Payload Scaling for Adaptive Steganography: An Empirical Study

Quentin Giboulot and Jessica Fridrich, *Fellow, IEEE*

***Abstract*—Payload-scaling laws of imperfect steganography inform the steganographer about how the size of secret payload should grow with cover size for constant statistical detectability. In this paper, we carry out an empirical study for the case when the steganographer and the steganalyst operate at a game-theoretic equilibrium. We first explore the possibility to leverage a generalization of the square root law to content-adaptive steganography due to Ker. Since this result does not appear to be tight enough for realistic cover sizes, we instead work with a detectability limited sender in image sources with a forced model as well as real images in both spatial and JPEG domain. The scaling is observed in practice when the images are carefully cropped to preserve the distribution of costs across scales.**

***Index Terms*—Steganography, steganalysis, square root law, payload scaling, adaptive embedding**

I. Introduction

Payload-scaling laws of imperfect steganography [7], [12], [4], [8], [9], [10], example of which is the square root law [9], relate the statistical detectability of embedding changes with the size of the embedded payload as the number of cover elements, n, tends to infinity. In particular, they often specify a *critical payload size*, sublinear in n, at which constant statistical detectability is observed. These theoretical results assume that the steganography is *imperfect* – it fails to preserve the statistical properties of the cover source – since the secure payload of perfect steganography [2] increases linearly with n [16].

Scaling laws are typically derived for a specific cover model and embedding operation. Early forms of the law [7], [4], [8] assumed independent and identically distributed (i.i.d.) cover elements and 1st-order Markov chains. Recent extensions include models with dependent cover elements – Markov random fields, kth-order Markov chains, Ising models, and hidden-layer models [9].

As for the embedding process, it has traditionally been assumed that the same stochastic operation is applied independently to each cover element, which inherently limits such results to non-adaptive steganographic schemes that do not use source coding. It has been conjectured in [9] that source coding makes the critical payload size asymptotically proportional to $\sqrt{n}\log n$ instead of $\sqrt{n}$. This conjecture was formulated in an adversarial setting when the detector has the knowledge of the embedding change probabilities π_i at each cover element i. Considering steganography as a zero-sum game between the steganographer and the steganalyst, at equilibrium the sender should select π_i that minimize the statistical detectability, which is asymptotically directly linked to the so-called deflection coefficient

$$\delta^2 \propto \sum_{i=1}^{n} \pi_i^2 c_i, \tag{1}$$

where c_i is the cost of applying the embedding operation at cover element i [11] (the steganographic Fisher information). A payload scaling law in this setting of *equilibrium embedding* and source coding was derived in [10].

The square root law was for the first time experimentally validated in 2008 [12]. It was shown to hold robustly for both the spatial and JPEG domain with the steganographic schemes of the time, which were not content adaptive and the source coding (for nsF5) was only simulated. Continuing the spirit of these experiments, the goal of this paper is to investigate whether payload-scaling laws for modern content-adaptive steganography are observed in practice when the steganographer is not guaranteed to embed at equilibrium due to adoption of idealized models, images at different scales are prepared with heuristic rules, and when the steganalyst uses empirical detectors.

II. Payload scaling for content-adaptive steganography

Given a cover with n elements and costs c_i, a Payload Limited Sender (PLS) minimizes (1) under the payload constraint $M = \sum_{i=1}^{n} \mathrm{H}(\pi_i)$, H being the binary entropy function. This minimum is achieved when $\lambda c_i = \frac{\mathrm{H}'(\pi_i)}{\pi_i}$ for all i, where λ is a Lagrange multiplier determined from the payload constraint. The solution to the PLS has no closed form and must be either approximated or computed numerically. This means that there is no simple relationship between δ and M.[1] To study the critical payload size in practice, we need to fix δ for different values of n. This leaves two possible approaches :

1) Use an asymptotic relationship between δ and M [10].
2) Use a Detectability Limited Sender (DLS)[2] to fix the detectability (1) across different scales (values of n).

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

Quentin Giboulot is with the Laboratory for System Modelling and Dependability, ICD, UMR 6281 CNRS, Troyes University of Technology, Troyes, France. This work has been done while he was a visiting scholar at Binghamton University. Email: quentin.giboulot@utt.fr.

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

[1] This is due to the use of optimal coding in the PLS; when no coding is used, $M = 2\sum_i \pi_i$, and it is routine to show that $M = 2\delta\sqrt{n}$.

[2] A DLS is the dual problem to the PLS in which the sender maximizes M given a bound on δ.

The advantage of the second approach is that it is non-asymptotic but it prevents us from working with "cost-based" steganography that minimizes a heuristically defined distortion $D = \sum_{i=1}^{n} \pi_i c_i$ instead of deflection. The first approach gives flexibility in this regard at the cost of having to rely on asymptotic approximations, which may not be sufficiently accurate for realistic image sizes as will become apparent in the next section.

In [10], the author derived the following asymptotic formula for the deflection as $n \to \infty$:

$$\delta^2_{asymp} \approx \frac{r^2}{\frac{1}{n}\sum_{i=1}^{n} c_i} \tag{2}$$

as long as the payload size $M = r\sqrt{n}\log_2(\sqrt{n})$, where r is the equivalent of the concept of relative payload, the root rate. This result was formally established under the assumption that the costs are ergodic and satisfy the "no free bits" and "no deterministic pixels" conditions. The reader is referred to [10] for more details.

In practice, we need to know how fast the approximation (2) converges to the true deflection (1) as $n \to \infty$. Furthermore, the "no free bits" and "no determinism" conditions, while not very restrictive in practice since free bits are somewhat rare in natural images and the number of deterministic pixels can be controlled (e.g., by choosing an appropriate crop of the image), their presence indicates that the accuracy of the approximation will depend on the distribution from which the costs are sampled. Finally, it should be noted that the proof relies on a few approximations that might not be accurate enough for typical image sizes n and root rates r of practical interest.

To test the impact of n and r on the approximation (2), we sampled n costs from a fixed cost distribution and computed the probabilities π_i associated with this sequence of costs c_i by solving the PLS problem given $M = r\sqrt{n}\log_2(\sqrt{n})$. We then compared (2) with the true deflection (1).

Several distributions of costs were tested: the Poisson, Beta, and distributions with only two possible costs. Similar trends were observed as long as the dynamic range of the costs (the ratio of the largest finite cost to the smallest non-zero cost $\Omega(\mathbf{c}) = \max(c_i)/\min(c_i)$) was not too large, e. g., $\Omega < 30$, for distributions with two costs, and $\Omega < 150$ for a scaled Poisson(5) distribution. Figure 1 (top) shows an example of the ratio of the asymptotic approximation (2) and the true deflection (1) for Poisson(5) + 1. The approximation is accurate irrespective of the scale for $r \approx 2.5$; the approximation overestimates the deflection for larger r and underestimates it for smaller r. As $n \to \infty$, nevertheless, the curves are guaranteed to approach a flat line at 1 for any r. This convergence is, however, very slow and depends on the dynamic range of costs. This makes the approximation (2) impractical for typical image sizes as the range of usable values of r is quite narrow. For 512×512 images, $r \approx 2.5$ corresponds to payloads close to 0.05 bits per pixels (bpp), which is relatively small when compared to values usually selected in steganographic benchmarks.

If these deviations did not depend on the distributions of costs, they could be easily rectified, if only empirically, for example by adding a correction term polynomial in r. However, as hinted above, these deviations depend heavily on the distribution of the costs themselves, the major factor being their dynamic range. To illustrate this impact, we repeated the simulations by sampling c_i from a uniform distribution $U(1, \Omega)$ (with no free bits, the costs can always be scaled to have the smallest cost equal to 1) and gradually increasing the dynamic range Ω. The results are summarized in Figure 1 (bottom) showing a clear relationship between Ω and the slope of the deviation: the larger the dynamic range, the larger the slope. For $\Omega < 10$, the curves look like the one observed in Figure 1 (bottom), however, the deviation quickly increases as soon as $\Omega > 10$. We would like to emphasize that encountering $\Omega > 100$ in a natural image is quite common, especially when using MiPOD [14] where the Fisher information (the reciprocal square of local pixel variance) often ranges from 10^{-4} to 10^4. Thus, one cannot dismiss such cases as rare or pathological. Similar behavior can be observed irrespectively of the distribution used, the only difference being the minimum dynamic range when the deviation from the curves in Figure 1 (top) is observed.

It should be noted that (2) is but a corrolary of a more general relationship in [10], which in its more precise form can be written as $\delta^2 \approx M/K^{-1}(\sum_{i=1}^{n} c_i/M)$, $K(x) = x/(\log_2(x))^2$. However, even when computing the exact inverse of K, this approximation was far from being satisfied in our tests with the above cost distributions.

This short study shows that the asymptotic result (2) is not tight enough for typical image sizes and current adaptive embedding schemes. Instead, we follow the second option and rely on a DLS to fix the deflection for each image across the scales in our experiments.

III. Experimental investigation

We investigate whether constant empirical detectability is observed when one embeds for constant detectability of the optimal detector and when

1) embedding is executed at an equilibrium,
2) using empirical detectors instead of the optimal ones,
3) heuristically adding or removing pixels (empirical source preservation).

To fulfill Condition 1 and use a DLS as argued in Section II, we use MiPOD as our embedding scheme. Schemes minimizing a heuristic distortion, such as the UNIWARD family [6] or HILL [13], cannot be used because a meaningful DLS is not available for them – there is no general approach that would link a heuristically derived distortion to detectability (in terms, for example of P_E).

Even with MiPOD, however, we still rely on a model, namely that the deflection based on this model accurately captures the statistical impact of embedding, which hinges upon a modeling assumption that each pixel is an independent

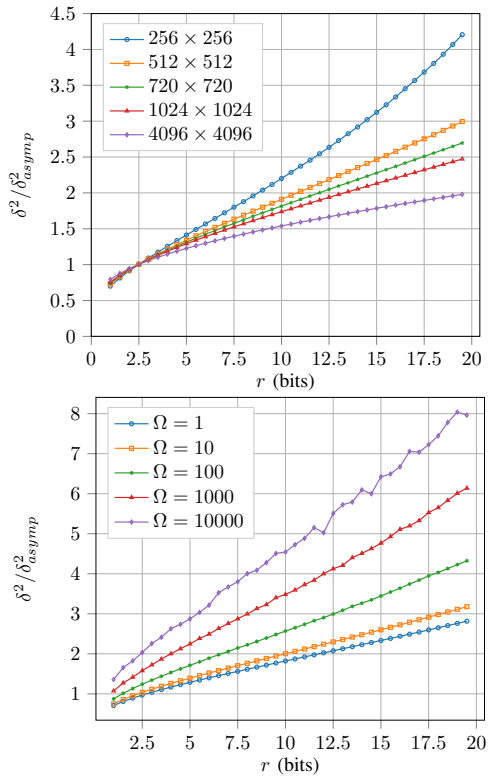

Figure 1: Top: Ratio of the asymptotic approximation of the deflection and the true deflection when sampling the costs from Poisson(5)+1 for different n with $\Omega = 21$. The behavior is almost identical for other cost distributions as long as Ω is not too large. Bottom: The ratio when sampling $n = 512 \times 512$ costs from $U(1, \Omega)$. Each curve corresponds to a different value of Ω.

realization of a Gaussian random variable with a known mean and variance that depends on local content complexity. There is no reason to believe that this model holds everywhere in natural images, especially in textured regions. To solve this difficulty, we first use artificial sources that follow MiPOD's model exactly. This will ensure that the only possible discrepancy between theoretical and empirical detectability will necessarily come from Conditions 2 and 3.

The rest of this section describes our datasets, including the the algorithm for "source-preserving" cropping, the embedding across scales, and the classifiers used for detection.

A. Dataset construction

Our datasets were constructed from RAW BOSSbase images converted to grayscale without downscaling or cropping. To force MiPOD's model, we used the methodology from [1]. First, MiPOD's variance estimator was used to estimate the variance $\sigma^2_{i,j}$ of each each pixel. The image was then denoised and its dynamic range linearly scaled (and rounded) to $[15, 240]$ (this is the "pixel mean" $\mu_{i,j}$). The variance $\sigma^2_{i,j}$ was further adjusted to $\underline{\sigma}^2_{i,j}$ so that, after adding to $\mu_{i,j}$ a sample from $\mathcal{N}(0, \underline{\sigma}^2_{i,j})$, the probability of getting out of the 8-bit dynamic range is equal to the probability of a one-sided 5σ Gaussian outlier. The cost (Fisher information) at pixel (i, j) is $I_{i,j} = \underline{\sigma}^{-4}_{i,j}$.

In JPEG domain, the dataset was constructed using a similar method. The denoised image was JPEG compressed to obtain the means in the JPEG domain. The (k, l)th DCT coefficient was noisified by adding to it a Gaussian sample from $\mathcal{N}(0, (\sigma^{(a,b)}_{k,l})^2)$, where $(\sigma^{(a,b)}_{k,l})^2 = \sum_{i,j=0}^{7} (f^{k,l}_{i,j})^2 \cdot (\sigma^{(a,b)}_{i,j})^2/q^2_{l,k}$, where (a, b) denotes the (a, b)th DCT block, (i, j) the indices of the (i, j)th pixel in that block, (k, l) the indices of the (k, l)th DCT coefficient in that block, $q_{k,l}$ is the quantization step for the (k, l)th DCT mode, $f^{k,l}_{i,j} = \frac{1}{4} w_k w_l \cos\frac{\pi k(2i+1)}{16} \cos\frac{\pi l(2j+1)}{16}$, and $w_0 = \frac{1}{\sqrt{2}}$, $w_k = 1$ for $k > 0$. The variance was clipped at 0.01, $(\underline{\sigma}^{(a,b)}_{l,k})^2 = \max\left\{0.01, (\sigma^{(a,b)}_{l,k})^2\right\}$ to allow us to compute the Fisher information for MiPOD in the fine quantization limit $I^{(a,b)}_{k,l} = \left(\underline{\sigma}^{(a,b)}_{k,l}\right)^{-4}$. Each DCT coefficient is thus an independent realization of a Gaussian random variable with a known mean and variance.

Once the full size datasets were produced, we generated five datasets by cropping the images to 256×256, 512×512, 1024×1024, and 2000×2000 in both the spatial and JPEG domain. To approximately preserve the cover source across scales, for each scale we selected the crop that minimized the L1 norm between the normalized histogram of the Fisher information of the crop and of the full size image. Histograms are first constructed using the image in its original size using 50 bins of equal size, the same binning is then used accross other scales.

B. Embedding

Each image was embedded using MiPOD modified to use binary instead of ternary embedding. First, a fixed relative (base) payload α in bpp was embedded into each image of the 512×512 datasets using a PLS. These images then serve as a baseline to which we match the deflection for other scales using the DLS $\delta^2_{MiPOD} = \sum_{i=1}^{n} \pi_i^2 \underline{\sigma}_i^{-4}$. For other scales, a DLS was used to determine the payload size to be embedded in each image in such a way that each crop coming from the same full size image has the same deflection across all scales. This ensures that each embedded dataset has exactly the same distribution of deflections.

C. Classification

The Low Complexity Linear Classifier (LCLC) [3] trained on the Spatial Rich Model (SRM) [5] and Gabor Filter Residual [15] was used as the empirical detector for the spatial and JPEG domain, respectively. While selection-channel aware versions of both features would follow the equilibrium condition more closely, our experiments showed that this did not impact the results. We used 5000 images for training, 5000 for testing, and five-fold cross-validation to select the tolerance parameter in LCLC. The performance is measured using $P_{\mathrm{E}} = \min_{P_{\mathrm{FA}}} \frac{1}{2}(P_{\mathrm{MD}} + P_{\mathrm{FA}})$, where P_{FA} and P_{MD} are the false-alarm and missed-detection rates.

α / Size	256×256	512×512	1024×1024	2000×2000
0.2	36.19	35.64	36.48	36.88
0.3	30.86	29.12	31.52	31.41
0.4	28.08	25.61	25.98	26.15
0.6	24.97	20.98	20.41	20.25

(a) Spatial domain.

α / Size	256×256	512×512	1024×1024	1920×1920
0.1	29.19	29.14	30.48	31.20
0.2	15.34	14.66	15.34	16.05
0.4	7.60	6.69	6.21	6.06

(b) JPEG domain quality 98.

Table I: P_{E} when forcing MiPOD's model on images. α corresponds to the base relative payload used on 512×512 images which is then scaled accordingly for other scales to match the deflection of 512×512 images using a DLS.

α / Size	256×256	512×512	1024×1024	1920×1920
0.10	25.46	28.97	32.61	35.18
0.20	11.88	13.26	17.37	21.12

(a) Spatial domain.

α / Size	256×256	512×512	1024×1024	2000×2000
0.15	22.34	22.74	22.85	24.26
0.25	11.15	11.88	9.55	13.22

(b) JPEG domain quality 98.

Table II: P_{E} on natural images.

IV. Results

For a forced MiPOD's model in the spatial domain s(Table Ia), P_{E} is approximately constant accross scales except for 256×256 images where it is by $\approx 3\%$ larger than the 512×512 baseline when the base payload exceeds 0.3bpp. These discrepancies are likely due to the fact that MiPOD's deflection δ^2_{MiPOD} has been derived for small payloads, while in our experiments, due to the scaling across image sizes, the relative payload embedded in 256×256 images is close to 1 bpp as soon as the base payload reaches 0.4bpp.

For the forced MiPOD's model in the JPEG domain (Table Ib), P_{E} similarly stays approximately constant ($\pm 1.5\%$) across the scales. This indicates that matching histograms of the Fisher Information is a viable strategy for source-preserving cropping for empirical detectors as long as the histograms are matched **almost exactly across all scales** and the relative payload is not too large.

A natural question to ask is how effective this strategy fairs on natural images, when the deflection is computed from a model that does not necessarily capture real images well. To this end, we repeated our experiments on full size grayscale BOSSbase images, with no further processing or model forcing applied. The results are summarized in Table II. In the spatial domain, P_{E} increases with the number of pixels despite the deflection being constant across scales. On the other hand, in the JPEG domain P_{E} never deviates more than 2% from the 512×512 baseline across all tested payloads. We hypothesize that this is due to the multi-variate Gaussian model of DCT coefficients being closer to reality in the JPEG domain thanks to the central limit theorem while the Gaussianity assumption imposed on pixels is comparatively less accurate in the spatial domain..

V. Conclusion

This work investigates payload scaling for constant statistical detectability with content-adaptive steganography that embeds at an equilibrium of a zero-sum game played by the steganographer and the steganalyst. In this setting, both players optimize their activity for the same objective function, which is the detectability or deflection coefficient determined by the steganographic Fisher information at each cover element and the payload / detectability, depending on the type of the sender. It is far from obvious that the theoretically derived scaling is observed in practice because practical embedding algorithms are not guaranteed to embed at equilibrium, the steganalyst does not use optimal detectors but empirical ones, and finally, heuristic rules need to be used to prepare datasets at different scales. To address the first concern, we use MiPOD rather than cost-based steganography and scale the payload to guarantee the same deflection (the detectability-limited sender) across datasets of different image scales prepared to approximately match the histogram of cover elements' Fisher information. Using detectors built as classifiers with rich models, the expected scaling is observed in sources with a forced cover model in both spatial and JPEG domains. For real images, the scaling was observed for JPEG images but not in the spatial domain, possibly due to modeling mismatch.

We also considered using the asymptotic relationship between deflection and payload derived by Ker for this work but concluded that the result is not tight enough for cover sizes of practical significance.

References

[1] M. Boroumand, J. Fridrich, and R. Cogranne. Are we there yet? In *Proc. IS&T, Electronic Imaging, Media Watermarking, Security, and Forensics 2019*, 2019.

[2] C. Cachin. An information-theoretic model for steganography. *Information and Computation*, 192(1):41–56, July 2004.

[3] R. Cogranne, V. Sedighi, J. Fridrich, and T. Pevny. Is ensemble classifier needed for steganalysis in high-dimensional feature spaces? In *2015 IEEE International Workshop on Information Forensics and Security (WIFS)*. IEEE, nov 2015.

[4] 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, Media Forensics and Security*, volume 7254, pages 08 1–11, San Jose, CA, January 18–21, 2009.

[5] J. Fridrich and J. Kodovsky. Rich models for steganalysis of digital images. *IEEE Transactions on Information Forensics and Security*, 7(3):868–882, jun 2012.

[6] 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.

[7] A. D. Ker. Batch steganography and the threshold game. In E. J. Delp and P. W. Wong, editors, *Proceedings SPIE, Electronic Imaging, Security, Steganography, and Watermarking of Multimedia Contents IX*, volume 6505, pages 04 1–13, San Jose, CA, January 29–February 1, 2007.

[8] 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.

[9] A. D. Ker. The square root law of steganography. In *Proceedings of the 5th ACM Workshop on Information Hiding and Multimedia Security IHMMSec2017*. ACM Press, 2017.

[10] A. D. Ker. On the relationship between embedding costs and steganographic capacity. In *Proceedings of the 6th ACM Workshop on Information Hiding and Multimedia Security - IHMMSec2018*. ACM Press, 2018.

[11] A. D. Ker, T. Pevny, and P. Bas. Rethinking optimal embedding. In *Proceedings of the 4th ACM Workshop on Information Hiding and Multimedia Security - IHMMSEC 2016*. ACM Press, 2016.

[12] A. D. Ker, T. Pevný, J. Kodovský, and J. Fridrich. The square root law of steganographic capacity. In *Proceedings of the 10th ACM workshop on Multimedia and security - IHMMSEC 2008*. ACM Press, 2008.

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

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

[15] X. Song, F. Liu, C. Yang, X. Luo, and Y. Zhang. Steganalysis of adaptive JPEG steganography using 2d gabor filters. In *Proceedings of the 3rd ACM Workshop on Information Hiding and Multimedia Security - IHMMSec2015*. ACM Press, 2015.

[16] 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.