The Square Root Law Requires a Linear Key

Andrew D. Ker
Oxford University Computing Laboratory
Parks Road
Oxford OX1 3QD, UK
adk@comlab.ox.ac.uk

ABSTRACT

We extend the *square root law of steganographic capacity*, for the simplest case of iid covers, in two ways. First, we show that the law still holds under a more realistic embedding assumption, where the payload is of fixed length (instead of, in the classic result, independent embedding at each location). Second, we consider the case of nonuniform embedding paths, which is forced when the stegosystem's secret key is of limited size: we show that the secret key must be of length at least linear in the payload size, if a square root law is to hold. The latter is parallel to Shannon's perfect cryptography bound.

Categories and Subject Descriptors

D.2.11 [**Software Engineering**]: Software Architectures—*information hiding*; H.1.1 [**Models and Principles**]: Systems and Information Theory—*information theory*

General Terms

Security, Algorithms

Keywords

Steganographic Capacity, Square Root Law, Steganography, Steganalysis

1. INTRODUCTION

For a given cover object and steganographic embedding method, the *capacity* is the largest payload which does not exceed a particular risk of detection. There is now a body of literature proving that, under certain conditions, the capacity of a cover of size n is proportional to $\sqrt{n}$. These *square root laws* were first conjectured for the case of multiple, independent covers [10], and proved to hold under certain conditions about steganalysis of individual objects [11, 13]. Later, a square root law for individual covers of size n was proved, under the assumption that the cover source is a suitably well-behaved Markov chain [6].

MM&Sec'09, September 7–8, 2009, Princeton, New Jersey, USA.

In this paper we consider the simplest type of square root law for individual objects – the cover source is assumed to produce independent, identically distributed (iid) elements – and extend it in two ways concerning the embedding key. Although iid samples are a poor model for the practice of steganography in digital media, this case illustrates the square root law without the complicated analysis required in [6], and we expect that these results will eventually be proved for the Markov chain case too. Here, our aim is to quantify how the size of the embedding key relates to the security of the stegosystem.

We next briefly outline our notation (Subsect. 1.1) and state a useful inequality (Subsect. 1.2). In Sect. 2 we re-state the classic square root law, then extend it to a more realistic form of embedding in which the payload size is fixed (Subsect. 2.2), and then prove a theorem about the minimum size of shared secret key (Subsect. 2.3). Sect. 3 considers the significance of the new results, and examines ways in which they could be improved or extended. Finally, Sect. 4 briefly concludes the work.

1.1 Notational Conventions

In this paper we will use uppercase Roman letters for random variables, probability measures, and sets; lowercase and Greek letters are for realisations of random variables, constants, and functions. Expectation of a random variable is written $\mathrm{E}[X]$ (with the probability measure for X implicit). $X \sim \mathrm{Bi}(n,p)$ indicates that X has the binomial distribution: a sum of n independent Bernoulli random variables each with probability p.

Vectors will be written $(x_1,\ldots,x_n)$, or equivalently x_1^n, and the subsequence $(x_i,\ldots,x_j)$ will be denoted by x_i^j. The term $O(\psi(n))$ indicates a function of n bounded asymptotically by $\psi(n)$: $\phi(n) = O(\psi(n))$ if there is a constant C such that $\phi(n) \leq C\psi(n)$ for sufficiently large n. All logs are to natural base.

1.2 Hoeffding's Inequality

We will need to bound the tail probabilities for various combinations of binomial distributions, and for reasons of uniformity will use the same bound – Hoeffding's inequality [9] – throughout. In fact, some of our results do not require an exponential bound (e.g. Chebychev's inequality will often suffice, as in [6]) but Hoeffding's inequality can still be tidier because it does not require us to compute the variance of the random variable whose tail probabilities are to be bounded. We state the inequality in the form we will use it:

LEMMA 1 (HOEFFDING'S INEQUALITY). *Let X be a sum of n independent, not necessarily identically distributed, random variables each bounded in $[0,1]$. For $t > 0$,*

$$\Pr[X \geq \mathrm{E}[X] + nt] \leq \exp(-2nt^2), \quad \textit{and}$$
$$\Pr[X \leq \mathrm{E}[X] - nt] \leq \exp(-2nt^2).$$

The inequality applies immediately to random variables with the binomial distribution, which are independent sums of Bernoulli variables. It also applies to sums of independent binomially distributed random variables.

2. SQUARE ROOT LAWS FOR IID COVERS

Square root laws relate a maximum secure payload size (bounded by an acceptable risk of detection) to the size of the cover. Generally they are asymptotic, and the square root laws of [6, 11] are of the following form: if the payload size increases asymptotically faster than the square root of the cover size then probability of detection tends to one; if the payload size increases asymptotically slower than the square root of the cover size then probability of detection tends to zero.

It is easy to find embedding which violates the law, for example padding the payload to match the cover size, or to construct cover sources where the law does not hold, for example those producing a single sample repeated endlessly. In this work, our assumption about embedding will be natural, but we make the strong (and, for practical purposes, not realistic) assumption that the covers consist of independent samples which we will call "pixels", though they could be another representation of the cover such as transform domain coefficients. For us, an n pixel cover $(X_1, \ldots, X_n)$ is a realisation of n independent and identically distributed random variables each with mass function $p(x)$. We tacitly assume that the alphabet – the set of possible values for the pixels – is finite.

On embedding, some of the pixels are altered. It will not matter, for our purposes, exactly what embedding function is used, as long as the change at each embedding location is independent of the others. This applies to common embedding such as bit replacement or additive noise. We will denote the mass function of payload pixels $q(x)$ and we will need some weak assumptions about p and q.

Throughout this section, we refer to the *null hypothesis*, H_0, for the situation when there is no embedding, and the *alternative hypothesis*, H_1, for the case when a particular length payload is embedded. A false positive detection is a type-I error, whose probability is conventionally denoted α, and a false negative result is a type-II error with probability denoted β.

2.1 Classic Result

Historically, the first square root law was proved under an independence assumption [11] but in the context of *batch steganography* (multiple cover objects), in which case independence is quite plausible. A more sophisticated square root law, aimed at individual covers from a Markov chain source, was developed in [6]. To our knowledge, the simplest square root law is for single covers whose pixels are iid, and this classic result has certainly been known to researchers for a year or two, but it seems to have been "skipped" in the literature so we reproduce it here. In any case, it is useful to compare with the novel results which follow.

THEOREM 1. *Suppose that the cover consists of n pixels $(X_1, \ldots, X_n)$, independent and identically distributed each with mass function $p(x)$. Suppose that a payload of size m causes each pixel to be replaced, independently of each other and the cover, with probability $\lambda = m/n$, and that replaced pixels are independent each with mass function $q(x)$. Finally, suppose that for all x, $p(x) \neq 0$ and $q(x) \neq 0$, and there exists y such that $p(y) \neq q(y)$.*

(i) If $m/\sqrt{n} \to \infty$ then, for sufficiently large n, covers and stego objects can be distinguished with arbitrarily low error rate.

(ii) If $m/\sqrt{n} \to 0$ then, for sufficiently large n, any detector must have arbitrarily high error rate.

PROOF. For (i), we construct a detector with arbitrarily low false positive rate, and for which the false negative rate tends to zero as $n \to \infty$.

We write $p = p(y)$ and $q = q(y)$ for convenience; without loss of generality, we may assume that $p < q$. Our detector simply counts how many pixels take the value y – define $Y = |\{\, i \mid X_i = y\}|$ – and rejects the null hypothesis H_0 if $Y > y^*$, where y^* is the critical threshold

$$y^* = np + c\sqrt{n}$$

(c is a positive constant to be determined later).

We compute the probability of a false positive detection: under H_0, $Y \sim \mathrm{Bi}(n, p)$, so

$$\begin{aligned} \alpha &= \Pr[Y > y^*] \\ &= \Pr[Y - \mathrm{E}[Y] > c\sqrt{n}] \\ &\leq \exp(-2c^2), \end{aligned}$$

using Hoeffding's inequality. This can be made arbitrarily small by suitable choice of c.

For the probability of a false negative observe that, under H_1, $Y \sim \mathrm{Bi}(n, (1-\lambda)p + \lambda q)$, so

$$\begin{aligned} \beta &= \Pr[Y \leq y^*] \\ &= \Pr[Y - \mathrm{E}[Y] \leq c\sqrt{n} - (q-p)m] \\ &\leq \exp\bigl(-2(q-p)^2 m^2/n + O(m/\sqrt{n})\bigr) \\ &\to 0 \end{aligned}$$

as $n \to \infty$, again using Hoeffding's inequality, so β will be arbitrarily small for sufficiently large n.

Moving to (ii), consider the distributions of entire cover objects and stego objects; let us write P and Q for the respective probability measures. By independence, $P(X_1^n = x_1^n) = \prod p(x_i)$ and $Q(X_1^n = x_1^n) = \prod (1-\lambda)p(x_i) + \lambda q(x_i)$. We bound the performance of any detector by computing the Kullback-Leibler (KL) divergence from P to Q[1]:

$$\begin{aligned} D_{\mathrm{KL}}(P \parallel Q) &= -\sum_{i=1}^{n} \int p(x_i) \log\Bigl(\tfrac{(1-\lambda)p(x_i) + \lambda q(x_i)}{p(x_i)}\Bigr)\, \mathrm{d}x_i \\ &= -n \int p(x) \log\Bigl(1 + \lambda\bigl(\tfrac{q(x) - p(x)}{p(x)}\bigr)\Bigr)\, \mathrm{d}x. \end{aligned}$$

[1]We use the familiar integral symbol in the definition of KL divergence, though under our assumptions the alphabet is finite. The integration is properly over a discrete measure.

Now recall that $\log(1+z) \geq z - z^2$ for (at least) $z > -\frac{1}{2}$, and by assumption $\frac{q(x)-p(x)}{p(x)}$ is bounded, so for sufficiently large n we will have λ small enough that

$$D_{\mathrm{KL}}(P \parallel Q) \leq n\lambda \int p(x) - q(x)\,\mathrm{d}x + n\lambda^2 \int \tfrac{(q(x)-p(x))^2}{p(x)}\,\mathrm{d}x;$$

the first term is zero since both p and q must integrate to unity, and the second term is a constant multiple of m^2/n (note that $(q(x)-p(x))^2/p(x)$ is bounded and nonnegative, by the assumptions on p and q), and therefore tends to zero. Now, by the well-known corollary to the data processing theorem [1], we have lower bounds on the false positive and negative probabilities α and β:

$$\alpha \log \frac{\alpha}{1-\beta} + (1-\alpha)\log\frac{1-\alpha}{\beta} \leq D_{\mathrm{KL}}(P \parallel Q)$$

and $D_{\mathrm{KL}}(P \parallel Q) \to 0$ forces $\alpha \to 1-\beta$, i.e. detector performance tends to random, as $n \to \infty$. □

The proof of a square root law for Markov covers, in [6], is similar in concept, but the analysis is much more involved.

One may immediately ask: what if $\sqrt{m}/n \to r$, where r is a positive constant known as the *root rate*? For the purposes of this paper we will ignore this case, but it can be addressed by means similar to (ii), as demonstrated in [6]. The outcome is to bound $D_{\mathrm{KL}}(P \parallel Q)$ (and hence put lower bounds on α and/or β) by a multiple of r^2; the constant of multiplicity is related to *Fisher information*, and estimation of Fisher information for real-world cover sources is the subject of current research [5, 14].

We should clarify that the square root law applies to the number of embedding changes, not necessarily to the payload size itself. In some cases, these quantities need not be proportional. If the stego system includes an asymptotically optimal adaptive choice of source coding, for example matrix embedding using Hamming codes [8], then c changes in a cover of size n can convey payload of order $c\log(n/c)$ (this is an upper bound) and so capacity in terms of payload size becomes of order $\sqrt{n}\log n$. Nonetheless, we prefer to separate the source coding from the embedding function, and continue to identify payload size with (something proportional to) the number of embedding changes.

Note our conditions on p and q, which seem unavoidable. If the mass functions $p(x)$ and $q(x)$ are identical, then so are cover and stego objects: perfect steganography is achievable simply by overwriting the entire cover, and a *linear* law applies to capacity. (Indeed, a number of naive embedding schemes manage to preserve first-order statistics of covers, but the practice of steganalysis is rather different to the iid theory considered here, and inevitably such embedding is detected using higher-order statistics. We believe that perfect steganography is not realisable in practical circumstances.)

If there exists x such that $p(x) = 0$ but $q(x) \neq 0$ then the symbol x is a certain indicator that payload is present: a detector with zero false positives can be constructed based solely on x. As $n \to \infty$, the probability that x is observed in a stego object tends to 1. Similarly, if there exists x with $q(x) = 0$ but $p(x) \neq 0$ then the symbol x is a certain indicator that payload is not present, and a similarly asymptotically perfect detector can be constructed.

2.2 Fixed-Length Payload

Although the main weakness of the model in the previous result is the simplicity of the covers, one can also criticize the assumption about payload: when steganography is performed, there is usually a fixed-length payload to be embedded. It is *not* the case that each pixel is affected independently because changes cease after enough have been made to carry the payload. In practice, we would expect that the sender and recipient share a secret key which determines exactly m locations (or, in the case of source coding, some number of locations proportional to m) to be used. Effectively, exactly m locations, chosen uniformly at random from all n, will be overwritten by symbols with the alternative distribution $q(x)$.

We now show that the square root law still holds in such a case.

THEOREM 2. *Suppose that the cover consists of n pixels $(X_1, \ldots, X_n)$, independent and identically distributed each with mass function $p(x)$. Suppose that a payload of size m causes exactly m pixels to be replaced with mass function $q(x)$, and that this pixel selection is made uniformly from all $\binom{n}{m}$ possibilities. Finally, suppose that for all x, $p(x) \neq 0$ and $q(x) \neq 0$, and there exists y such that $p(y) \neq q(y)$.*

(i) If $m/\sqrt{n} \to \infty$ then, for sufficiently large n, covers and stego objects can be distinguished with arbitrarily low error rate.

(ii) If $m/\sqrt{n} \to 0$ then, for sufficiently large n, any detector must have arbitrarily high error rate.

PROOF. The structure is similar to Theorem 1 and (i) is not much altered, but we have to work particularly hard for (ii) because the stego pixels are no longer independent.

For (i), we use exactly the same detector as in the proof of Theorem 1: reject the null hypothesis H_0 if $Y = |\{\, i \mid X_i = y\}| > y^*$, where

$$y^* = np + c\sqrt{n}.$$

The null hypothesis is exactly as before, and so the false positive probability α still meets the bound $\alpha \leq \exp(-2c^2)$ and can be made arbitrarily small.

This time, under H_1, we know that exactly m pixels have distribution $q(x)$ and n have distribution $p(x)$, so

$$Y \sim \mathrm{Bi}(m, q) + \mathrm{Bi}(n-m, p),$$

where the sum is of independent distributions. This random variable has the same mean as its counterpart in Theorem 1, so exactly the same application of Hoeffding's inequality gives $\beta \leq \exp\bigl(-2(q-p)^2m^2/n + O(m/\sqrt{n})\bigr) \to 0$. Again, β will be arbitrarily small for sufficiently large n.

For (ii), we again show that $D_{\mathrm{KL}}(P \parallel Q) \to 0$. We expand the KL divergence using the chain rule [3, §2.5]:

$$D_{\mathrm{KL}}(P(X_1^n) \parallel Q(X_1^n)) = \sum_{k=1}^{n} D_k$$

where D_k is the conditional divergence defined by

$$\begin{aligned} D_k &= D_{\mathrm{KL}}(P(X_k \mid X_1^{k-1}) \parallel Q(X_k \mid X_1^{k-1})) \\ &= \mathrm{E}_P\left[-\int P(X_k = x \mid X_1^{k-1}) \log\Bigl(\tfrac{Q(X_k=x\mid X_1^{k-1})}{P(X_k=x\mid X_1^{k-1})}\Bigr)\,\mathrm{d}x \right] \\ &= \mathrm{E}_P\left[-\int P(X_k = x) \log\Bigl(\tfrac{Q(X_k=x\mid X_1^{k-1})}{P(X_k=x)}\Bigr)\,\mathrm{d}x \right]. \end{aligned}$$

At the last step, we used the independence of X_1^n under P, but the same is not true under Q. The expectation is taken over the random variables X_1^{k-1}, with the probability measure P.

Let us write E_k for the event that location X_k is used for embedding, $\overline{E}_k$ for the complement, and denote the conditional probability

$$e_k = Q(E_k \mid X_1^{k-1}).$$

Lemma 2 in Appendix A establishes a key property of e_k: for sufficiently large n and all k it is bounded by a multiple of the unconditional probability $Q(E_k) = m/n$.

Conditional on E_k, X_k *is* independent of X_1^{k-1} under Q: if E_k then X_k has distribution q, and if $\overline{E}_k$ then X_k has distribution p, so

$$\begin{aligned} D_k &= \mathrm{E}_P\Big[-\int P(X_k = x) \cdot \\ &\qquad \log\Big(\tfrac{e_k Q(X_k=x|E_k, X_1^{k-1}) + (1-e_k) Q(X_k|\overline{E}_k, X_1^{k-1})}{P(X_k=x)}\Big)\,\mathrm{d}x\Big] \\ &= \mathrm{E}_P\Big[-\int p(x) \log\Big(\tfrac{e_k q(x) + (1-e_k)p(x)}{p(x)}\Big)\,\mathrm{d}x\Big] \\ &= \mathrm{E}_P\Big[-\int p(x) \log\Big(1 + e_k\big(\tfrac{q(x)-p(x)}{p(x)}\big)\Big)\,\mathrm{d}x\Big]. \end{aligned}$$

Now we can continue as in Theorem 1: because $e_k \le C\frac{m}{n} \to 0$ the inequality $\log(1+z) \ge z - z^2$ can be used for large enough n, so

$$\begin{aligned} D_{\mathrm{KL}}(P(X_1^n) \parallel Q(X_1^n)) &= \sum_{k=1}^n D_k \\ &\le \sum_{k=1}^n e_k \int p(x) - q(x)\,\mathrm{d}x + e_k^2 \int \tfrac{(q(x)-p(x))^2}{p(x)}\,\mathrm{d}x \\ &\le \sum_{k=1}^n C' \frac{m^2}{n^2}, \end{aligned}$$

for a constant C'. This tends to zero, so as before we deduce that any detector must have performance tending to purely random as $n \to \infty$. □

2.3 On Minimum Key Size

In practice, the hypothesis that the payload location is chosen uniformly from all $\binom{n}{m}$ possibilities is unlikely to be realised. Supposing that the embedding function hides one bit per used location (e.g. LSB embedding), the sender and recipient would have to agree on the location of each payload bit, requiring them to share $\log_2 n!/(n-m+1)!$ bits of secret information for their covert communication to succeed. If $m = O(\sqrt{n})$, this means of order $m \log m$ bits of information: the sender and recipient must share a secret key longer than their covert payload! (They would not do better to re-use the same key for many communications, either, because we know that this is unsound, see e.g. [12] or [15]). Note that we use the word *key*, here, to indicate the shared secret information by which the sender and recipient agree the location of the payload. This need not be related to any encryption key and knowledge of this embedding key would not necessarily be sufficient for their enemy to decode the hidden message, though we would certainly expect it to improve the efficiency of detectors.

What typically occurs, then, is that the sender and recipient agree on a shorter secret key, which generates the embedding path pseudorandomly. But recall Kerckhoffs' Principle: prudent security analysis assumes that the enemy knows the entire system, and only the secret key itself is unknown. The detector should be granted knowledge of the connection between secret keys and embedding paths, and they may be able to exploit this knowledge (see e.g. [7]). So now we examine what happens to the square root law if the number of embedding keys is not large enough. It turns out that the secret key size cannot be sublinear in the payload size, else asymptotically perfect detection can be achieved:

THEOREM 3. *Suppose that the cover consists of n pixels $(X_1, \ldots, X_n)$, independent and identically distributed each with mass function $p(x)$. Suppose a payload of size m which will cause exactly m pixels to be replaced with mass function $q(x)$. Suppose that the sender, recipient, and attacker share knowledge of a set K of secret keys, each of which generates a path of length m determining the payload locations, but only the sender and recipient know which key is used. Finally, suppose that there exists y such that $p(y) \ne q(y)$.*

If $(\log |K|)/m \to 0$, as $m \to \infty$, and $m \to \infty$ as $n \to \infty$, then, for sufficiently large n, covers and stego objects can be distinguished with arbitrarily low error rate.

PROOF. As before, write $p = p(y)$ and $q = q(y)$ and assume that $p < q$. For each $k \in K$, write L_k for the set of locations selected by key k, and write k^* for the true key used for embedding.

The detector will look exhaustively at each possible embedding path and see whether the proportion of y symbols is high enough, in any one of them, to indicate a payload. Define

$$Y_k = |\{\, i \in L_k \mid X_i = y\}|$$

and reject H_0 if $\max_k Y_k \ge y^*$, where the critical threshold this time is

$$y^* = qm - c\sqrt{m},$$

with c a positive constant to be determined later. We show that the detector has, for sufficiently large n and with a suitably chosen c, arbitrarily small probability of error.

The false negative rate can easily be bounded because, under H_1, $Y_{k^*} \sim \mathrm{Bi}(m, q)$, so

$$\begin{aligned} \beta &= \Pr[\text{all } Y_k < y^*] \\ &\le \Pr[Y_{k^*} < y^*] \\ &= \Pr[Y_{k^*} - \mathrm{E}[Y_{k^*}] < -c\sqrt{m}] \\ &\le \exp(-2c^2), \end{aligned}$$

using Hoeffding's inequality. By choosing c sufficiently large, any nonzero bound on β can be met.

For the false positive rate α, note that $Y_k \sim \mathrm{Bi}(m, p)$ for all keys k, but Y_k are not independent because some embedding paths will cross. However, let us enumerate the keys $K = \{k_1, \ldots, k_{|K|}\}$, then

$$\begin{aligned} 1 - \alpha &= \Pr[\text{all } Y_k < y^*] \\ &= \prod_{i=1}^{|K|} \Pr[Y_{k_i} < y^* \mid Y_{k_1} < y^* \wedge \cdots \wedge Y_{k_{i-1}} < y^*] \\ &\ge \Pr[Y_1 < y^*]^{|K|} \end{aligned}$$

$$
\begin{aligned}
&= \Pr[Y_1 - \mathrm{E}[Y_1] < m(q-p) - c\sqrt{m}]^{|K|} \\
&\geq \left(1 - \exp(-2m(q-p)^2 + O(\sqrt{m}))\right)^{|K|} \\
&\geq \exp\left(-2\exp(\log|K| - 2m(q-p)^2 + O(\sqrt{m}))\right) \\
&\to 1
\end{aligned}
$$

The first inequality is because, conditional on some Y_k's being less than y^*, the probability that any other Y_k is less than y^* is only increased (or at least equal) because of possible overlaps. This can be proved rigorously, but is so intuitively obvious that we will omit to do so. The second inequality is Hoeffding's. The final inequality follows from $1-\exp(x) \geq \exp(-2\exp(x))$ (for x sufficiently small). Given that $(\log|K|)/m \to 0$ as $n \to \infty$, the second exponent must tend to $-\infty$, hence the result. □

3. DISCUSSION

Theorem 2 presents a simple change to the *independent embedding* model typically used in square root laws. We have argued that it is more natural to consider a fixed number of affected locations because it corresponds to the practice of steganographic embedding by bit replacement, (mod k)-matching, and most JPEG embedding operations. The proof of the theorem is mostly very similar to that of Theorem 1, but it requires a substantial additional component (Lemma 2). The key is that the conditional probability $Q(E_k \mid X_1^{k-1})$ behaves asymptotically like the unconditional probability $Q(E_k)$, i.e. that the condition accounts for at worst a constant factor. The author thinks it likely that a simpler proof of Lemma 2 exists.

Theorem 3 is of different character. $\log_2|K|$ is the number of bits of information needed for the sender and recipient to agree on the embedding path, so if the number of secret key bits is asymptotically less than linear in the payload size, we do *not* have a square root law of capacity. Indeed, Theorem 3 shows that m cannot even tend to infinity with n, so capacity reduces to a *constant*. This makes sense when we consider the detector constructed in the proof, which does not involve n.

Another result relating minimum key size to information hiding is described in [2], where it is shown that a linear key is necessary for a certain kind of watermarking security, essentially equivalent to *perfect* steganography, to be possible. That result is different to ours, and inherits simply from classical information theory. Perfect steganography is completely unlike the cases we considered here, since the square root law does not apply to it; we would argue that it represents an unachievable singularity.

Our theorem asks more questions than it answers, and we believe it should be a springboard for further research. Most importantly, it lacks a second clause parallel to (ii) of Theorems 1 and 2: we have not yet shown that a linear key is sufficient for the square root law to reappear. Proving such a result ought to be analogous to the other theorems, and the challenge will be to show an analogue of Lemma 2 bounding the conditional embedding probabilities. However, this appears to be difficult. It certainly requires some constraints on the set K: if, for example, the $|K|$ embedding paths almost or entirely overlap then there will not be enough uncertainty to get the required result. And it is difficult to formalise such constraints, depending as they do on n. One possibility is to demand that the set K be chosen uniformly at random from all embedding paths, but then we must be careful about the interpretation of error "probabilities": they would apply over all possible sets K, but for some specific sets K the bounds might be broken. Alternatively, we could perhaps set a bound on the amount of overlap between the paths in K. Either way, it appears unlikely that the calculations of Appendix A can be brought to bear on the limited-key problem.

That key size must be (at least) linear in payload size parallels Shannon's classic perfect cryptography bound [16]: for perfect cryptography to be possible, the communicating parties must share a secret key at least linear in the size of the plaintext. In that case, the minimum constant of proportionality is the entropy rate of the source. In the case of steganography, we may be able to determine the constant if an asymptotically optimal detector is constructed. Certainly the detector used in the proof of Theorem 3 is far from optimal, making use of the frequency of only a single symbol. Performance can be improved by using a chi-square test, and perhaps this will turn out to be optimal because of the connection between chi-square and likelihood ratio tests. Likewise, our use of Hoeffding's inequality is profligate, and perhaps the Chernoff bound would be tighter (there is also a potential connection with KL divergence).

The value of the constant is crucial. If greater than one (and if the result remains in the presence of adaptive source coding), it would imply that steganography is asymptotically pointless because the sender and recipient must agree a longer secret than their payload. If less than one, more and more secret information can be bootstrapped from short keys.

Apart from extending the results in the iid case, further research is needed to transfer them to the Markov chain cover case in [6]. Some new insights may be needed to make the analysis tractable.

Finally, we should note that the idea of exhaustively testing all possible embedding keys, which is used in the proof of Theorem 3, was proposed in [7]. It is not clear whether such a search is really practicable because it requires effort proportional to $|K|$, analogous to breaking a password by brute force. Restricted key lengths reduce the theoretical secure capacity of stego systems, but perhaps the practical secure capacity remains high given computational constraints on the detector. There has been relatively little work on such complexity aspects of steganalysis.

4. CONCLUSIONS

Until now, square root laws have worked under the *independent embedding assumption*, that each pixel (sample) is affected independently. Some important properties of independent embedding can be found in [4], but we have argued that it is slightly unrealistic: in practice, a fixed-size payload is embedded and only as many embedding changes as necessary will be made, thus breaking the independence assumption. In this work, concentrating only on the iid cover model, we have investigated what happens to the square root law under a fixed-length payload assumption.

We have shown that the square root law is unaltered given a uniform choice of embedding path, but agreement on such a path requires too large a secret stego key. If the secret key is not at least linear in the payload size then capacity is not even potentially infinite, let alone of square root order. Important questions remain unanswered, including whether

a linear key is even sufficient, and the constant of linearity if so.

We believe that these questions are of vital importance to both theory and practice of steganography. The relationship between cover size, payload size, and key size would be the foundation of a theory of hidden information.

5. ACKNOWLEDGMENTS

The author is a Royal Society University Research Fellow. Thanks are due to Jessica Fridrich, who helped this paper come to completion under difficult circumstances.

6. REFERENCES

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

[2] C. Cayre, C. Fontaine and T. Furon. Watermarking Security: Theory and Practice. *IEEE Transactions on Signal Processing*, 53(10):3976–3987, 2005.

[3] T. Cover and J. Thomas. *Elements of Information Theory*. Wiley, 1991.

[4] T. Filler and J. Fridrich. Complete characterization of perfectly secure stego-systems with mutually independent embedding operation. In *Proc. IEEE International Conference on Acoustics, Speech, and Signal Processing*, 2009.

[5] T. Filler and J. Fridrich. Fisher Information determines capacity of ϵ-secure steganography. To appear in *Proc. 11th Information Hiding Workshop*, 2009.

[6] T. Filler, A. Ker, and J. Fridrich. The square root law of steganographic capacity for Markov covers. In *Media Forensics and Security XI*, volume 7254 of *Proc. SPIE*, pages 0801–0811, 2009.

[7] J. Fridrich, M. Goljan, and D. Soukal. Searching for the stego key. In *Security, Steganography, and Watermarking of Multimedia Contents VI*, volume 5306 of *Proc. SPIE*, pages 70–82, 2004.

[8] J. Fridrich and D. Soukal. Matrix embedding for large payloads. *IEEE Transactions on Information Forensics and Security*, 1(3):390–394, 2006.

[9] W. Hoeffding. Probability inequalities for sums of bounded random variables. *Journal of the American Statistical Association*, 58(301):13–30, 1963.

[10] A. Ker. Batch steganography and pooled steganalysis. In *Proc. 8th Information Hiding Workshop*, volume 4437 of *Springer LNCS*, pages 265–281, 2006.

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

[12] A. Ker. Locating steganographic payload via WS residuals. In *Proc. 10th ACM Workshop on Multimedia and Security*, pages 27–31, 2008.

[13] A. Ker. Steganographic strategies for a square distortion function. In *Security, Forensics, Steganography and Watermarking of Multimedia Contents X*, volume 6819 of *Proc. SPIE*, pages 0401–0413, 2008.

[14] A. Ker. Estimating Steganographic Fisher Information in real images. To appear in *Proc. 11th Information Hiding Workshop*, 2009.

[15] A. Ker and I. Lubenko. Feature reduction and payload location with WAM steganalysis. In *Media Forensics and Security XI*, volume 7254 of *Proc. SPIE*, pages 0A01–0A13, 2009.

[16] C. Shannon. Communication theory of secrecy systems. *Bell System Technical Journal*, 28:656–715, 1949.

APPENDIX

A. EMBEDDING PROBABILITY LEMMA

We establish the key property of conditional embedding probabilities, needed for Theorem 2.

LEMMA 2. *Let e_k be defined as in Theorem 2 and continue to assume that $m/\sqrt{n} \to 0$ as $n \to \infty$. Then there exists a constant C such that, for sufficiently large n, $e_k \leq C\frac{m}{n}$ for all k*

PROOF. First, $e_1 = m/n$, by uniformity of the embedding path selection. For $k \geq 2$,

$$\begin{aligned} e_k &= Q(E_k \mid X_1^{k-1}) \\ &= \frac{Q(E_k)Q(X_1^{k-1} \mid E_k)}{Q(X_1^{k-1})} \\ &= \frac{m}{n}\frac{Q(X_1^{k-1} \mid E_k)}{Q(X_1^{k-1})} \end{aligned}$$

(noting that, unconditionally, $Q(E_k) = \frac{m}{n}$, again by uniformity of the embedding paths).

Let us consider first $Q(X_1^{k-1})$. We can write it as the sum

$$Q(X_1^{k-1} = x_1^{k-1}) = \sum_{i=0}^{m} \pi_i r_i(x_1^{k-1})$$

where π_i is the probability that exactly i of the first $k-1$ locations are used for embedding (given that a total of m out of n are) and $r_i(x_1^{k-1})$ is the probability of observing the sequence x_1^{k-1}, given that exactly i of the first $k-1$ locations are used for embedding. We have

$$\begin{aligned} \pi_i &= \tfrac{m(m-1)\cdots(m-i+1)(n-m)(n-m-1)\cdots(n-m-k+i+2)}{n(n-1)\cdots(n-k+2)} \\ &= \frac{m!(n-m)!(n-k+1)!}{(m-i)!(n-m-k+i+1)!n!} \end{aligned}$$

for $\max(0, k+m-n-1) \leq i \leq \min(k-1, m)$, otherwise $\pi_i = 0$. Also,

$$r_i(x_1^{k-1}) = \sum_{\substack{S \subseteq \{1,\ldots,k-1\}, \\ |S|=i}} \prod_{i \notin S} p(x_i) \prod_{i \in S} q(x_i).$$

Now turn to $Q(X_1^{k-1} \mid E_k)$; this is exactly the same, except that one embedding location is fixed at k and so there remain $m-1$ payload locations to spread amongst $n-1$ cover locations;

$$Q(X_1^{k-1} = x_1^{k-1} \mid E_k) = \sum_{i=0}^{m-1} \pi_i' r_i(x_1^{k-1})$$

where

$$\pi_i' = \frac{(m-1)!(n-m)!(n-k)!}{(m-i-1)!(n-m-k+i+1)!(n-1)!}$$

for $\max(0, k+m-n-1) \le i \le \min(k-1, m-1)$, otherwise $\pi_i = 0$.

We find some bounds on ratios between π_i and π'_i terms, and also between $r_i(-)$ terms. First, cancelling factorials we compute

$$\frac{\pi'_i}{\pi_i} = \frac{(m-i)n}{m(n-k+1)}$$

and

$$\frac{\pi'_i}{\pi_{i+1}} = \frac{(n-m-k+i+2)n}{m(n-k+1)}.$$

Next, recall that $p(x)$ and $q(x)$ are both nonzero so there exists a positive constant c such that, for all x, $p(x) \le cq(x)$. Then consider

$$\begin{aligned} r_{i+1}(x_1^k) &= \sum_{\substack{S \subseteq \{1,\ldots,k\}, \\ |S|=i+1}} \prod_{i \notin S} p(x_i) \prod_{i \in S} q(x_i) \\ &= \sum_{\substack{S' \subseteq \{1,\ldots,k\}, \\ |S'|=i}} \prod_{i \notin S'} p(x_i) \prod_{i \in S'} q(x_i) \sum_{j \notin S'} q(x_j)/p(x_j) \\ &\ge \frac{k-i}{c}\, r_i(x_1^k). \end{aligned}$$

Now we can bound e_k. First, we may assume that n is large enough that $1 \le m \le n/4$. We split into two cases, when $k \le 3n/4$ or $k > 3n/4$.

If $k \le 3n/4$ then $k+m \le n$ so we can be sure that $\pi_0 \ne 0$ and $\pi'_0 \ne 0$. We use

$$\begin{aligned} e_k &= \frac{m}{n} \frac{\sum_{i=0}^{m-1} \pi'_i r_i(x_1^{k-1})}{\sum_{i=0}^{m} \pi_i r_i(x_1^{k-1})} \\ &\le \frac{m}{n} \max_i \frac{\pi'_i}{\pi_i} \\ &= \frac{m}{n} \max_i \frac{(m-i)n}{m(n-k+1)} \\ &= \frac{m}{n-k+1} \\ &\le 4\frac{m}{n} \end{aligned}$$

If $k > 3n/4$ then certainly $k > m$ so we can be sure that $\pi_m \ne 0$ and $\pi'_{m-1} \ne 0$. We use

$$\begin{aligned} e_k &= \frac{m}{n} \frac{\sum_{i=0}^{m-1} \pi'_i r_i(x_1^{k-1})}{\sum_{i=0}^{m} \pi_i r_i(x_1^{k-1})} \\ &\le \frac{m}{n} \max_i \frac{\pi'_i}{\pi_{i+1}} \max_i \frac{r_i(x_1^{k-1})}{r_{i+1}(x_1^{k-1})} \\ &\le \frac{m}{n} \max_i \frac{(n-m-k+i+2)n}{m(n-k+1)} \max_i \frac{c}{k-i} \\ &\le \frac{m}{n} \frac{n}{m} \frac{c}{k-m} \\ &\le 2c\frac{m}{n} \end{aligned}$$

We have determined a constant $C = \max(4, 2c)$ such that, at least for $n \ge 4m$, $e_k \le C\frac{m}{n}$ for all k. $\square$