# The Square Root Law in Stegosystems with Imperfect Information

Andrew D. Ker

Oxford University Computing Laboratory, Parks Road, Oxford OX1 3QD, England
adk@comlab.ox.ac.uk

**Abstract.** Theoretical results about the capacity of stegosystems typically assume that one or both of the adversaries has perfect knowledge of the cover source. So-called perfect steganography is possible if the embedder has this perfect knowledge, and the Square Root Law of capacity applies when the embedder has imperfect knowledge but the detector has perfect knowledge. The epistemology of stegosystems is underdeveloped and these assumptions are sometimes unstated. In this work we consider stegosystems where the detector has imperfect information about the cover source: once the problem is suitably formalized, we show a parallel to the Square Root Law. This answers a question raised by Böhme.

## 1 Introduction

The most important ingredient of a stegosystem is the cover source, and the relevant epistemology – who knows what about the covers – is sometimes neglected in the literature. But knowledge about the covers is fundamental to the problem. It is known that perfect steganography is possible, conveying an amount of payload linear in the cover size with zero risk of detection, as long as the embedder has perfect knowledge of the cover distribution [1]. This is so because the embedder can ensure that all the statistics of the cover source are preserved by their embedding function[1]. On the other hand, we have a completely different situation when the embedder does not know, and fails to preserve, all the statistics of the covers: this leads to the Square Root Law [3–5] stating that the information conveyed can grow at most of order $\sqrt{n}$, where $n$ represents the size of the cover, or face eventual certain detection.

The latter result assumes that the *detector* has perfect knowledge of the distribution of covers, so to compare potential stegotexts against it. This attack model is defensible: taking the role of the embedder, it is conservative to assume that the opponent has perfect knowledge, since we cannot then make the error of underestimating them. However, it is arguable that for steganography in real objects, be they digital images, audio, or even plain text, no perfect model of the source exists [6] and therefore the opponent cannot have such perfect knowledge

[1] In fact, perfectly secure steganography remains theoretically possible even if the embedder does not know about the cover source, as long as they have an inexhaustible cover supply from which to sample [2], but such embedding is practically infeasible.

any more than can the embedder. In practice, steganalysts create models of cover objects by examining finitely many examples of genuine covers, which leads to an imperfect level of knowledge about the source. What is the consequence of this imperfect knowledge for a detector, and can the capacity rule be exceeded? This question was raised by Böhme [7] who pointed out

> "The square root law is supported with evidence for fixed cover models of the adversary... so far it does not anticipate adversaries who refine their cover models adaptively."

Modifying the square root law to this circumstance, in the simplest and most abstract setting, is our goal in this paper.

In Sect. 2 we describe and justify our framework for a stegosystem with imperfect information, and in Sect. 3 state and prove a modified square root law for this situation. There are new statistical challenges in this setting and we require a discussion of the statistical property of *unbiasedness*, along with some analysis. The significance of the theorem is discussed in Sect. 4: we shall see that the embedder remains limited to payloads of order $\sqrt{n}$ as long as the detector has access to a linear number of examples of covers from which to learn about their distribution. A superlinear number of covers conveys no asymptotic advantage but a sublinear number of covers allows the embedder to improve on the order of capacity $\sqrt{n}$. We briefly extend the result to the case when the embedder also learns about the cover source from examples, and adapts their embedding accordingly. We also discuss many avenues for further research.

We will use the following notation: $f(n) = O(g(n))$ indicates that $f$ grows asymptotically no faster than $g$, i.e. $|f(n)| \leq c|g(n)|$ for $n \geq N$, for some $c$ and $N$; $f(n) \sim g(n)$ means that $f$ and $g$ are asymptotically equal, i.e. $f(n)/g(n) \to 1$ as $n \to \infty$; $f(n) = o(g(n))$ means that $f$ grows strictly slower than $g$, i.e. $f(n)/g(n) \to 0$ as $n \to \infty$. $\Omega$ is the converse to $O$: $f(n) = \Omega(g(n))$ if $g(n) = O(f(n))$. We will use Knuth's notation [8] for the falling Pochhammer symbol $n^{\underline{k}} = n(n-1)\cdots(n-k+1)$, so that the binomial coefficients are $\binom{n}{k} = n^{\underline{k}}/k!$. $\mathrm{E}[X]$ and $\mathrm{Var}[X]$ will denote the mean and variance of the random variable $X$, and $X \sim \mathrm{Bi}(n,p)$ that $X$ takes the binomial distribution with size parameter $n$ and probability parameter $p$.

## 2 Stegosystems with Imperfect Information

In a stegosystem with perfect information, both parties know the exact distribution of the covers. In this scenario it has already been shown that the embedder can communicate completely undetectably at a nonzero (linear) rate by preserving all the statistics of the covers [1]. However, no such system, for embedding in real-world cover media, has ever been constructed. That is because of the distinction, made clear by Böhme in [6,7], between *artificial* covers – mathematical structures such as random variables or Markov chains – and *empirical* covers which are

> "digital representations of parts of reality... They have to be obtained through observation of reality... All kinds of digitised media data, such as images, audio or video files, belong to this class."

Böhme argues that the true distribution of real-world covers is incognisable. Given this, we should focus on *stegosystems with imperfect information* where the distribution can only be obtained approximately, by observation of the source.

It is consideration of imperfect embedding, where some statistic of the cover source is not preserved by embedding, that leads to the Square Root Law of capacity [3–5]. As the cover size increases the embedding rate must diminish, lest evidence in abnormal statistics build to eventual certain detection. But these results still require the detector to have perfect knowledge of the cover source, arguably an equally unrealistic situation to that of perfect embedding. Böhme comments that

> "security in empirical covers depends on the steganalyst's knowledge of the cover source and the amount of evidence available to distinguish any abnormality"

and we aim to quantify how secure capacity – the maximum payload which cannot be detected with nontrivial error rates – depends on this steganalyst's knowledge and the size of the stego object.

Proving a theoretical result about capacity requires us to remain in the world of artificial covers, and in Sect. 3 we will use the very simplest mathematical model, of i.i.d. binary sequences, but we can reflect something of the incognisability of cover distributions by saying that the exact value of the parameter(s) of the model are not known to either party. This is our model for a stegosystem with imperfect information. To an extent, this contradicts Kerckhoffs' Principle: we are assuming that the "enemy" (the steganalyst) does not have full knowledge of the "system", if we take the system to include the covers; in view of the preceding discussion, this demonstrates that Kerckhoffs' Principle should not be applied blindly to the steganography problem. The covers are, in reality, external to the system.

For our result, most of the assumptions apart from knowledge of the covers will not be relevant: it will not matter whether the detector knows the embedding algorithm or the size of embedded payload, though it is important that they not know which locations in the cover have been used for payload. The latter information is part of the secret embedding key, but for more on this assumption see [5].

There remains the question of how we measure the embedder's and detector's uncertainty about the parameter(s) of the artificial cover model. As in the classical square root law setting, we will assume that the embedder is not adaptive and proceeds with some fixed method which replaces symbols in the cover by symbols with a non-identical distribution. For the detector, we will be guided by the practice of steganalysis, in which researchers most commonly train a detector using sets of genuine covers. We will assume that the detector has access to some genuine covers from an independent but identically-distributed source

to that of the embedder, i.e. limited access to a cover oracle. It will turn out (Subsect. 4.1) that any finite limit on the size or number of covers leads to a trivial situation, so we will suppose a limit of $m_n$ accesses to the oracle when the cover size is $n$. (Briefly, in Subsect. 4.3, we will consider the case when the embedder also makes use of a cover oracle. The result is rather different.)

We should ask whether it is plausible that the detector has access to the embedder's cover source. Of course it depends on the application, but for example one could imagine that a well-motivated steganalyst can at least determine forensically the model of camera used to take cover images, and then purchase their own. Or, once the steganographer comes under suspicion, seize the camera itself. A similar detector model, using limited access to a cover oracle, is found in [9], although no capacity result is proved. The limitation in that work, inspired by computational complexity bounds, is of polynomially-many accesses to the oracle; in our model we shall see (Subsect. 4.1) that linearly-many accesses suffice to recover a square root law.

Taking the role of the embedder, it is conservative to assume that the opponent has perfect knowledge of the cover source. It is also rather pessimistic, and the work in this paper is motivated by the desire to relax the condition. We, the embedder, may feel happier about transmitting information at a rate *above* the capacity predicted by the square root law if we could, for example, prove that our opponent needed to spend an exponential amount of time learning about the cover source in order to catch us. Sadly for steganographers, it will turn out that this is not the case.

## 3 The Square Root Law for Imperfect Detectors

To explain the difficulties in constructing a modified square root law for imperfect detectors, and to contrast with new results, we begin by restating the simplest square root law for the classical system when the detector has perfect knowledge of the cover source.

**Theorem 1.** *Suppose that cover objects consist of sequences of symbols which are Bernoulli random variables with parameter $p$; we exclude the pathological cases $p = 0, 1$. Suppose that the embedder replaces the cover symbols, independently with probability $\gamma$, by stego symbols which are Bernoulli with parameter $q \neq p$. The steganalyst wishes to distinguish the cases $\gamma = 0$ and $\gamma > 0$. Let $n$ be size of the cover, i.e. the number of symbols.*

*(1) If $\gamma\sqrt{n} \to \infty$ as $n \to \infty$ then, for sufficiently large $n$, the steganalyst can create an arbitrarily accurate detector.*

*(2) If $\gamma\sqrt{n} \to 0$ as $n \to \infty$ then, for sufficiently large $n$, every detector has arbitrarily low accuracy.*

We interpret the theorem to mean that $\gamma = O(1/\sqrt{n})$ is the critical rate: unless $\gamma$ diminishes at least this fast, large enough $n$ will result in certain detection. If the embedding involves a simple substitution (no source coding by

the embedder) then the payload size is $\gamma n$ and this must grow slower than $\sqrt{n}$, hence the name Square Root Law.

Theorem 1 has been known for a few years but the first published proof is found in [5]. We briefly sketch the techniques used, for comparison with the result about imperfect stegosystems below. For (1), the simple detector which compares the proportion of observed "1" symbols with $p$, and rejects the null hypothesis $\gamma = 0$ if the difference is significant, can be analyzed using tail inequalities for the binomial distribution. It can be shown that this detector has false positive and negative rates which tend to zero as $n \to \infty$ as long as $\gamma\sqrt{n} \to \infty$. For (2), the Kullback-Leibler divergence between the distribution of the observations when $\gamma = 0$ and $\gamma = \gamma_1 > 0$ is computed: it can be shown that this tends to zero as long as $\gamma_1\sqrt{n} \to 0$, which means that any hypothesis test for $\gamma = 0$ against $\gamma > 0$ must have power tending to size (i.e. the error rate tends to that of a purely random decision).

Some of the apparent limitations in this result can be avoided. First, we have stated it here in the context of binary alphabets, but this is not essential and the same is true for arbitrary finite alphabets [5], for a reason we shall discuss in Subsect. 4.2. Second, the symbols in the cover must be independent. This is a severe condition not likely to be satisfied by real cover media. However, the result still holds when there is non-pathological dependence between the symbols [4]. A square root relationship between cover size and capacity has been verified empirically, for contemporary steganography and steganalysis methods, in [3]. Finally, the theorem does not address the critical case when $\gamma = c/\sqrt{n}$, in which case the value of $c$ determines a maximum possible detection accuracy. This is arguably the most important situation because it gives the embedder a genuine "secure" capacity for their embedding; this is related to Steganographic Fisher Information which has recently been examined in [10, 11]. One other minor limitation is that formalising the embedding as affecting each symbol independently with probability $\gamma$ is not quite right for embedding a fixed-length message; this is addressed in [5].

However there is one limitation that remains: the opponent is assumed to have knowledge of $p$. They do not need to know $q$ or $\gamma$: the detector constructed for (1) does not depend on these quantities, and the bound in (2) applies even if we grant the detector this knowledge anyway. But if we wish to prove a similar result for a stegosystem with imperfect information, as described in Sect. 2, the detector must not know $p$, instead learning something about its value through a limited number of observations of true cover bits from an oracle.

There is no great difficulty in adapting part (1) of Th. 1 to such a situation. But part (2) is simply untrue, because there *does* exist a detector which distinguishes cover and stego objects, with the same condition on $\gamma$ as in Th. 1 (2) and without using the oracle at all: it is simply the detector which makes use of a fixed value of $p$ "hardwired" into it, in the case when it just happens to have the correct $p$. It is difficult to use Kullback-Leibler divergence to bound the performance of a detector subject to the constraint that the detector does not

know a certain piece of information: an accurate detector *does* exist, even if it requires a very lucky guess to pick the right value of $p$ to begin with.

Inspired by the idea of a "lucky guess," we tried to model the true value of $p$ as random, uniformly on $[0, 1]$: placing a uniform prior on unknown parameters seems a reasonable way to reflect absence of information about them. Unfortunately this model does not function as desired because the probability of observing a "1" in stego sequences is $p + \gamma(q - p)$ which is *not* uniformly distributed: it is biased towards either $q$ (if $q$ is known) or $1/2$ (if $q$ is also given a uniform prior). Computing the KL divergence between the cover and stego sequences when $p$ has a uniform prior is algebraically challenging because, unconditionally, the bits are no longer independent; the calculations are far too long to include here but the author has outlined them in a technical report [12]. It turns out that this KL divergence is always positive, even when the cover oracle is completely absent – but the detector was supposed to have no knowledge of the cover source! This means that the stego signal leaks information about the presence of payload even when the cover oracle is disregarded. Sadly, placing a uniform prior on $p$ did not properly reflect a lack of knowledge of $p$, and, crucially, did not allow us to use Kullback-Leibler divergence to bound the accuracy of ignorant detectors. For more details of this argument, see [12].

Instead, we seek to rule out those detectors which have knowledge of $p$, using the statistical concept of unbiasedness. A test for the null hypothesis class $H_0 : \theta \in \Theta_0$ against an alternative class $H_1 : \theta \in \Theta_1$ is called *unbiased* if, whenever the true value of $\theta$ is in $\Theta_1$, the probability of rejection of $H_0$ is never less than the probability of rejection of $H_0$ when $\theta \in \Theta_0$. In the language of detectors, this is to say that the probability of a true positive is always at least the probability of a false positive.

There is comprehensive information about the theory and application of unbiased hypothesis tests in [13, Chs. 4-5]: it is popular to restrict attention to unbiased tests for many reasons, including the existence of uniformly most powerful (UMP) tests within this class when general UMP tests do not exist. In the application we consider in this paper, we can justify restricting attention to unbiased detectors for two reasons. First, any detector with more false positives than true positives is not going to be much use in practice. Second, forbidding a test biased towards any particular value of $p$ reflects a lack of knowledge of $p$, without even placing a prior distribution on it, which was exactly our aim. Having restricted to unbiased tests, the detector with a "hardwired" value of $p$ is inadmissible because it is too likely to give a false negative when the true value of $q$ is actually the hardwired value of $p$. We can then make use of literature on UMP unbiased tests for exponential families to prove a modified square root law. It links secure embedding rates with both the cover size and the level of imperfect cover information at the detector, defined in terms of a number of true cover samples which the detector receives from an oracle.

We will retain the simplicity of Theorem 1 in that the covers will still be independent bit strings; when the detector learns about the cover source through

$m_n$ bits from a cover oracle (we would expect that $m_n$ is an increasing function of $n$), and if restricted to unbiased detectors, we then have the following result.

**Theorem 2.** *Suppose that cover objects consist of sequences of symbols which are Bernoulli random variables with parameter $p$; we exclude the pathological cases $p = 0, 1$. Suppose that the embedder replaces a randomly-selected proportion $\gamma$ of the cover symbols by stego symbols which are Bernoulli with parameter $q \neq p$. Let $n$ be size of the cover, i.e. the number of symbols.*

*The steganalyst wishes to distinguish the cases $\gamma = 0$ and $\gamma > 0$, but they have no knowledge of $p$ or $q$, instead they have access to $m_n$ independently-generated cover symbols from which they may learn about the cover source.*

*(1) If*

$$\frac{\gamma}{\sqrt{\frac{1}{m_n} + \frac{1}{n}}} \to \infty$$

*as $n \to \infty$ then, for sufficiently large $n$, the steganalyst can create an arbitrarily accurate detector.*

*(2) If*

$$\frac{\gamma}{\sqrt{\frac{1}{m_n} + \frac{1}{n}}} \to 0$$

*as $n \to \infty$ then, for sufficiently large $n$, every* unbiased *detector has arbitrarily low accuracy.*

The proof follows. We will interpret the result in Subsect. 4.1, and consider its limitations, paralleling those of Th. 1, in Subsect. 4.2.

### 3.1 Proof of Theorem 2 (1)

This half of the proof is the easier, using techniques similar to that of Th. 1 (1). Write $X$ for the number of 1 bits in the cover stream, and $Y$ for the number in the object to be classified. Then $X \sim \mathrm{Bi}(m_n, p)$ and $Y \sim \mathrm{Bi}(n, p + \gamma(q - p))$; under the null hypothesis that the object is a cover, $\gamma = 0$, otherwise $\gamma > 0$.

A suitably asymptotically powerful detector can be constructed, without knowledge of $p$, $q$, or $\gamma$, using the obvious statistic which measures the difference in proportion of ones between the unknown signal and the cover bits:

$$T = \frac{X}{m_n} - \frac{Y}{n} \tag{1}$$

and we will reject the null hypothesis, giving a positive detection of steganography, if $|T|$ exceeds a critical threshold $c\sqrt{\frac{1}{m_n} + \frac{1}{n}}$ for some positive constant $c$.

The variance of a binomial distribution $\mathrm{Bi}(k, p)$ is bounded by $\frac{1}{4}k$, regardless of $p$, so

$$\mathrm{Var}[T] \leq \frac{1}{4}\left(\frac{1}{m_n} + \frac{1}{n}\right).$$

Under the null hypothesis, when the unknown object is a cover, $\mathrm{E}[T] = 0$, so the probability of a false positive $\alpha$ satisfies

$$\alpha = \Pr\left(\left|T - \mathrm{E}[T]\right| > c\sqrt{\tfrac{1}{m_n} + \tfrac{1}{n}}\right) \leq \frac{\mathrm{Var}[T]}{c^2\left(\frac{1}{m_n} + \frac{1}{n}\right)} \leq \frac{1}{4c^2}$$

(the first inequality is Chebyshev's). This can be made arbitrarily small by suitable choice of $c$.

Under the alternative hypothesis, when the unknown object is a stego object, $\mathrm{E}[T] = \gamma r$ where $r = p - q \neq 0$, so the probability of missed detection $\beta$ satisfies

$$\begin{aligned}
\beta &= \Pr\left(\left|T - \mathrm{E}[T] + \gamma r\right| \leq c\sqrt{\tfrac{1}{m_n} + \tfrac{1}{n}}\right) \\
&\leq \Pr\left(\left|T - \mathrm{E}[T]\right| > \left|\gamma|r| - c\sqrt{\tfrac{1}{m_n} + \tfrac{1}{n}}\right|\right) \\
&\leq \frac{\frac{1}{4}\left(\frac{1}{m_n} + \frac{1}{n}\right)}{\gamma^2 r^2 - 2\gamma|r|c\sqrt{\frac{1}{m_n} + \frac{1}{n}} + c^2\left(\frac{1}{m_n} + \frac{1}{n}\right)} \\
&\to 0
\end{aligned}$$

as $n \to \infty$, for any positive $c$, because of the assumption that $\frac{\gamma}{\sqrt{\frac{1}{m_n} + \frac{1}{n}}} \to \infty$ and again using Chebyshev's inequality. We have shown that, for sufficiently large $n$, $T$ distinguishes cover and stego objects with arbitrarily high accuracy.

### 3.2 Proof of Theorem 2 (2)

For simplicity we will omit the subscript in $m_n$. The condition

$$\frac{\gamma}{\sqrt{\frac{1}{m_n} + \frac{1}{n}}} \to 0$$

is equivalent to

$$\gamma^2 \frac{mn}{m+n} \to 0. \tag{2}$$

Let $X$ and $Y$ be as in the previous subsection. A test of whether $\gamma = 0$ or $\gamma > 0$ amounts to a hypothesis test for whether the probability parameter, in the two binomial distributions $X \sim \mathrm{Bi}(m, p)$ and $Y \sim \mathrm{Bi}(n, p + \gamma(q - p))$, is equal or not. The problem of comparing two binomials has been studied in the statistics literature and an optimal (UMP) unbiased test can be found in [13]. To summarise the result in [13, §4.5], the optimal unbiased critical region (i.e. values for which we reject the null hypothesis and give a positive detection) is given by $Y \notin [C_1(X+Y), C_2(X+Y)]$ , where $C_1$ and $C_2$ are functions which, for each value of $X+Y$, give the desired size (lead to the desired false positive rate). We can imagine a suite of detectors, one for each observed value of $X+Y$, each optimal in the sense of the Neyman-Pearson Lemma and giving a positive result

for too-large or too-small values of $Y$. As long as these detectors all have the same false positive probability then collectively they form an optimal unbiased detector: for any given false positive rate, the false negative rate will be minimal amongst all unbiased detectors.

It is simple to verify that $Y \notin [c_1, c_2]$, for constants $c_1$ and $c_2$, is the optimal unbiased detector, given that $X + Y = t$ for some fixed $t$, by using the result in [13, §4.2]. But the apparently simple conclusion is quite deep because, in general, one cannot expect that optimal conditional tests will together form an optimal unconditional test. It holds in this case because the joint distribution of $X$ and $Y$ forms an exponential family with $p$ as the nuisance parameter, and by applying Th. 4.4.1 from [13]. Moreover, this optimal unbiased detector is not easy to construct in practice because all the conditional tests must have *exactly* the same false positive rate for their combination to be optimal, yet the underlying distribution is discrete and unlikely to oblige with such exact probabilities; we may require randomisation at the detector.

Thankfully, none of these details need concern us. We now know that, for the lowest possible false negative rate at any given false positive rate, the optimal unbiased detector depends only on the value of $Y$, conditional on $X + Y$. So we can analyze its behaviour using Kullback-Leibler divergence. All we need to do is to show that the KL divergence, between the cases $\gamma = 0$ and otherwise, tends to zero under assumption (2), although the analysis is rather tricky.

Let $P(y;\gamma)$ be the conditional distribution of $Y$, given $X + Y = t$. We have

$$P(y;\gamma) = \frac{\Pr(Y = y \wedge X = t - y)}{\sum_{y'=0}^{t} \Pr(Y = y' \wedge X = t - y')} = \frac{\rho(\gamma)^y \binom{m}{t-y}\binom{n}{y}}{\sum_{y'=0}^{t} \rho(\gamma)^{y'} \binom{m}{t-y'}\binom{n}{y'}}$$

where $\rho(\gamma)$ is the odds ratio between the events in $Y$ and $X$,

$$\rho(\gamma) = \frac{\big(p + \gamma(q-p)\big)/\big(1 - p - \gamma(q-p)\big)}{p/(1-p)} = \frac{1 + \gamma(\frac{q-p}{p})}{1 - \gamma(\frac{q-p}{1-p})}.$$

Without loss of generality we may assume that $q > p$, so that $\rho(\gamma) \geq 1$ with equality at $\gamma = 0$. Furthermore, $\rho'(\gamma) = (q-p)(1-p)/p\big(1 - p - \gamma(q-p)\big)^2$, a bounded function, so by the mean value theorem

$$0 \leq \rho(\gamma) - 1 \leq C\gamma, \tag{3}$$

where $C$ is a positive constant. When $\gamma = 0$, the conditional distribution of $Y$ is hypergeometric, therefore

$$\frac{P(y;0)}{P(y;\gamma)} = \frac{\sum_{y'=0}^{t} \rho(\gamma)^{y'} \binom{m}{t-y'}\binom{n}{y'} / \binom{m+n}{t}}{\rho(\gamma)^y};$$

where the numerator is the expectation of $\rho(\gamma)^Y$ when $Y$ has a hypergeometric distribution. Taking the expectation over a random variable $Y$ with this distri-

bution, we have

$$\begin{aligned}
0 &\le D_{\mathrm{KL}}(P(y;0) \,\|\, P(y;\gamma)) \\
&= \mathrm{E}_Y\left[\log\left(\frac{P(Y;0)}{P(Y;\gamma)}\right)\right] \\
&= \mathrm{E}\Big[\log\Big(\mathrm{E}\big[\rho(\gamma)^Y\big]\Big) - \log\rho(\gamma)^Y\Big] \\
&= \log\Big(\mathrm{E}\big[(1+\rho(\gamma)-1)^Y\big]\Big) - \mathrm{E}[Y]\log\rho(\gamma) \\
&\overset{(a)}{=} \log\left(\sum_{k=0}^{n}(\rho(\gamma)-1)^k\frac{\mathrm{E}[Y^{\underline{k}}]}{k!}\right) - \mathrm{E}[Y]\log\rho(\gamma) \\
&\overset{(b)}{=} \log\left(\sum_{k=0}^{n}(\rho(\gamma)-1)^k\frac{n^{\underline{k}}\,t^{\underline{k}}}{(m+n)^{\underline{k}}\,k!}\right) - \frac{nt}{m+n}\log\rho(\gamma) \\
&\overset{(c)}{\le} \log\left(\sum_{k=0}^{t}\Big(\tfrac{n}{m+n}(\rho(\gamma)-1)\Big)^k\frac{t^{\underline{k}}}{k!}\right) - \frac{nt}{m+n}\log\rho(\gamma) \\
&\overset{(d)}{=} t\left(\log\Big(1+\tfrac{n}{m+n}(\rho(\gamma)-1)\Big) - \frac{n}{m+n}\log\rho(\gamma)\right) \\
&\overset{(e)}{\le} \frac{tnm}{2(n+m)^2}(\rho(\gamma)-1)^2 \\
&\overset{(f)}{\le} \frac{nm}{2(n+m)}C^2\gamma^2 \\
&\to 0
\end{aligned}$$

We justify the steps as follows. (a) is the binomial expansion; note that we may take the sum to $n$ because $Y$ is an integer, so any summand with $k > Y$ will be zero. (b) is from the following property of the hypergeometric distribution:

$$\begin{aligned}
\mathrm{E}[Y^{\underline{k}}] &= \sum_{i=k}^{n} i^{\underline{k}}\tbinom{n}{i}\tbinom{m}{t-i}/\tbinom{m+n}{t} = \sum_{i=k}^{n} n^{\underline{k}}\tbinom{n-k}{i-k}\tbinom{m}{(t-k)-(i-k)}t^{\underline{k}}/(m+n)^{\underline{k}}\tbinom{m+n-k}{t-k} \\
&= \frac{n^{\underline{k}}\,t^{\underline{k}}}{(m+n)^{\underline{k}}}.
\end{aligned}$$

(c) is because $n^{\underline{k}}/(m+n)^{\underline{k}} \le (n/(m+n))^k$; we may take the sum to $t$ because the summands are zero when $k > n$ or $k > t$. (d) is another binomial expansion. (e) follows from

$$\log(1+ax) - a\log(1+x) \le \tfrac{1}{2}a(1-a)x^2$$

for $a \in [0,1]$ and $x \ge 0$, which may easily be verified by differentiating the difference. (f) is because $t \le m+n$, and using (3). The final limit is just (2).

Since the Kullback-Leibler divergence tends to zero, the information processing theorem [14] implies that the performance of any decision based on $Y$ given $X+Y=t$ must tend to that of a purely random decision. This is true whatever the value of $t$ so the same holds for the collection of detectors, one for each value of $X+Y$, which together form the optimal unbiased detector: no other unbiased detector can do better. Therefore every unbiased detector has asymptotically random output: in the language of Cachin [15], the system is $\epsilon$-secure, for arbitrarily small $\epsilon$, if $n$ is large enough; in terms of detectors, the false positive probability $\alpha$ tends to $1-\beta$, the true positive probability.

## 4 Discussion

We conclude with a discussion of the significance of the result as regards secure payload size, its limitations, how it might be extended, and briefly consider the case when the embedder also learns about the cover source through an oracle.

### 4.1 Interpretation

We consider the consequences of Th. 2 as regards secure embedding capacities. Let us first assume that the embedder requires a fixed average number of changes per bit of payload conveyed (for example in the simplest case of overwriting pseudorandomly-selected locations in the cover with payload bits, whether or not the payload is compressed prior to embedding). Then the payload transmitted $M$ is proportional to $\gamma n$ and we can deduce the following corollaries to Th. 2.

**Corollary 1.** *If $m_n = \Omega(n)$ then the situation, up to asymptotic order, is the same as in the perfect knowledge case: $M = o(\sqrt{n})$ for asymptotically perfect security.*

The classical square root exponent cannot be reduced, no matter how large $m_n$. We know this because, of course, the classical Square Root Law tells us that the square root cannot be beaten even when the detector has perfect knowledge of $p$. We also now know that a linear number of accesses to the cover oracle is sufficient for the detector: more accesses might reduce the secure capacity by a constant multiple, but do not affect its order of growth.

**Corollary 2.** *If $m_n = o(n)$ then $M = o(n/\sqrt{m_n})$ and the classical square root rate can be beaten. For example, if $m_n \sim n^e$ with $e < 1$, then $M = o(n^{1-2e})$ for asymptotically perfect security.*

This tells us that a linear number of cover examples is necessary for the detector, otherwise their knowledge about the cover source grows too slowly and their opponent can exceed the square root law. In particular,

**Corollary 3.** *If $m_n = O(1)$ then the embedder can achieve any sublinear payload rate $M = o(n)$ with asymptotically perfect security.*

In the case when the detector only has finitely much information about the cover source, i.e. their information does not grow with the cover size, they will always have some uncertainty about the true value of $p$. By decreasing the embedding rate, no matter how slowly, the embedder can ensure that the perturbation to the frequency of 0s and 1s falls within this uncertainty.

In fact, of course, all that is required is for $\gamma$ to be sufficiently small to ensure that the risk of detection is sufficiently low. Quantifying this, over and above the asymptotic relationship, is a problem related to Steganographic Fisher Information (SFI) [10, 11], and a direction for future research. It should not be infeasible to estimate the SFI and hence find a concrete capacity bound, in terms of $m_n$ and $p$, given a KL divergence limit on the risk of detection.

Finally, we turn to the case when the steganographer uses not a simple substitution but an adaptive source-coding at the embedding stage, e.g. matrix embedding [16]. Effectively, the location of the changes, as well as their content, conveys information to the recipient, allowing the payload transmitted to be (slightly) superlinear in $\gamma$.

**Corollary 4.** *With optimal source coding, the payload size $M$ is bounded by $M = O(\gamma n \log(1/\gamma))$, which is achievable (asymptotically) using simple syndromes from the Hamming code.*

*If $m_n = \Omega(n)$ then $M = o(\sqrt{n} \log n)$, otherwise $M = o(\log m_n n/\sqrt{m_n})$, for asymptotic perfect security.*

Care must be taken to ensure that the embedding locations remain unpredictable by the detector, and that the code does not introduce predictable dependencies between the symbols embedded. This could be achieved by using a random codebook, but the secret key parameterising it might need to be large (note a parallel result in [5]). The equivalent problem for the perfect knowledge detector is examined in [17], where it is shown that the dependencies introduced by a certain type of matrix embedding do not grant, asymptotically, any extra evidence. It is for further work to consider the imperfect information case.

### 4.2 Limitations and Extensions

Theorem 2 shares the limitations of Th. 1: the version proved applies only to an abstract mathematical structure which is not a good model for any realistic digital media.

One limitation can be removed immediately: we can extend the binary i.i.d. case to any finite alphabet i.i.d. case by the following argument. If the alphabet is $\Sigma = \{z_1, \ldots, z_N\}$ then we can consider the $N$ binary hypothesis tests which, for each $i$, count only $z_i$ against all other symbols. If stego objects perturb the frequency of any one of those symbols, one of the hypothesis tests has asymptotically negligible error, and the combination of $N$ hypothesis tests (with the rule that rejecting one leads to rejecting the ensemble) will have asymptotic error a factor of $N$ higher, but still tending to zero. For the converse, it can be shown that whenever the KL divergence between all pairs tends to zero, so does the KL divergence of the entire set.

The most obvious avenue for future research is to remove the i.i.d. condition, for example extending to Markov chains as in [4]. But an involved analysis was required to prove that result, and the difficulty will be compounded by existence of a cover oracle. And we would prefer to go even further, to properly two dimensional objects with arbitrary correlation structures. Perhaps the method just described, which lifts results from the binary to arbitrary finite symbol case, can provide a simplification. Whether the technique we used to prove the main result of this paper can be adapted to more complex cover models depends on whether a sufficiently simple UMP unbiased detector exists. Some of our current research indicates that a square root law can hold even in the case of nonstationary cover sources (perhaps contrary to intuition), and there may be further extensions for an imperfect knowledge detector.

Another direction for further research is to compute concrete capacities in the case when $\gamma \sim c\sqrt{\frac{1}{m_n} + \frac{1}{n}}$. The definition of capacity, in the context of a restriction to unbiased detectors, needs some care, but in principle it should not be too difficult to compute a Steganographic Fisher Information quantity for the model of Th. 2.

It is, of course, important that the steganalyst's cover oracle matches exactly the distribution of the covers used by the embedder. Even the smallest deviation means that the detector will give false positive results with asymptotic probability one. And notice that the detector constructed in Subsect. 3.1 does not need to know $q$. This is analogous to a steganalyst who is not sure of the embedding method used by their adversary. They require that $p \neq q$ – the embedding method does change the distribution of covers – and if not then their detector will produce asymptotically no false positives, but no true positives either. A further extension to the information model is considered next.

### 4.3 Embedding with Learning

Our imperfect information stegosystem was not "fair" to the steganographer: we assumed that the embedding method was fixed, causing an alteration to the bit probabilities in a stego object, whereas we allowed the detector access to a cover oracle to learn about the distribution of covers. Briefly, we examine the situation when both embedder and detector have access to cover oracles, and both adapt their behaviour according to the information they learn.

First, how does the embedder make use of the oracle? Assuming that the embedder is still constrained to overwrite symbols in the cover, their optimal behaviour is to estimate the symbol distribution and then adjust their embedding so that the overwritten symbols have the same distribution. We will not concern ourselves with how this is achieved, but note that it is fairly simple, at least in the i.i.d. cover bit scenario, to do using a randomised arithmetic coding (a similar idea appears in [18]).

We have yet to prove a general counterpart to Th. 2, but can make a strong conjecture. Suppose that the embedder has $l_n$ accesses to cover oracle bits, of which $Z$ turn out to take value 1. They estimate $p$ from the ratio $Z/l_n$ – here

$Z \sim \mathrm{Bi}(l_n, p)$ – so that the conditional distribution of $Y$ given $Z$ is $\mathrm{Bi}(n, p + \gamma(Z/l_n - p))$. If $T$ is as in (1), one can compute $\mathrm{Var}[T] \sim l_n(\frac{1}{m_n} + \frac{1}{n})$. On the other hand, the constant $C$ in (3) is proportional to $Z/l_n - p$, so $C^2$ is of average order $1/l_n$, so the KL divergence in Subsect. 3.2 probably tends to zero if $\frac{nm}{(n+m)l_n}\gamma^2 \to 0$ (this is a long way from a proper proof!). Together these suggest:

**Conjecture 3** *In the i.i.d. binary sequence model of Th. 2, in the case when the embedder estimates the distribution of covers from $l_n$ cover symbols generated independently of the detector's information, the critical rate of $\gamma$ is $\sqrt{l_n(\frac{1}{m_n} + \frac{1}{n})}$.*

We mean "critical rate" in the sense that if $\gamma$ exceeds it asymptotically this gives eventual certain detection, and below it gives asymptotic perfect security.

Although the conjecture has yet to be proved, we can deal with a special case, although lack of space prevents the inclusion of the proof here. Since the embedder already has a source of covers – the one they embed in – they can use this for linearly many examples. (Note that, if $l_n = \Omega(n)$, $m_n$ would be irrelevant.) Even accounting for the dependencies thus introduced,

**Theorem 4.** *In the i.i.d. binary sequence model of Th. 2, if the embedder learns from the cover they embed in and replaces cover symbols with bits distributed according to their estimate of $p$, even if the detector has exact knowledge of $p$, the critical rate of $\gamma$ is 1 in the sense that, if $\gamma \to 0$ as $n \to \infty$, the embedding is asymptotically perfectly secure.*

It appears that the embedder "wins" this contest, since they can learn enough information about the covers to keep the detector uncertain, no matter how slowly their embedding rate tends to zero. (Of more interest is the case when $\gamma$ is constant, and the consequential bounds on detection accuracy. This is for further research.) It makes the square root law redundant. But, although it seems fair to allow both embedder and detector to learn about the cover source, this is not what happens in practice: steganalysts constantly refine their cover models, but current steganography algorithms are not sophisticated enough to learn about the cover source[2]. It remains to be seen whether the same results hold in more realistic cover models where dependence between symbols is permitted.

### 4.4 Conclusions

We have proved a result which shows, amongst other things, that linearly many accesses to a cover oracle suffice for the detector to restrict the embedder to a square root capacity law; this is a significantly weaker condition than for the classical square root law for stegosystems with perfect information. We have illustrated the difficulty in reasoning about a *lack* of knowledge on the part of

[2] Although some embedding methods try to preserve statistical properties of the individual cover object used for embedding, they do not use information from any larger sample of covers and so cannot reduce learning error asymptotically.

the detector: placing a uniform prior on the unknown parameters is not the same as having no knowledge. The statistical property of unbiasedness has provided the solution in this case, and the literature on UMP unbiased hypothesis tests has been a useful resource.

We also briefly considered the case when the embedder learns from a cover oracle. Although we omitted the proof of the technical result, we can show that an adaptive embedder who learns from the cover source can come arbitrarily close to a linear law of capacity, even if their opponent is granted complete knowledge of the cover source. All results, however, are in the context of a highly simplified cover model with i.i.d. bits and there remains much further research to expand them to, at least, correlated cover symbols.

What both these results emphasise, however, is that the epistemology of steganography needs more study. Indeed, the situation is more complex than the simple classification of stegosystems into perfect and imperfect information: to say that "the detector needs linearly many cover examples to restrict their opponent to a square root law" is not quite the whole story. What matters is that *the embedder knows that* the detector has, or might have, linearly many cover examples. In a similar vein, for the classical square root law *the detector knows that the embedder does not know* all the statistics of the cover source, and once *the embedder knows that the detector knows this* they are forced to a square root capacity law. In a similar vein, if the detector's cover oracle has a slightly different distribution to the true covers used by the embedder then it is practically useless and a linear law is recovered, but it requires *the embedder to know that the detector has faulty information* about the covers.

These considerations once again illustrate that steganography requires a more subtle information asymmetry than cryptography, and Kerckhoffs' Principle is not suitable to provide the whole framework. We expect that there will be a number of different applications which drive problems all currently named "steganography", with different capacities and solutions, depending on different levels of knowledge amongst the actors.

**Acknowledgements**

The author is a Royal Society University Research Fellow. This line of research was prompted by conversations with Rainer Böhme at the last Information Hiding conference.

## References

1. Wang, Y., Moulin, P.: Perfectly secure steganography: Capacity, error exponents, and code constructions. IEEE Transactions on Information Theory **54**(6) (2008) 2706–2722
2. Hopper, N., Langford, J., von Ahn, L.: Provable secure steganography. In: Proc. of CRYPTO. Volume 2442 of Springer LNCS. (2002) 77–92

3. Ker, A., Pevný, T., Kodovský, J., Fridrich, J.: The square root law of steganographic capacity. In: Proc. 10th ACM Workshop on Multimedia and Security. (2008) 107–116
4. Filler, T., Ker, A., Fridrich, J.: The square root law of steganographic capacity for Markov covers. In: Media Forensics and Security XI. Volume 7254 of Proc. SPIE. (2009) 0801–0811
5. Ker, A.: The Square Root Law requires a linear key. In: Proc. 11th ACM Workshop on Multimedia and Security. (2009) 85–92
6. Böhme, R.: Improved Statistical Steganalysis using Models of Heterogeneous Cover Signals. PhD thesis, Technische Universität Dresden (2008)
7. Böhme, R.: An epistemological approach to steganography. In: Proc. 11th Information Hiding Workshop. Volume 5806 of Springer LNCS. (2009) 15–30
8. Graham, R., Knuth, D., Patashnik, O.: Concrete Mathematics: A Foundation for Computer Science. Addison-Wesley (1994)
9. Katzenbeisser, S., Petitcolas, F.: Defining security in steganographic systems. In: Security and Watermarking of Multimedia Contents IV. Volume 4675 of Proc. SPIE. (2002) 50–56
10. Ker, A.: Estimating steganographic Fisher Information in real images. In: Proc. 11th Information Hiding Workshop. Volume 5806 of Springer LNCS. (2009) 73–88
11. Filler, T., Fridrich, J.: Fisher Information determines capacity of $\epsilon$-secure steganography. In: Proc. 11th Information Hiding Workshop. Volume 5806 of Springer LNCS. (2009) 31–47
12. Ker, A.: The Uniform Prior and Zero Information: A Technical Note. Oxford University Computing Laboratory Research Report CS-RR-10-06 (2010)
13. Lehmann, E., Romano, J.: Testing Statistical Hypotheses. Third edn. Springer (2005)
14. Cover, T., Thomas, J.: Elements of Information Theory. Wiley (1991)
15. Cachin, C.: An information-theoretic model for steganography. Information and Computation **192**(1) (2004) 41–56
16. Fridrich, J., Lisoněk, P., Soukal, D.: On steganographic embedding efficiency. In: Proc. 8th Information Hiding Workshop. Volume 4437 of Springer LNCS. (2008) 282–296
17. Ker, A.: The Square Root Law does not require a linear key. Submitted for publication, 2010
18. Sallee, P.: Model-based methods for steganography and steganalysis. International Journal of Image and Graphics **5**(1) (2005) 167–189