i,j} + b_2 - b_1). \tag{4}$$

Assuming that the observed photo-site is close to its expectation, $\mu_{i,j} \approx x_{i,j}^{(1)}$, the photo-site of the stego image is distributed as:

$$\begin{aligned} Y_{i,j} &\sim \mathcal{N}(\mu_{i,j}, a_1\mu_{i,j} + b_1 + (a_2 - a_1)\mu_{i,j} + b_2 - b_1) \\ &\sim X_{i,j}^{(2)}. \end{aligned} \tag{5}$$

The distribution of the stego signal in the continuous domain takes into account the statistical model of the shot noise estimated for two ISO settings, ISO_1 and ISO_2, using the procedure described in [4]. The work presented in [5, 4] shows that for monochrome sensors, this model in the spatial domain can be used to derive the distribution of the stego signal in the spatial domain after quantization, gamma correction, and image downsampling using bilinear kernels. We next study the properties of the acquisition noise after DCT.

Model in the DCT domain

We now compute the joint-distribution of the heteroscedastic noise in the DCT domain. This mathematical derivation can be used for a specific practical scenario of image development when a RAW image coming from a monochrome sensor is directly transformed into a JPEG image. To a certain extent, the derivations are also valid when gamma correction is performed before the DCT transform.

Given an 8×8 block of shot noise in the spatial domain, $\mathbf{S}$, its block 8×8 DCT transform can be written as the following matrix multiplication [26]:

$$\mathrm{DCT}(\mathbf{S}) = \mathbf{A}(\mathbf{A}\mathbf{S}^t)^t = \mathbf{A}\mathbf{S}\mathbf{A}^t, \tag{6}$$

where

$$\mathbf{A} = \begin{bmatrix} a & a & a & a & a & a & a & a \\ b & d & e & g & -g & -e & -d & -b \\ c & f & -f & -c & -c & -f & f & c \\ d & -g & -b & -e & e & b & g & -d \\ a & -a & -a & a & a & -a & -a & a \\ e & -b & g & d & -d & -g & b & -e \\ f & -c & c & -f & -f & c & -c & f \\ g & -e & d & -b & b & -d & e & -g \end{bmatrix} \tag{7}$$

$$\begin{bmatrix} a \\ b \\ c \\ d \\ e \\ f \\ g \end{bmatrix} = \frac{1}{2} \begin{bmatrix} \cos(\pi/4) \\ \cos(\pi/16) \\ \cos(\pi/8) \\ \cos(3\pi/16) \\ \cos(5\pi/16) \\ \cos(3\pi/8) \\ \cos(7\pi/16) \end{bmatrix}. \tag{8}$$

Note that the multiplication by $\mathbf{A}$ and $\mathbf{A}^t$ comes from first transforming the columns and then the rows of matrix $\mathbf{A}$.

In order to compute the covariance matrix of the stego signal $\mathbf{S}$, it is convenient to use vector notation by transforming the matrix $\mathbf{S} \in \mathbb{R}^{8\times 8}$ into a vector $\mathbf{s} \in \mathbb{R}^{64}$ by concatenating the columns. The transpose operation $\mathbf{S}^t$ is then equivalent to the multiplication $\mathbf{T}\mathbf{s}$, by $\mathbf{T}$ given by:

$$\mathbf{T} = \begin{bmatrix} 1 & 0 & \dots & & & & & & \\ 0 & \dots & \dots & 1 & 0 & \dots & & & \\ & & & & & & 1 & 0 & \dots \\ \vdots & \vdots & \vdots & \vdots & \vdots & \vdots & \vdots & \vdots & \\ 0 & 1 & \dots & & & & & & \\ & & & 0 & 1 & \dots & & & \\ & & & & & & 0 & 1 & \dots \\ \vdots & \vdots & \vdots & \vdots & \vdots & \vdots & \vdots & \vdots & \end{bmatrix}. \tag{9}$$

Consequently, the 8×8 matrix $\mathbf{A}$ is transformed into a 64×64 matrix $\mathbf{A}_v$ given by:

$$\mathbf{A}_v = \begin{bmatrix} \mathbf{A} & 0 & \cdots & & & & & \\ 0 & \mathbf{A} & 0 & \cdots & & & & \\ \vdots & 0 & \mathbf{A} & 0 & \cdots & 0 & & \\ & \cdots & 0 & \mathbf{A} & 0 & \vdots & & \\ & & \vdots & 0 & \mathbf{A} & 0 & \vdots & \\ & & 0 & \cdots & 0 & \mathbf{A} & 0 & \vdots \\ & & & & \cdots & 0 & \mathbf{A} & 0 \\ & & & & & \cdots & 0 & \mathbf{A} \end{bmatrix}. \quad (10)$$

The vector form of the DCT (6) finally becomes

$$\mathrm{DCT}_v(\mathbf{s}) = \mathbf{A}_v\mathbf{T}\mathbf{A}_v\mathbf{T}\mathbf{s} = \mathbf{B}\mathbf{s}, \quad (11)$$

where $\mathbf{B} = \mathbf{A}_v\mathbf{T}\mathbf{A}_v\mathbf{T}$.

Since the stego signal in the spatial domain follows the normal distribution 4 and since the DCT is linear, the stego signal in the DCT domain, S_{DCT}, follows a 64-dimensional multivariate normal distribution

$$S_{\mathrm{DCT}} \sim \mathcal{N}(\mathbf{0}, \mathbf{C}_{\mathrm{DCT}}), \quad (12)$$

where

$$\mathbf{C}_{\mathrm{DCT}} = \mathrm{E}[\mathbf{B}SS^t\mathbf{B}^t] = \mathbf{B}\,\mathrm{Cov}(S)\,\mathbf{B}^t, \quad (13)$$

and $\mathrm{Cov}(S)$ denotes a diagonal matrix with diagonal elements equal to $\mathrm{Var}(S_{i,j}) = (a_2 - a_1)x_{i,j} + b_2 - b_1$.

Discussion

Even though the stego signal (the sensor noise) is independent in the spatial domain, it follows a general multivariate normal distribution in the DCT domain. Thus, ideally the embedding should take into account dependencies that exist between DCT modes within each 8×8 block. Note that in this setting, no dependencies exist between DCT blocks. This model consequently enables us to explicitly compute the variance of the stego signal for each DCT mode and the covariance between DCT modes.

To better understand the nature of the dependencies between DCT coefficients, we sample the stego signal directly in the DCT domain and observe the dependencies before and after JPEG quantization.

In Figure 1, we visually compare sampled blocks before (even columns) and after quantization (odd columns) with the standard JPEG quantization matrix corresponding to quality factor 95. Note that the quantization process is here $\mathrm{quant}(x) = \Delta \times \mathrm{round}(x/\Delta)$, where Δ is the quantization step.

For different spatial cover blocks represented in the first row, blocks of stego signals are sampled in the DCT domain (the second row, S) using (12) or in the spatial domain (the third row, S^s) using (4) and then transformed.

While the first two spatial blocks with horizontal/vertical directions produce vertical/horizontal correlations in the DCT domain, neither the checkerboard or the constant block produce significant correlations (for the constant block, the signal must be i.i.d. since it is the DCT of an i.i.d. signal). The diagonal blocks produce slightly correlated stego signals which are more pronounced for the minor-diagonal block. Comparing the second and third rows enables us to verify that the sampling either in the spatial domain or directly in the DCT domain exhibits similar dependencies. The odd columns illustrate the effect of JPEG quantization, which tends to reduce the dependencies between coefficients by nullifying high frequencies.

This experiment also shows that the dependencies in the DCT domain, contrary to the spatial domain (see [8, 29]), heavily depend on the cover block content. However, we shall see in Section "Experiments" that for large quantization regimes not taking into account the dependencies does not significantly impact the detectability of embedding.

Overview of the algorithms

We remind the reader that our goal is to develop a NS method capable of embedding messages in JPEG images by utilizing a cover source switch from ISO_1 to a larger ISO_2. The first step in building such a steganographic method is to estimate the parameters of the heteroscedastic sensor noise for the specific camera that will be used for communication and for both ISO settings: (a_1, b_1) and (a_2, b_2). This has been executed by taking images of a gray gradient as explained in [5, 4] and in Section "Shot noise distributions". Having estimated these four parameters, from Eqs. (2)–(5) the stego photo-site is obtained from the cover photo-site $X^{(1)}_{i,j} \sim \mathcal{N}(\mu_{i,j}, a_1\mu_{i,j} + b_1)$ by adding to it a realization of $S_{i,j} \sim \mathcal{N}(0, \sigma^2_{i,j})$, where $\sigma^2_{i,j} = (a_2 - a_1)\mu_{i,j} + b_2 - b_1 \approx (a_2 - a_1)x_{i,j} + b_2 - b_1$.

In order to perform a cover-source switch on a raw pre-cover image, we adopt special rules for photo-sites saturated at 2^r, where r is the dynamic range of the sensor, typically 12 or 14 bits. Our strategy is similar to the one presented in [33]. The photo-site value $y^{(2)}_{i,j}$ after the cover-source switch mimicking sensitivity ISO_2 is:

$$y^{(2)}_{i,j} = \begin{cases} 2^r & \text{if } x^{(1)}_{i,j} = 2^r \text{ or } x^{(2)}_{i,j} > 2^r \\ 0 & \text{if } x^{(2)}_{i,j} < 0, \\ x^{(2)}_{i,j} & \text{else.} \end{cases}, \quad (14)$$

To obtain a better insight into the role of our simplifying assumptions and the effect of estimating the noise in the DCT domain and to establish upper bounds on the detection error, we investigate five different approaches explained below. The security of these approaches is evaluated by building a classifier distinguishing between the images from both classes after JPEG compression. To get closer to a practical scheme, we added the case when the developer is treated as a black box and the noise distribution is estimated from quantized DCT coefficients in the developed domain using Monte-Carlo (MC) sampling.

1. **Pure, unmodified images**. As a baseline experiment, no modifications were introduced to the ISO_1 images and a classifier was trained to recognize between these and ISO_2 images after JPEG compression.
2. **Simulated noise**. Before any developing, we added a simulated heteroscedastic noise to the raw images at ISO_1 using

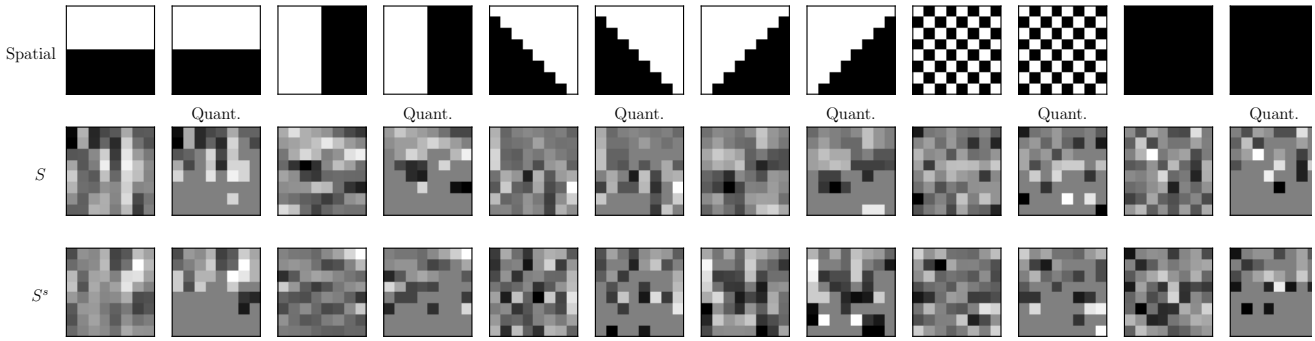


Figure 1: First row: spatial 8×8 blocks. The second and third rows are samples where, for the purpose of comparison, the signal S is sampled directly in the DCT domain by sampling a 64-dimensional multivariate Gaussian distribution while S^s is sampled in the spatial domain and then converted to DCT coefficients.

the clipping rule 14. Provided the parameters of the heteroscedastic noise at each ISO setting were precisely estimated and under the assumption that we steganalyze using the best possible detector, this result could serve as an upper bound on the security of the method – the detection error. Note, however, that it does not correspond to any practical embedding.

3. **Monte-Carlo estimate of the variance**. We add 300 independent realizations of the heteroscedastic noise estimated as in Approach 2 to the raw pre-cover image acquired at ISO_1, $x_{i,j}^{(1)}$, employing again 14, develop the images, and apply the DCT. Then, we independently estimate the mean and the variance of each DCT coefficient from the MC samples. To obtain the stego JPEG file, we add independent realizations of such random variables to the unquantized DCT coefficients of the developed pre-cover and round to integers to obtain the JPEG DCT coefficients of the stego image.
4. **Monte-Carlo estimate of the pmf**. To remove the Gaussianity assumption, we use the MC samples to directly estimate the probability mass function of each *rounded* DCT coefficient. Then, we sampled from this distribution for each coefficient to obtain the final quantized DCT coefficient from the stego image.
5. **SI-UNIWARD**. For comparison with the current state of the art, we embedded all images also with SI-UNIWARD [23] with the same average embedding rate or lower if the embedding rate was over 1 bit per DCT coefficient, the maximal payload of SI-UNIWARD.

We now proceed with a formal description of the NS method that hides messages in the JPEG file given a pre-cover in a RAW format. The sender basically uses the pre-cover in the RAW format to *estimate* the Gaussian variance from MC samples (Approach 3) and then *compute* the pmf of the quantized stego DCT coefficient, or to *directly estimate the pmf* of each quantized DCT coefficient from the stego image (Approach 4). The advantage of Approach 4 is that it can be applied for realistic (i.e., complicated) developers that output more complex (non-Gaussian) shot noise distribution.

Denoting the pmf of a fixed quantized stego DCT coefficient as q_k, the payload that could embed at this coefficient is

$$-\sum_k q_k \log_2 q_k \text{ bits}, \tag{15}$$

the entropy of the pmf. For Approach 3, given the variance ω^2 of a specific unquantized stego DCT coefficient with quantization step Δ, q_k corresponds to the kth bin in a quantized Gaussian distribution $\mathcal{N}(0, \omega^2/\Delta^2)$:In contrast, in Approach 4 the pmf q_k is estimated directly from the 300 MC samples by computing an empirical histogram.

The actual message embedding can be implemented in practice using the multi-layered version of syndrome-trellis codes [12], which essentially allow embedding payload close to the entropy (15) at each DCT coefficient. We would also like to stress that the total payload that can be embedded is determined by the two ISO values and is equal to the sum of entropies (15) over all DCT coefficients in the JPEG image. The payload size also depends on the JPEG quality factor and the content of the image. Should the sender need to embed a shorter payload, the message could be padded with random bits. Alternatively, the sender could also switch to a smaller value of ISO_2. On the contrary, if the payload to be embedded is larger than the admissible payload offered by the cover-source switch, the sender would have to use a larger value of ISO_2 or split the payload across multiple images.

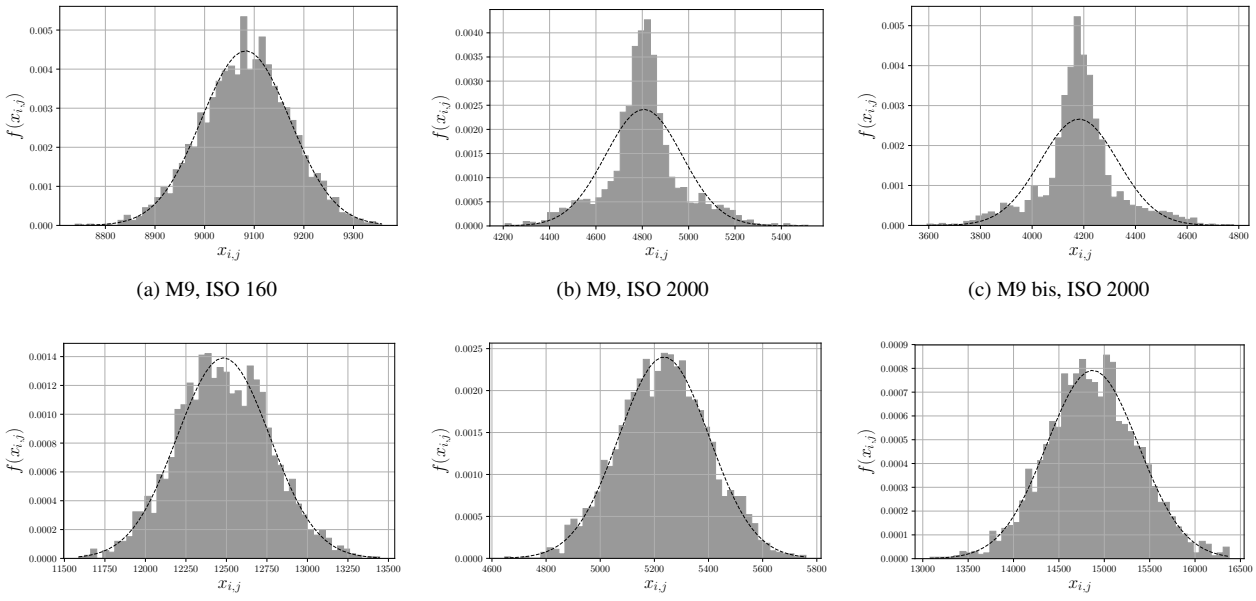
Note that the proposed NS method may, depending on the ISO settings, embed more bits in an image than SI-UNIWARD. For a fair comparison, for SI-UNIWARD we therefore embedded the same relative payload in each image (in bits per pixel rather than per non-zero AC DCT coefficient) obtained by averaging the payload embedded by Approach 3 over the whole database.

Database acquisition and shot noise distribution

In this section, we describe how we acquired the image databases needed to benchmark Natural Steganography, and discuss the statistical properties of the photonic noise distribution for different sensors.

Acquisition process

In contrast to the widely used BOSSBase [1] used in steganography and steganalysis, for benchmarking Natural Steganography the datasets need to be built with special care. Because the goal of the embedding is to mimic a shot noise at ISO_2 from images captured at sensitivity $ISO_1 < ISO_2$, two sets of images have to be acquired: one set at ISO_1 that will be used for



(a) M9, ISO 160 (b) M9, ISO 2000 (c) M9 bis, ISO 2000

(d) Mono, ISO 4000 (e) M262, ISO 2000 (f) E1, ISO 1600

Figure 2: Comparison between distributions of shot noise coming from different sensors. Histograms are computed from the photo-site values of one given channel (for color sensors) on a uniform patch. Dash lines represent Gaussian distributions with the same mean and variance as the histogram.

the embedding and another set at ISO_2 that will represent the set of cover images. The steganalyst will then compare stego images coming from the set at ISO_1 and cover images acquired at ISO_2. We assume here that the sender will modify or remove the ISO setting from the stego images because the steganalyst could potentially utilize the discrepancy between the noise level in stego images and the ISO setting.

It is important to mention that in order to build a classifier that will only detect the steganographic embedding, the two sets of images have to represent identical content.

During our acquisition campaign, we consequently paid attention to use constant acquisition parameters: the same focus, the same scene with the use of a tripod, the same white balance, and the same aperture to have only the sensitivity and the exposure time fluctuating. We realized the importance of this step when at one point we slightly modified the focus between the two sets, which resulted in increased classification accuracy due to the ability of the classifier to distinguish between content sharpness rather than the steganographic changes.

To alleviate the labor associated with the acquisition of these databases, we took around 200 raw images[1] at each setting and subsequently cropped each picture to non-overlapping 512×512 images to generate around 10,000 crops in each set. The development of the raw image was done using the `'dcraw'` Linux command line with the parameters `'-k 0'` to obtain the same darkness level for each set, `'-g 1 1'` to disable gamma correction, `'-W'` to obtain the same white balance for each set and `'-6'` to generate 16-bit ppm/pgm images instead of 8-bit images.

We ran these acquisition campaigns on three sensors: one monochrome CCD sensor from the Leica M Monochrome Type 230 camera, one color CCD sensor from the Leica M9 camera, and one CMOS sensor from the Z CAM E1 action camera. Note that the Leica M Monochrome and Leica M9 cameras have identical sensors but the Monochrome does not have a Bayer CFA. The databases built from Leica cameras are images from different scenes, shot using a tripod at different ISO settings (320, 1000, and 1250 for the Monochrome), the databases from the E1 sensor have been captured using a rotating platform in a room filled with different objects.

Shot noise distributions

We were very surprised to notice that, using exactly the switch 14, the detectability of the M9 sensor was extremely high compared to the detectability of the Monochrome sensor. After some investigations, we noticed that the shot noise on the M9 sensor does not have a Gaussian distribution at high ISO sensitivities.

This phenomenon is illustrated in Figure 2, where we compare the shot noise distributions for both different sensors. To estimate the distribution of the sensor noise, we used here a simple but robust technique: we shot a white wall at a distance of 1m from the sensor and out of focus in order to obtain an image with average constant illumination. We then computed the histogram from the RAW image, using the photo-site values[2] of a 100×100 patch centered on the image to avoid vignetting for one given color channel (we checked that our observations were consistent for all channels and for different patches). While the histogram of the noise taken at ISO 160 (Figure 2a) corresponds to a Gaussian signal, as soon as the ISO sensitivity is increased, the shot noise distribution becomes strongly non-Gaussian. For

[1] The exact number depends of the sensor resolution.

[2] They are, for example, easily accessible using the Rawkit Python module [30].

example, at ISO 2000 (Figure 2b) the distribution has heavy tails and the Gaussian assumption is rejected. The sensor capture from another M9 camera (see Figure 2c) shows that this artifact does not come from the specific camera. It is also not specific to the manufacturer since Figure 2e depicts a Gaussian distribution for the Leica Type M262 CMOS sensor. What is even more surprising is the fact that the Leica M Monochrome Type 230 camera (an M9 version without Bayer CFA) does not exhibit this artifact (Figure 2d).

In the end, we selected two sensors:

1. Leica M Monochrome camera to directly acquire grayscale images because this sensor does not have demosaicking applied during the development process and it also exhibits normally distributed shot noise,
2. Z CAM E1 to acquire color images because of a Gaussian shot noise distribution at high ISO (Figure 2f). This action camera has a time-lapse mode, which enables fast acquisition of new images.

Note that the M9 Leica was the only camera for which we observed a rather peculiar non-Gaussian shot noise.

Experiments

In this section, we subject the proposed natural steganography algorithms to tests on images taken with the CCD sensor from the Leica M Monochrome Type 230 and the CMOS sensor from the Z CAM E1 action camera. Images coming from the monochrome sensor are named MonoBase and are composed of 10,320 512×512 images in 16-bit PGM format developed using the command `"dcraw -k 0 -6 -W -g 1 1"`. Since there is no demosaicking on this sensor, this format is very close to the RAW format. Images coming from the E1 sensor are named E1Base and are generated from 200 DNG images that are developed and cropped to provide 10,800 512×512 images. Both E1Base and MonoBase can be downloaded from [3] and [6].

The switches used for MonoBase are from ISO 320 to ISO 1000 and for E1Base from ISO 100 to ISO 200. Because the E1 sensor is smaller and cheaper, the power of the stego signal for both sensors is of the same order of magnitude. The parameters (a,b) used to realize the switches are $(4.3, 3801)$ for the MonoBase and $(0.9, -800.0)$ for the E1Base. Note, however, MonoBase images use values coded between $[0; 2^{16}-1]$ due to the PGM format while E1Base values are between $[0; 2^{14}-1]$ due to the sensor dynamic range. We believe that the negative value of b for E1Base is due to a bias correction that is ISO dependent and coded inside the chipset.

The detection error is evaluated as the minimal total classification error probability under equal priors, $P_{\mathrm{E}} = \min_{P_{\mathrm{FA}}} \frac{1}{2}(P_{\mathrm{FA}} + P_{\mathrm{MD}})$, with P_{FA} and P_{MD} standing for the false-alarm and missed-detection rates, using a low complexity linear classifier [7]. The JPEG images were steganalyzed with the SRM [19], GFR [34], DCTR [22] and cc-JRM [28] feature sets. For improved readability, we report only the best detection (lowest error) over these four feature sets. All reported errors are averaged over ten different splits of the database into equal sized training and testing sets. The largest measured standard deviation over the ten splits was 0.0097.

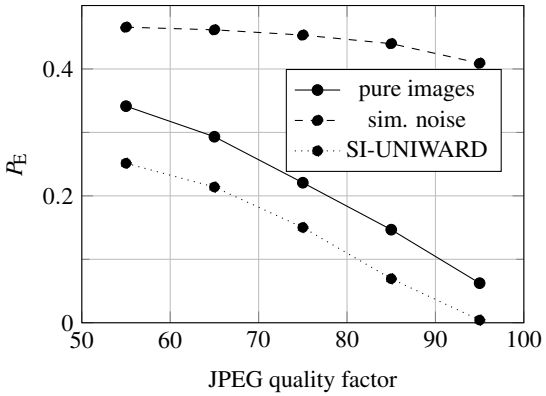


Figure 3: Detection error P_{E} for the pure, simulated noise, and SI-UNIWARD (Approaches 1, 2, and 6) for a switch from ISO 320 to ISO 1000 on MonoBase as a function of the JPEG quality factor. Approaches 3–5 exhibit security that is approximately equal to that of Approach 2 (simulated noise), see Table 2.

Results on MonoBase

In Figure 3, we show the detection errors for Approach 1, 2, and 5 (Pure, Simulated, and SI-UNIWARD) as a function of the JPEG quality factor. SI-UNIWARD embeds into each image the same payload obtained as the average payload over the whole database for Approach 4 using Eq. (15) (see Table 1).

In Table 2, we list the detection errors for all five approaches. The fact that the detection errors for Approach 3 and 4 are very close to the errors of Approach 2, the simulated acquisition noise that should preserve all dependencies among DCT coefficients, validates the simplifications of ignoring the dependencies during embedding used by Approach 3 and 4.

Our method clearly has a great promise, particularly w.r.t. the current state of the art in side-informed steganography, the SI-UNIWARD. For a JPEG QF of 95, the practical security of NS using MC-pmf (Approach 4) leads to $P_{\mathrm{E}} \simeq 40\%$ for an average embedding rate of 2.36 bpnzac when $P_{\mathrm{E}} \simeq 0\%$ for SI-UNIWARD. Note, however, that when making this comparison, it should be taken into account that NS needs the RAW file while SI-UNIWARD only needs the non-rounded values of DCT coefficients that are computed from the developed image, which is a substantially less extensive side-information.

It is also interesting to note that taking into account the dependencies between DCT coefficients within the same block has virtually no impact on the empirical security for this particular sensor. This is probably due to the fact that the DCT tends to generate uncorrelated coefficients whose dependencies are rather weak and/or not captured by the employed steganalysis features in this case. The analysis performed in Section "Model in the DCT domain" also shows that dependencies have to be taken into account only when the content inside a block is structured (edges, patterns). Such blocks will not be as common in high-resolution images investigated in this paper.

Another important point is that Approach 4, which treats the developing process as a black box and has access only to rounded DCT coefficients, is basically as secure as the other approaches. This indicates a path that can be taken for other, more advanced and more realistic developers for this sensor.

QF	Average embedding rate (bpp)	Average embedding rate (bpnzac)
55%	0.0158	0.2315
65%	0.0277	0.3474
75%	0.0462	0.4734
85%	0.1127	0.8607
95%	0.5300	2.3650

Table 1: Average payload in bits per pixel per MonoBase image embedded by Approach 3.

QF	Pure	Sim.	MC var.	MC pmf	SI-UNI
App	1	2	3	4	5
55%	.3414	.4659	.4672	.4716	0.2514
65%	.2932	.4617	.4610	.4601	0.2139
75%	.2207	.4534	.4511	.4486	0.1502
85%	.1467	.4399	.4449	.4438	0.0694
95%	.0624	.4090	.4112	.4093	0.0042

Table 2: Minimum detection error when steganalyzing the NS in MonoBase with SRM, GFR, DCTR, and cc-JRM feature sets for five different approaches and a range of JPEG quality factors for a switch from ISO 320 to ISO 1000.

Results on E1Base

We now evaluate the empirical security of NS in the JPEG domain for images coming from a color sensor. In contrast to monochrome sensors, after development the stego signal becomes dependent due to demosaicking.

Table 3 contains the detection results for all five embedding approaches for the E1 sensor. Compared to images from the monochrome sensor, the empirical security of Approach 2 (Simulated Noise) decreased by about 10% but the P_{E} remained above 30% for all QFs. However, the security of Approaches 3 and 4 ('MC var' and 'MC pmf'), is much lower especially for high QFs. We recommend using NS with Approaches 3 and 4 only for quality factors lower than 65 for which $P_{\mathrm{E}} \geq 25\%$ while the average embedding rate is still high with 2.5 bpnzac, see Table 4. The fact that the empirical security of 'MC var' is slightly larger than for 'MC pmf' is probably due to the fact that the number of samples used during Monte Carlo sampling (300), it not enough to accurately estimate the theoretical pmfs.

Note that Approach 1 (pure images) is also more detectable than for the monochrome sensor despite the gap between the two ISO sensitivities being similar. This is likely due to the dependencies introduced by demosaicking for images from the E1 sensor.

Comparing the embedding rates in bpnzac for the two databases (see Tables 1 and 4), while for MonoBase the rates increase from 0.2 to 2.36 bpnzac with increasing QF, they are nearly constant for the E1Base and always larger than 2 bpnzac. This can be explained by the fact that the demosaicking applied to E1 images increases the number of small DCT coefficients before quantization, especially in high frequencies. Thus, after quantization the number of non-zero coefficients is larger for MonoBase than for E1Base and the rate in bpnzac is correspondingly smaller.

Discussion

In this section, we attempt to explain the striking difference in empirical security of NS (Approach 4) when applied to the monochrome sensor and the color sensor. As analyzed in Section "Model in the DCT domain", depending on the block content, intra-block dependencies exist between DCT coefficients of the stego-signal. Furthermore, inter-block dependencies also exist between DCT coefficients from neighboring blocks due to the demosaicking process. Note, however, that the natural dependencies among neighboring pixels do not impact *per se* the dependency of the stego signal since the shot noise is independent from the photo-site values.

QF	Pure	Sim.	MC var.	MC pmf	SI-UNI
App	1	2	3	4	5
65%	0.1168	0.3426	0.2433	0.1920	0.1168
75%	0.0937	0.3385	0.1757	0.1473	0. 0957
85%	0.0732	0.3350	0.0881	0.0715	0.0752
95%	0.0595	0.3077	0.0056	0.0023	0.0032

Table 3: Detection error P_{E} when steganalyzing the NS in E1Base with SRM, GFR, DCTR, and cc-JRM feature sets for different approaches and JPEG quality factors for ISO switch 100 to 200. Note that the embedding capacity of SI-UNIWARD is limited to 1 bpnzac.

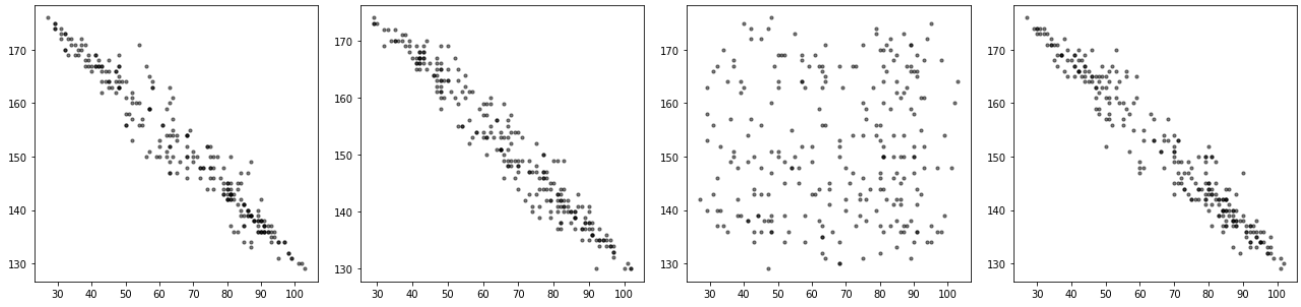
QF	Average embedding rate (bpp)	Average embedding rate (bpnzac)
65	0.0330	2.5556
75	0.0618	2.2849
85	0.1493	2.3336
95	0.5671	2.8488

Table 4: Average embedding rate for Approach 4 (MC pmf), E1Base.

In order to determine whether the loss of security is due to not preserving intra or inter-block dependencies among DCT coefficients, we conducted two experiments:

Experiment 1: We generated the stego images to preserve intra-block dependencies of the stego noise in each each DCT block. In particular, each block came from one specific realization of Approach 2 but different blocks came from different realizations. Calling this strategy 'Sim block-wise', its practical security is compared with Approach 2 in Table 5, which shows that the empirical security is even lower than the security of 'MC-pmf' (Approach 4). This means that the loss of security of 'MC-pmf' and 'MC-var' must be due to violating inter-block dependencies rather than not preserving intra-block dependencies.

Experiment 2: To confirm this hypothesis, we next used a synthetic RAW image with all photo-site values from even columns equal to 8192 and the values of odd columns equal to 5461 (see [20]). This content was selected purposely with harsh high-frequency discontinuities in order to magnify the errors the interpolation algorithm will introduce. The demosaicking has to predict the missing color components. After adding stego noise with arbitrary (a,b) parameters, we then apply both Approach 2 and Approach 'Sim block-wise' without considering the JPEG quantization step. Since co-occurrence matrices are sensitive to steganographic embedding – they are for example the basis of SPAM or SRM feature sets [19, 31] – we plot in Figure 4 the co-occurrence of the red color component of adjacent pixels after development. These sets of pairs of adjacent pixels are either lo-



(a) (b) (c) (d)

Figure 4: Experiment with a synthetic RAW image: co-occurrences of pixel pairs of adjacent pixels belonging either to adjacent blocks (a), to the same block (b), to adjacent blocks (c) or same block (d) after simulating noise that preserves dependencies only at the block-wise level.

cated on the boundaries of two adjacent DCT blocks for Approach 2 (Figure 4a) and for Approach 'Sim block-wise' (Figure 4c), or in the middle of one DCT block for Approach 2 (Figure 4b) and for Approach 'Sim block-wise' (Figure 4d). Note that the demosaicking process has a profound effect on inter-block dependencies. After Approach 2, which does not violate the demosaicking step, the co-occurrences are nearly identical for different pixel locations but if we compare Approach 2 and Approach 'Sim block-wise' for pixels located across the boundaries of DCT blocks (Figure 4a vs Figure 4c) the co-occurrences become very different because 'Sim block-wise' only preserves intra-block dependencies.

QF	Sim.	MC pmf	Sim block-wise
65	0.3426	0.1920	0.1511
75	0.3385	0.1473	0.1093
85	0.3350	0.0715	0.0274
95	0.3077	0.0023	0.0005

Table 5: Comparison between Approach 2 (simulated noise), which preserves intra-block dependencies, Approach 4 (independent embedding at each DCT coefficient), and simulated noise sampled independently for each DCT block.

We conclude from these two experiments that the low empirical security of Approach 'MC pmf' is due to the fact that it does not preserve inter-block dependencies between DCT coefficients. This conclusion is supported by the fact that preserving intra-block dependencies but not inter-block dependencies does not improve security (Experiment 1), and also by the fact that discrepancies form in co-occurrences of adjacent pixels from neighboring blocks (see Experiment 2).

Conclusions and perspectives

Natural steganography is an embedding paradigm in which sensor noise is added to a RAW (cover) image capture to embed the secret message, making thus the stego image look as if it was acquired at a higher ISO setting. The novel idea explored in this paper is extending NS to allow embedding of the message in quantized DCT coefficients in a JPEG file and with more complex RAW format developers. The most promising embedding algorithms studied in this paper estimate the distribution of quantized stego DCT coefficients using Monte-Carlo sampling by adding sensor noise to the RAW cover capture, developing the images, and then JPEG compressing. This approach is free of any modeling assumptions on the distribution of stego image DCT coefficients and can also be used with more complex (e.g., more realistic) developers.

Our findings can be summarized as follows:

- For images acquired by monochrome sensors, such as the Leica M Monochrome Type 230, when adopting a linear development, NS can embed large payloads (more than 2 bpnzac) with high empirical security ($P_{\mathrm{E}} > 0.4$) for a wide range of JPEG quality factors. We experimentally verified that making independent embedding changes to DCT coefficients does not significantly impact the security.
- When the same strategy (independent embedding in each DCT coefficient, linear development) is applied to images from a color sensor, the empirical security of NS becomes low. Further analysis showed that this loss of security can be attributed to the failure of the embedding algorithm to preserve inter-block dependencies between DCT coefficients introduced by the demosaicking process.

In our future work, we plan to address the problem of statistically modeling and better preserving inter-block dependencies between DCT coefficients for color sensors and move towards more advanced development pipelines. To this end, generative models, such as the PCA or the Generative Adversarial Networks using strategies similar to [35], could be used.

Acknowledgments

The work on this paper was partially supported by NSF grant No. 1561446 and 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.

Patrick Bas thanks Francesco Corona for the experiments he made with his M9 camera. This work was also partially supported by the French ANR DEFALS program (ANR-16-DEFA-0003).

References

[1] BOSSbase 1.01. http://agents.fel.cvut.cz/stegodata/.

[2] R. Anderson. Stretching the limits of steganography. In R. J. Anderson, editor, *Information Hiding, 1st International Workshop*, volume 1174 of Lecture Notes in Computer Science, pages 39–48, Cambridge, UK, May 30–June 1, 1996. Springer-Verlag, Berlin.

[3] P. Bas. Monobase. http://patrickbas.ec-lille.fr/MonoBase/, July 2016.

[4] P. Bas. Steganography via Cover-Source Switching. IEEE Workshop on Information Forensics and Security (WIFS), 2016.

[5] P. Bas. An embedding mechanism for Natural Steganography after down-sampling. IEEE ICASSP, 2017.

[6] P. Bas. E1base. http://patrickbas.ec-lille.fr/E1Base/, January 2018.

[7] R. Cogranne, V. Sedighi, T. Pevný, and J. Fridrich. Is ensemble classifier needed for steganalysis in high-dimensional feature spaces? In *IEEE International Workshop on Information Forensics and Security*, Rome, Italy, November 16–19 2015.

[8] T. Denemark and J. Fridrich. Improving steganographic security by synchronizing the selection channel. In A. Alattar, J. Fridrich, N. Smith, and P. Comesana Alfaro, editors, *The 3rd ACM Workshop on Information Hiding and Multimedia Security*, Portland, OR, June 17–19, 2015.

[9] T. Denemark and J. Fridrich. Side-informed steganography with additive distortion. In *IEEE International Workshop on Information Forensics and Security*, Rome, Italy, November 16–19 2015.

[10] T. Denemark and J. Fridrich. Steganography with multiple JPEG images of the same scene. *IEEE Transactions on Information Forensics and Security*, 12(10):2308–2319, October 2017.

[11] T. Denemark and J. Fridrich. Steganography with two JPEGs of the same scene. In *IEEE ICASSP*, New Orleans, March 5–9 2017.

[12] 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.

[13] E. Franz. Steganography preserving statistical properties. In F. A. P. Petitcolas, editor, *Information Hiding, 5th International Workshop*, volume 2578 of Lecture Notes in Computer Science, pages 278–294, Noordwijkerhout, The Netherlands, October 7–9, 2002. Springer-Verlag, New York.

[14] E. Franz. Embedding considering dependencies between pixels. In E. J. Delp, P. W. Wong, J. Dittmann, and N. D. Memon, editors, *Proceedings SPIE, Electronic Imaging, Security, Forensics, Steganography, and Watermarking of Multimedia Contents X*, volume 6819, pages D 1–D 12, San Jose, CA, January 27–31, 2008.

[15] E. Franz and A. Schneidewind. Pre-processing for adding noise steganography. In M. Barni, J. Herrera, S. Katzenbeisser, and F. Pérez-González, editors, *Information Hiding, 7th International Workshop*, volume 3727 of Lecture Notes in Computer Science, pages 189–203, Barcelona, Spain, June 6–8, 2005. Springer-Verlag, Berlin.

[16] 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.

[17] J. Fridrich and M. Goljan. Digital image steganography using stochastic modulation. In E. J. Delp and P. W. Wong, editors, *Proceedings SPIE, Electronic Imaging, Security and Watermarking of Multimedia Contents V*, volume 5020, pages 191–202, Santa Clara, CA, January 21–24, 2003.

[18] J. Fridrich, M. Goljan, and D. Soukal. Perturbed quantization steganography using wet paper codes. In J. Dittmann and J. Fridrich, editors, *Proceedings of the 6th ACM Multimedia & Security Workshop*, pages 4–15, Magdeburg, Germany, September 20–21, 2004.

[19] 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.

[20] Q. Giboulot, R. Cogranne, and P. Bas. Steganalysis into the wild: How to define a source? In A. Alattar and N. D. Memon, editors, *Proceedings IS&T, Electronic Imaging, Media Watermarking, Security, and Forensics 2016*, San Francisco, CA, January 29–February 2, 2018.

[21] L. Guo, J. Ni, and Y. Q. Shi. Uniform embedding for efficient JPEG steganography. *IEEE Transactions on Information Forensics and Security*, 9(5):814–825, May 2014.

[22] V. Holub and J. Fridrich. Low-complexity features for JPEG steganalysis using undecimated DCT. *IEEE Transactions on Information Forensics and Security*, 10(2):219–228, February 2015.

[23] 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.

[24] F. Huang, J. Huang, and Y.-Q. Shi. New channel selection rule for JPEG steganography. *IEEE Transactions on Information Forensics and Security*, 7(4):1181–1191, August 2012.

[25] A. D. Ker. A fusion of maximal likelihood and structural steganalysis. In T. Furon, F. Cayre, G. Doërr, and P. Bas, editors, *Information Hiding, 9th International Workshop*, volume 4567 of Lecture Notes in Computer Science, pages 204–219, Saint Malo, France, June 11–13, 2007. Springer-Verlag, Berlin.

[26] S. Kim and W. Sung. Optimum wordlength determination of 8x8 IDCT architectures conforming to the IEEE standard specifications. In *Signals, Systems and Computers, 1995. 1995 Conference Record of the Twenty-Ninth Asilomar Conference on*, volume 2, pages 821–825. IEEE, 1995.

[27] Y. Kim, Z. Duric, and D. Richards. Modified matrix encoding technique for minimal distortion steganography. In J. L. Camenisch, C. S. Collberg, N. F. Johnson, and P. Sallee, editors, *Information Hiding, 8th International Workshop*, volume 4437 of Lecture Notes in Computer Science, pages 314–327, Alexandria, VA, July 10–12, 2006. Springer-Verlag, New York.

[28] J. Kodovský and J. Fridrich. Steganalysis in high dimensions: Fusing classifiers built on random subspaces. In A. Alattar, N. D. Memon, E. J. Delp, and J. Dittmann, editors, *Proceedings SPIE, Electronic Imaging, Media Watermarking, Security and Forensics III*, volume 7880, pages OL 1–13, San Francisco, CA, January 23–26, 2011.

[29] 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.

[30] Rawkit package. `https://rawkit.readthedocs.io/en/latest/`. Python package.

[31] T. Pevný, P. Bas, and J. Fridrich. Steganalysis by subtractive pixel adjacency matrix. In J. Dittmann, S. Craver, and J. Fridrich, editors, *Proceedings of the 11th ACM Multimedia & Security Workshop*, pages 75–84, Princeton, NJ, September 7–8, 2009.

[32] V. Sachnev, H. J. Kim, and R. Zhang. Less detectable JPEG steganography method based on heuristic optimization and BCH syndrome coding. In J. Dittmann, S. Craver, and J. Fridrich, editors, *Proceedings of the 11th ACM Multimedia & Security Workshop*, pages 131–140, Princeton, NJ, September 7–8, 2009.

[33] V. Sedighi and J. Fridrich. Effect of saturated pixels on security of steganographic schemes for digital images. In *Image Processing (ICIP), 2016 IEEE International Conference on*, pages 2747–2751. IEEE, 2016.

[34] X. Song, F. Liu, C. Yang, X. Luo, and Y. Zhang. Steganalysis of adaptive JPEG steganography using 2D Gabor filters. In A. Alattar, J. Fridrich, N. Smith, and P. Comesana Alfaro, editors, *The 3rd ACM Workshop on Information Hiding and Multimedia Security*, IH&MMSec '15, Portland, OR, June 17–19, 2015.

[35] W. Tang, S. Tan, B. Li, and J. Huang. Automatic steganographic distortion learning using a generative adversarial network. *IEEE Signal Processing Letters*, 24(10):1547–1551, 2017.

[36] C. Wang and J. Ni. An efficient JPEG steganographic scheme based on the block–entropy of DCT coefficents. In *Proc. of IEEE ICASSP*, Kyoto, Japan, March 25–30, 2012.

Author Biography

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

Patrick Bas received a Ph.D. degree in signal and image processing from Institut National Polytechnique de Grenoble, France, in 2000. He has co-organized the 2nd Edition of the BOWS-2 contest on watermarking in 2007, and the first edition of the BOSS contest on steganalysis in 2010. From 2013 to 2016 Patrick Bas was associate editor of IEEE Transactions of Information Forensics and Security (IEEE TIFS). Patrick Bas is the current group leader of the team working on Signal and Images.

Jessica Fridrich is Distinguished 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 forensics. Since 1995, she has received 20 research grants totaling over $11 mil that lead to more than 180 papers and 7 US patents.