Steganographic Strategies for a Square Distortion Function

Andrew D. Ker

Oxford University Computing Laboratory, Parks Road, Oxford OX1 3QD, England

ABSTRACT

Recent results on the information theory of steganography suggest, and under some conditions prove, that the detectability of payload is proportional to the square of the number of changes caused by the embedding. Assuming that result in general, this paper examines the implications for an embedder when a payload is to be spread amongst multiple cover objects. A number of variants are considered: embedding with and without adaptive source coding, in uniform and nonuniform covers, and embedding in both a fixed number of covers (so-called batch steganography) as well as establishing a covert channel in an infinite stream (sequential steganography, studied here for the first time). The results show that steganographic capacity is sublinear, and strictly asymptotically greater in the case of a fixed batch than an infinite stream. In the former it is possible to describe optimal embedding strategies; in the latter the situation is much more complex, with a continuum of strategies which approach the unachievable asymptotic optimum.

Keywords: Steganographic Capacity, Batch Steganography, Information Theory, Source Coding

1. INTRODUCTION

Although there is now much literature on efficient methods of embedding a secret payload in cover media, and for detecting the hidden content, it is usual to consider only single cover objects. This paper is concerned with embedding in a finite or infinite stream of objects, deriving capacity bounds and optimal methods. The problem of embedding in a fixed number of covers was posed in Ref. 1, where it was called the *batch steganography* problem, and the question is now also extended to infinite streams; we call this *sequential steganography*.

A key assumption, here, will be that the detectability of payload in a single object is proportional to the *square* of the number of changes caused by the embedding. Results of this nature have recently arisen in a number of steganalysis papers.[2–4] Assuming that the same holds in general, we examine the implications for an embedder when a large payload is to be spread amongst multiple cover objects. The choice of how to split payload between multiple covers is called an *embedding strategy* and the aim is to find the optimal strategies implied by the square law. Related work is found in Refs. 1 and 5, where optimal embedding strategies are found but only in the context of highly restricted detection frameworks; in this paper we do not assume knowledge of the steganalyst's behaviour.

The paper is structured thus: in the immediately following subsections we present the problem of batch steganography and sequential steganography, summarise the theory which predicts locally square detectability in many cases, and relate some recent work on source coding which bounds payload size in terms of embedding changes.[6,7] Section 2 then applies the theory to the batch steganography problem and derives optimal embedding strategies and maximum undetectable payload for three embedding scenarios:

(i) simple embedding (no source coding) in uniform covers,

(ii) simple embedding in mixed covers, and

(iii) matrix embedding (a form of source coding) in uniform covers.

We will not consider matrix embedding in nonuniform covers, for which the algebra would be rather untidy. Section 3 considers corresponding versions of the sequential steganography problem; there is no optimal strategy in this case, but bounds and good strategies can be derived. It is shown that the asymptotic payload, as a function of the number of covers, must be strictly lower in the sequential than the batch setting. An illustration of the practical steganographic capacity of large numbers of digital images, implied by recent detectors, is also included. Finally, Section 4 considers the significance of the results and asks whether the mathematical assumptions which produced them are reasonable.

Further author information: E-mail: adk@comlab.ox.ac.uk, Telephone: +44 1865 283530, Fax: +44 1865 273839

1.1. Batch Steganography and Sequential Steganography

It is rather plausible to suppose that a steganographer has access to multiple covers among which the payload can be spread, and that a steganalyst is presented with a large number of objects for steganalysis. In Ref. 1 we formulated the competing aims of *batch steganography*, in which it is assumed that a fixed number N of covers is available to a steganographer who spreads payload amongst some or all of them, and *pooled steganalysis*, in which a steganalyst attempts to pool the evidence of N objects to determine whether some payload is present (without knowing which or how many do contain payload). Only the former will concern us here: we want to determine, subject to some assumptions about accumulation of evidence and a maximum acceptable risk of detection, what the optimal strategies are for the steganographer, and how much payload can be embedded.

We also tackle for the first time a more difficult problem, termed *sequential steganography* in Ref. 1 but not studied there. In the sequential setting we no longer suppose that either the number of covers N or the payload size M is fixed in advance of embedding (this differs materially from the batch problem, because optimal strategies require advance knowledge of M and N). In the sequential setting, we want to establish a strategy for an infinite stream of communications, with transmission of as much payload as possible over time. We will see that, although the steganographer is forced to reduce the payload *rate* over time, an infinite payload can still be transmitted in an infinite amount of time. However it will be shown that there is a tension between transmitting information sooner and transmitting asymptotically faster as $N \to \infty$. Further, we shall see that the steganographer must be asymptotically less efficient in sequential embedding than in the batch setting.

We will not, in this work, ask how the intended recipient of the payload is to recombine the parts extracted from all the objects: we assume that knowledge of the size and order of the payload segments is determined by a secret key already shared between the communicating parties.

1.2. Locally Square Distortion

As in Refs. 1 and 5, we will suppose that the steganalyst is applying some detector to individual objects in the batch or stream of those transmitted by the steganographer, and pooling their evidence in some way. This is plausible because, at present, steganalysis methods only work on individual objects.

The steganalyst wants to decide whether any payload is present: a hypothesis testing scenario. Rather than attempt to describe the complete performance profile of a steganalysis detector for individual objects – even if we could do so, it would very likely lead to intractable optimization problems – we will make use of some statistical theory to simplify the problem. For this work we will assume:

(1) that evidence of the presence of steganography, from multiple objects, is additive, and

(2) that the evidence is, in individual objects, proportional to the *square* of the number of embedding changes made.

Although these assumptions certainly represent a simplified abstraction of the problem, we argue that they do bear some relation to reality, at least as far as asymptotic behaviour (as the number of covers grows) is concerned.

Consider the distribution of cover objects and stego objects; we suppose that there is a sequence of covers to be used, and that the distribution of object i when p_i (randomly located) embedding changes are made is $X_i^{p_i}$. It is important that payload is measured by the number of embedding changes induced: although payload size might seem to be the more natural measure, it is only the changes which can be detected by a steganalyser. But (since it is only a parameter) it will not matter whether p_i indicates the absolute number of embedding changes or whether they are measured as a proportion of the available embedding locations; most of the steganalysis literature tends to take the latter approach, but we will use the former.

In particular, then, (X_i^0) is the sequence of random variables corresponding to the covers. It is convenient to write $\boldsymbol{X} = (X_i^0)$ and $\boldsymbol{X^p} = (X_i^{p_i})$. Any detector – binary classifier for the presence or absence of payload in the sequence as a whole – is deciding whether a sequence of objects is a realisation of $\boldsymbol{X}$ or $\boldsymbol{X^p}$. By the information processing theorem, any detector must have false positive probability α and false negative probability β satisfying

$$\alpha \log \tfrac{\alpha}{1-\beta} + (1-\alpha) \log \tfrac{1-\alpha}{\beta} \leq D_{\mathrm{KL}}(\boldsymbol{X}, \boldsymbol{X^p}),$$

where D_{KL} represents the Kullback-Leibler (KL) divergence. In this sense, the worst-case risk to the steganographer is bounded by $D_{\mathrm{KL}}(\boldsymbol{X}, \boldsymbol{X}^{\boldsymbol{p}})$. Measuring "evidence" by KL divergence is a standard idea, first applied to steganography in Ref. 8 and now widely adopted.

If we assume that the observations of $(X_i^{p_i})$, are independent for fixed (p_i), then

$$D_{\mathrm{KL}}(\boldsymbol{X}, \boldsymbol{X}^{\boldsymbol{p}}) = \sum_i D_{\mathrm{KL}}(X_i^0, X_i^{p_i});$$

in this sense, evidence is additive, justifying (1). As long as the stream of cover objects come from a sensible source (a random selection from an image library, for example, *not* something like consecutive frames from a video camera) it is plausible to assume such independence.

Second, the square distortion assumption (2): a simple observation is that this is exactly true if $X_i \sim \mathrm{N}(p_i, \sigma_i^2)$ for some finite variances σ_i^2. But for other distributions – remember that the X_i are distributions of the whole objects, so are likely to be complex random vectors – we may still make a similar approximation. We appeal to a theorem of Kullback,[9] which says that (under some regularity conditions) KL divergence of a one-parameter family is locally square in perturbations of the parameter. We will not repeat this argument, but refer the reader to Ref. 3. As $p_i \to 0$, $D_{\mathrm{KL}}(X_i^0, X_i^{p_i}) \to Q_i p_i^2$ for some constant Q_i which is called the *Q-factor* for the i-th cover. Furthermore, it is sensible to assume that $p_i \to 0$ as $i \to \infty$ – this too is argued in Ref. 3 on the grounds that embedding at a rate which does not diminish is a surefire way for the steganographer to get caught. Hence, at least eventually, the "evidence" provided by cover i is proportional to p_i^2, although the constant of proportionality depends on the nature of cover i. Indeed it is a consequence of the regularity conditions that $D_{\mathrm{KL}}(X_i^0, X_i^{p_i})$, if always finite and positive for $p_i > 0$, is bounded above and below on $p_i \in [0, 1]$ by multiples of p_i^2.

Of course the assumption (2) does depend on the regularity conditions of Kullback's theorem, but they are satisfied by very many distributions if the parameterization is suitable: the parameter should have an asymptotically linear effect on the distribution it determines. This seems a reasonable property for the effect of embedding changes on a distribution of covers.

1.3. Bounds on Embedding Efficiency

The final component for our analysis is to relate the size of an embedded payload to the number of embedding changes necessary to insert it into a cover. Under a simple embedding scheme such as LSB replacement, on average $\frac{1}{2}$ cover samples must be altered for the embedding of each payload bit. But, when there is excess capacity, we can do better using a source coding method called *Matrix Embedding*. This was first noted by Crandall in an unpublished manuscript, and a good survey can be found in Ref. 6. In this paper, as in the survey, we will assume that the cover objects consist of a number of locations, each of which can store one bit of information. In grayscale images this might correspond to the least significant bit, for example. (Generalizing to the case of q-ary alphabets is not difficult, but we aim to avoid additional complications in this paper.)

We now quote a well-known result about source coding, which is explicitly applied to steganographic embedding in Ref. 7 but has been known for decades[10]:

LEMMA 1. *If a payload of m bits is to be embedded in n cover locations, the number of embedding changes c must satisfy*

$$c \geq nH^{-1}(\tfrac{m}{n}) \tag{1}$$

where $H : (0, \frac{1}{2}] \to (0, 1]$ is the binary entropy function

$$H(x) = -x \log_2 x - (1 - x) \log_2 (1 - x)^*.$$

Furthermore, we now know of coding methods which can get extremely close to this bound, for example the Low Density Generator Matrix (LDGM) codes found in Ref. 7, at least when n is a few thousand in size. We consider this bound to be, in practice, almost achievable and so it will make a good approximation for subsequent analyses.

[*] H is conventionally extended to $H(0) = 0$.

2. EMBEDDING STRATEGIES FOR BATCH STEGANOGRAPHY

We are ready to tackle the batch steganography problem, under the assumption that the approximations in Subsects. 1.2 and 1.3 are exact. The number of covers N is assumed fixed, and let us suppose that the steganographer will embed payload of size m_i in object number i. We write $\boldsymbol{m}$ for the vector (m_i) and call the choice of $\boldsymbol{m}$ an *embedding strategy*. It is useful to write $M = \sum_{i=1}^{N} m_i$ for the total payload size. For simplicity we assume that the m_i can take fractional values, although this will be revisited in Subsect. 3.4. Depending on the embedding scenario –

(i) simple embedding in uniform covers,
(ii) simple embedding in mixed covers,
(iii) matrix embedding in uniform covers, subject to the bound (1)

– an embedding strategy produces a vector of the number of embedding changes in each object $\boldsymbol{p} = (p_i)$, and we can derive different formulae for the steganographer's *risk*, assuming that the risk is well-modelled by the KL divergence $D_{\mathrm{KL}}(\boldsymbol{X}, \boldsymbol{X}^{\boldsymbol{p}})$.

There are two possible optimization problems involving choice of embedding strategy: to maximize the total payload size M subject to a bound on the acceptable risk

$$\textbf{Maximize } \sum m_i \quad \textbf{s.t.} \quad D_{\mathrm{KL}}(\boldsymbol{X}, \boldsymbol{X}^{\boldsymbol{p}}) \leq D,$$

or, for a given payload size M to minimize the risk

$$\textbf{Minimize } D_{\mathrm{KL}}(\boldsymbol{X}, \boldsymbol{X}^{\boldsymbol{p}}) \quad \textbf{s.t.} \quad \sum m_i \geq M.$$

These problems are completely dual, because of the following (very simple) observation.

LEMMA 2. *If $\phi : \mathbb{R}^n \to \mathbb{R}$ and $\psi : \mathbb{R}^n \to \mathbb{R}$ are both strictly increasing in every argument, and the maximization problem*

$$\textbf{Maximize } \phi(x_1, \ldots, x_n) \quad \textbf{s.t.} \quad \psi(x_1, \ldots, x_n) \leq \alpha$$

has solution $\boldsymbol{x}^*$ and objective maximum β, then the dual problem

$$\textbf{Minimize } \psi(x_1, \ldots, x_n) \quad \textbf{s.t.} \quad \phi(x_1, \ldots, x_n) \geq \beta$$

has the same solution $\boldsymbol{x}^*$ and objective minimum α.

(For reasons of space, full proofs are omitted from this paper; a one-line sketch will follow each result. The above lemma is a simple proof by contradiction.)

The result applies to the batch steganography problem because the number of embedding changes, in any cover, is strictly increasing in the payload size, and under the hypotheses of Subsect. 1.2 the total KL divergence is strictly increasing in the number of embedding changes in any object.

We now consider each of the embedding scenarios (i), (ii), (iii), in turn. The following result will be useful.

LEMMA 3. *Let χ be a convex function, and all α_i positive constants. Then the solution to the optimization problem*

$$\textbf{Maximize } \sum m_i \quad \textbf{s.t.} \quad \sum \alpha_i \chi(m_i) \leq D$$

is such that $\alpha_i \chi'(m_i)$ is constant (if such m_i can be found).

(The result is proved by the straightforward application of a Lagrange multiplier.)

2.1. Optimal Batch Embedding (i)

In the simplest case we assume that the covers are *uniform*: not only the same size, but also of the same character so that payload is equally detectable in each of them. More precisely, the steganalyst's detector has the same response distributions to a fixed payload in any of the covers. Further, we assume that the embedding stores a fixed number of payload bits per embedding change (this ratio $e = m_i/p_i$ is called the *embedding efficiency*). It is not necessary that source coding be unused, merely that the code is not adaptive to the payload size M.

Because of uniformity, there is a single Q-factor and, under the assumptions of Subsect. 1.2, the total KL divergence is additive across objects and equal to Qp_i^2 in object i, where p_i is the number of embedding changes. This gives:

THEOREM 4 (UNIFORM COVERS; NO ADAPTIVE CODING). *If all covers have an identical Q-factor Q, and each embedding change transmits e payload bits, the optimization problem is*

$$\textbf{Maximize } \sum m_i \quad \textbf{s.t.} \quad \tfrac{Q}{e^2} \sum m_i^2 \leq D$$

and the solution is

$$m_i = \sqrt{\tfrac{De^2}{NQ}} \text{ for each } i, \quad \text{so that} \quad M = \sqrt{\tfrac{De^2 N}{Q}}.$$

In this case, $M = \Theta(\sqrt{N})$.

(The proof is immediate from Lemma 3.)

Implict in the result is that the payload embedded in each object must be within the objects' capacity; for sufficiently large N this is guaranteed.

Recall that $f(N) = \Theta(g(N))$ means $\exists c > 0, d, N_0 . \forall N \geq N_0, cg(N) \leq f(N) \leq dg(N)$ i.e. $f(N)$ grows asymptotically as fast as $g(N)$. Analogous results, that total capacity grows with the square root of the number of covers, are also found in Refs. 2 and 5; they are compared in Sect. 4.

2.2. Optimal Batch Embedding (ii)

Suppose instead that the covers are not uniform: they might be of different size, or different character affecting the difficulty of steganalysis in each (e.g. some might be more noisy than others). One would expect that larger amounts of payload should be embedded into those covers best able to disguise it, and we seek to analyse the relationship.

It would be enormously valuable to understand the full relationship between cover properties and difficulty of payload detection, but this is known for practically no steganalysis methods; the literature contains some empirical investigations of some factors[11,12] and for one particular steganalysis method an almost-complete answer is derived in Ref. 4 but it is not likely that such questions can be answered in general. However we can model nonuniformity of covers by assuming that each cover has a different steganalysis Q-factor: larger payloads are more readily detectable in those with larger Q-factors.

Still assuming a fixed embedding efficiency, we then have:

THEOREM 5 (NONUNIFORM COVERS; NO ADAPTIVE CODING). *If cover i has Q-factor Q_i, and each embedding change transmits e payload bits, the optimization problem is*

$$\textbf{Maximize } \sum m_i \quad \textbf{s.t.} \quad \tfrac{1}{e^2} \sum Q_i m_i^2 \leq D$$

and the solution is

$$m_i = \sqrt{\tfrac{De^2}{\sum_j Q_j^{-1}}} \, \tfrac{1}{Q_i} \quad \text{so that} \quad M = \sqrt{De^2 \sum Q_i^{-1}}$$

As long as the quantities Q_i are bounded above, and below away from zero (i.e. there are constants c and $\epsilon > 0$ such that $c > Q_i \geq \epsilon$ for all i), $M = \Theta(\sqrt{N})$.

(This is another application of Lemma 3.)

Again, we assume that this gives each m_i less than the capacity of cover i; for sufficiently large N, this is guaranteed.

We see that there is a harmonic relationship between the Q-factor for object i and the payload to be placed in it. Although the multiplicative constant is different – it depends on the harmonic mean of the Q-factors of the covers in the batch – the rate of growth of capacity is not changed by nonuniform covers. This is to be expected. Of course, were Q_i unbounded below, for example if the individual covers became larger and larger whereby $Q_i \to 0$, the same result would not hold, but that does not seem to be a realistic situation.

2.3. Optimal Batch Embedding (iii)

When N is large, the amount of payload embedded in each cover is small. The steganographer can do better if they use matrix embedding to reduce the number of embedding changes, and adapt the choice of source coding so that the embedding efficiency e is as high as possible, given the constraints of individual cover capacity. This should improve the overall capacity.

In order to keep the algebra tidy, we return to the supposition that the covers are uniform, with a single Q-factor describing the steganalysis risk. Further, we suppose that the relationship between embedding changes and transmitted information is given by equality in (1); recall that this is achievable asymptotically and close approximations can be made in practice. Then we have:

THEOREM 6 (UNIFORM COVERS; OPTIMAL ADAPTIVE CODING). *If all covers can contain n embedding changes and have identical Q-factor Q, and a matrix embedding scheme which attains the bound (1) is used, the optimization problem is*

$$\textbf{Maximize } \sum m_i \quad \textbf{s.t.} \quad Qn^2 \sum \Big(H^{-1}\big(\tfrac{m_i}{n}\big)\Big)^2 \leq D$$

and the solution is

$$m_i = nH\Big(\sqrt{\tfrac{D}{NQn^2}}\Big) \text{ for each } i, \quad \text{so that} \quad M = nNH\Big(\sqrt{\tfrac{D}{NQn^2}}\Big).$$

In this case, $M = \Theta(\sqrt{N}\log N)$.

(This is a more complicated application of Lemma 3, and uses the fact that $H(x) \sim x \log \frac{1}{x}$ as $x \to 0$.)

Adaptive source coding has increased the asymptotic capacity by a factor of $\log N$. However, capacity remains substantially sublinear in N. This is in contrast to capacity results for noisy channels, where information transmitted is always linear in the number of symbols sent.

We may also wish to consider the effect of increasing n, but this relationship cannot be analysed unless we know the effect of cover size on Q. In some cases it seems likely that Q will be inversely proportional to n (for reasons beyond the scope of this work), in which case $M = \Theta(\sqrt{nN}\log nN)$.

3. EMBEDDING STRATEGIES FOR SEQUENTIAL STEGANOGRAPHY

In the preceding section it was vital that the number of covers N was fixed in advance. Subject to a fixed total acceptable risk D or total payload M, the optimal strategies involve N, so this is not applicable to an endless stream of covers. Although we have phrased capacity asymptotically as N "grows", in fact N did not grow.

Now we consider a different problem, when the steganographer wants to establish a communication channel with their recipient. We suppose that there is an infinite stream of covers, in which payload can be embedded, and the steganographer aims to embed as much as possible subject to a bound on the risk. This time the distortion bound is subtly different: we want

$$D_{\mathrm{KL}}\big((X_1^0, \ldots, X_N^0), (X_1^{p_1}, \ldots, X_N^{p_n})\big) \leq D$$

for all N. Since KL divergence is additive and nonnegative, this is equivalent to

$$\sum_{i=1}^{\infty} D_{\mathrm{KL}}(X_i^0, X_i^{p_i}) \leq D.$$

It is important to understand where this bound comes from: the steganographer is worried about a steganalyst who makes a *single* hypothesis test for the presence or absence of payload, based on the objects transmitted up that point, but does not know when that hypothesis test is going to take place. If this seems overly restrictive on the steganalyst, note that it would be suboptimal to make two (or more) hypothesis tests because this would simply compound the probability of false positives: at the point of the second (or last) test, all the information available to earlier tests is still present, so nothing can be gained by repeating hypothesis tests.

We continue to write $M = \sum_1^N m_i$, but now M is a variable which grows with N, and it now makes sense to discuss the asymptotic behaviour of M as N "grows". The first aim is to make sure that M grows without bound, so that the steganographic channel does not completely dry up, and the second is to have M grow asymptotically as fast as possible.

3.1. Sequential Embedding (i)

If all covers have an identical Q-factor Q, and each embedding change transmits e payload bits, the distortion bound is

$$\sum_{i=1}^{\infty} m_i^2 \leq \frac{De^2}{Q}. \tag{2}$$

Immediately we can see a tension between transmitting payload early and transmitting a larger payload: if the steganographer sends the most-possible information in the first object, $m_1 = \sqrt{De^2/Q}$, they have used up all their distortion budget and cannot send any more information at all. On the other hand, if they spread all the distortion over the first N objects, the total transmitted is $M = \sqrt{DNe^2/Q}$, exactly as in Subsect. 2.1. By varying N, arbitrarily large payload can be sent, but this does not establish a true covert channel because after a certain point the transmission must stop.

In an effort to use all of the infinite stream of covers, the steganographer might attempt *geometric embedding*:

$$m_i = \sqrt{\frac{De^2}{Q2^i}}.$$

This uses half of the distortion budget in the first cover, one quarter in the second, and so on. Unfortunately, the total payload transmitted $M = \frac{\sqrt{De^2/Q}}{\sqrt{2}-1}$ is finite, so all this has achieved is to take an infinite amount of time to send a finite amount of information.

However, it *is* possible to transmit an infinite total payload. The simplest scheme is *harmonic embedding*:

$$m_i = \tfrac{c}{i},$$

can meet (2) for a suitable constant c, while $\sum m_i = \infty$. As a function of N, the total payload transmitted after N objects grows without bound, but only asymptotically as fast as $\log N$.

Now the problem becomes clearer. The steganographer must pick a sequence (cm_i) such that $\sum m_i^2$ converges so the distortion bound can be met by a suitable choice of c, but $\sum m_i$ diverges as fast as possible. But for $\sum m_i^2$ to converge, the m_i terms must diminish sufficiently fast. It is possible to prove a result about the order of growth of $\sum m_i$:

THEOREM 7 (EMBEDDING BOUND). *If the distortion bound is of the form (2) then* $M = o(\sqrt{N})$.

(This proof is a bit tricky: the Cauchy-Schwartz inequality is used, and the infinite series $(\sum_{i=1}^N m_i)/\sqrt{N}$ must be separated into the correct choice of initial sequence and tail, each bounded separately.)

Recall that $M = o(\sqrt{N})$ means that $M/\sqrt{N} \to 0$ as $N \to \infty$; this is stronger than $M = O(\sqrt{M})$. M must grow *strictly* more slowly than $\sqrt{N}$, so the asymptotic order of payload growth of the batch setting cannot be matched in the sequential setting. However, we can get arbitrarily close using the following class of embedding strategies.

THEOREM 8 (ZETA EMBEDDING). *Suppose that $m_i = ci^{-\nu}$ for constants c and ν.*

(i) If $\nu \leq \frac{1}{2}$ then $\sum_1^\infty m_i^2$ diverges, so no distortion bound of the form (2) can be met.

(ii) If $\frac{1}{2} < \nu < 1$ then $c = \sqrt{\frac{De^2}{Q\zeta(2\nu)}}$, where ζ is the Riemann zeta function, is the largest constant to meet the bound (2). In this case, $M \sim \frac{N^{1-\nu}}{1-\nu}\sqrt{\frac{De^2}{Q\zeta(2\nu)}}$.

(iii) If $\nu = 1$ then $c = \sqrt{\frac{6De^2}{Q\pi^2}}$ is the largest constant to meet the bound (2). In this case, $M \sim \ln N \sqrt{\frac{6De^2}{Q\pi^2}}$.

(iv) If $\nu > 1$ then $\sum_1^\infty m_i$ converges, so that only a finite amount of information is ever transferred and no secret "channel" has been established.

(Proof: standard results on infinite series, see e.g. Ref. 13.)

Harmonic embedding, which we saw earlier, corresponds to $\nu = 1$ and is the worst of the zeta embedding strategies because it does not even achieve polynomial capacity in N: indeed, it is one of the most basic results of the theory of infinite series that $\sum_{i=1}^N i^{-1}$ only *just* diverges.

By picking $\nu = \frac{1}{2} + \epsilon$ and $c = \sqrt{\frac{De^2}{Q\zeta(1+2\epsilon)}}$ we allow M to grow asymptotically as $\sqrt{\frac{De^2}{Q\zeta(1+2\epsilon)(\frac{1}{2}-\epsilon)^2}}N^{\frac{1}{2}-\epsilon}$. But we have a dilemma: the larger the polynomial degree, the smaller the constant multiplier (which, as $\epsilon \to 0$, tends to $\sqrt{8\epsilon De^2/Q}$). Thus the tension which we saw at the beginning of this section, between transmitting more payload in any finite amount of time and maintaining the largest asymptotic capacity, exists for these infinite strategies too. The extent of this tradeoff will be explored in Subsect. 3.4.

3.2. Sequential Embedding (ii)

If we allow the covers to have different Q-factors, the problem is not greatly altered. The distortion bound becomes

$$\sum_{i=1}^{\infty} Q_i m_i^2 \leq De^2. \tag{3}$$

The same techniques can be applied as in the previous section with $\sqrt{Q_i}m_i$ for m_i, and the conclusions are identical: as long as Q_i are bounded above, and below away from zero, the embedding bound is still $M = o(\sqrt{N})$, and a simple modified zeta embedding $m_i = ci^{-(\frac{1}{2}+\epsilon)}Q_i^{-\frac{1}{2}}$ achieves $M = \Theta(N^{\frac{1}{2}-\epsilon})$. The multiplicative constant is more difficult to compute: it depends on the harmonic mean of $\sqrt{Q_i}$. There seems no significant new interest in this scenario so we will not study it further.

3.3. Sequential Embedding (iii)

Finally, return to the supposition that the covers are uniform and allow the use of adaptive source coding. This is particularly useful in embedding strategies where the distortion-per-object diminishes, because ever more efficient source coding (more payload per embedding change) becomes possible, so the rate of reduction in payload-per-object can be slowed. Assuming that (1) is an equality, and the covers can contain n embedding changes, we have the distortion bound:

$$\sum_{i=1}^{\infty} H^{-1}\left(\tfrac{m_i}{n}\right)^2 \leq \frac{D}{Qn^2}. \tag{4}$$

Now the steganographer's aim is to find a sequence (m_i) such that $\sum H^{-1}\left(\frac{m_i}{n}\right)^2$ converges, but $\sum m_i$ diverges as fast as possible. Again, there is a bound on the speed of the latter:

THEOREM 9 (EMBEDDING BOUND). *If the distortion bound is of the form (4) then $M = o(\sqrt{N}\log N)$.*

(Proof: a modification of that of Theorem 7.)

Once again, we see a capacity bound involving $\sqrt{N}\log N$, and that the sequential setting forces a payload to be asymptotically strictly lower than for batch steganography. And we can modify the details of zeta embedding to come polynomially close to this order of growth:

THEOREM 10 (MODIFIED ZETA EMBEDDING). *Suppose that the cover size is n and $m_i = nH\left(\frac{ci^{-\nu}}{n}\right)$ for constants c and ν.*

(i) If $\nu \leq \frac{1}{2}$ then no distortion bound of the form (4) can be met.

(ii) If $\frac{1}{2} < \nu < 1$ then $c = \sqrt{\frac{D}{Q\zeta(2\nu)}}$, where ζ is the Riemann zeta function, is the largest constant to meet the bound (4). In this case, $M \sim (\log_2 N) N^{1-\nu} \frac{\nu}{1-\nu} \sqrt{\frac{D}{Q\zeta(2\nu)}}$.

(iii) If $\nu = 1$ then $c = \sqrt{\frac{6D}{Q\pi^2}}$ is the largest constant to meet the bound (4). In this case, $M \sim (\log_2 N) \ln N \sqrt{\frac{3D}{2Q\pi^2}}$.

(iv) If $\nu > 1$ then only a finite amount of information is ever transferred.

(Proof: more theory of infinite series; the integral test is a useful tool here.)

In each case the source coding has produced an extra logarithmic factor in the capacity, and the constant multiplier is changed. By taking ν close to $\frac{1}{2}$, the asymptotic capacity approaches the bound of Theorem 9, but the multiplicative constant tends to zero.

3.4. Illustration of Sequential Embedding

We now select some realistic parameters, and examine how close practical steganography might come to these theoretical bounds. This will quantify the effect of source coding, and illustrate the tradeoffs implied by choice of ν for zeta embedding.

We have in mind the transmission of payload as the least significant bits (LSBs) of a grayscale digital image, hence the following parameters.

- The cover size n is 10^6, corresponding to a 1 megapixel image.
- When there is no adaptive source coding, the embedding efficiency e is 2, corresponding to LSB matching (the harder-to-detect variant of LSB replacement).[14]
- We suppose that the steganalyst bases their decision on the result of steganalysis of each image using the so-called Calibrated HCF COM detector,[14] which a few years ago was the state-of-art for detection of LSB matching. According to Ref. 3, a Q-factor of 10^{-10} is a realistic order of magnitude for the output of this detector.
- The distortion bound D is 1. This means that the steganographer is prepared to accept a risk corresponding to a KL divergence of 1 nat: this represents a moderate risk for, if the steganalyst were to equalize false positive and false negative rates, the bound implies that there would have to be at least a 17.6% chance of both.
- Experiments will be repeated with between 1 and 10^7 covers, both batch and sequential embedding.

Figure 1 shows the payloads embedded using zeta embedding for five different choices of ν, as well as the capacity of an individual cover (subject to the same acceptable KL divergence risk) and the capacity achievable using optimal batch steganography strategies. One chart corresponds to no source coding, and the other to hypothesising a source coding method satisfying the bound (1) exactly. Observe that lower values of ν lead to eventually-larger payloads, but with the penalty of shorter payloads at the beginning of the embedding stream. And, although the asymptotic order of growth of zeta embedding approaches that of optimal batch steganography, the multiplicative constants are low enough that quite a substantial gap remains. Also observe that almost an extra order of magnitude of capacity is achieved by adaptive source coding. Note that the accuracy of these capacity limits is contingent on correctness of the square distortion assumption in Subsect. 1.2. Some experiments have shown this to be *approximately* correct for HCF COM steganalysis, but the capacity levels should be regarded as indicative, not precise. And, of course, if a more effective steganalysis were used, with a larger Q-factor, the safe capacity would be lower.

However, there remains another reason to be skeptical about these results. All of the analysis in this paper has made the (quite false!) assumption that payload size is a real number. Of course in practice payload is not infinitely divisible, and cannot be transmitted in quanta smaller than one bit. In sequential embedding, the payloads transmitted must tend to zero and so, eventually, this limitation might become significant.

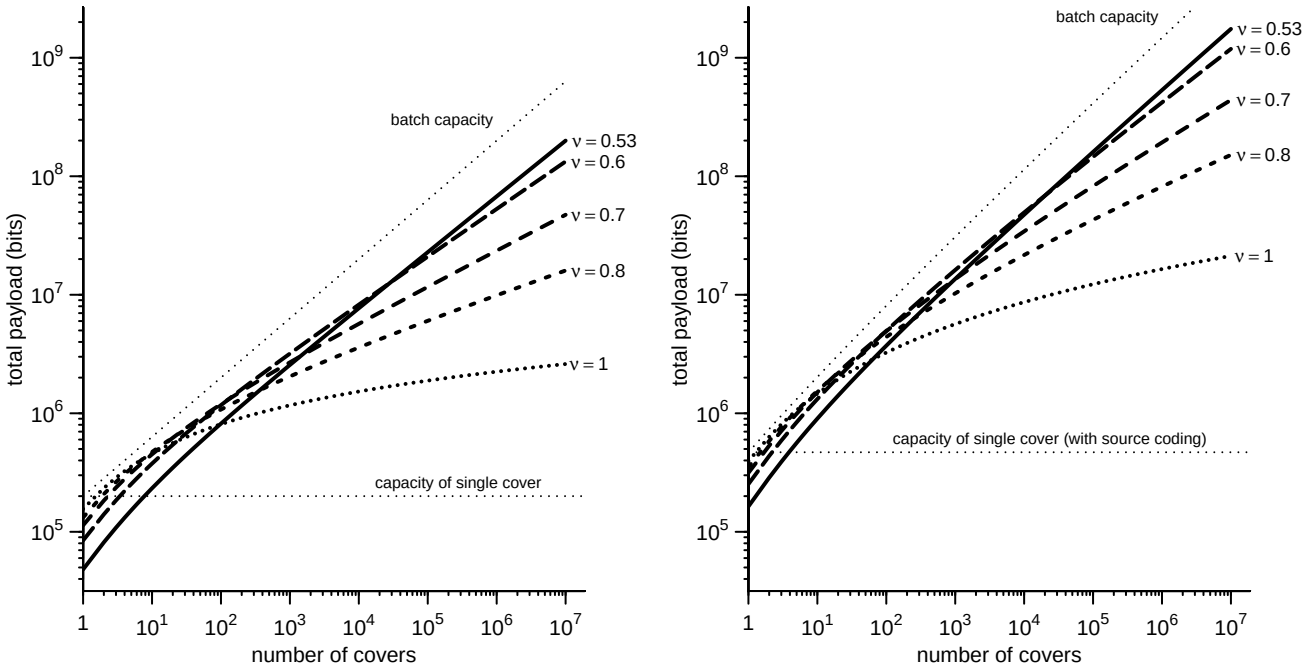

Figure 1. Total payload for the zeta embedding strategies (five choices of ν), as a function of the number of covers, if the parameters are as in Subsect. 3.4. Both axes are displayed on a log scale. On the left, a fixed embedding efficiency of 2 payload bits per cover change; on the right, adaptive source coding according to the bound (1). The capacity of a single cover, and the capacity of batch steganography when N is fixed in advance of all embedding, are also displayed.

To investigate whether the atomicity of the bit substantially effects steganographic capacity, we repeated the experiments in Fig. 1 with the additional constraint that payloads must be integral. In the absence of source coding, this amounts to bounding the permissible total KL divergence in the first j covers

$$d_j = \sum_{i=1}^{j} (ci^{-\nu})^2,$$

with c such that $d_j \to D$, and choosing each m_j to be the largest integer such that $\sum_{i=1}^{j} Q\left(\frac{m_i}{e}\right)^2 \leq d_j$.

On the left of Fig. 2, for three choices of ν, we display capacities achieved, compared with the same capacities without enforcing integrality of the payload in each cover. Except for $\nu = 1$, the difference is so small as to be invisible. To see whether any difference exists at all, some of the same data are displayed numerically in Table 1. In no cases is the capacity loss very substantial, and in a few cases capacity is very slightly increased by forcing integral payload sizes: this is because the greedy allocation method slightly levels out the payload stored in each object, compared with pure zeta embedding.

We can perform a similar experiment for modified zeta embedding with matrix coding. But let us go even further in aligning the theoretical results with reality by removing two more slightly dubious assumptions.

First, instead of claiming that the coding bound (1) is exact, we will use a genuine source coding algorithm: the very simple family of binary Hamming codes. Their application to steganography is described thoroughly in Ref. 6; all we need to know is that, for any integral parameter p, they allow us to embed p payload bits in $2^p - 1$ cover locations by making zero (with probability 2^{-p}) or one (with probability $1 - 2^{-p}$) embedding changes. We can concatenate instances of the Hamming codes with suitable choices of p to make use of as much of the cover as possible and embed as large a payload as possible within a limit on the number of changes. For example, if we are permitted 100 embedding changes in a cover of size 10^6, we may concatenate 23 instances of the code with $p = 14$, 76 with $p = 13$, and one with $p = 9$, embedding 1319 bits of payload using 999836 cover locations,

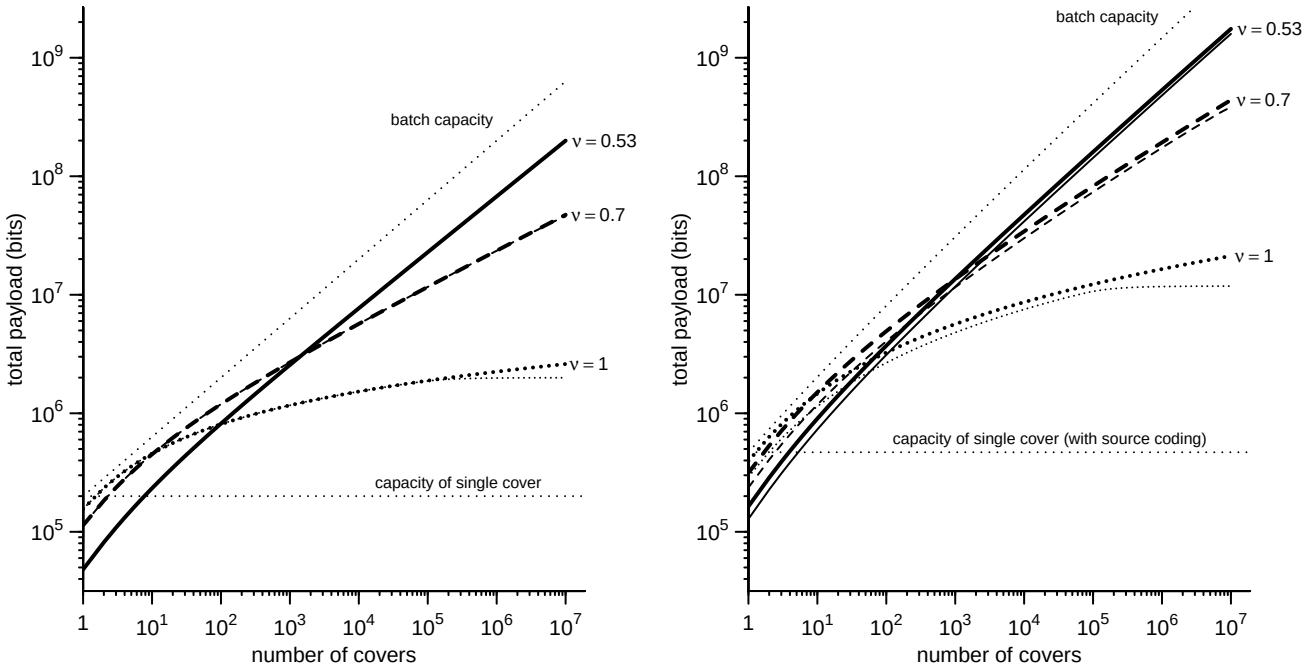

Figure 2. Comparison of the theoretical results, which allow fractional bit payloads and suppose that the coding bound (1) is achieved *(heavy lines)*, and more practical embedding which forces integral payload sizes and, on the right only, uses concatenations from the Hamming code family *(light lines)*. On the left, a fixed embedding efficiency of 2 payload bits per cover change; on the right, with matrix embedding.

and guarantee that there are at most 100 changes. It is demonstrated in Ref. 6 that the Hamming codes are far from optimal in practice so this choice of code is extremely conservative. (However for fixed payload size and as cover size tends to infinity, Hamming codes do approach the bound (1) asymptotically).

Second, we have been conflating the *expected* number of embedding changes with the true number: without source coding, treating $\frac{m_i}{e}$ as a fixed number of changes when in fact it is only true on average. This time we will assume the worst case, that every instance of the Hamming code involves one embedding change, never zero.

As with the previous experiment, we use zeta embedding to set a limit on the total distortion in the first j objects, and use a greedy algorithm to determine the allowable number of changes in each cover. The size of payloads thus embedded are displayed on the right of Fig. 2, compared with the theoretical performance of modified zeta embedding. We see that some capacity is lost by forcing payloads to be integral and using a suboptimal embedding scheme, but the difference is not very substantial. To make the comparison precise, Table 2 quantifies the difference between theoretical and practical capacities for a few values of N. The gap would grow for very large N, but it is hard to imagine realistic scenarios with so many millions of covers.

We conclude that the unrealistic assumptions about fractional payload, and use of the coding bound (1), do not invalidate the conclusions for realistic parameters.

4. DISCUSSION AND CONCLUSIONS

Three other papers deal with the batch steganography problem[1, 2, 5] and draw conclusions, of different strength, about steganographic capacity. They all agree that, in the absence of adaptive source coding, the capacity of a batch of N objects is of order $\sqrt{N}$; Theorem 4 agrees with this conclusion. However we should be clear that this paper's results are distinct from the others: Ref. 1 applies to particular steganalysis methods, Ref. 2 assumes a linear relationship between payload and steganalysis output (but goes further in providing an asymptotically optimal detection strategy), and Ref. 5 is only for a particular sort of evidence pooling. In this paper we have also

Table 1. Comparison of theoretical capacities when bits embedded are allowed to be fractional, or forced to be integral. Very little capacity is lost by forcing integral payloads per cover, and in some cases capacity is slightly increased due to partial levelling of payloads across images. Total payloads in bits.

	Payload	Number of covers N			
		10	10^3	10^5	10^7
$\nu = 0.53$	No source coding (fractional bits)	233065.9	2557917.4	22861917.6	199707237.2
	No source coding (integral bits)	233066	2557920	22861722	199485010
	Percentage capacity achieved	$> 100\%$	$> 100\%$	99.999%	99.889%
$\nu = 0.7$	No source coding (fractional bits)	450682.0	2690095.7	11647684.3	47310192.2
	No source coding (integral bits)	450682	2690100	11647150	45813134
	Percentage capacity achieved	100%	$> 100\%$	99.995%	96.836%
$\nu = 1$	No source coding (fractional bits)	456741.4	1167279.5	1885329.7	2603456.2
	No source coding (integral bits)	456744	1167286	1875198	1995568
	Percentage capacity achieved	$> 100\%$	$> 100\%$	99.462%	76.651%

Table 2. Comparison of theoretical capacities according to the source coding bound (1), allowing fractional embedding changes, against the simple Hamming code family with integral embedding changes. Even with this inefficient source code, relatively little capacity is lost. Total payloads in bits.

	Payload	Number of covers N			
		10	10^3	10^5	10^7
$\nu = 0.53$	Source coding bound (fractional bits)	901843.3	13614933.4	160329101.8	1749551325.7
	Hamming codes (integral bits)	730733	11699126	142923560	1589498281
	Percentage capacity achieved	81.026%	85.929%	89.144%	90.852%
$\nu = 0.7$	Source coding bound (fractional bits)	1505503.3	13558785.0	82544135.0	440119284.2
	Hamming codes (integral bits)	1182561	11562120	73631740	381707645
	Percentage capacity achieved	78.549%	85.274%	89.203%	86.728%
$\nu = 1$	Source coding bound (fractional bits)	1465984.8	5658968.6	12269231.1	21265636.6
	Hamming codes (integral bits)	1142788	4801454	10690787	11834302
	Percentage capacity achieved	77.954%	84.847%	87.135%	55.650%

gone further in explicitly dealing with adaptive source coding. The effect is to increase steganographic capacity by a logarithmic factor, but it remains a sublinear quantity. We have also considered nonuniform embedding.

Although the batch problem is convincing from the steganalyst's point of view – at the time of steganalysis, they have a certain number of objects whose evidence they wish to pool – it is perhaps less so for the steganographer. The latter is unlikely to know when the steganalyst will seize/monitor their communications, so must proceed under the assumption that communications might be examined at any time. Then the sequential steganography problem applies, and we have shown here that sequential steganographic capacity has some similarities to, but is not the same as, the batch problem. In particular a) steganographic capacity is asymptotically strictly lower, and b) zeta embedding is a family of sequential strategies which approach the optimum asymptotic payload rate, but the multiplicative constants are less favourable than optimal batch strategies.

We believe that these results are the first of their kind, and particularly notable is the conclusion that steganographic capacity, whether batch or sequential, with or without adaptive source coding, is sublinear. However we should also question the assumptions on which the results are founded. Some – the coding bound

(1), that embedding changes and payload sizes can take non-integral values, and the use of expected embedding changes disregarding randomness – have been investigated empirically in Subsect. 3.4 and shown to make a difference which is not very significant.

The use of KL divergence as a measure of evidence assumes that the steganalyst knows exactly the distribution of the source objects, or at least the response of a steganalysis method to them. More seriously, it is also implicit that the steganalyst knows the potential allocation of payload amongst the cover objects. This is probably not truly realistic, although it is appropriate to grant ones opponent extra knowledge if preparing for the worst case. It is much more difficult to reason about detection performance when the detector does not know the exact distributions they are observing, for example if there are unknown parameters.

The most important assumption is that of square distortion. Some experiments reported in Ref. 3 seem to confirm that KL divergence is (at least asymptotically) square in the number of embedding changes for some real steganalysis methods, but there is no guarantee that it applies universally. It would be of significance if a steganalysis method could be found which produces KL divergence which grows at a rate faster than the square of the number of embedding changes. Another implicit assumption is that all embedding changes are equally detectable. This is probably not the case in practice, but experience has shown that adaptive embedding methods can defeat their own aims by making the embedding locations more predictable.

ACKNOWLEDGMENTS

The author is a Royal Society University Research Fellow. Theorems 7 and 9 were proved with the assistance of Michael Collins and Roger Heath-Brown.

REFERENCES

1. A. Ker, "Batch steganography and pooled steganalysis," in *Proc. 8th Information Hiding Workshop*, *Springer LNCS* **4437**, pp. 265–281, 2006.
2. A. Ker, "A capacity result for batch steganography," *IEEE Signal Processing Letters* **14**(8), pp. 525–528, 2007.
3. A. Ker, "The ultimate steganalysis benchmark?," in *Proc. 9th ACM Workshop on Multimedia and Security*, pp. 141–148, 2007.
4. A. Ker, "Derivation of error distribution in least-squares steganalysis," *IEEE Transactions on Information Forensics and Security* **2**(2), pp. 140–148, 2007.
5. A. Ker, "Batch steganography and the threshold game," in *Security, Steganography and Watermarking of Multimedia Contents IX*, E. J. Delp III and P. W. Wong, eds., *Proc. SPIE* **6505**, pp. 0401–0413, 2007.
6. J. Fridrich and D. Soukal, "Matrix embedding for large payloads," *IEEE Transactions on Information Forensics and Security* **1**(3), pp. 390–394, 2006.
7. J. Fridrich and T. Filler, "Practical methods for minimizing embedding impact in steganography," in *Security, Steganography and Watermarking of Multimedia Contents IX*, E. J. Delp III and P. W. Wong, eds., *Proc. SPIE* **6505**, pp. 0201–0215, 2007.
8. C. Cachin, "An information-theoretic model for steganography," *Information and Computation* **192**(1), pp. 41–56, 2004.
9. S. Kullback, *Information Theory and Statistics*, Dover, New York, 1968.
10. G. Cohen, "A nonconstructive upper bound on covering radius," *IEEE Transactions on Information Theory* **29**(3), pp. 352–353, 1983.
11. R. Böhme, "Assessment of steganalytic methods using multiple regression models," in *Proc. 7th Information Hiding Workshop*, *Springer LNCS* **3727**, pp. 278–295, 2005.
12. R. Böhme and A. Ker, "A two-factor error model for quantitative steganalysis," in *Security, Steganography and Watermarking of Multimedia Contents VIII*, E. J. Delp III and P. W. Wong, eds., *Proc. SPIE* **6072**, pp. 59–74, 2006.
13. W. L. Ferrar, *A Text-Book of Convergence*, Oxford: The Clarendon Press, 1938.
14. A. Ker, "Steganalysis of LSB matching in grayscale images," *IEEE Signal Processing Letters* **12**(6), pp. 441–444, 2005.