Effect of Cover Quantization on Steganographic Fisher Information

Jessica Fridrich, *Member, IEEE*

***Abstract*—The square-root law of imperfect steganography ties the embedding change rate and the cover length with statistical detectability. In this article, we extend the law to consider the effects of cover quantization. Assuming the individual cover elements are quantized i.i.d. samples drawn from an underlying continuous-valued 'precover' distribution, the steganographic Fisher information scales as $\triangle^s$, where $\triangle$ is the quantization step and s is determined jointly by the smoothness of the precover distribution and the properties of the embedding function. This extension is relevant for understanding the effects of the pixel color depth and the JPEG quality factor on the length of secure payload.**

I. Introduction

Perfectly secure stegosystems are statistically undetectable in the sense that they preserve the statistical distribution of covers. Granting a complete knowledge of the cover distribution to both the embedder and the Warden, perfectly secure stegosystems exist and they are capable of communicating a positive payload per cover sample (channel use) – their steganographic capacity is positive. Another way of stating this is that the secure payload is linearly proportional to the cover length. This result has been established under fairly general conditions for the case when the actions of the embedder are power-restricted (distortion–limited embedder) and the Warden is active and her actions power-restricted as well [25], [26].

When using empirical objects for steganography, such as digital images or audio, the cover model is never known completely and perfect security becomes unachievable [2]. In fact, so far all stegosystems designed for digital media have been shown detectable because the Warden seems always able to find such a representation of the media (feature space) in which the distribution of covers is not preserved. The square-root law (SRL) of steganography is an asymptotic scaling result concerning the length of secure payload for stegosystems that are imperfect. It allows quantifying how long a message can be embedded in a cover of a certain size for a given non-zero level of statistical detectability. The critical quantity here is the product of the square root of the cover size, N, and the change rate β with which the steganographer modifies cover samples. If $\beta\sqrt{N} \to 0$ with $N \to \infty$, the law guarantees asymptotic perfect security, while when $\beta\sqrt{N} \to \infty$, the Warden wins as asymptotically perfect detection becomes possible. The SRL becomes readily apparent when expanding the KL divergence between cover and stego distributions at $\beta = 0$ using Taylor series – the leading quadratic term is $\frac{1}{2}N\beta^2 I(0)$, where $I(0)$ is the steganographic Fisher information.[1]

The effects of the law have been suspected[2] and observed by practitioners long before it was discovered [12], formulated, formally established [7], and verified [17] – the same relative payload was observed to be more easily detectable in larger covers than in small covers. This made the relative payload unsuitable for quantifying the secure payload and lead to an alternative – the so-called root rate tightly connected to the steganographic Fisher information [6], [15].

The law manifests when the Warden is computationally unbounded and granted the full knowledge of the cover source and the embedding method. For an ignorant Warden who needs to *learn* the cover source, the law may not manifest in the above form depending on the knowledge available to the Warden [16].[3] In this paper, we constrain ourselves to a fully-informed and computationally unbounded Warden.

For digital media, the individual cover elements are typically obtained by quantizing an analog precover signal with scalar quantizer with step $\triangle$,[4] such as photon counts registered at pixel wells on a sensor or non-rounded transform coefficients when compressing an image to the JPEG format. The quantization step $\triangle$ directly affects the statistical properties of covers and thus the asymptotic scaling laws through changes in $I(0)$.[5] Investigation of this

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.

The author would like to thank Jan Kodovský and Tomáš Filler for useful comments.

The author is with the Department of Electrical and Computer Engineering, Binghamton University, NY, 13902, USA. Email: fridrich@binghamton.edu.

[1]The first mentioning of Fisher information in steganography is due to Ker [14].

[2]The fact that the secure payload length may be sublinear in cover size appeared for the first time in [1].

[3]Recently, game theory was proposed as an alternative and appealing possibility to formally capture the sender's and Warden's ignorance [3].

[4]The concept of precover is due to Ker [13].

[5]The importance of quantizers in the theoretical analysis of steganographic security was studied in [28].

effect is the main topic of this paper. The main result is a theoretical understanding of the mechanism through which the quantized precover distribution and the embedding operation affect the steganographic Fisher information. The original SRL is extended as constant statistical detectability is now obtained when $\frac{1}{2}N\beta^2\triangle^s = const.$, where s is the scaling exponent of the Fisher information determined jointly by the precover distribution and the embedding operation. Since the Fisher information is a multiplicative factor in the error exponent, quantization strongly affects statistical detectability.

Unlike the SRL, which is quite robust and observable in practice even when the source is empirical and the Warden uses feature-based classifiers instead or optimal detectors, the scaling due to quantization is fragile as it sensitively depends on the precover distribution. This prevents applying the results for quantifying the scaling in empirical cover sources with empirical measures of security (e.g., relating the classifier error to color bit depth or JPEG quality factor) although some qualitative conclusions appear to be in agreement with experiments and prior art.

The paper starts in the next section with a formalization of basic concepts, such as the precover and cover source, embedding operation, steganographic Fisher information, and two types of continuous distributions used to model elements of digital media files. In Section III, we establish the main result, which is the scaling of the Fisher information w.r.t. the quantization step for smooth precover distributions as well as distributions with a singularity. For better readability, the proofs of all theorems are in appendices. The theory is applied to four common embedding operations and two precover models in Section IV. Extension of this work to the case when the embedding is realized in the pixel domain but the model is built from pixel residuals is outlined in Section V. The theoretical results are interpreted in Section VI, where they are also related to practical steganographic schemes in digital media. In particular, the derived scaling appears to be in qualitative agreement with experiments reported in prior art. Section VII contains an experiment with a database of grayscale images sampled at varying bit depths that demonstrates the strong effect of cover quantization on security. It also discusses the limitations of observing the theoretical scaling in practice when using sampled data. Finally, Section VIII summarizes the contribution.

A condensed version of this paper appeared at the 2012 IEEE Workshop on Information Forensic and Security.

II. Preliminaries

Throughout the paper, a_{ij} denotes a (potentially infinite) two-dimensional array with elements a_{ij}, $i,j \in \mathbb{Z}$. Calligraphic font is reserved for sets, while capital letters with their corresponding lower-case letters are used for random variables and their realizations. The symbol $\lceil x \rceil$ stands for rounding up and $\Gamma(x)$ for the gamma function. For real functions g, h, we define $g(x) = \Theta(h(x))$ at $x = a$ if $G_1 h(x) \le g(x) \le G_2 h(x)$ for $G_1, G_2 > 0$ on some neighborhood of $a \in \{\mathbb{R}, -\infty, \infty\}$. We write $g \approx h$ whenever $\lim_{x\to a} g(x)/h(x)$ exists and is positive. We also use the standard Landau big-O and little-o notation. The lth derivative of a real function F will be denoted $F^{(l)}$.

Given a countable set of scalar bin centroids, $\mathcal{M} = (m_j)$, $m_j < m_{j+1}$, a scalar quantizer is a mapping $Q_{\mathcal{M}} : \mathbb{R} \to \mathcal{M}$, defined as $Q_{\mathcal{M}}(x) = \arg\min_{m_j \in \mathcal{M}} |x - m_j|$. In this paper, we will assume that $Q_{\mathcal{M}}$ is uniform, $m_j = j\triangle$, $j \in \mathbb{Z}$, where $\triangle > 0$ is the bin width. A uniform quantizer with bin width $\triangle$ will be denoted $Q_\triangle$.

A. Precover and cover source

An N-element precover source will be represented using a random variable $Z \triangleq (Z_1, \dots, Z_N)$ where Z_k are mutually independent and identically distributed (i.i.d.) continuous-valued random variables $Z_k \sim f(x)$, where $f(x)$ is a probability density function. For convenience, we will assume that f is even and that its domain can be continuously extended to $\mathbb{R}$.

A cover source X corresponding to a precover source Z is obtained by applying $Q_\triangle$ to each element of Z, $X = (X_1, \dots, X_N) = (Q_\triangle(Z_1), \dots, Q_\triangle(Z_N))$ with $X_k \sim p$, a probability mass function on $\mathcal{M}$:

$$p_j(f, \triangle) = \Pr(X_k = m_j) = \int_{m_j - \triangle/2}^{m_j + \triangle/2} f(x)\mathrm{d}x. \quad (1)$$

B. Embedding operation

Since the specific details of the embedding (and extraction) algorithms are not important for our study, we only model the probabilistic impact of embedding. In particular, we narrow our study to the so-called mutually independent embedding[6] that modifies every cover element X_k independently to a corresponding element of the stego object Y_k with probability

$$\Pr(Y_k = m_j | X_k = m_i) \triangleq b_{ij}(\beta) = \begin{cases} 1 + \beta c_{ii} & \text{if } i = j \\ \beta c_{ij} & \text{otherwise,} \end{cases} \quad (2)$$

for some constants $c_{ij} \ge 0$ for $i \ne j$. Since $\sum_j b_{ij} = 1$, we must have $c_{ii} = -\sum_{j \ne i} c_{ij}$ for each i. The scalar parameter $\beta \ge 0$ will typically be the change rate. Also, (2) implies that $c_{ij} < C$ for some $C > 0$ and all i, j.

Note that the domain and range of the embedding operation depend on $\triangle$. Since we will let $\triangle \to 0$, the number of bins (indices i and j) will be unbounded. In this paper, we restrict ourselves to embedding with changes from a limited range: $c_{ij} = 0$ whenever $|i - j| > L$ for some fixed positive integer L. This restriction is quite reasonable since most practical steganographic schemes modify cover elements to a few neighboring values, such as by ± 1. Having said this, it is possible (and perhaps also meaningful) to study embedding operations whose range L increases with $\triangle \to 0$.

[6] The concept of mutually independent embedding was introduced and studied for the first time in [7].

C. Stego distribution and Fisher information

Given the embedding operation above, the stego object is an i.i.d. sequence of random variables $Y \triangleq (Y_1, \dots, Y_N)$ with $Y_k \sim q(f, \triangle, \beta)$. For better readability, we will drop the arguments $\triangle$ and f from p and q.

Assumption (2) leads to the following relationship between p_j and $q_j(\beta)$:

$$q_j(\beta) = \Pr(Y_k = m_j) = \sum_i b_{ij} p_i = p_j + \beta \sum_i c_{ij} p_i. \quad (3)$$

In this article, we study four specific embedding operations chosen as representative examples of today's popular stego algorithms. They all hide message bits by changing the Least Significant Bit (LSB) of cover elements. In LSB replacement (LSBR), the cover LSBs are *replaced* with message bits, while in LSB matching (LSBM), the cover LSB is *matched* with the message bit by randomly adding or subtracting 1 from the cover value. In F5 [27], the absolute value of the cover element is decreased if the LSB needs to be changed; F5 does not embed in cover elements that are equal to zero. Finally, symmetrized Jsteg (symJsteg) is a symmetrized version [18] of Jsteg [23] in which cover values are exchanged within the following pairs of values: $\dots, \{-4,-3\}, \{-2,-1\}, \{1,2\}, \{3,4,\dots\}$. Again, zeros are not used for embedding.

Assuming that the above embedding operations are executed at change rate β at randomly selected cover elements, the expected values of the stego distributions are for LSBM, LSBR, F5, and symJsteg, respectively:

$$q_j(\beta) = (1-\beta)p_j + \frac{1}{2}\beta(p_{j+1} + p_{j-1}) \quad \text{for all } j, \quad (4)$$

$$q_j(\beta) = (1-\beta)p_j + \beta p_{j+(-1)^j} \quad \text{for all } j, \quad (5)$$

$$q_j(\beta) = \begin{cases} (1-\beta)p_j + \beta p_{j+\mathrm{sign}(j)} & \text{for } j \neq 0, \\ p_0 + \beta(p_1 + p_{-1}) & \text{for } j = 0, \end{cases} \quad (6)$$

$$q_j(\beta) = \begin{cases} (1-\beta)p_j + \beta p_{j+\mathrm{sign}(j)} & \text{for } j \text{ odd}, \\ (1-\beta)p_j + \beta p_{j-\mathrm{sign}(j)} & \text{for } j \text{ even}, \\ p_j & \text{for } j = 0. \end{cases} \quad (7)$$

In accordance with the information-theoretic definition of steganographic security by Cachin [4], we measure security using the Kullback–Leibler (KL) divergence between the cover and stego distributions, $D_{\mathrm{KL}}(p||q(\beta))$. By expanding it using Taylor series at $\beta = 0$, the following standard result (see, e.g., [8]) is obtained:

$$D_{\mathrm{KL}}(p||q) \triangleq \sum_j p_j \log \frac{p_j}{q_j(\beta)} = \frac{1}{2}\beta^2 I_\triangle(0) + O(\beta^3), \quad (8)$$

where, using (3),

$$I_\triangle(0) \triangleq \sum_j \frac{1}{p_j} \left(\frac{\mathrm{d}q_j(\beta)}{\mathrm{d}\beta}\Big|_{\beta=0} \right)^2 = \sum_j \frac{1}{p_j} \left(\sum_i c_{ij} p_i \right)^2, \quad (9)$$

is the steganographic Fisher information (FI), which encapsulates the effect of embedding. Larger values of $I_\triangle(0)$ lead to larger D_{KL} and thus lower the steganographic security. The KL divergence is the scaling exponent that controls the probability of missed detection, P_{MD}, for a fixed false alarm rate,[7] P_{FA}, in Neyman–Pearson hypothesis testing, $P_{\mathrm{MD}} \approx e^{-N D_{\mathrm{KL}}(p||q)}$ for large N (Chernoff–Stein Lemma, Sec. 12.8 in [5]). Therefore, even a small change in the FI has a pronounced effect on the detection error. For example, twice as large FI changes P_{MD} to P_{MD}^2.

Note that, formally, $I_\triangle(0)$ depends on f and $\triangle$ through (1). It is precisely this relationship that is of interest in this paper.

D. Common distributions

Two continuous densities commonly used for modeling the distribution of digital media elements, such as pixel differences or transform coefficients, are the generalized Gaussian distribution (GGD) and the generalized Cauchy distribution (GCD). Both depend on three parameters: the mean μ, the shape parameter $\alpha > 0$ ($\tau > 1$), and the parameter controlling the width of the distribution, $b > 0$:

$$f_{\mathrm{GG}}(x) = \frac{\alpha}{2b\Gamma(1/\alpha)} \exp\left(-\frac{|x-\mu|^\alpha}{b} \right), \quad (10)$$

$$f_{\mathrm{GC}}(x) = \frac{\tau - 1}{2b} \left(1 + \frac{|x-\mu|}{b} \right)^{-\tau}. \quad (11)$$

Besides the case when $\alpha \geq 2$, $\alpha \in \mathbb{Z}$, the GGD has a singularity at $x = \mu$ as its derivatives become unbounded there, starting with the $\lceil \alpha \rceil$th derivative. In contrast, all one-sided derivatives of the GCD are bounded but do not exist at $x = \mu$.

III. Scaling due to quantization

In this section, we analyze the effects of cover quantization on the Fisher information. First, a general result is derived for smooth precover densities and then extended to densities with singularities to cover the GGD as well as the GCD.

For $\triangle > 0$ and $x \in \mathbb{R}$, we define

$$F_\triangle(x) \triangleq \int_{x-\triangle/2}^{x+\triangle/2} f(t)\mathrm{d}t. \quad (12)$$

Thus, $p_i = F_\triangle(i\triangle)$ and (9) becomes:

$$I_\triangle(0) = \sum_j \frac{\left(\sum_i c_{ij} F_\triangle(i\triangle)\right)^2}{F_\triangle(j\triangle)}. \quad (13)$$

The sum in the numerator is a discrete filter with (a generally non stationary) kernel $c_{\cdot j}$ applied to $F_\triangle$ sampled at $j\triangle$. It is shown in this section that the scaling exponent s in $I_\triangle(0) \propto \triangle^s$ depends on the leading order, k, of

$$\sum_i c_{ij} F_\triangle(i\triangle) \propto \triangle^k, \quad (14)$$

which is jointly determined by f and the embedding operation.

To obtain the scaling of the FI (13) w.r.t. $\triangle$, we will typically divide the set of all real numbers into a union

[7] False alarm corresponds to identifying a cover image as stego.

of finitely many disjoint intervals, $[m_1, m_2]$, $m_1, m_2 \in \mathbb{R} \cup \{\infty, -\infty\}$, and establish the scaling separately for each *partial sum* defined as

$$s_{m_1,m_2}(\triangle) \triangleq \sum_{m_1 \le j\triangle \le m_2} \frac{\left(\sum_i c_{ij} F_\triangle(i\triangle)\right)^2}{F_\triangle(j\triangle)}. \tag{15}$$

A. f differentiable

From the first mean value theorem for integration:

$$F_\triangle(j\triangle) = \triangle f(u_j) \tag{16}$$

for some $u_j \in (j\triangle - \triangle/2, j\triangle + \triangle/2)$. For $i \neq j$, $|i-j| \le L$, we expand $F_\triangle(i\triangle) = F_\triangle(j\triangle + (i-j)\triangle)$ at $j\triangle$ using Taylor series with Lagrange remainder. Assuming $F_\triangle(x)$ is $k \ge 0$ times continuously differentiable

$$F_\triangle(i\triangle) = \sum_{l=0}^{k-1} \frac{F_\triangle^{(l)}(j\triangle)}{l!} \triangle^l (i-j)^l + \frac{F_\triangle^{(k)}(\xi_{ij})}{k!} \triangle^k (i-j)^k, \tag{17}$$

where $\xi_{ij} \in (j\triangle, i\triangle)$ or $(i\triangle, j\triangle)$, depending on whether $i > j$ or $j > i$. Therefore,

$$\sum_i c_{ij} F_\triangle(i\triangle) = \sum_{l=0}^{k-1} \frac{F_\triangle^{(l)}(j\triangle)}{l!} \triangle^l w_{jl} + \frac{\triangle^k}{k!} \sum_i c_{ij} (i-j)^k F_\triangle^{(k)}(\xi_{ij}) \tag{18}$$

$$= \sum_{l=0}^{k-1} \frac{f^{(l)}(\phi_{jl})}{l!} \triangle^{l+1} w_{jl} + \frac{\triangle^{k+1}}{k!} \sum_i c_{ij} (i-j)^k f^{(k)}(\tilde{\phi}_{ij}), \tag{19}$$

where

$$\phi_{jl} \in (j\triangle - \triangle/2, j\triangle + \triangle/2), \tag{20}$$

$$\tilde{\phi}_{ij} \in (\xi_{ij} - \triangle/2, \xi_{ij} + \triangle/2) \subset \mathcal{I}_j(L, \triangle), \tag{21}$$

$$\mathcal{I}_j(L, \triangle) \triangleq [(j - L - 1/2)\triangle, (j + L + 1/2)\triangle], \tag{22}$$

and the weights

$$w_{jl} = \sum_i c_{ij} (i-j)^l, \tag{23}$$

depend only on the embedding operation but not on the density f. Equation (19) follows from $F_\triangle^{(l)}(x) = f^{(l-1)}(x + \triangle/2) - f^{(l-1)}(x - \triangle/2) = \triangle f^{(l)}(\phi)$ for $1 \le l \le k+1$ and some $\phi \in (x - \triangle/2, x + \triangle/2)$.

The scaling of the FI w.r.t. $\triangle$ depends on the scaling of the sum (14) for each bin j, which from (19) depends on the first non-zero element in the sequence $w_{j0}, w_{j1}, w_{j2}, \ldots$. This justifies the following definition.

Definition 1. The leading order $k^\star$ of the sum (14) at $x = j\triangle$ is defined as the largest k for which $w_{jl} = 0$ for all $l < k$. If $w_{jl} = 0$ for all $l \ge 0$, we set $k^\star = \infty$ (as is the case of symJsteg for $j = 0$). Note that (23) implies that in general $k^\star \in \{0, 1, 2, \infty\}$. Furthermore, if $|w_{jk^\star}| = w_{k^\star}$ for all j, such that $j\triangle \in \mathcal{I} \subset \mathbb{R}$, we say that the embedding operation is bin-invariant on $\mathcal{I}$.

TABLE I
LEADING ORDER $k^\star$ FOR FOUR EMBEDDING OPERATIONS. FOR F5, THE NUMBERS IN BRACKETS ARE AT $j = 0$.

Embedding	$k^\star$	$w_{k^\star}$
LSBM	2	1
LSBR	1	1
F5	1 (0)	1 (2)
symJsteg	1	1
$c_{j-m,j} = c_{j+m,j}$	2	$2\sum_{m=1}^{L} m^2 c_{j+m,j}$

To obtain more insight on how the leading order depends on the embedding operation, we now run through the four embedding operations described in Section II-C. For LSBR and symJsteg, $w_{j0} = \sum_i c_{ij} = 0$ for all j. The condition $w_{j0} = 0$ for all j is equivalent with b_{ij} being doubly stochastic (the sum of rows and columns is equal to 1). For LSBM, $w_{j0} \neq 0$ only at the boundary of the dynamic range of cover elements, which cannot happen in our formulation with an unbounded dynamic range. Thus, the only embedding operation with non-zero w_{j0} is F5 for $j = 0$ where $w_{00} = 2$ because $c_{-1,0} = c_{1,0} = 1$, and $c_{00} = 0$ (see (6)).

Example 1. Hypothetically, one could construct embedding operations with $w_{j0} \neq 0$ for all j, such as a scheme that does not embed in bins $3j$ and always changes $3j-1$ and $3j+1$ into $3j$.

Continuing our discussion of the weights and the leading order, note that $w_{j1} = 0$ for embedding operations that are "symmetrical" in the sense that they modify each value of the cover by $\pm(i-j)$ with equal probabilities, e.g., for LSBM. For LSBR, symJsteg, and F5 (at $j \neq 0$), $|w_{j1}| = 1$. Finally, in general $w_{j2} > 0$ for all embedding operations except when the embedding does not embed in bin j ($c_{ij} = 0$ whenever $i \neq j$). In particular, for all four embedding operations $w_{j2} = 1$ for all j with the exception of F5, where $w_{02} = 2$, and symJsteg where $w_{0k} = 0$ for all $k \ge 0$.

Table I summarizes the leading order for all embedding operations considered so far. The leading order for F5 is 1 for $j \neq 0$ and it is equal to 0 for $j = 0$. LSBR and LSBM are bin-invariant on $\mathbb{R}$ while F5 and symJsteg are bin-invariant on $\mathbb{R} - \{0\}$ as they both apply a different embedding rule at $j = 0$.

We are now ready to state the first scaling theorem for precover densities $f(x)$ satisfying the following regularity conditions for some sufficiently large $M > 0$:

R1. $|f^{(k)}(x)|$ is monotone decreasing for $0 \le k \le k^\star$ and $x > M$.

R2. There exists $\delta_0 > 0$ such that $\int_M^\infty \frac{(f^{(k^\star)}(x))^2}{f(x+\delta)} \mathrm{d}x$ is convergent in Riemann sense for all $\delta \in (0, \delta_0]$.

Assumption R2 essentially guarantees that $f(x)$ does not fall to zero too quickly. An example of a density that satisfies R2 for $\delta = 0$ but not for any $\delta > 0$ is the double exponential, $\exp(-\exp(x^2))$. Assumptions R1–R2

are easily established for both the GGD and GCD for all $k \geq 0$, $k \in \mathbb{Z}$.

Theorem 1. Let the embedding operation be bin-invariant with leading order $k^\star$ everywhere. Assuming that $f(x)$ is $k^\star+1$ times continuously differentiable and satisfies the regularity conditions R1–R2, the Fisher information scales as

$$\lim_{\triangle \to 0^+} \frac{I_\triangle(0)}{\triangle^{2k^\star}} = \frac{w_{k^\star}^2}{k^\star!^2} \int_{-\infty}^{\infty} \frac{(f^{(k^\star)}(x))^2}{f(x)} \mathrm{d}x. \tag{24}$$

The theorem is proved in Appendix A by writing $I_\triangle(0)$ as a sum of three partial sums, $I_\triangle(0) = s_{-\infty,-M} + s_{-M,M} + s_{M,\infty}$, and analyzing each sum separately. After substituting (19) and (16) into the numerator and denominator, respectively, the infinite partial sums are shown to be $\triangle^{2k^\star} \times o(M)$ due to the restrictions at ∞ imposed on f, while $s_{-M,M} \approx \triangle^{2k^\star}$.

Remark 1. By analyzing the technical approach in the proof, it can be seen that Theorem 1 can be easily extended to non-even precover densities by imposing the regularity condition at both $\pm\infty$. It can also be generalized to embedding operations that are not bin-invariant on the entire range. Quite often, special embedding rules are adopted at $j = 0$. For example, the F5 embedding operation has $w_{00} \neq 0$ in which case, $I(0) \propto \triangle^1$ since at most $2L + 1$ bins (a number which does not depend on $\triangle$) are affected. In general, modifying the embedding rule so that the leading order is $k' < k^\star$ for finitely many bins j changes the scaling from $\triangle^{2k^\star}$ to $\triangle^{2k'+1}$. Finally, note that $I(0) \propto \triangle^0$ for the operation from Example 1 since $k^\star = 0$ for all j.

Remark 2. For the GGD (10), the integral $R(f) = \int_{\mathbb{R}} (f^{(k^\star)}(x))^2 / f(x) \mathrm{d}x$ in Theorem 1 can be evaluated analytically:

$$R(f) = \begin{cases} \frac{\alpha^2}{b^{2/\alpha}} \frac{\Gamma(2-1/\alpha)}{\Gamma(1/\alpha)} & \text{for } k^\star = 1, \alpha \geq 1/2, \\ \frac{\alpha^2(\alpha-1)(3\alpha-4)}{b^{4/\alpha}} \frac{\Gamma(2-3/\alpha)}{\Gamma(1/\alpha)} & \text{for } k^\star = 2, \alpha \geq 3/2. \end{cases} \tag{25}$$

B. f with singularity

We limit our analysis to densities with only one singularity located at $x = 0$. Since the scaling of the partial sums (15) on closed intervals $[m_1, m_2]$ not containing the singularity can be carried out as for smooth densities, we only need to address the scaling *near* or *at* the singularity. (The meaning of both terms will be specified shortly.) In particular, we restrict our study to the case when the embedding operation and precover density satisfy the following two assumptions:[8]

S1. The embedding operation is bin-invariant on $\mathbb{R} - \{0\}$ with leading order $k^\star \geq 1$.

S2. $f(x)$ is continuous on $\mathbb{R}$ and has a singularity at $x = 0$ such that on some neighborhood of zero, $f^{(k^\star)}(x) = g(x)|x|^{-n}$, $n > 0$ for a continuous $g(x)$ with $g(0) \neq 0$.

[8]Both assumptions hold for the GGD and GCD and all four embedding operations studied in this paper.

Under these assumptions, a fairly general result (Theorem 2 below) can still be obtained for partial sums on the "immediate neighborhood" of the singularity – intervals $[(L+2)\triangle, \epsilon]$ for a fixed $\epsilon > 0$. Since many steganographic algorithms adjust the embedding rule at the singularity, the scaling of the remaining $2L+3$ terms j for $|j| \leq L+1$ is carried out in less generality only for specific embedding schemes in Section III-B2.

1) Immediate neighborhood of singularity:

Theorem 2. Under Assumptions S1–S2, for all sufficiently small $\epsilon > 0$ and $\triangle$ such that $(L+2)\triangle < \epsilon$,

$$s_{(L+2)\triangle,\epsilon} = \begin{cases} \Theta(\triangle^{2k^\star+1-2n}) & \text{when } n \geq 1/2, \\ \Theta(\triangle^{2k^\star}) & \text{when } n < 1/2. \end{cases} \tag{26}$$

The proof, which appears in Appendix B, starts with rewriting (19) using the fact that $\sum_i \gamma_i f(x_i) = f(\overline{x}) \sum_i \gamma_i$, $\overline{x} \in [\min_i x_i, \max_i x_i]$, for any f continuous when all γ_i are of the same sign. Then, thanks to the assumption on the $k^\star$th derivative, the partial sum $s_{(L+2)\triangle,\epsilon}$ can be squeezed between two integrals that are shown to exhibit the same scaling.

Remark 3. Since for the GGD (10), $f'(x) \approx |x|^{\alpha-1}$ and $f''(x) \approx |x|^{\alpha-2}$ at $x = 0$, the scaling of the partial sum near the singularity depends only on the shape parameter α. Table (II) summarizes Theorem 2 for the GGD.

Remark 4. For the GCD, the scaling is much simpler than for the GGD. In fact, whenever the singularity is such that f is not differentiable at $x = 0$ but the one-sided derivatives exist and are bounded (which holds for the GCD but not for the GGD), the same approach as in the proof of Theorem 1 can be used to show that $s_{(L+2)\triangle,\epsilon} \propto \triangle^{2k^\star}$.

2) At the singularity: In this section, we explain how to obtain the scaling of the remaining $2L + 3$ terms of the partial sum – the cases when $|j| \leq L + 1$. As shown in Appendix C, Assumptions S1–S2 guarantee the following form of the precover density on some neighborhood of the singularity: $f(x) = a_0 - a_1|x|^\lambda + o(|x|^\lambda)$ for some $\lambda > 0$, $a_0 > 0$, and $|a_1| > 0$. Since f is even, $p_i = p_{-i}$ for all i, and straightforward integration gives:

$$p_0 = 2 \int_0^{\triangle/2} a_0 - a_1 x^\lambda + o(x^\lambda) \mathrm{d}x \tag{27}$$

$$= a_0\triangle - \frac{a_1 \triangle^{\lambda+1}}{2^\lambda(\lambda+1)} + o(\triangle^{\lambda+1}), \tag{28}$$

$$p_i = \int_{(i-1/2)\triangle}^{(i+1/2)\triangle} a_0 - a_1 x^\lambda + o(x^\lambda) \mathrm{d}x \tag{29}$$

$$= \left[a_0 x - \frac{a_1 x^{\lambda+1}}{\lambda+1} \right]_{(i-1/2)\triangle}^{(i+1/2)\triangle} + o(\triangle^{\lambda+1}), \tag{30}$$

$$= a_0\triangle - \frac{a_1\triangle^{\lambda+1}}{\lambda+1} \left((i+1/2)^{\lambda+1} - (i-1/2)^{\lambda+1} \right) + o(\triangle^{\lambda+1}). \tag{31}$$

TABLE II
SCALING OF THE SUM $s_{(L+2)\triangle,\epsilon}$ NEAR THE SINGULARITY FOR GGD DEPENDS ONLY ON THE SHAPE PARAMETER α.

$k^\star$	$s_{(L+2)\triangle,\epsilon}$	α
1	$\triangle^{1+2\alpha}$	$\alpha \le 1/2$
	$\triangle^2$	$\alpha > 1/2$
2	$\triangle^{1+2\alpha}$	$\alpha \le 3/2$
	$\triangle^4$	$\alpha > 3/2$

Once the cover distribution p_i is obtained using these formulas, it can be substituted into (9) for a given embedding operation to compute all $2L+3$ remaining terms of the partial sum at the singularity.

IV. Scaling for specific distributions

We now combine the results from Sections III-A and III-B to obtain the scaling of the Fisher information for four embedding operations and two precover distributions for the entire range of their parameters.

First, we work out the scaling of the partial sums at the singularity for all four embedding operations. For LSBM, equations (28) and (31) can be used to compute

$$\frac{\mathrm{d}q_1}{\mathrm{d}\beta} = \frac{p_2+p_0}{2} - p_1 = -\frac{a_1\triangle^{\lambda+1}}{2^{\lambda+2}(\lambda+1)} \times (4+5^{\lambda+1}-3^{\lambda+2}) + o(\triangle^{\lambda+1}), \quad (32)$$

$$\frac{\mathrm{d}q_0}{\mathrm{d}\beta} = \frac{p_{-1}+p_1}{2} - p_0 = p_1 - p_0 = -\frac{a_1\triangle^{\lambda+1}}{2^{\lambda+1}(\lambda+1)}(3^\lambda - 1) + o(\triangle^{\lambda+1}). \quad (33)$$

Since $p_j \approx \triangle$, we obtain $\frac{1}{p_j}\left(\frac{\mathrm{d}q_j}{\mathrm{d}\beta}\right)^2 \approx \triangle^{1+2\lambda}$ for $j = -1, 0, 1$. Similarly, for LSBR, since $\frac{\mathrm{d}q_1}{\mathrm{d}\beta} = p_0 - p_1 = -\frac{\mathrm{d}q_0}{\mathrm{d}\beta}$, using the result above, $\frac{1}{p_j}\left(\frac{\mathrm{d}q_j}{\mathrm{d}\beta}\right)^2 \approx \triangle^{1+2\lambda}$ for $j = 0, 1$. For F5, $\frac{\mathrm{d}q_0}{\mathrm{d}\beta} = p_{-1} + p_1 = 2p_1$, which means that $\frac{\mathrm{d}q_0}{\mathrm{d}\beta} \approx \triangle$ and thus $\frac{1}{p_0}\left(\frac{\mathrm{d}q_0}{\mathrm{d}\beta}\right)^2 \approx \triangle^1$. Finally, for symJsteg, the bin $j = 0$ is invariant to embedding and thus the scaling for this algorithm is only influenced by scaling in the immediate neighborhood and at the points of smoothness, which is the same as for LSBR.

A. Generalized Gaussian

The analysis now splits depending on the precover distribution. Starting with the GGD, since $f_{\mathrm{GG}}(x) \approx \exp(-|x|^\alpha/b) = 1 - |x|^\alpha/b + o(|x|^\alpha)$, $\lambda = \alpha$ for the expansion at zero. Combining the results from the previous paragraph with Table II, we obtain the second column of Table III graphically rendered in Figure 1. To verify the results, the figure contains simulations obtained by evaluating p_j (1) in the sum (9) by numerical integration.

TABLE III
SCALING OF THE FISHER INFORMATION $I_\triangle(0)$ W.R.T. THE QUANTIZATION STEP $\triangle$ FOR FOUR EMBEDDING OPERATIONS FOR THE GGD AND GCD; $\alpha > 0$ IS THE SHAPE PARAMETER OF GGD. THE SCALING IS INVARIANT TO THE SHAPE PARAMETER τ OF THE GCD.

Embedding	GGD	GCD
LSBM	$\triangle^{\min\{4,1+2\alpha\}}$	$\triangle^3$
LSBR	$\triangle^{\min\{2,1+2\alpha\}}$	$\triangle^2$
F5	$\triangle$	$\triangle$
symJsteg	$\triangle^{\min\{2,1+2\alpha\}}$	$\triangle^2$

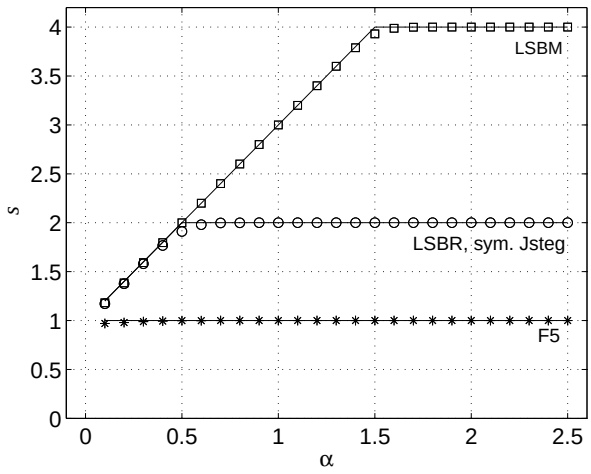

Fig. 1. Scaling exponent s versus the parameter α for the generalized Gaussian precover model and four embedding operations. Solid lines show the theoretical result; the markers are from numerical simulations.

B. Generalized Cauchy

The GCD is smooth everywhere with bounded derivatives. At zero, the derivatives do not exist but the one-sided ones do and are bounded. Thus, from Remark 4, the scaling is determined solely by the terms at the singularity. Since the expansion of (11) at zero is

$$f_{\mathrm{GC}}(x) = \frac{\tau-1}{b}\left(1 - \frac{\tau}{b}|x| + \frac{\tau(\tau+1)}{b^2}a^2x^2 + \cdots\right), \quad (34)$$

$\lambda = 1$ and is independent of the shape parameter τ. Since all bins with the exception of $-1, 0, 1$ scale as $\triangle^4$, the resulting scaling for the GCD and LSBM is $\triangle^{1+2\lambda} = \triangle^3$. For LSBR and symJsteg, the scaling at smooth points is $\triangle^2$ while the scaling at bin 0 is $\triangle^3$, giving the final scaling $\triangle^2$ for both algorithms for all $\tau > 0$. Finally, for the F5 operation, the scaling is determined by the zero bin, which scales as $\triangle$ for all τ. The scaling for the GCD is summarized in the third column of Table III.

V. Extension to noise residuals

Modern spatial-domain steganalysis algorithms represent images with co-occurrences of their noise residuals (see [29], [22], [9] and the references therein) obtained by applying a high-pass filter to the image. This extensive body of literature provides an empirical justification for modeling just the noise component of images – since

steganographic embedding changes typically manifest as a high-frequency noise, the SNR between the image and the stego signal is increased. Moreover, the residual has a narrower distribution that can be better modeled and also represented with a lower-dimensional feature vector.

In this section, we analyze the case when the embedding occurs in the pixel domain, yet the steganalysis utilizes models based on the noise-residual representation. The derivations are carried out for the simplest case of the noise residual – the difference between two adjacent pixels, which is the basis of the SPAM model [22]. For simplicity, we also limit ourselves to a one-dimensional co-occurrence, which is the histogram of pixel differences. The author hopes that extending the analysis to more complicated residuals should be apparent.

We start with a joint pdf for two neighboring precover pixels, $f(x,y)$, and define

$$F_{\triangle}(i\triangle, j\triangle) = \int_{(i-1/2)\triangle}^{(i+1/2)} \int_{(j-1/2)\triangle}^{(j-1/2)\triangle} f(x,y)\mathrm{d}x\mathrm{d}y. \tag{35}$$

Using the more compact notation, $p_{ij} \triangleq F_{\triangle}(i\triangle, j\triangle)$, we have for the histogram of pixel differences:

$$p_d = \Pr(j-i=d) = \sum_i p_{i,i+d}. \tag{36}$$

For the stego image:

$$\begin{aligned} q_{ij}(\beta) &= \sum_{u,v} p_{uv} b_{ui} b_{vj} && (37)\\ &= p_{ij}(1+c_{ii}\beta)(1+c_{jj}\beta) \\ &\quad + \sum_{v\neq j} p_{iv}(1+c_{ii}\beta)\beta c_{vj} \\ &\quad + \sum_{u\neq i} p_{uj}(1+c_{jj}\beta)\beta c_{ui} + O(\beta^2), && (38)\\ q_d(\beta) &\triangleq \sum_i q_{i,i+d}(\beta), && (39) \end{aligned}$$

which implies, after simplification:

$$\begin{aligned} \left.\frac{\mathrm{d}q_{ij}(\beta)}{\mathrm{d}\beta}\right|_{\beta=0} &= \sum_v p_{iv}c_{vj} + p_{vj}c_{vi} && (40)\\ \left.\frac{\mathrm{d}q_d(\beta)}{\mathrm{d}\beta}\right|_{\beta=0} &= \sum_i \sum_{v=i+d-L}^{i+d+L} p_{vi}c_{v,i+d} \\ &\quad + \sum_i \sum_{v=i-L}^{i+L} p_{v,i+d}c_{vi} && (41)\\ &= \sum_i \sum_{v=-L}^{L} (p_{i+d+v,i}c_{i+d+v,i+d} \\ &\quad + p_{i+v,i+d}c_{i+v,i}). && (42) \end{aligned}$$

To further simplify the matters, we adopt the following two assumptions:

C1. $p_{ij} = p_{ji}$ for all i,j (symmetry of natural images).
C2. $c_{ij} = c_{i+d,j+d}$ for all i,j and d.

Assumption C2 postulates invariance of the embedding operation. Although this invariance seems to exclude parity-based operations, such as those used in LSBR or symJsteg, the impact of such operations on pixel differences is essentially identical to that of parity-blind operations, such as the LSBM.

Assumptions C1–C2 together with (39) allow us to simplify (42):

$$\left.\frac{\mathrm{d}q_d(\beta)}{\mathrm{d}\beta}\right|_{\beta=0} = \sum_{v=-L}^{L} (p_{d+v} + p_{d-v})c_{v,0}. \tag{43}$$

The next step is to expand p_{d+z} using Taylor series at d and convert the entire analysis to the one-dimensional case of Section III. To this end, formally, we need to require f to have continuous partial derivatives up to order k. For $|d-z| \le L$,

$$p_{d+z} = \sum_{l=0}^{k-1} \frac{(z\triangle)^l}{l!} p_d^{(l)} + \frac{(z\triangle)^k}{k!} p_{\xi_{dk}}^{(k)}, \tag{44}$$

where $\xi_{dk} \in (d, d+z)$ and

$$p_d^{(l)} \triangleq \left.\frac{\mathrm{d}^{(l)} p_{d+z}}{\mathrm{d}z^l}\right|_{z=0}. \tag{45}$$

The derivatives for $l \ge 1$ are obtained from (35) and (36):

$$\begin{aligned} p_d^{(l)} &= \sum_i \int_{(i-1/2)\triangle}^{(i+1/2)\triangle} \left.\frac{\partial f^{(l-1)}}{\partial u^{l-1}}\right|_{(u,(i+d-1/2)\triangle)}^{(u,(i+d+1/2)\triangle)} \mathrm{d}u && (46)\\ &= \triangle \sum_i \int_{(i-1/2)\triangle}^{(i+1/2)\triangle} \frac{\partial f^{(l)}}{\partial u^l}(u, \phi_{idl})\mathrm{d}u, && (47) \end{aligned}$$

where $\phi_{idl} \in ((i+d-1/2)\triangle, (i+d+1/2)\triangle)$. The scaling for a given joint precover density f can now be obtained by substituting (47) into (44) and (43) and following the same steps as in Section III.

VI. Discussion

It is not possible to directly relate the results of this paper to empirical cover sources because real digital images are complicated mixtures that are not iid signals. Additionally, since modern steganalysis works with low-dimensional representations of images (features), this processing decreases the KL divergence between cover and stego features. Nevertheless, some interesting qualitative conclusions could still be reached. With finer quantization (smaller quantization step $\triangle$), the FI decreases as $\triangle^s$, where the scaling exponent s is determined jointly by the precover distribution smoothness and the embedding function. In general, s (and thus the secure payload) is larger for smoother distributions and for embedding operations that act as low-pass filters of the first-order statistic of cover samples. Operations that tend to make the singularity "sharper" (e.g., F5 or the operation from Example 1) have a lower leading order $k^\star$ and thus a lower scaling exponent s.

The singularity in the precover distribution has no effect on scaling (Theorem 2 and Section III-B1) as long as the singularity is "not sharp enough" (formally, $f^{(k^\star)}(x) \approx 1/|x|^n$ with $n < 1/2$). For "sharper" singularities, formally $n \geq 1/2$, the scaling exponent starts decreasing, which in turn increases the FI and makes the embedding scheme more detectable. This can be understood on an intuitive basis in the following manner. Since steganographic embedding changes are similar to adding noise, embedding acts as a low-pass filter on the first-order statistic (histogram) of cover elements. This smoothing will impact sharper singularities more. Theorem 2 is a more precise formulation of this intuitive statement.

For JPEG images, larger quality factors correspond to smaller $\triangle$ and thus smaller FI. At the same time, the width of the distribution b increases (c.f., (25)). Both mechanisms will allow the steganographer to embed larger payloads for a given level of security than what one would expect from the square root law only.

Theorems 1 and 2 show that the scaling exponent increases with increasing $k^\star$. It is thus desirable for the steganographer to choose embedding operations with a larger leading order $k^\star$ to decrease the FI and thus increase security. LSBM with its leading order of 2 provides better security than LSBR with order 1, a fact already known in steganalysis for other reasons. Furthermore, the presence of bins with lower leading order increases the FI and thus lowers security. From this point of view, F5 might become more secure if it allowed changing zero coefficients to non-zero to increase the leading order of bin 0. In fact, some recently proposed JPEG-domain steganographic algorithms [11], [24] that allow changes of zero coefficients indeed exhibit lower empirical detectability.

For the GG precover, equation (25) informs us that the FI increases with decreasing b or, equivalently with decreasing precover variance. This is natural as detection of steganography in covers with lower entropy should indeed be easier. Table III tells us that the precover singularity impacts GC-distributed precovers very differently than precovers with the GG distribution and this impact depends strongly on the embedding operation. The singularity decreases the scaling exponent (and thus lowers the security) for the GG model. This negative impact is larger for more "sharp" distributions, which are more likely to occur in images with smooth content. Curiously, for F5 the scaling exponent is the same (and the lowest) for both models, which is due to the properties of the embedding operation at zero.

Unquantized discrete cosine and wavelet coefficients of digital images are often modeled using the GGD as well as GCD (see, e.g., the comparison in [21] and the references therein). While both GGD and GCD are often "good enough" models for other signal processing applications, such as source coding, their secure payload scales very differently w.r.t. the quantization step. It appears that adopting a model for estimating steganographic security may require model validation that is different from traditional approaches, such as those based on the least-square or maximum likelihood fits.

VII. Practical issues

A. Scaling in practice

The original SRL is quite robust in the sense that even though it has been established only for artificial sources, such as Markov chains, it has been verified to manifest quite robustly for empirical cover sources despite the fact that practitioners build detectors using machine learning rather than as likelihood ratio tests and that pixels (transform coefficients) are quite heterogeneous and non stationary.

The scaling of the FI w.r.t. the quantization step undoubtedly manifests in practice as well. However, the specific scaling strongly depends on the precover distribution (see Table III). Furthermore, relating the results of experiments on real images with the theoretical results of this paper derived for i.i.d. sources is likely not possible for reasons listed in the previous section.

It is true, however, and our analysis of artificial sources confirms this, that quantization must have an extremely strong impact on statistical detectability because the Fisher information appears in the error exponent that controls detector errors through the Chernoff–Stein Lemma as discussed in Section II-C. To assess this effect, the following experiment was carried out on a database of grayscale images represented with a varying bit depth.

A total of 5,000 images were taken with Canon EOS 550D equipped with an 18 megapixel sensor. The images were stored in the CR2 format with 12-bit per pixel and then converted from CR2 to the 48-bit TIFF color format using dcraw with default parameters. Subsequently, the images were converted to grayscale using the `rgb2gray` command in Matlab and stored as 16-bit grayscale. To speed up the experiments, all images were further downsampled by a factor of four by selecting every fourth pixel to avoid introducing resampling artifacts.

The steganographic technique that was tested was LSBM simulated at a given change rate β (changes per pixel). Steganalysis was performed using the second-order SPAM feature vector [22] with threshold $T = 3$ (dimension 686). The classifier was the ensemble [20] run with the automatic choice of the subspace dimensionality and the number of base learners.[9] The database was randomly split into two halves, one used for training and the other for testing. This was repeated ten times and the minimal total detection error under equal priors, $P_{\rm E} = (P_{\rm MD} + P_{\rm FA})/2$, was averaged over the ten database splits (denoted $\overline{P}_{\rm E}$).

The experimental results are shown in Figure 2. The curves were interpolated using cubic splines. Only the bit depths at $B \leq 8$ are shown as the detectability for all bit depths larger than 8 was very close to random guessing. The figure confirms that statistical detectability strongly depends on the quantization. In fact, in this particular source statistical undetectability is reached at any payload

[9]The ensemble code is available from http://dde.binghamton.edu/download/ensemble/

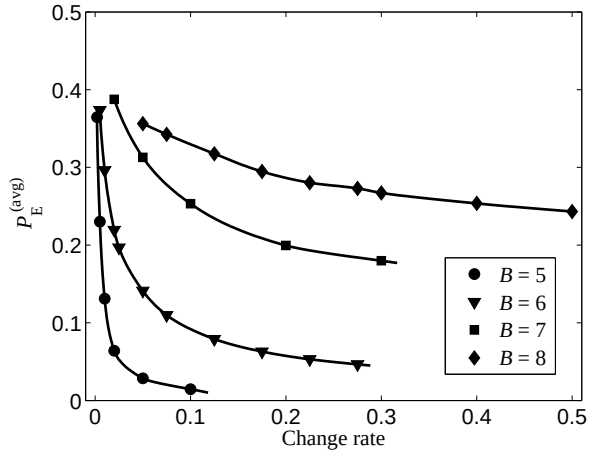

Fig. 2. The average detection error, $\overline{P}_{\rm E}$, as a function of change rate β for LSBM and cover grayscale images represented using B bits.

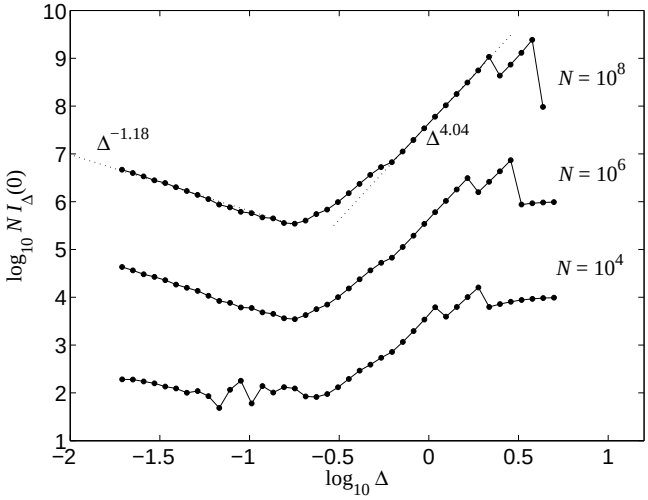

Fig. 3. Fisher information for LSBM vs. the quantization step $\triangle$ for $N = 10^8, 10^6, 10^4$ samples from a GGD with $\alpha = 1.5$, $b = 1$, and $\mu = 0$. Note the region of over-sampling for small $\triangle$, where $I_\triangle(0) \propto \triangle^{-1}$, and under-sampling for large $\triangle$. The range of $\triangle$ where $s \doteq \max\{4, 1+2\alpha\} = 4$, as predicted by the theory, becomes smaller with decreasing N. The dotted lines are linear fits in their corresponding ranges of $\triangle$.

(!) for images with 9 or more bits per pixel.[10] The secure change rate at detectability fixed to $\overline{P}_{\rm E} = 0.25$ is $\beta = 0.0044$, 0.0152, 0.1038, and 0.4322 changes per pixel for bit depths 5, 6, 7, and 8, respectively. This approximately corresponds to scaling of the FI by $\triangle^2$ (the quantization step $\triangle$ decreases by a factor of 2 with each bit level added). Even though it is tempting to say that this indicates a GG model with $\alpha = 1/2$, one should stay away from such a claim due to the reasons outlined above (and the effect of finite samples discussed below).

B. Effect of over/undersampling

In practice, working with a finite number of samples may prevent us to observe the correct scaling even when observations do follow the model. Oversampling (too small $\triangle$) will produce a noisy estimate of the cover distribution, p_j, which will completely change the scaling. On the other hand, $\triangle$ that is too large will not yet exhibit the limiting behavior when $\triangle \to 0$.

Next, we provide a crude qualitative analysis of the impact of over-sampling. Given a cover object with N elements from a finite range $\mathcal{R}$, the number of samples in the jth bin, P_j, is a random variable whose binomial distribution will be approximated with a Gaussian:

$$P_j \sim \mathrm{N}\left(p_j, p_j(1-p_j)/N\right). \tag{48}$$

If P_j were independent, the expected value of

$$\left(\sum_i c_{ij} P_i\right)^2 = \left(\sum_i c_{ij} p_i + \sum_i c_{ij}(P_i - p_i)\right)^2, \tag{49}$$

would be

$$\mu_j^2 = \left(\sum_i c_{ij} p_i\right)^2 + \sum_i c_{ij}^2 p_i(1-p_i)/N. \tag{50}$$

Since in practice, there will be finitely many bins, $j\triangle \in \mathcal{R} \subset [-R, R]$, where R is the dynamic range of X,

$$E\left[\sum_{j\in\mathcal{R}} \frac{\left(\sum_i c_{ij} P_i\right)^2}{P_j}\right] \doteq \sum_{j\in\mathcal{R}} \frac{\mu_j^2}{p_j} \tag{51}$$

$$= I_\triangle(0) + \frac{1}{N}\sum_{j\in\mathcal{R}} \frac{\sum_i c_{ij}^2 p_i(1-p_i)}{p_j} \tag{52}$$

$$\propto \frac{1}{N\triangle}. \tag{53}$$

because $I_\triangle(0) \propto \triangle^s$ with $s \geq 0$ and $|\mathcal{R}|$ is inversely proportional to the quantization step $\triangle$.

To assess the effects of over/undersampling in practice, we generated $N = 10^4 - 10^8$ i.i.d. samples from a GGD with $\alpha = 1.5$ and $b = 1$ and computed $I_\triangle(0)$ for a range of $\triangle$ with LSBM as the stego algorithm. The result shown in Figure 3 confirms the crude analysis of over-sampling with the scaling exponent $s \doteq -1$ for small $\triangle$. The scaling matches the theoretical result, $s \doteq 1 + 2\alpha = 4$, for midrange values and breaks up when $\triangle$ becomes too large. The range of proper scaling gets smaller with decreasing N.

VIII. Conclusion

The square root law of imperfect steganography connects statistical detectability with the cover size and the change rate in the asymptotic limit of large covers and

[10] We note that the detectability for this particular source was markedly worse than what has been reported elsewhere for images of approximately the same size or smaller [10]. We attribute this difference to the subsampling, which in our case did not introduce any artifacts that could aid the steganalysis.

small change rates. The current paper extends this law to the case when the cover object is obtained by quantizing a precover that follows a continuous-valued distribution. In this case, constant statistical detectability is obtained when $N\beta^2\triangle^s = const.$, where N is the cover size, β the change rate, and s the scaling exponent w.r.t. the quantization step $\triangle$ that tends to zero. The scaling exponent is determined jointly by the embedding operation and the precover distribution. In general, s is larger for smoother distributions and for embedding operations that act as low-pass filters of the first-order statistic of cover samples. Numerous qualitative consequences and other implications of the scaling for practitioners are discussed in Section VI.

The scaling has been worked out in detail for two precover distributions commonly used in signal modeling – the generalized Gaussian and Cauchy distributions. While both distributions apparently lead to comparable results, for example, in source coding [21], the scaling of secure payload in steganography depends rather sensitively on the precover model. The singularity of the generalized Gaussian is "sharper," which has the effect of decreasing the scaling exponent and thus the secure payload length. This decrease is larger for smoother images with a smaller shape parameter. Thus, the scaling w.r.t. the quantization step may be a more relevant criterion for model fitting for steganography instead of the more traditional fitting approaches.

This work is relevant for understanding the effect of color bit depth on security of schemes that embed in the spatial domain and the effect of the quality factor on security for stegosystems that embed data in the JPEG domain. Experiments on real images confirm the strong effect of quantization on statistical detectability due to the fact that the factor $\triangle^s$ multiplies the error exponent. However, unlike the original square root law, which is quite robust, the scaling w.r.t. $\triangle$ strongly depends on the precover distribution, which makes its interpretation in practice rather hard due to the necessity to find a sufficiently accurate model. When the number of cover samples, N, is small, the theoretical scaling is observable only in a rather small midrange of quantization steps due to negative effect of under- and over-sampling.

The value of this work is primarily in shedding light on the fundamental connection between statistical detectability and the complex interplay between the precover distribution and the embedding after quantizing the precover. Although the main result has been derived for artificial (i.i.d.) covers, an extension has been presented to the case when the cover is modeled as an i.i.d. sequence of pixel groups (pairs of pixels). This is relevant because many modern steganalytic methods represent images with histograms of groups of neighboring pixels/DCT coefficients or noise residuals [9], [19].

Appendix A
Proof of Theorem 1

We write the FI (13) as a sum of three partial sums, $I_\triangle(0) = s_{-\infty,-M} + s_{-M,M} + s_{M,\infty}$, and study their behavior when $\triangle \to 0$. Since $f(x)$ is even, it is sufficient to consider only one unbounded interval and the closed interval.

A.1 Unbounded interval $[M, \infty)$

Using $k = k^\star$ in (19),

$$\left(\sum_i c_{ij} F_\triangle(i\triangle)\right)^2 \le \frac{\triangle^{2k^\star+2}}{k^\star!^2} \tilde{w}^2_{jk^\star} \max_{x\in\mathcal{I}_j} (f^{(k^\star)}(x))^2, \quad (54)$$

where $\tilde{w}_{jk^\star} = \sum_i c_{ij}|i-j|^{k^\star} \le C(2L+1)L^{k^\star} \triangleq C'$. The continuity of $f^{(k^\star)}(x)$ implies that the maximum is reached at some $\theta_j \in \mathcal{I}_j$. Since $F_\triangle(j\triangle) = \triangle f(u_j)$,

$$s_{M,\infty}(\triangle) \le \frac{\triangle^{2k^\star}}{k^\star!^2} C'^2 \underbrace{\sum_{j\triangle\ge M} \frac{\left(f^{(k^\star)}(\theta_j)\right)^2}{f(u_j)}\triangle}_{r_M}. \quad (55)$$

For $j\triangle - (L+1)\triangle \ge M$, from the monotonicity of $|f^{(k)}(x)|$ at ∞ and the range of θ_j and u_j,

$$\frac{\left(f^{(k^\star)}(\theta_j)\right)^2}{f(u_j)} \le \frac{\left(f^{(k^\star)}(j\triangle - (L+1)\triangle)\right)^2}{f(j\triangle + \triangle/2)}. \quad (56)$$

Thus, for $(L+3/2)\triangle \le \delta_0$, r_M can be bounded with a Riemann sum:

$$\begin{aligned} r_M &\le \sum_{j\triangle \ge M-(L+1)\triangle} \frac{\left(f^{(k^\star)}(j\triangle)\right)^2}{f(j\triangle + (L+3/2)\triangle)}\triangle \\ &\to \int_M^\infty \frac{(f^{(k^\star)}(x))^2}{f(x+\delta)}\mathrm{d}x \end{aligned} \quad (57)$$

as $\triangle \to 0$. Assumption R2 finally implies that $r_M \to 0$ with $M \to \infty$.

A.2 Closed interval $[-M, M]$

The continuity of derivatives up to order $k^\star+1$ guarantees their boundedness on $[-M, M]$. Thus, using $k = k^\star+1$ in (19):

$$\begin{aligned} \left(\sum_i c_{ij} F_\triangle(i\triangle)\right)^2 &= \frac{(f^{(k^\star)}(\phi_{jk^\star}))^2}{k^\star!^2}\triangle^{2k^\star+2} w^2_{k^\star} \\ &\quad + O(\triangle^{2k^\star+3}). \end{aligned} \quad (58)$$

The continuity of f implies the existence of $f_0 > 0$ such that $f(x) \ge f_0 > 0$ on $[-M, M]$. Furthermore, $f(\phi_{jk^\star}) - D_1\triangle \le f(u_j) \le f(\phi_{jk^\star}) + D_1\triangle$, where $D_1 > 0$ bounds the first derivative on $[-M, M]$. Therefore,

$$\frac{1}{\triangle f(u_j)} = \frac{\kappa(\triangle)}{\triangle f(\phi_{jk^\star})}, \quad (59)$$

where, for $\triangle < f_0/D_1$,

$$\frac{1}{1 + D_1\triangle/f_0} \le \kappa(\triangle) \le \frac{1}{1 - D_1\triangle/f_0}. \quad (60)$$

Riemann integrability of $(f^{(k^\star)}(x))^2/f(x)$ on $[-M,M]$ together with (58), (59), and (60) imply:

$$\frac{s_{-M,M}}{\triangle^{2k^\star}} = \frac{w_{k^\star}^2\kappa(\triangle)}{k^\star!^2}\left[\sum_{-M\le j\triangle\le M}\left(\frac{(f^{(k^\star)}(\phi_{jk^\star}))^2}{f(\phi_{jk^\star})}\triangle + O(\triangle^2)\right)\right] \to \frac{w_{k^\star}^2}{k^\star!^2}\int_{-M}^{M}\frac{(f^{(k^\star)}(x))^2}{f(x)}\mathrm{d}x, \quad (61)$$

as $\triangle\to 0$. Combining (57) and (61) proves the theorem. **Q.E.D.**

Appendix B
Proof of Theorem 2

We first carry out the proof for $k^\star$ even and then deal with the case of $k^\star$ odd.

B.1 Even $k^\star = 2$

We will use the fact that for f continuous and γ_i all of the same sign,

$$\sum_i \gamma_i f(x_i) = f(\bar{x})\sum_i \gamma_i,\ \bar{x}\in[\min_i x_i, \max_i x_i]. \quad (62)$$

Using (19) with $k=k^\star$, by (62):

$$\sum_i c_{ij}F_\triangle(i\triangle) = \frac{\triangle^{k^\star+1}}{k^\star!}\sum_i c_{ij}(i-j)^{k^\star}f^{(k^\star)}(\tilde{\phi}_{ij}) \quad (63)$$

$$= \frac{\triangle^{k^\star+1}}{k^\star!}w_{k^\star}f^{(k^\star)}(\phi_j), \quad (64)$$

$\tilde{\phi}_{ij}\in(j\triangle-\triangle/2, i\triangle+\triangle/2)$, when $i>j$, and $\tilde{\phi}_{ij}\in(i\triangle-\triangle/2, j\triangle+\triangle/2)$, when $j>i$,

$$\phi_j\in\mathcal{I}_j\subset\left[\frac{3\triangle}{2}, \epsilon+\left(L+\frac{1}{2}\right)\triangle\right]\triangleq\mathcal{I}_{L,\epsilon}. \quad (65)$$

Note that (64) is valid as long as $c_{ij}(i-j)^{k^\star}\ge 0$ for all i, which holds for $k^\star$ even. Combining (16) and (64) with Assumption S1 leads to:

$$\frac{s_{(L+2)\triangle,\epsilon}k^\star!^2}{w_{k^\star}^2\triangle^{2k^\star+1}} = \sum_{(L+2)\triangle\le j\triangle\le\epsilon}\frac{\left(f^{(k^\star)}(\phi_j)\right)^2}{f(u_j)} \quad (66)$$

$$= \sum_{(L+2)\triangle\le j\triangle\le\epsilon}\frac{g^2(\phi_j){\phi_j}^{-2n}}{f(u_j)}. \quad (67)$$

Since $\phi_j\in\mathcal{I}_{L,\epsilon}$ (65) for all j,

$$\frac{\min_{x\in\mathcal{I}_{L,\epsilon}}g^2(x)}{\max_{x\in\mathcal{I}_{L,\epsilon}}f(x)}\le\frac{g^2(\phi_j)}{f(u_j)}\le\frac{\max_{x\in\mathcal{I}_{L,\epsilon}}g^2(x)}{\min_{x\in\mathcal{I}_{L,\epsilon}}f(x)}. \quad (68)$$

The continuity of both f and g at zero, together with the fact that $g(0)\neq 0$ and $f(0)>0$, imply that for $\epsilon>0$ sufficiently small the scaling of (67) when $\triangle\to 0$ (in the sense of $\Theta(x)$ symbolics) depends solely on $\sum\phi_j^{-2n}$. Since $\phi_j\in\mathcal{I}_j$ (22), this sum can be bounded from below and above by

$$\sum_{j=L+2}^{\epsilon/\triangle}\left(\frac{1}{(j+L+1/2)\triangle}\right)^{2n}\ge\triangle^{-2n}\int_{2L+5/2}^{\epsilon/\triangle+L+1/2}x^{-2n}\mathrm{d}x$$

$$= \frac{\triangle^{-2n}}{-2n+1}\left[\left(\frac{\epsilon}{\triangle}+L+\frac{1}{2}\right)^{-2n+1}-\left(2L+\frac{5}{2}\right)^{-2n+1}\right] \quad (69)$$

and

$$\sum_{j=L+2}^{\epsilon/\triangle}\left(\frac{1}{(j-L-1/2)\triangle}\right)^{2n}\le\triangle^{-2n}\int_{1/2}^{\epsilon/\triangle-L-3/2}x^{-2n}\mathrm{d}x$$

$$= \frac{\triangle^{-2n}}{-2n+1}\left[\left(\frac{\epsilon}{\triangle}-L-3/2\right)^{-2n+1}-\left(\frac{1}{2}\right)^{-2n+1}\right], \quad (70)$$

respectively. Both bounds scale as $\triangle^{-2n}$ when $-2n+1\le 0$ and as $\triangle^{-1}$ when $-2n+1>0$.

In summary:

$$s_{(L+2)\triangle,\epsilon} = \begin{cases}\Theta(\triangle^{2k^\star+1-2n}) & \text{when } n\ge 1/2\\ \Theta(\triangle^{2k^\star}) & \text{when } n<1/2.\end{cases} \quad (71)$$

B.2 Odd $k^\star = 1$

Here, we first obtain an upper bound using (62)

$$\left|\sum_i c_{ij}F_\triangle(i\triangle)\right| \le \frac{\triangle^{k^\star+1}}{k^\star!}\sum_i c_{ij}|i-j|^{k^\star}f^{(k^\star)}(\tilde{\phi}_{ij}^{(1)}) \quad (72)$$

$$\le \frac{\triangle^{k^\star+1}}{k^\star!}C(2L+1)L^{k^\star}f^{(k^\star)}(\phi_j^{(1)}), \quad (73)$$

where $\phi_j^{(1)}\in\mathcal{I}_j$, leading to

$$\frac{s_{(L+2)\triangle,\epsilon}k^\star!^2}{\triangle^{2k^\star+1}C^2(2L+1)^2L^{2k^\star}}\le\sum_{(L+2)\triangle\le j\triangle\le\epsilon}\frac{\left(f^{(k^\star)}(\tilde{\phi}_j)\right)^2}{f(u_j)}. \quad (74)$$

Getting the lower bound is a little more involved. We first introduce

$$\gamma_j^- = \frac{\triangle^{k^\star+1}}{k^\star!}\sum_{i>j}c_{ij}(i-j)^{k^\star}\ge 0, \quad (75)$$

$$\gamma_j^+ = \frac{\triangle^{k^\star+1}}{k^\star!}\sum_{i<j}c_{ij}(j-i)^{k^\star}\ge 0, \quad (76)$$

and split (63) into a sum over $i>j$ and over $i<j$. Both sums can be simplified using (62) to:

$$\left|\sum_i c_{ij}F_\triangle(i\triangle)\right| = \left|\gamma_j^+f^{(k^\star)}(\phi_j^+)-\gamma_j^-f^{(k^\star)}(\phi_j^-)\right|, \quad (77)$$

for some $\phi_j^+,\phi_j^-\in\mathcal{I}_j$. If either $\gamma_j^-=0$ or $\gamma_j^+=0$, we could continue exactly the same way as for $k^\star$ even. Thus, we will assume that $\gamma_j^+>0$ and $\gamma_j^->0$. Note that both γ_j^-,γ_j^+ are bounded independently of j:

$$\gamma_j^\pm\le\frac{\triangle^{k^\star+1}}{k^\star!}CL^{k^\star+1}. \quad (78)$$

Since $|\gamma_j^- - \gamma_j^+| = |w_{k^\star}|\frac{\triangle^{k^\star+1}}{k^\star!}>0$, we have $\gamma_j^-\neq\gamma_j^+$. Assuming, for example, that $\gamma_j^-<\gamma_j^+$ (the other case is

analogical):

$$\frac{\gamma_j^-}{\gamma_j^+} = \frac{\gamma_j^-}{\gamma_j^- + |\gamma_j^- - \gamma_j^+|} \tag{79}$$

$$= \frac{1}{1 + \frac{|w_{k^\star}|}{\gamma_j^-}\frac{\triangle^{k^\star+1}}{k^\star!}} \tag{80}$$

$$\leq \frac{1}{1 + \frac{|w_{k^\star}|}{CL^{k^\star+1}}} \triangleq \rho < 1, \tag{81}$$

and

$$\gamma_j^+ \geq \gamma_j^+ - \gamma_j^- = |w_{k^\star}|\frac{\triangle^{k^\star+1}}{k^\star!} > 0. \tag{82}$$

Since $\phi_j^-, \phi_j^+ \in \mathcal{I}_{L,\epsilon}$ and $g(x)$ is continuous at zero with $g(0) \neq 0$, we can choose $\epsilon > 0$ such that $r_j = g(\phi_j^-)/g(\phi_j^+)$ is arbitrarily close to 1 whenever $j\triangle < \epsilon$. In particular, we can choose it such that

$$\left(\frac{1+\rho}{2\rho r_j}\right)^{1/n} \geq \eta > 1, \tag{83}$$

for some η and all j with $j\triangle < \epsilon$.

If j additionally satisfies

$$j \geq \frac{\eta+1}{\eta-1}(L+1/2) \triangleq L', \tag{84}$$

which is equivalent with $\eta \geq \frac{j+L+1/2}{j-L-1/2}$, we obtain using (83):

$$1 - \rho r_j \left|\frac{j+L+1/2}{j-L-1/2}\right|^n \geq 1 - \rho r_j \eta^n \tag{85}$$

$$\geq 1 - \rho\frac{1+\rho}{2\rho r_j} r_j \tag{86}$$

$$= \frac{1-\rho}{2}. \tag{87}$$

Using (77),(81), and (87), we can write whenever $j\triangle \in (jL', \epsilon)$:

$$\frac{\left|\sum_i c_{ij} F_\triangle(i\triangle)\right|}{|\gamma_j^+ f^{(k^\star)}(\phi_j^+)|} \geq 1 - \rho\frac{f^{(k^\star)}(\phi_j^-)}{f^{(k^\star)}(\phi_j^+)} \tag{88}$$

$$= 1 - \rho\frac{|\phi_j^+|^n}{|\phi_j^-|^n}\frac{g(\phi_j^-)}{g(\phi_j^+)} \tag{89}$$

$$\geq 1 - \rho r_j \left|\frac{j+L+1/2}{j-L-1/2}\right|^n \tag{90}$$

$$= \frac{1-\rho}{2}, \tag{91}$$

which gives us a lower bound:

$$s_{(L+2)\triangle,\epsilon} = \sum_{(L+2)\triangle \leq j\triangle \leq \epsilon} \frac{\left|\sum_i c_{ij} F_\triangle(i\triangle)\right|^2}{\triangle f(u_j)} \tag{92}$$

$$\geq \sum_{L'\triangle \leq j\triangle \leq \epsilon} \frac{\left(\gamma_j^+ \frac{1-\rho}{2} f^{(k^\star)}(\phi_j^+)\right)^2}{\triangle f(u_j)} \tag{93}$$

$$\geq |w_{k^\star}|^2 \frac{\triangle^{2k^\star+1}}{k^\star!^2}\frac{(1-\rho)^2}{4} \times \sum_{L'\triangle \leq j\triangle \leq \epsilon} \frac{\left(f^{(k^\star)}(\phi_j^+)\right)^2}{f(u_j)}. \tag{94}$$

Combining (74) and (94), the scaling is established using the same arguments as for the case $k^\star$ even. **Q.E.D.**

Appendix C
Form of $f(x)$ at singularity

Here, we show that Assumptions S1–S2 guarantee the following form for the density f: $f(x) = a_0 - a_1|x|^\lambda + o(|x|^\lambda)$ at $x = 0$, for $\lambda > 0$, $a_0 > 0$, and $|a_1| > 0$. The arguments are carried out for $k^\star = 1$ as the extension to the case of $k^\star = 2$ can be obtained simply by applying one more integration.

We remind that for $k^\star = 1$, from S2, $f'(x) = g(x)/|x|^n$ on some neighborhood of zero. Let $\delta_0 > 0$ be such that, WLOG, $g(x) > 0$ on $[0, \delta_0]$. Furthermore, let $\underline{g}(z) \triangleq \min_{t\in[0,z]} g(t)$, $\overline{g}(z) \triangleq \max_{t\in[0,z]} g(t)$, and $L(u,v) \triangleq \int_u^v f'(t)\mathrm{d}t = f(v) - f(u)$. Since for $0 < x < \delta_0$

$$\underline{g}(\delta_0)\frac{\delta_0^{1-n} - x^{1-n}}{1-n} \leq L(x,\delta_0) \leq \overline{g}(\delta_0)\frac{\delta_0^{1-n} - x^{1-n}}{1-n}, \tag{95}$$

if $1 - n < 0$, $f(x)$ would not be continuous at zero. Thus, $1 - n \geq 0$ and we have

$$\underline{g}(x)\frac{x^{1-n}}{1-n} \leq L(0,x) \leq \overline{g}(x)\frac{x^{1-n}}{1-n}, \tag{96}$$

which implies

$$\lim_{x\to 0^+} \frac{f(x) - f(0)}{x^{1-n}} = \frac{g(0)}{1-n}. \tag{97}$$

With a similar argument for the left neighborhood of zero, we can write

$$f(x) = f(0) + \frac{g(0)}{1-n}|x|^{1-n} + o(|x|^{1-n}). \tag{98}$$

References

[1] 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.

[2] R. Böhme. *Advanced Statistical Steganalysis*. Springer-Verlag, Berlin Heidelberg, 2010.

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

[4] C. Cachin. An information-theoretic model for steganography. In D. Aucsmith, editor, *Information Hiding, 2nd International Workshop*, volume 1525 of Lecture Notes in Computer Science, pages 306–318, Portland, OR, April 14–17, 1998. Springer-Verlag, New York.

[5] T. M. Cover and J. A. Thomas. *Elements of Information Theory*. New York: John Wiley & Sons, Inc., 1991.

[6] T. Filler and J. Fridrich. Fisher information determines capacity of ϵ-secure steganography. In S. Katzenbeisser and A.-R. Sadeghi, editors, *Information Hiding, 11th International Conference*, volume 5806 of Lecture Notes in Computer Science, pages 31–47, Darmstadt, Germany, June 7–10, 2009. Springer-Verlag, New York.

[7] T. Filler, A. D. Ker, and J. Fridrich. The Square Root Law of steganographic capacity for Markov covers. In N. D. Memon, E. J. Delp, P. W. Wong, and J. Dittmann, editors, *Proceedings SPIE, Electronic Imaging, Security and Forensics of Multimedia XI*, volume 7254, pages 08 1–08 11, San Jose, CA, January 18–21, 2009.

[8] J. Fridrich. *Steganography in Digital Media: Principles, Algorithms, and Applications*. Cambridge University Press, 2009.

[9] 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.

[10] V. Holub and J. Fridrich. Optimizing pixel predictors for steganalysis. In A. Alattar, N. D. Memon, and E. J. Delp, editors, *Proceedings SPIE, Electronic Imaging, Media Watermarking, Security, and Forensics of Multimedia XIV*, volume 8303, pages 09–1–09–13, San Francisco, CA, January 23–26, 2012.

[11] F. Huang, Y. Q. Shi, and J. Huang. New JPEG steganographic scheme with high security performance. In H.-J. Kim, Y. Q. Shi, and M. Barni, editors, *Proc. 9th International Workshop on Digital Watermarking*, volume 6526 of Lecture Notes in Computer Science, pages 189–201, Seoul, Korea, October 1–3, 2010. Springer-Verlag, New York.

[12] A. D. Ker. A capacity result for batch steganography. *IEEE Signal Processing Letters*, 14(8):525–528, 2007.

[13] 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.

[14] A. D. Ker. The ultimate steganalysis benchmark? In J. Dittmann and J. Fridrich, editors, *Proceedings of the 9th ACM Multimedia & Security Workshop*, pages 141–148, Dallas, TX, September 20–21, 2007.

[15] A. D. Ker. Estimating steganographic Fisher information in real images. In S. Katzenbeisser and A.-R. Sadeghi, editors, *Information Hiding, 11th International Conference*, volume 5806 of Lecture Notes in Computer Science, pages 73–88, Darmstadt, Germany, June 7–10, 2009. Springer-Verlag, New York.

[16] A. D. Ker. The square root law in stegosystems with imperfect information. In R. Böhme and R. Safavi-Naini, editors, *Information Hiding, 12th International Conference*, volume 6387 of Lecture Notes in Computer Science, pages 145–160, Calgary, Canada, June 28–30, 2010. Springer-Verlag, New York.

[17] A. D. Ker, T. Pevný, J. Kodovský, and J. Fridrich. The Square Root Law of steganographic capacity. In A. D. Ker, J. Dittmann, and J. Fridrich, editors, *Proceedings of the 10th ACM Multimedia & Security Workshop*, pages 107–116, Oxford, UK, September 22–23, 2008.

[18] J. Kodovský and J. Fridrich. Quantitative structural steganalysis of Jsteg. *IEEE Transactions on Information Forensics and Security*, 5(4):681–693, December 2010.

[19] J. Kodovský and J. Fridrich. Steganalysis of JPEG images using rich models. In A. Alattar, N. D. Memon, and E. J. Delp, editors, *Proceedings SPIE, Electronic Imaging, Media Watermarking, Security, and Forensics of Multimedia XIV*, volume 8303, pages 0A–1–0A–13, San Francisco, CA, January 23–26, 2012.

[20] 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.

[21] V. K. Nath, , and D. Hazarika. Comparison of generalized Gaussian and Cauchy distributions in modeling of dyadic rearranged 2D DCT coefficients. In *3rd National Conference on Emerging Trends and Applications in Computer Science (NCETACS)*, pages 89–92, March 2012.

[22] T. Pevný, P. Bas, and J. Fridrich. Steganalysis by subtractive pixel adjacency matrix. *IEEE Transactions on Information Forensics and Security*, 5(2):215–224, June 2010.

[23] D. Upham. Steganographic algorithm JSteg. Software available at http://zooid.org/~paul/crypto/jsteg.

[24] 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.

[25] Y. Wang and P. Moulin. Steganalysis of block-structured stegotext. In E. J. Delp and P. W. Wong, editors, *Proceedings SPIE, Electronic Imaging, Security, Steganography, and Watermarking of Multimedia Contents VI*, volume 5306, pages 477–488, San Jose, CA, January 19–22, 2004.

[26] Y. Wang and P. Moulin. Perfectly secure steganography: Capacity, error exponents, and code constructions. *IEEE Transactions on Information Theory, Special Issue on Security*, 55(6):2706–2722, June 2008.

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

[28] 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.

[29] D. Zou, Y. Q. Shi, W. Su, and G. Xuan. Steganalysis based on Markov model of thresholded prediction-error image. In *Proceedings IEEE, International Conference on Multimedia and Expo*, pages 1365–1368, Toronto, Canada, July 9–12, 2006.

Jessica Fridrich holds the position of Professor of Electrical and Computer Engineering at Binghamton University (SUNY). She has received her PhD in Systems Science from Binghamton University in 1995 and MS in Applied Mathematics from Czech Technical University in Prague in 1987. Her main interests are in steganography, steganalysis, digital watermarking, and digital image forensic. Dr. Fridrich's research work has been generously supported by the US Air Force and AFOSR. Since 1995, she received 19 research grants totaling over $9.4mil for projects on data embedding and steganalysis that lead to more than 140 papers and 7 US patents. Dr. Fridrich is a member of IEEE and ACM.