The Square Root Law of Steganographic Capacity for Markov Covers

Tomáš Filler[a], Andrew D. Ker[b] and Jessica Fridrich[a]

[a]Department of Electrical and Computer Engineering, SUNY Binghamton
Binghamton, NY 13902-6000, USA
[b]Oxford University Computing Laboratory, Parks Road, Oxford OX1 3QD, England

ABSTRACT

It is a well-established result that steganographic capacity of perfectly secure stegosystems grows linearly with the number of cover elements—secure steganography has a positive rate. In practice, however, neither the Warden nor the Steganographer has perfect knowledge of the cover source and thus it is unlikely that perfectly secure stegosystems for complex covers, such as digital media, will ever be constructed. This justifies study of secure capacity of imperfect stegosystems. Recent theoretical results from batch steganography, supported by experiments with blind steganalyzers, point to an emerging paradigm: whether steganography is performed in a large batch of cover objects or a single large object, there is a wide range of practical situations in which secure capacity rate is vanishing. In particular, the absolute size of secure payload appears to only grow with the square root of the cover size. In this paper, we study the square root law of steganographic capacity and give a formal proof of this law for imperfect stegosystems, assuming that the cover source is a stationary Markov chain and the embedding changes are mutually independent.

Keywords: Steganography, steganographic capacity, square root law, Markov source

1. INTRODUCTION

In steganography, the sender communicates with the receiver by hiding her messages inside innocuous looking (cover) objects. Most practical steganographic methods embed messages by slightly modifying individual elements of the cover, obtaining thus the modified stego object that conveys the hidden message. The goal here is to make the stego objects statistically indistinguishable from covers. By this we mean that a passive warden inspecting the traffic cannot construct a detector of stego objects that would work better than an algorithm that makes random guesses. The assumption is that, up to a secret shared key, the Warden is familiar with all details of the steganographic scheme: this is the so-called Kerckhoffs' principle, which we also interpret to mean that the Warden has complete knowledge of the distribution of covers.

Statistical detectability of embedding changes depends on their character and extent. Intuitively, it should be possible to send short messages with lower risk of being detected than long messages. From a practical point of view, the sender needs to know how long a message she can embed for a chosen risk—she needs to know the steganographic capacity of the stegosystem. Unfortunately, determining the steganographic capacity analytically for real digital media objects, such as digital images, is very difficult even for the simplest steganographic paradigms, such as LSB embedding. The reason is the lack of accurate statistical models.

One may intuitively expect the steganographic capacity to be linear in the size of the cover object by referring to a similar result for capacity of noisy communication channels. This is, indeed, valid if the stegosystem is perfectly secure, since there is no possible detector.[1,2] In view of the absence of provably secure steganographic methods for real digital media, it makes sense to investigate steganographic capacity of imperfect embedding methods for which detectors exist and inquire about the largest payload that can be embedded using their ϵ-secure versions in the sense of Cachin.[3]

Further author information:
T. Filler: E-mail: tomas.filler@binghamton.edu; http://dde.binghamton.edu/filler
A. D. Ker: E-mail: adk@comlab.ox.ac.uk, Telephone: +44 1865 283530
J. Fridrich: E-mail: fridrich@binghamton.edu, Telephone: +1 607 777 6177, Fax: +1 607 777 4464

Media Forensics and Security, edited by Edward J. Delp III, Jana Dittmann, Nasir D. Memon, Ping Wah Wong, Proc. of SPIE-IS&T Electronic Imaging, SPIE Vol. 7254, 725408 · © 2009 SPIE-IS&T · CCC code: 0277-786X/09/$18 · doi: 10.1117/12.805911

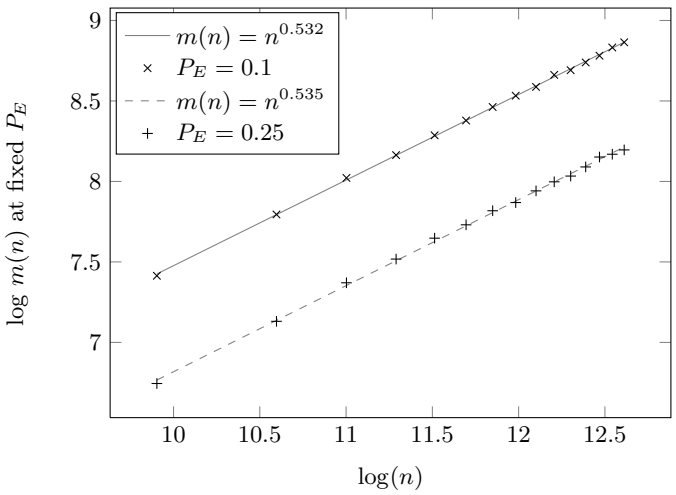

Figure 1. The largest payload $m(n_i)$ embedded using the embedding operation of F5 that produces a fixed steganalyzer error, P_E, for images with n_i non-zero DCT coefficients. The straight lines are corresponding linear fits. The slope of the lines is 0.53 and 0.54, which is in good agreement with the square root law.

The fact that steganographic capacity is most likely sublinear was already suspected by Anderson[4] in 1996:

> "Thanks to the Central Limit Theorem, the more covertext we give the Warden, the better he will be able to estimate its statistics, and so the smaller the rate at which [the Steganographer] will be able to tweak bits safely. The rate might even tend to zero..."

Recent analysis of batch steganography and pooled steganalysis by Ker[5] tells us that steganographic capacity of imperfect stegosystems only grows as the square root of the number of communicated covers. This result could be interpreted as the square root capacity law for a single image by dividing it into smaller blocks. The capacity result, however, was obtained with the assumption that the individual images (blocks) form a sequence of independent random variables, which is clearly false not only for images but also other digital media files. The main goal of this paper is to study the steganographic capacity for the simplest form of dependence that enables analytical reasoning—we assume that individual elements of the cover follow stationary Markov chain. The reason why we expect that the square root capacity law will hold is its experimental verification[6] for various embedding methods in both spatial and DCT domains. In particular, the maximal payload that leads to the same fixed detection accuracy of the steganalyzer is proportional to the square root of the cover size. A sample result of these experiments on JPEG images for the embedding operation of F5 is reprinted in Figure 1. There, the accuracy of the detector is represented using the minimal average classification error under equal prior probabilities of cover and stego images, P_E. For each set of images with a given number of non-zero DCT coefficients, n, the largest payload, $m(n)$, was iteratively found for which the steganalyzer obtained a fixed value of P_E. A linear fit through the experimental data displayed in a log-log plot confirms the square root law.

This paper is organized as follows. In the next section, we introduce notation, definitions, and basic assumptions under which we will operate. The square root law of steganographic capacity (SRL) is formulated and proved in Section 3. The result is interpreted and discussed in Section 4. To improve the flow of arguments, two important but rather technical results needed to prove the SRL are formulated as lemmas and moved to an appendix.

2. BASIC ASSUMPTIONS

In this section, we introduce notation and formulate and discuss three basic assumptions under which we prove the SRL. The first assumption concerns the impact of embedding. We postulate that the stego object is obtained by applying a mutually independent embedding operation to each cover element. This type of embedding can be found in majority of practical embedding methods (see, e.g., Ref. 7 and the references therein). The second

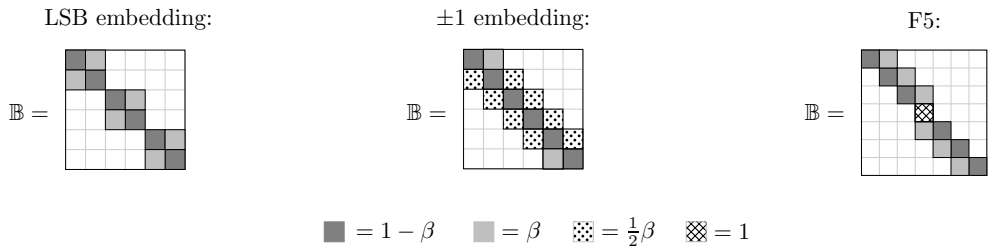

Figure 2. Examples of several embedding methods in the form of a functional matrix $\mathbb{B}$.

assumption is our model of covers. We require the individual cover elements to form a first-order Markov chain because this model is analytically tractable while allowing study of more realistic cover sources with memory. Finally, the third assumption essentially states that the steganographic method is not perfectly secure.

Throughout the paper, we use $\mathbb{A} = (a_{ij})$ to denote a matrix with elements a_{ij}, calligraphic font ($\mathcal{X}$) to denote sets, and capital letters (X, Y) to denote random variables, both vector and scalar. If y is a vector with components $y = (y_1, \ldots, y_n)$, y_k^l denotes the subsequence $y_k^l = (y_k, \ldots, y_l)$. If $Y = (Y_1, \ldots, Y_n)$ is a random vector with underlying probability distribution P, then $P(Y_k^l = y_k^l)$ denotes the marginal probability $P(Y_k = y_k, Y_{k+1} = y_{k+1}, \ldots, Y_l = y_l)$.

An n-element cover source will be represented using a random variable $X_1^n \triangleq (X_1, \ldots, X_n)$ distributed according to some general distribution $P^{(n)}$ over $\mathcal{X}^n$, $\mathcal{X} \triangleq \{1, \ldots, N\}$. A specific cover object is a realization of X_1^n and will be denoted with the corresponding lower case letter $x_1^n \triangleq (x_1, \ldots, x_n) \in \mathcal{X}^n$. A stegosystem, with covers of fixed size n, is a triple $S_n = (X_1^n, \Phi^{(n)}, \Psi^{(n)})$ consisting of the random variable describing the cover source, embedding mapping $\Phi^{(n)}$, and extraction mapping $\Psi^{(n)}$. The embedding mapping $\Phi^{(n)}$ applied to X_1^n induces another random variable $Y_1^n \triangleq (Y_1, \ldots, Y_n)$ with probability distribution $Q_\beta^{(n)}$ over $\mathcal{X}^n$. Specific realizations of Y_1^n are called stego objects and will be denoted $y_1^n \triangleq (y_1, \ldots, y_n)$. Here, $\beta \geq 0$ is a scalar parameter of embedding whose meaning will be explained shortly. The specific details of the embedding (and extraction) mappings are immaterial for our study. We only need to postulate the probabilistic *impact* of embedding.

Assumption 1: [Mutually independent embedding] The embedding algorithm visits every cover element X_k and, independently of all other elements, modifies it to a corresponding element of the stego object Y_k with probability

$$Q_\beta(Y_k = j | X_k = i) \triangleq b_{i,j}(\beta) = \begin{cases} 1 + \beta c_{i,i} & \text{if } i = j \\ \beta c_{i,j} & \text{otherwise,} \end{cases} \tag{1}$$

for some constants $(c_{i,j})$ with $c_{i,j} \geq 0$ for $i \neq j$. Note that because $\sum_{j \in \mathcal{X}} b_{i,j} = 1$, we must have $c_{i,i} = -\sum_{j \neq i} c_{i,j}$ for each $i \in \mathcal{X}$. The matrix $(c_{i,j})$ reflects the inner workings of the embedding algorithm, while the parameter β captures the *extent* of embedding changes. It will be useful to think of β as the relative number of changes (change rate) or some function of the change rate. Also note that we can find sufficiently small β_0 such that $b_{i,i}(\beta) > 0$ for $\beta \in [0, \beta_0]$ and all $i \in \mathcal{X}$.

Because the matrix $\mathbb{B}_\beta \triangleq (b_{i,j}(\beta))$ does not depend on $k \in \{1, \ldots, n\}$ or the history of embedding changes, one can say that the stego object is obtained from the cover by applying to each cover element a Mutually Independent embedding operation (we speak of *MI embedding*). The independence of embedding modifications implies that the conditional probability of stego object given the cover object can be factorized, i.e., $Q_\beta^{(n)}(Y_1^n | X_1^n) = \prod_{i=1}^n Q_\beta(Y_i | X_i)$.

Many embedding algorithms across different domains use MI embedding. Representative examples are LSB embedding, ±1 embedding, stochastic modulation, Jsteg, MMx, and various versions of the F5 algorithm.[7] Examples of the matrix $\mathbb{B}_\beta$ for three selected embedding methods are shown in Figure 2.

Next, we formulate our assumption on the cover source.

Assumption 2: [Markov cover source] We assume that the cover source X_1^n is a first-order stationary Markov Chain over $\mathcal{X}$, which we will often abbreviate as simply Markov Chain (MC). This source is completely

described by its stochastic transition probability matrix $\mathbb{A} \triangleq (a_{ij}) \in \mathbb{R}^{N\times N}$, $a_{ij} = Pr(X_k = j|X_{k-1} = i)$, and by the initial distribution $Pr(X_1)$. The probability distribution induced by the MC source generating n-element cover objects satisfies $P^{(n)}(X_1^n = x_1^n) = P^{(n-1)}(X_1^{n-1} = x_1^{n-1})a_{x_{n-1}x_n}$, where $P^{(1)}(X_1)$ is the initial distribution. We further assume that the transition probability matrix of the cover source satisfies $a_{ij} \geq \delta > 0$, for some δ and thus the MC is irreducible. The stationary distribution of the MC source is a vector $\pi \triangleq (\pi_1, \ldots, \pi_N)$ satisfying $\pi\mathbb{A} = \pi$. In this paper, we will always assume that the initial distribution $P^{(1)}(X_1) = \pi$, which implies $P^{(n)}(X_k) = \pi$ for every n and k. This assumption simplifies the analysis without loss of generality because the marginal probabilities $P^{(n)}(X_k)$ converge to π with exponential rate w.r.t. k (see Doob,[8] equation (2.2) on page 173). In other words, MCs are "forgetting" their initial distribution with exponential rate.

Under the above assumption and the class of MI embedding, the source of stego objects no longer exhibits the Markov property and forms a Hidden Markov Chain (HMC) instead.[9] The HMC model is described by its hidden states (cover elements) and output transition probabilities (MI embedding). Hidden states are described by the cover MC and the output probability transition matrix $\mathbb{B}$ is taken from the definition of MI embedding.

Unless stated otherwise, in the rest of this paper $Q_\beta^{(n)}$ denotes the probability measure induced by the HMC source embedded with parameter β into n-element MC cover objects. By the stationarity of the MC source, the marginal probabilities $P^{(n)}(X_k^{k+1}) = P^{(2)}(X_1^2)$ and $Q_\beta^{(n)}(Y_k^{k+1}) = Q_\beta^{(n)}(Y_1^2)$ for all k. Sometimes we will omit the number of elements, n, and denote as P and Q_β the probability distribution over cover and stego objects, respectively.

The third assumption we formulate concerns the entire stegosystem S_n. Because it is known[1,2] that steganographic capacity of perfectly secure stegosystems is linear in n, the SRL can only apply to imperfect stegosystems.

Assumption 3: [**FI condition**] We assume that the stegosystem $S_n = (X_1^n, \Phi^{(n)}, \Psi^{(n)})$ is not perfectly secure in the sense of Cachin[3] (the KL divergence $D_{KL}(P^{(n)}||Q_\beta^{(n)}) > 0$). For our special case of Markov cover sources X_1^n and MI embedding $\Phi^{(n)}$, this assumption can be equivalently stated in two different forms:

1. The pair $(P^{(2)}, Q_\beta^{(2)})$ *does not* satisfy so called *Fisher Information condition,*

$$\forall y_1^2 \in \mathcal{X}^2 \quad \Big(P^{(2)}(X_1^2 = y_1^2) > 0\Big) \;\Rightarrow\; \Big(\frac{d}{d\beta}Q_\beta^{(2)}(y_1^2)\big|_{\beta=0} = 0\Big). \tag{2}$$

2. There exists a pair of states (i, j) such that

$$P(X_1^2 = (i,j)) \neq Q_\beta(Y_1^2 = (i,j)) \text{ for all } \beta > 0. \tag{3}$$

The complete proof of equivalence of each condition and perfect security appears in [10, Cor. 7]. Here, we only provide a few brief arguments. First of all, perfectly secure stegosystems must satisfy (2) because the Fisher information

$$I(0) = E_P\left[\left(\frac{d}{d\beta}\ln Q_\beta^{(2)}(y_1^2)\big|_{\beta=0}\right)^2\right]$$

appears as a coefficient in front of β^2 in the Taylor expansion of KL divergence $D_{KL}(P^{(2)}||Q_\beta^{(2)})$ w.r.t. β and thus $\frac{d}{d\beta}Q_\beta^{(2)}(y_1^2)\big|_{\beta=0}$ must be zero whenever $P^{(2)}(y_1^2) > 0$. The opposite implication (zero Fisher information implies zero KL divergence) is not valid in general but holds for MI embedding as shown in Ref. 10. The second condition follows from the fact that second-order marginal statistics fully describe the first-order MC process and thus if (3) does not hold, then both cover and stego distributions are the same for all n (the stegosystem is perfectly secure).

Finally, we would like to stress that Assumptions 1–3 are not overly restrictive and will likely be satisfied for all practical steganographic schemes in some appropriate representation of the cover. For example, in digital images it is unlikely that the distribution of each pixel depends only on its neighbor, but the dependency is likely to be spatially-limited. Then the image can be modeled as a Markov chain made up of overlapping pixel groups. Furthermore, if a stegosystem preserves the first-order statistics of a cover source, it is likely to be detectable by considering higher-order dependencies: the apparently-perfect stegosystem becomes imperfect when the cover is represented by pairs or groups of pixels, coefficients, or some other derived quantities.

3. THE SQUARE ROOT LAW OF STEGANOGRAPHIC CAPACITY

In this section, we formulate and prove the main result of this paper, which states that the steganographic capacity of imperfect stegosystems with Markov covers and mutually independent embedding operation only grows with the square root of the number of cover elements. This finding has some fundamental implications in steganography and steganalysis. Probably the most remarkable one is that steganographic capacity exhibits quite different properties when compared with capacity of noisy channels or lossless compression. For example, while a mismatch in source model decreases the compression gain by a constant (the KL divergence between the source model and true source distribution), a cover model mismatch in steganography leads to vanishing capacity.

For the formulation of the SRL Theorem, we borrow the term used in Ref. 5. We will say that the Steganographer is *at risk* (w.r.t. some fixed tuple (P^*_{FA}, P^*_{MD}), with $0 < P^*_{FA} < 1$ and $0 < P^*_{MD} < 1 - P^*_{FA}$) if the Warden has a detector with probability of false alarms and missed detection P_{FA}, P_{MD} satisfying $P_{FA} < P^*_{FA}$ and $P_{MD} < P^*_{MD}$.

Theorem 1: [The square root law of steganography for Markov covers] *For a sequence of stegosystems $(S_n)_{n=1}^{\infty}$ satisfying Assumptions 1–3, the following holds:*

1. *If the sequence of embedding parameters $\beta(n)$ increases faster than $1/\sqrt{n}$ in the sense that $\lim_{n\to\infty} \frac{\beta(n)}{1/\sqrt{n}} = \infty$, then, for sufficiently large n, the Steganographer is at risk for arbitrary tuple (P^*_{FA}, P^*_{MD}).*
2. *If $\beta(n)$ increases slower than $1/\sqrt{n}$, $\lim_{n\to\infty} \frac{\beta(n)}{1/\sqrt{n}} = 0$, then the stegosystem can be made ε-secure for any $\varepsilon > 0$ for sufficiently large n. This implies that the Steganographer is not at risk, for any tuple (P^*_{FA}, P^*_{MD}).*
3. *Finally, if $\beta(n)$ grows asymptotically as fast as $1/\sqrt{n}$, $\lim_{n\to\infty} \frac{\beta(n)}{1/\sqrt{n}} = \epsilon$ for some $0 < \epsilon < \infty$, then the stegosystem is asymptotically $C\epsilon^2$-secure for some constant C.*

Proof: We prove each part of the theorem separately. We remind the reader that, under our interpretation of Kerckhoffs' principle, the Warden knows the distribution of cover objects $P^{(n)} = Q_0^{(n)}$.

Part 1 [Steganographer at risk] Here, we prove that the Steganographer is at risk w.r.t. any (P^*_{FA}, P^*_{MD}) for all sufficiently large n. This means that we need to construct a sequence of detectors, D_n, for the following composite binary hypothesis testing problem

$$\begin{aligned} \mathrm{H}_0 &: \beta = 0 \\ \mathrm{H}_1 &: \beta > 0 \end{aligned}$$

based on observing one stego object (one realization of a random sequence with distribution $Q_\beta^{(n)}$). The error probabilities of these detectors are required to satisfy $P_{FA} < P^*_{FA}$ and $P_{MD} < P^*_{MD}$ for all sufficiently large n. We now describe the test statistic for each detector D_n.

Equation (3) in Assumption 3 guarantees the existence of pair of states (i, j) such that $P(X_1^2 = (i, j)) \neq Q_\beta(Y_1^2 = (i, j))$ for all $\beta > 0$. Thus, we define the test statistic $\nu_{\beta,n}$ for detector D_n as

$$\nu_{\beta,n} = \sqrt{n}\left|\frac{1}{n-1}h_\beta[i,j] - P\big(X_1^2 = (i,j)\big)\right|, \tag{4}$$

where $\frac{1}{n-1}h_\beta[i, j]$ is the relative count of the number of consecutive pairs (i, j) in an n-element stego object embedded using parameter β (In terms of indicator functions*, $h_\beta[i,j] = \sum_{k=1}^{n-1} \mathbb{I}_{\{Y_k=i, Y_{k+1}=j\}}$). Note that due to stationarity of the cover source, $E\left[\frac{1}{n-1}h_\beta[i,j]\right] = Q_\beta(Y_1^2 = (i,j))$ for all β.

We prove the following for the difference between the means of $\nu_{\beta,n}$ under both hypotheses

$$\lim_{n\to\infty} E[\nu_{\beta,n}] - E[\nu_{0,n}] = \infty \text{ when } \sqrt{n}\beta \to \infty. \tag{5}$$

*For any two statements A, B, $\mathbb{I}_{\{A,B\}} = 1$ if A and B are true, otherwise $\mathbb{I}_{\{A,B\}} = 0$.

Suppose, for a contradiction, that there exists $K > 0$, and a strictly increasing sequence of integers $(n_m)_{m=1}^{\infty}$ for which

$$|E[\nu_{\beta,n_m}] - E[\nu_{0,n_m}]| < K \text{ for all } m. \tag{6}$$

If $\limsup_{m\to\infty} \beta(n_m) = \beta_0 > 0$, then there exists a subsequence of $(n_m)_{m=1}^{\infty}$, which we denote the same to keep the notation simple, such that $\lim_{m\to\infty} \beta(n_m) = \beta_0$. For this subsequence, however, the difference

$$E[\nu_{\beta,n_m}] - E[\nu_{0,n_m}] \quad = \quad \sqrt{n_m}\Big|Q_\beta\big(Y_1^2 = (i,j)\big) - P\big(X_1^2 = (i,j)\big)\Big|$$

tends to ∞ with $m \to \infty$ because by (3) the absolute value converges to a positive value independent of m. This is, however, a contradiction with (6).

If $\lim_{m\to\infty} \beta(n_m) = 0$, we find the contradiction in a different manner. By the FI condition from Assumption 3, there must exist states (i,j) such that $\frac{d}{d\beta}Q_{\beta=0}(Y_1^2 = (i,j)) \neq 0$. From the Taylor expansion† of $Q_\beta(Y_1^2 = (i,j))$ at $\beta = 0$ with Lagrange remainder and $0 < \xi < 1$

$$E[\nu_{\beta,n_m}] - E[\nu_{0,n_m}] \quad = \quad \sqrt{n_m}\beta\left|\frac{d}{d\beta}Q_{\beta=0}\big(Y_1^2 = (i,j)\big) + \frac{1}{2}\beta\frac{d^2}{d\beta^2}Q_{\xi\beta}\big(Y_1^2 = (i,j)\big)\right|, \tag{7}$$

which tends to ∞ as $m \to \infty$ when $\sqrt{n_m}\beta \to \infty$, which is again a contradiction with (6). We summarize that $E[\nu_{\beta,n}] - E[\nu_{0,n}] \to \infty$ holds for any sequence $\beta(n)$ for which $\sqrt{n}\beta(n) \to \infty$.

Lemma 1 in the Appendix shows that exponential forgetting of Markov chains guarantees that

$$Var[\nu_{\beta,n}] < C \tag{8}$$

for some constant C independent of n and β. Equations (5) and (8) are all we need to construct detectors D_n that will put the Steganographer at risk for all sufficiently large n. The detector D_n has the following form

$$\begin{aligned} \nu_{\beta,n} > T &\quad \text{decide stego } (\beta > 0) \\ \nu_{\beta,n} \leq T &\quad \text{decide cover } (\beta = 0), \end{aligned}$$

where T is a fixed threshold. We now show that T can be chosen to make the detector probability of false alarms and missed detections satisfy

$$\begin{aligned} P_{FA} &< P_{FA}^* \\ P_{MD} &< P_{MD}^* \end{aligned}$$

for sufficiently large n. The threshold $T(P_{FA}^*)$ will be determined from the requirement that the probability of the right tail, $x \geq T(P_{FA}^*)$, under H_0 is at most P_{FA}^*. Using Chebyshev's inequality,

$$P_{FA} = Pr(\nu_{0,n} \geq T) \leq Pr(|\nu_{0,n}| \geq T) \quad \leq \quad \frac{Var[\nu_{0,n}]}{T^2} < \frac{C}{T^2}.$$

Setting $T = \sqrt{C/P_{FA}^*}$ gives us $P_{FA} < P_{FA}^*$.

Because of the growing difference between the means (5), we can find n large enough so that the probability of the left tail, $x \leq T(P_{FA}^*)$, under H_1 is less than or equal to P_{MD}^*. Again, we use the Chebyshev's inequality with the bound on the variance of $\nu_{\beta,n}$ to prove this

$$\begin{aligned} P_{MD} &= Pr\big(\nu_{\beta,n} < T(P_{FA}^*)\big) = Pr\big(\nu_{\beta,n} - E[\nu_{\beta,n} - \nu_{0,n}] < T(P_{FA}^*) - E[\nu_{\beta,n} - \nu_{0,n}]\big) \\ &\leq Pr\big(|\nu_{\beta,n} - E[\nu_{\beta,n} - \nu_{0,n}]| > E[\nu_{\beta,n} - \nu_{0,n}] - T(P_{FA}^*)\big) < \frac{C}{\big(E[\nu_{\beta,n} - \nu_{0,n}] - T(P_{FA}^*)\big)^2}, \end{aligned}$$

†The Taylor expansion is valid since by its form the function $Q_\beta(Y_k^{k+1} = (i,j))$ is analytic.

which can be made arbitrarily small for sufficiently large n because $E[\nu_{\beta,n}] - E[\nu_{0,n}] \to \infty$. This establishes the first part of the square root law.

Part 2 [Asymptotic undetectability] Now we prove that when $\sqrt{n}\beta \to 0$, then the KL divergence between the distributions of cover and stego objects tends to zero,

$$d_n(\beta) = D_{KL}\Big(P^{(n)}||Q_\beta^{(n)}\Big) = \sum_{y_1^n \in \mathcal{X}^n} P^{(n)}(X_1^n = y_1^n) \lg \frac{P^{(n)}(X_1^n = y_1^n)}{Q_\beta^{(n)}(Y_1^n = y_1^n)} \to 0, \tag{9}$$

which will establish that the steganography is ε-secure for any $\varepsilon > 0$ for sufficiently large n. By the well-known connection between hypothesis testing and KL divergence,[3] no nontrivial upper bound on false alarms and missed detections will be met, for large enough n.

Using Taylor expansion of $d_n(\beta)$ with Lagrange remainder at $\beta = 0$ we have $d_n(\beta) = d_n(0) + d_n'(0)\beta + \frac{d_n''(\upsilon\beta)}{2!}\beta^2$, where $0 < \upsilon < 1$. This step is valid since under our assumptions all derivatives of (normalized) KL divergence are continuous w.r.t. β, for proof see Ref. 11. The term $d_n(0)$ is zero because both distributions are the same when $\beta = 0$. The term $d_n'(0)$ is also zero because

$$\begin{aligned} d_n'(0) &= \lim_{\beta\to 0} d_n'(\beta) = \lim_{\beta\to 0} \frac{-1}{\ln 2} \sum_{y_1^n} P^{(n)}(X_1^n = y_1^n) \frac{\frac{d}{d\beta} Q_\beta^{(n)}(Y_1^n = y_1^n)}{Q_\beta^{(n)}(Y_1^n = y_1^n)} = \frac{-1}{\ln 2} \sum_{y_1^n} \frac{d}{d\beta} Q_{\beta=0}^{(n)}(Y_1^n = y_1^n) \\ &= \lim_{\beta\to 0} \frac{-1}{\ln 2} \frac{d}{d\beta} \Big(\underbrace{\sum_{y_1^n} Q_\beta^{(n)}(Y_1^n = y_1^n)}_{=1} \Big) = 0. \end{aligned}$$

Finally, by Lemma 2 in Appendix there exists a constant $\tilde{C}$, such that $\frac{1}{n} d_n''(\beta) < \tilde{C}$ for $\beta \in [0, \beta_0]$ and all n. Thus, $d_n(\beta) \le \frac{1}{2}\tilde{C} n\beta^2 \to 0$ when $\sqrt{n}\beta \to 0$.

Part 3 [Asymptotic $C\epsilon^2$-security] To prove the third part of the square root law, we again expand the KL divergence $d_n(\beta)$ at $\beta = 0$ up to the third order with the Lagrange form of the remainder

$$d_n(\beta) = \frac{1}{2!}\Big(\frac{d_n''(0)}{n}\Big) n\beta^2 + \frac{1}{3!}\Big(\frac{d_n'''(\upsilon\beta)}{n}\Big) n\beta^3 \tag{10}$$

for some $0 < \upsilon < 1$. According to Ref. 11, both normalized derivatives of the KL divergence, $\frac{1}{n} d_n''(0)$ and $\frac{1}{n} d_n'''(\upsilon\beta)$, are upper bounded by the same finite constant $\tilde{C}$ for all $\beta \in [0, \beta_0]$. Since $\beta(n)\sqrt{n} \to \epsilon$ with $n \to \infty$, $\beta(n) \to 0$ and thus the expansion is valid. By the same reason, the second term in (10) converges to zero as $n \to \infty$. From this result, we obtain the asymptotic bound on KL divergence in the form $d_n(\beta) \le \frac{1}{2}\tilde{C}\epsilon^2$ as was to be shown. Q.E.D.

4. CONCLUSION

A general theme is now emerging in steganography literature: whether steganography is performed in a large batch of cover objects or a single large object, there is a wide range of situations in which *secure capacity grows according to the square root of the cover size*. Such results will likely hold for all stegosystems that are not perfectly secure in the sense of Cachin. It appears that the theory of *hidden* information is quite unlike the traditional theory of information.

The result presented in this paper is the first to allow dependence between the components of the cover. We have proved the square root law of steganographic capacity for single covers under essentially three conditions: that they can be represented as a Markov chain, that the embedding operation can be modeled as independent substitutions of one state for another, and that the embedding scheme does not preserve all statistical properties of the cover. This applies to a very wide range of popular steganographic algorithms, in spatial and transform domains. The last condition is important because it is known that perfectly secure steganography, conveying information at a linear rate, can always be constructed if the cover source is perfectly understood.[12] However, we

have argued that digital media cover sources will never be perfectly understood. We should explain, therefore, why it makes sense to assume that the Warden has perfect knowledge of the covers (necessary for construction of the detectors in Part 1 of the proof). It can be justified by the cautious nature of Kerckhoffs' principle: although the Steganographer may believe that the Warden does not know the complete cover distribution, there is always the risk that the Warden knows more than the Steganographer does. For example, she might know more about dependencies between cover elements. Since the Steganographer cannot say for certain how much the Warden will know, they must assume the worse case: complete knowledge.

The formulation of our theorem parallels that of Ref. 5: embedding at a rate faster than $\sqrt{n}$ leads to eventual detection, whereas embedding at a rate slower than $\sqrt{n}$ leads to eventual ε-security. At rates $A\sqrt{n}$, the stegosystem is ϵ-secure. We clarify, though, that this does *not* refer to embedding at a diminishing rate in a single cover object (which would be a different theorem, and is an avenue for further research). The quantity $\beta(n)$ describes a constant embedding rate throughout an object of size n: one could think of the function β as giving a strategy describing how much data can be hidden in an object of each size. The SRL says that over-ambitious strategies lead to easier detection in larger objects, cautious strategies lead to more difficult detection in larger objects, and quantifies the boundary.

The square root law has some important implications in steganography and steganalysis. Most significantly, the simple fact that capacity is sublinear means that a true steganographic channel, with positive rate, cannot be constructed unless the cover source is known perfectly. For steganalysis, the SRL explains in part why the same relative payload can be detected more accurately in large images (however it is not the whole explanation: larger images tend to have more smooth gradients and less noise, and these factors also influence detection accuracy). Thus, when benchmarking steganography, the distribution of image sizes in the database influences the reliability of steganalysis and makes it more difficult to compare the results on two different databases. To resolve this issue, one might switch to measuring the payload in bits per square root of pixel.

Finally, we emphasize that the square root law of capacity relates to the number of changes caused by the embedding process, and not to the size of the information transmitted. With adaptive source coding methods (even simple matrix embedding based on Hamming codes will do) the number of bits of information which can be conveyed, by making at most c changes in n cover locations, is $O(c\log(n/c))$. Therefore the SRL implies an asymptotic information capacity which is $O(\sqrt{n}\log n)$ (i.e., still sublinear), in the absence of perfect steganography.

ACKNOWLEDGMENTS

The work on this paper was supported by Air Force Office of Scientific Research under the research grant number FA9550-08-1-0084. 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. Andrew Ker is a Royal Society University Research Fellow.

APPENDIX

In this appendix, we include two auxiliary lemmas needed in the proof of the SRL in Section 3.

Lemma 1: Let $\nu_{\beta,n}$ be the random variable defined in (4) for a fixed value of the parameter β and cover size n. The variance of this random variable can be bounded by a constant C for every value of β and n

$$\exists C, \forall\beta, \forall n \quad Var[\nu_{\beta,n}] \leq C.$$

Proof: From the definition of $\nu_{\beta,n}$

$$\begin{aligned}\frac{(n-1)^2}{n}Var[\nu_{\beta,n}] &= E\Big[\Big(\sum_{k=1}^{n-1}\mathbb{I}_{\{Y_k^{k+1}=(i,j)\}}\Big)^2\Big] - E\Big[\sum_{k=1}^{n-1}\mathbb{I}_{\{Y_k^{k+1}=(i,j)\}}\Big]^2 \leq \sum_{k=1}^{n-1}Var[\mathbb{I}_{\{Y_k^{k+1}=(i,j)\}}] + \\ &\quad +2\Big[\sum_{k+1<\hat{k}}E[\mathbb{I}_{\{Y_k^{k+1}=(i,j)\}}\mathbb{I}_{\{Y_{\hat{k}}^{\hat{k}+1}=(i,j)\}}] - E[\mathbb{I}_{\{Y_k^{k+1}=(i,j)\}}]E[\mathbb{I}_{\{Y_{\hat{k}}^{\hat{k}+1}=(i,j)\}}]\Big] + 2n. \quad (11)\end{aligned}$$

In the last sum, we bounded all terms for $k = \hat{k} - 1$ by 1 and thus obtained the term $2n$ in the last inequality. For any event A,

$$Var[\mathbb{I}_A] = Pr(A) - Pr(A)^2 = \tfrac{1}{4} - \big(Pr(A) - \tfrac{1}{2}\big)^2 \leq \tfrac{1}{4}$$

so the first term is bounded by $\frac{n-1}{4}$.

Finally, we find an upper bound on the sum in (11) in the form of $C_2 n$ for some positive constant C_2. This will give us the proof because $Var[\nu_{\beta,n}] \leq \frac{n}{(n-1)^2}((n-1)\frac{1}{4} + 2C_2 n + 2n) \leq 4(\frac{1}{4} + 2C_2 + 2)$, and $\frac{n^2}{(n-1)^2} \leq 4$ for $n \geq 2$. Thus $C = 8C_2 + 9$.

We start by showing that

$$\begin{aligned} &Q_\beta\big(Y_k^{k+1} = (i,j), Y_{\hat{k}}^{\hat{k}+1} = (i,j)\big) - Q_\beta\big(Y_k^{k+1} = (i,j)\big) Q_\beta\big(Y_{\hat{k}}^{\hat{k}+1} = (i,j)\big) \\ &= \Big[\underbrace{Q_\beta\big(Y_{\hat{k}}^{\hat{k}+1} = (i,j) \big| Y_k^{k+1} = (i,j)\big) - Q_\beta\big(Y_{\hat{k}}^{\hat{k}+1} = (i,j)\big)}_{\leq N^2 \rho^{\hat{k}-k-2}}\Big] Q_\beta\big(Y_k^{k+1} = (i,j)\big) \leq N^2 \rho^{\hat{k}-k-2}, \end{aligned} \tag{12}$$

for some $0 \leq \rho < 1$ and $k+1 < \hat{k}$ (N is the number of all possible states of the MC). In other words, the HMC is exponentially forgetting its initial condition. Then, we will be able to bound the sum in (11) by $N^2 \sum_{\hat{k}=3}^{n} \sum_{k=1}^{\hat{k}-2} \rho^{\hat{k}-k-2} = N^2 \sum_{\hat{k}=3}^{n} \frac{1-\rho^{\hat{k}-2}}{1-\rho} \leq N^2 \sum_{\hat{k}=3}^{n} \frac{1}{1-\rho} = N^2(n-2)\frac{1}{1-\rho} \leq \frac{N^2 n}{1-\rho}$. Thus, $C_2 = \frac{N^2}{1-\rho}$ because $Q_\beta\big(Y_k^{k+1} = (i,j)\big) \leq 1$.

The term $Q_\beta\big(Y_{\hat{k}}^{\hat{k}+1} = (i,j)\big)$ in (12) can be written as

$$Q_\beta\big(Y_{\hat{k}}^{\hat{k}+1} = (i,j)\big) = \sum_{(\hat{i},\hat{j})} Q_\beta\big(Y_{\hat{k}}^{\hat{k}+1} = (i,j) \big| X_{\hat{k}}^{\hat{k}+1} = (\hat{i},\hat{j})\big) P\big(X_{\hat{k}}^{\hat{k}+1} = (\hat{i},\hat{j})\big) = \sum_{(\hat{i},\hat{j})} b_{\hat{i},i} b_{\hat{j},j} P\big(X_{\hat{k}}^{\hat{k}+1} = (\hat{i},\hat{j})\big). \tag{13}$$

The term $Q_\beta\big(Y_{\hat{k}}^{\hat{k}+1} = (i,j) \big| Y_k^{k+1} = (i,j)\big)$ in (12) can be written as

$$\begin{aligned} Q_\beta\big(Y_{\hat{k}}^{\hat{k}+1} = (i,j) \big| Y_k^{k+1} = (i,j)\big) &= \frac{Q_\beta\big(Y_{\hat{k}}^{\hat{k}+1} = (i,j), Y_k^{k+1} = (i,j)\big)}{Q_\beta\big(Y_k^{k+1} = (i,j)\big)} \\ &= \frac{\sum_{(\hat{i},\hat{j})} \sum_{(\tilde{i},\tilde{j})} b_{\hat{i},i} b_{\hat{j},j} b_{\tilde{i},i} b_{\tilde{j},j} P\big(X_{\hat{k}}^{\hat{k}+1} = (\hat{i},\hat{j}), X_k^{k+1} = (\tilde{i},\tilde{j})\big)}{Q_\beta\big(Y_k^{k+1} = (i,j)\big)} = (\#). \end{aligned}$$

Finally, $P\big(X_{\hat{k}}^{\hat{k}+1} = (\hat{i},\hat{j}), X_k^{k+1} = (\tilde{i},\tilde{j})\big)$ can be factorized as $P\big(X_{\hat{k}}^{\hat{k}+1} = (\hat{i},\hat{j}) | X_k^{k+1} = (\tilde{i},\tilde{j})\big) P\big(X_k^{k+1} = (\tilde{i},\tilde{j})\big)$. Due to the Markov property of the random variable X_1^n, $P\big(X_{\hat{k}}^{\hat{k}+1} = (\hat{i},\hat{j}) | X_k^{k+1} = (\tilde{i},\tilde{j})\big) = P\big(X_{\hat{k}}^{\hat{k}+1} = (\hat{i},\hat{j}) | X_{k+1} = \tilde{j}\big)$. For each pair of indices $(\hat{i},\hat{j})$, we define index $\tilde{j}_{max} = \arg\max_{\tilde{j}} P\big(X_{\hat{k}}^{\hat{k}+1} = (\hat{i},\hat{j}) | X_{k+1} = \tilde{j}\big)$. Then

$$\begin{aligned} (\#) &\leq \frac{\sum_{(\hat{i},\hat{j})} b_{\hat{i},i} b_{\hat{j},j} P\big(X_{\hat{k}}^{\hat{k}+1} = (\hat{i},\hat{j}) \big| X_{k+1} = \tilde{j}_{max}\big) \sum_{(\tilde{i},\tilde{j})} b_{\tilde{i},i} b_{\tilde{j},j} P\big(X_k^{k+1} = (\tilde{i},\tilde{j})\big)}{Q_\beta\big(Y_k^{k+1} = (i,j)\big)} \\ &\overset{(13)}{=} \sum_{(\hat{i},\hat{j})} b_{\hat{i},i} b_{\hat{j},j} P\big(X_{\hat{k}}^{\hat{k}+1} = (\hat{i},\hat{j}) \big| X_{k+1} = \tilde{j}_{max}\big). \end{aligned} \tag{14}$$

Now, we can combine (13) and (14) to prove (12) as

$$\begin{aligned} &Q_\beta\big(Y_{\hat{k}}^{\hat{k}+1} = (i,j) \big| Y_k^{k+1} = (i,j)\big) - Q_\beta\big(Y_{\hat{k}}^{\hat{k}+1} = (i,j)\big) \\ &\overset{(13),(14)}{\leq} \sum_{(\hat{i},\hat{j})} b_{\hat{i},i} b_{\hat{j},j} \Big(P\big(X_{\hat{k}}^{\hat{k}+1} = (\hat{i},\hat{j}) \big| X_{k+1} = \tilde{j}_{max}\big) - P\big(X_{\hat{k}}^{\hat{k}+1} = (\hat{i},\hat{j})\big)\Big) \\ &= \sum_{(\hat{i},\hat{j})} b_{\hat{i},i} b_{\hat{j},j} P\big(X_{\hat{k}+1} = \hat{j} \big| X_{\hat{k}} = \hat{i}\big) \Big(P\big(X_{\hat{k}} = \hat{i} \big| X_{k+1} = \tilde{j}_{max}\big) - P\big(X_{\hat{k}} = \hat{i}\big)\Big) \leq N^2 \rho^{\hat{k}-k-2}. \end{aligned} \tag{15}$$

It is a well known result in MCs that the absolute value of the term $P\big(X_{\hat{k}} = \hat{i} \big| X_{k+1} = \tilde{j}_{max}\big) - P\big(X_{\hat{k}} = \hat{i}\big)$ in (15) can be bounded by $\rho^{\hat{k}-k-2}$ (exponential forgetting), for some constant $0 \leq \rho < 1$. This is because the MC is irreducible due to the assumption $a_{i,j} \geq \delta$ (see equation (2.2) on page 173 in Doob[8]). This bound does not depend on $\tilde{j}_{max}$. The final bound does not depend on β because $b_{\hat{i},i} \leq 1$ and $b_{\hat{j},j} \leq 1$. Q.E.D.

Lemma 2: Let $d_n(\beta) = D_{KL}(P^{(n)}||Q_\beta^{(n)})$ be the KL divergence between n-element cover and stego sources as defined in (9). Then,

$$\exists \beta_0, \exists C > 0, \forall \beta \in [0, \beta_0], \forall n \quad \frac{1}{n} d_n''(\beta) < \tilde{C}.$$

In other words, the second derivative of the normalized KL divergence, $\frac{1}{n} d_n''(\beta)$, can be bounded by a constant $\tilde{C}$ for each n and β. And this bound does not depend on n or β.

Proof: The problem of bounding normalized derivatives of KL divergence for the case of HMC was studied by Mevel et al.[13] Their results, namely Theorem 4.4 and Theorem 4.7, however, cannot be directly applied to our case because our assumptions are different. In particular, Assumption C on page 1124 is not satisfied because we allow zeros in matrix $\mathbb{B}$. Motivated by this work, we derive a more general result about the normalized KL divergence and its derivatives (see the report in Ref. 11). Intuitively, we can expect the normalized KL divergence to be arbitrarily smooth and bounded due to the smooth transition from P to Q_β and the fact that $d_n(0) = 0$. The main result of the report, formally stated in [11, Theorem 3], says that every derivative of $\frac{1}{n} d_n(\beta)$ w.r.t. β (and the function $\frac{1}{n} d_n(\beta)$ itself) is uniformly bounded and Lipschitz-continuous (or simply continuous) on $[0, \beta_0]$. These properties are independent of $n \geq 1$. From this fact, Lemma 2 can be obtained as a special case. This result also allows us to expand the KL divergence into a Taylor series with respect to β. Q.E.D.

REFERENCES

[1] Moulin, P. and Wang, Y., "New results on steganographic capacity," in [*Proceedings of the Conference on Information Sciences and Systems, CISS*], (March 17–19, 2004).

[2] Comesana, P. and Pérez-Gonzáles, F., "On the capacity of stegosystems," in [*Proceedings of the 9th ACM Multimedia & Security Workshop*], Dittmann, J. and Fridrich, J., eds., **1525**, 3–14 (September 20–21, 2007).

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

[4] Anderson, R., "Stretching the limits of steganography.," in [*Information Hiding, 1st International Workshop*], Anderson, R. J., ed., *Lecture Notes in Computer Science* **1174**, 39–48, Springer-Verlag (May 30 – June 1, 1996).

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

[6] Ker, A. D., Pevný, T., Kodovský, J., and Fridrich, J., "The square root law of steganographic capacity," in [*Proceedings of the 10th ACM Multimedia & Security Workshop*], Ker, A., Dittmann, J., and Fridrich, J., eds., 107–116 (September 22–23, 2008).

[7] Kodovský, J., Fridrich, J., and Pevný, T., "Statistically undetectable JPEG steganography: Dead ends, challenges, and opportunities," in [*Proceedings of the 9th ACM Multimedia & Security Workshop*], Dittmann, J. and Fridrich, J., eds., 3–14 (September 20–21, 2007).

[8] Doob, J. L., [*Stochastic processes*], Wiley, New York, 1st ed. (1953).

[9] Sidorov, M., "Hidden Markov models and steganalysis," in [*Proceedings of the 6th ACM Multimedia & Security Workshop*], Dittmann, J. and Fridrich, J., eds., 63–67 (September 20–21, 2004).

[10] Filler, T. and Fridrich, J., "Complete characterization of perfectly secure stego-systems with mutually independent embedding operation," in [*Proceedings IEEE, International Conference on Acoustics, Speech, and Signal Processing*], (April 19–24, 2009). To appear.

[11] Filler, T., "Important properties of normalized KL-divergence under HMC model," tech. rep., DDE Lab, SUNY Binghamton (2008). `http://dde.binghamton.edu/filler/kl-divergence-hmc.pdf`.

[12] Wang, Y. and Moulin, P., "Perfectly secure steganography: Capacity, error exponents, and code constructions," *IEEE Transactions on Information Theory, Special Issue on Security* **54**(6) (2008).

[13] Mevel, L. and Finesso, L., "Asymptotical statistics of misspecified hidden Markov models," *Automatic Control, IEEE Transactions on* **49**(7), 1123–1132 (2004).