Locally Square Distortion and Batch Steganographic Capacity

Andrew D. Ker
Department of Computer Science, Oxford University
Parks Road, Oxford OX1 3QD, UK

Contact information: E-mail: adk@cs.ox.ac.uk,
Telephone: +44 1865 283530, Fax: +44 1865 273839

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

Abstract

A fundamental question of the steganography problem is to determine the amount of data which can be hidden undetectably. Its answer is of direct importance to the embedder, but also aids a forensic investigator in bounding the size of payload which might be communicated. Recent results on the information theory of steganography suggest that the detectability of payload in an individual object is proportional to the square of the number of changes caused by the embedding. Here, we follow up the implications when a payload is to be spread amongst multiple cover objects, and give asymptotic results about the maximum secure payload. Two embedding scenarios are distinguished: embedding in a fixed finite batch of covers, and continuous embedding in an infinite stream. The steganographic capacity, as a function of the number of objects, is sublinear and strictly asymptotically lower in the second case. This work consolidates and extends our previous results on batch and sequential steganographic capacity.

A version of this paper was an invited article for the inaugural issue of the journal of *Digital Crime and Forensics*, which began in 2009. This document appears here for two reasons. First, the copyright holder does not permit authors to make the press versions of their articles available on their personal websites. Second, the author feels that the typesetters did a very bad job in this case, and mangled the equations. Thus this version of the paper, which is derived from a draft of the submitted article. Due to initial instructions from the publisher (which included a prohibition on square-root symbols, amongst others) some of the equations appear in an unusual format but at least they are correct and readable. If you like the paper, please purchase the article from `http://www.igi-global.com/article/international-journal-digital-crime-forensics/1590`**.**

We consider the following question: given a set of cover objects, how much data could be hidden in them? Although there is much literature on embedding and detection of steganographic payload, it is usual to consider only single cover objects, whereas this paper is concerned with embedding in a finite or infinite stream of objects, deriving capacity bounds and optimal methods. We posed the questions about embedding and detection in a fixed number of covers in Ker (2006), 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 (either exactly or locally for small payloads) proportional to the *square* of the number of changes caused by the embedding. Results of this nature have recently arisen in a number of theoretical steganalysis papers (Ker, 2007b, 2007c, 2007d) and the phenomenon has also been observed experimentally (Ker, Pevný, Kodovský, & Fridrich, 2008). 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. There is some recent related work (Ker, 2006, 2007a) where optimal embedding strategies were 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 structure of this paper is as follows. In the *Problem Formulation* section we will present the problems of batch steganography and sequential steganography; we will make and justify a series of assumptions about how steganalysis evidence accumulates. Evidence is not generated by payload itself – it is found as changes in the cover object, caused by the embedding process – so we must also relate embedding changes to payload transmitted and, with adaptive source coding methods, these are not always proportional (Fridrich & Soukal, 2006; Bierbrauer & Fridrich, 2008). In the *Analysis of the Batch Steganography Problem* section we will apply the theory to the batch steganography problem, deriving optimal embedding strategies and maximum undetectable payload, and in the *Analysis of the Sequential Steganography Problem* section to the sequential steganography problem; there is no optimal strategy in this case, but bounds can be derived, and strategies exist which come arbitrarily close to the bounds. It will be shown that the asymptotic payload, as a function of the number of covers, must be strictly lower in the sequential than the batch setting. Finally in the concluding *Discussion* section we will discuss the significance and limitations of the results.

An early version of some of this work has appeared in conference proceedings without any mathematical proofs (Ker, 2008b). In this work we have changed focus to concentrate on the embedding changes – this reduces the algebraic complexity – and are able to widen the applicability and weaken the assumptions. In particular, the square evidence law need hold only locally as payloads tend to zero.

Before continuing, we review some asymptotic notation. We write $f(n) = O\big(g(n)\big)$ if there are constants c and N such that $f(n) \leq cg(n)$ for all $n \geq N$. The analogous *strict* bound is $f(n) = o\big(g(n)\big)$, which means that $f(n)/g(n) \to 0$. We write $f(n) = \Theta\big(g(n)\big)$ if there are positive constants c, d and N such that $cg(n) \leq f(n) \leq dg(n)$ for all $n \geq N$. The most precise condition on growth is $f(x) \sim g(x)$, which means that $f(x)/g(x) \to 1$.

Problem Formulation

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. We formulated (Ker, 2006) the competing aims of *batch steganography*, in which it is assumed that a fixed set of N 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, how much payload can be embedded. In some cases we will also be able to identify the optimal strategies for the steganographer.

We also tackle a more difficult problem, dubbed *sequential steganography*. In the sequential setting we no longer suppose that the number of covers N is fixed in advance of embedding (this differs materially from the batch problem, because optimal strategies require advance knowledge of 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 payload segments extracted from the transmitted 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.

Distortion Bound

To determine the secure capacity of a set of covers we must choose a definition of *secure*, and the key is to measure the *evidence* available to the steganalyst. As in previous work (Ker, 2006, 2007a), 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.

Let us model the (finite or infinite) sequence of cover objects by a sequence of random variables $\boldsymbol{X} = (X_1, X_2, \ldots)$. These can represent entire cover objects or, more practically, a steganalyst's observation resulting from steganalysis of each object individually. Let us suppose that a sequence of stego objects, modelled by a sequence $\boldsymbol{Y} = (Y_1, Y_2, \ldots)$, is created with an embedding strategy causing $\boldsymbol{c} = (c_1, c_2, \ldots)$ embedding changes in the covers. (The distribution of Y_i depends, therefore, on c_i.) It is necessary 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. Later, we will relate payload size to number of changes.

Any detector – binary classifier for the presence or absence of payload in the sequence as a whole – must decide whether a sequence of objects is a realisation of $\boldsymbol{X}$ or $\boldsymbol{Y}$. By the information processing theorem (Cachin, 2004), any detector must have false positive probability α and false negative probability β satisfying

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

where D_{KL} represents the Kullback-Leibler (KL) divergence. In this sense, the worst-case risk to the steganographer is bounded by $D_{\mathrm{KL}}(\boldsymbol{X} \,\|\, \boldsymbol{Y})$ and we can say that *evidence* is, at least in this context of binary hypothesis testing, measured by this KL divergence. This is a standard idea, first applied to steganography by Cachin (2004) and now widely adopted.

Then the definition of a secure embedding strategy is one which does not exceed a certain risk (from the point of view of the embedder) or evidence level. Thus we make the assumption:

(A0) The steganographer's distortion bound is in terms of KL divergence, $D_{\mathrm{KL}}(\boldsymbol{X} \,\|\, \boldsymbol{Y}) \leq D$ for some positive D.

(Assumptions are numbered so that we may refer to the ones we require, later.) KL divergence has few attractive algebraic properties, but one is useful here. If we assume that the observations of Y_i are independent, then we can decompose the total evidence in N objects into a sum (Kullback, 1968, p. 23):

(A1) Evidence is additive: $D_{\mathrm{KL}}(\boldsymbol{X} \,\|\, \boldsymbol{Y}) = \sum_{i=1}^{N} D_{\mathrm{KL}}(X_i \,\|\, Y_i)$.

As long as the stream of cover objects come from a sensible source (a random selection from an image library, *not* consecutive frames from a video camera, say) it is plausible to assume such independence. Even if there is dependence between the cover objects it is not necessarily reflected in the steganalyst's observations, or is likely to be insignificant if the embedding process was chosen carefully.

Locally Square Distortion

Now we must relate the number of embedding changes c_i to the evidence found in object i, $D_{\mathrm{KL}}(X_i \,\|\, Y_i)$. We cannot expect to know the exact relationship (even if we knew everything about the cover source and the embedding method, it is likely to be intractable to compute the KL divergence exactly) but we can make some sensible approximations.

This paper is predicated on an assumption of *square distortion*:

(A2a) Evidence is a square law: for each i there is a positive constant Q_i such that $D_{\mathrm{KL}}(X_i \,\|\, Y_i) = Q_i c_i^2$.

The constants of proportionality Q_i are called the Q-factors (Ker, 2007d): note that the different cover objects are allowed different Q-factors reflecting different cover characteristics. (A2a) is a strong assumption, but it is true at least if X_i and Y_i have (possibly multivariate) normal distributions with mean shifted by a linear function of c_i.

For other distributions we argue that this still holds approximately. We appeal to a theorem of Kullback (1968, p. 26), 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 prior work (Ker, 2007d). Under these conditions, as $c_i \to 0$, $D_{\mathrm{KL}}(X_i \,\|\, Y_i) \sim Q_i c_i^2$ for a constant Q_i. We will later see that $c_i \to 0$ is forced, as the number of covers grows, if we are to meet a fixed evidence bound (this was also argued in Ker (2007d) 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 KL divergence evidence provided by cover i is proportional to c_i^2, although the constant of proportionality depends on the nature of cover i.

We codify this with the following assumption, weaker than (A2a).

(A2b) Evidence is *locally* a square law: for each i there is a positive constant Q_i such that $D_{\mathrm{KL}}(X_i \,\|\, Y_i) = \phi_i(c_i)$, with $\phi_i(0) = 0$, ϕ_i strictly increasing without bound, and $\phi_i''(x) \to 2Q_i$, uniformly in i, as $x \to 0^+$.

The condition in terms of ϕ_i'' is slightly stronger than $\phi_i(x) \sim Q_i x^2$, and it also guarantees a region of zero in which all the ϕ_i are convex. It can be proved by slightly stricter regularity conditions than in Kullback's theorem. The uniformity of the convergence is justifiable if we believe that the cover source is stationary.

Of course, correctness of assumption (A2b) still depends on regularity conditions, 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.

Cover Characteristics

For asymptotic results about capacity we also require some assumptions about the size and nature of the cover objects. The *size* of object i will be denoted n_i and measured by the number of possible embedding locations; we require only a very weak condition on the sequence of sizes. But it is well-established (Böhme, 2005; Böhme & Ker, 2006) that even similarly-sized covers can vary greatly in their capacity for secure payload: in images, factors including local variance, saturation and JPEG compression levels can have very significant impact on the rate at which cover changes produce evidence. These differences are reflected in the Q-factors of the covers, so for example we might expect that noisier covers have a lower value for Q_i. We need at least a weak assumption about the Q-factors too:

(A3a) The cover characteristics are bounded: there exist $\underline{n}$ and $\overline{n}$ such that $0 < \underline{n} \le n_i \le \overline{n}$ for all i, and there exist $\underline{Q}$ and $\overline{Q}$ such that $0 < \underline{Q} \le Q_i \le \overline{Q}$ for all i.

This assumption precludes the possibility of larger-and-larger, or ever-diminishing, cover objects, or unboundedly easier or more difficult covers to embed in. Such a situation would, of course, alter the asymptotic capacity. It is justified at least if we believe that the covers are from a stationary source.

It is also interesting to consider a more restricted case when it is only the cover size which varies. This would be plausible if, for example, cover objects are taken from the same

source and the embedding method cannot exploit any other differences between the covers. It is arguable that the Q factor should, all other things being equal, be inversely proportional to the cover size. This can be justified exactly if the cover consists of independent regions, and it is the subject of future work to prove that it holds even when there is (limited) dependence between different parts of the cover. Without justifying it further, we will allow this stronger assumption as an alternative to (A3a):

(A3b) The cover sizes are bounded: there exist $\underline{n}$ and $\overline{n}$ such that $0 < \underline{n} \leq n_i \leq \overline{n}$ for all i. Furthermore, the covers are of uniform character: there is a constant Q such that $Q_i = Q/n_i$.

Bounds on Embedding Efficiency

We have related the distortion bound to the number of embedding changes in each object, but both steganographers and forensic steganalysts are interested in payload size. The final component for our analysis is to relate these quantities. Recall that c_i is the number of embedding changes in object i, and let us write m_i for number of the payload bits that can always be conveyed by this many changes.

Under a simple embedding scheme there is a direct relationship between these, for example in simple least significant bit (LSB) replacement we have $m_i = c_i$ (note that we are using c_i as an upper bound on the number of changes: in LSB replacement on average only $\frac{1}{2}$ cover samples must be altered for the embedding of each payload bit, but in the worst case every payload bit requires one change). In similar cases, with fixed encodings, we may assume:

(A4a) The embedding code is fixed: for some positive constant E, $m_i = Ec_i$.

(We repeat that we are bounding the maximum *possible* number of changes, whereas some literature (Fridrich & Soukal, 2006; Fridrich, Lisonek, & Soukal, 2006) deals in the expected number of changes. Since our security model is about risk, it makes sense to take the pessimistic view and bound the maximal number of changes.)

But, when there is excess capacity, we can do better using a source coding method called *matrix embedding* (also known as *syndrome coding*), adapting the code to maximize the payload transmitted for a given number of locations and permitted changes. This technique was suggested by Crandall in an unpublished manuscript (Bierbrauer, 1998), and two works have been published surveying aspects of source coding for steganography (Fridrich & Soukal, 2006; Bierbrauer & Fridrich, 2008).

Following the literature, we will assume that the cover objects consist of a number of *locations*, each of which can carry an unconstrained q-ary symbol as payload: the embedding process may overwrite some or all of these symbols and each one overwritten is an *embedding change*. For example, under LSB embedding $q = 2$ and each pixel is a potential location: the symbol is just the LSB of each pixel value. Under ternary embedding $q = 3$, which allows a greater number of payload bits to be embedded in total, without changing the number of locations. We assume that q is fixed by the choice of embedding algorithm. For simplicity, we will also assume that q is a prime power, but in fact it would not affect any of the asymptotic conclusions were this not so.

Most literature focuses on relative embedding rates (payload bits per location) and embedding efficiency (payload bits per embedding change) but it is more convenient for us to consider absolute quantities. Let us define $\mu_q(n, c)$ to be the largest guaranteed payload size (measured in bits) which can be embedded in n locations using no more than c embedding changes. The complete function μ_q is not known, but we can bound it:

Lemma 1 *For any q, n, and c,*

$$c\log_2\left(\tfrac{n}{c}(q-1)\right) - c\log_2 q \leq \mu_q(n,c) \leq c\log_2\left(\tfrac{n}{c}(q-1)\right) + (n-c)\log_2\left(\tfrac{n}{n-c}\right).$$

Proof Both inequalities can be translated from Bierbrauer and Fridrich (2008), with extensions to q-ary alphabets as in Fridrich et al. (2006). The lower limit comes from using c repetitions of the $\left[\frac{q^p-1}{q-1}, \frac{q^p-1}{q-1}-p, 3\right]$ q-ary Hamming code, where $p = \left\lfloor \log_q\left(\frac{n}{c}(q-1)+1\right)\right\rfloor$: each repetition embeds p q-ary symbols in $\frac{q^p-1}{q-1}$ locations making at most one embedding change, and p is chosen to maximize the number of symbols. The upper limit derives from a sphere-packing bound from the theory of covering codes (e.g., Cohen, 1983). ■

Since, as $c/n \to 0$, the second term of both lower and upper bounds are dominated by the first, what is left is a concave function and we may make the simplification (valid for sufficiently large covers):

(A4b) Optimal adaptive source coding is used and $m_i = \chi_i(c_i)$, where χ_i is a strictly concave increasing function satisfying $\chi_i(x) \sim x\log_2\left(n_i(q-1)/x\right)$, uniformly in i, as $x \to 0^+$.

We highlight one further (implicit) assumption in this paper. When we come to optimization problems, we will not constrain c_i and m_i to be integers. In practice, of course, one cannot embed a fractional bit of payload nor make a fractional number of changes. However, because typical covers are very large, allowing the quantities to vary continuously is a reasonable approximation. Moreover, the problems of finding optimal batch and sequential steganography schemes would be much more difficult if restricted to integer domains. We will return, briefly, to this assumption – the only one which is strictly false – in the concluding section.

Analysis of the Batch Steganography Problem

We can formulate batch steganography, from the embedders point of view, as an optimization problem. Our aim is to derive the best embedding strategy and, hence, the maximal possible payload. Depending on which of the assumptions we select, our results will have to be asymptotic rather than exact.

This lemma, in which the conditions are stronger than necessary but fit well with our scenario, will be useful in what follows.

Lemma 2 *Suppose that, for $i = 1, \ldots, n$, $\phi_i : [0,\infty) \to [0,\infty)$ is continuous, convex, and strictly increasing without bound, $\chi_i : [0,\infty) \to [0,\infty)$ is continuous and strictly concave increasing, and $\phi_i(0) = \chi_i(0) = 0$. Then*

(1) for $D > 0$ the maximization problem

$$\textbf{Maximize } \sum \chi_i(x_i) \quad \textbf{s.t.} \quad \sum \phi_i(x_i) \leq D$$

has a unique solution determined by $\sum \phi_i(x_i) = D$ and $\chi_i'(x_i)/\phi_i'(x_i)$ constant, and

(2) if the objective maximum above is M then the dual optimization problem

$$\textbf{Minimize } \sum \phi_i(x_i) \quad \textbf{s.t.} \quad \sum \chi_i(x_i) = M$$

has the same solution, with objective minimum D.

Proof (1) By convexity of ϕ_i and $-\chi_i$, the problem is a one of convex optimization (Boyd & Vandenberghe, 2004). The feasible region is nonempty ($\phi_i(0) = \chi_i(0) = 0$, and $D > 0$, imply that $\boldsymbol{x} = \mathbf{0}$ is feasible) and compact (ϕ_i unboundedly increasing forces x_i to be bounded above). The objective function is strictly concave, so there exists a unique global minimum, at which the constraint is tight, and which may be determined by the method of Lagrange multipliers.

Writing $\Lambda = \sum \chi_i(x_i) - \lambda\big(\sum \phi_i(x_i) - D\big)$, we have $\frac{\partial \Lambda}{\partial x_i} = \chi_i'(x_i) - \lambda\phi_i'(x_i)$, so at the stationary point $\chi_i'(x_i)/\phi_i'(x_i) = \lambda$, a constant.

(2) This is just the standard duality theorem for strictly convex optimization. ■

The first part of the lemma will be used to solve the batch steganography optimization problem: maximize the payload transmitted $M = \sum m_i$, subject to the distortion bound $D_{\mathrm{KL}}(\boldsymbol{X} \parallel \boldsymbol{Y}) \leq D$. The second part of the lemma ensures that the solutions are the same as the alternative formulation: for a given payload size, minimize the KL divergence. The hypotheses of the lemma are covered by our assumptions about distortion and source coding.

There now follow a sequence of three theorems, applying Lemma 2 to versions of the batch steganography problem with different assumptions. We begin with the strongest assumptions and successively weaken them.

Theorem 3 *Suppose an exact square evidence law and fixed source coding, making assumptions (A0), (A1), (A2a), (A3a), and (A4a). Abbreviate $\tilde{Q} = \sum_{i=1}^{N} Q_i^{-1}$. Then*

(1) The optimal embedding strategy is $c_i = D^{1/2}\tilde{Q}^{-1/2}Q_i^{-1}$ and the total secure payload is $M = ED^{1/2}\tilde{Q}^{1/2}$. Asymptotically, $M = \Theta(N^{1/2})$.

(2) Under stronger assumption (A3b) (covers with uniform characteristics) the optimal strategy has $m_i = rn_i$ for a constant r, i.e. payload is embedded proportionally to cover size.

Proof (A0), (A1), (A2a), and (A4a) combine to give the following optimization problem:

$$\textbf{Maximize } \sum Ec_i \quad \textbf{s.t.} \quad \sum Q_i c_i^2 \leq D.$$

This can be solved using a variation of the Cauchy-Schwartz inequality, but it is just as simple to apply Lemma 2: the unique solution is given by

$$E/(2c_iQ_i) = k$$

where k is some constant, and substituting into $\sum Q_i c_i^2 = D$ gives $k = \frac{1}{2}ED^{-1/2}\tilde{Q}^{1/2}$. Hence $c_i = D^{1/2}\tilde{Q}^{-1/2}Q_i^{-1}$ and the formula for $M = \sum m_i = E\sum c_i$ follows immediately. By (A3a), $N\overline{Q}^{-1} \leq \tilde{Q} \leq N\underline{Q}^{-1}$, i.e. $\tilde{Q} = \Theta(N)$. This implies $M = \Theta(N^{1/2})$.

The second part is simple: observe that $m_i \propto c_i$, $c_i \propto Q_i^{-1}$ and, under (A3b), $Q_i \propto n_i^{-1}$. Overall, $m_i \propto n_i$. ■

Other results showing, under various different assumptions, that total steganographic capacity follows a square root law (in the overall size of the available cover) have arisen in the literature; we will consider them briefly in the *Discussion* section.

Now we weaken assumption (A2a) to (A2b), assuming only that the square evidence law holds locally to zero. We must be careful about the analytical details.

Theorem 4 *Suppose a local square evidence law and fixed source coding, making assumptions (A0), (A1), (A2b), (A3a), and (A4a). Again write* $\tilde{Q} = \sum_{i=1}^{N} Q_i^{-1}$. *Then*

(1) The optimal embedding strategy satisfies $c_i \sim D^{1/2}\tilde{Q}^{-1/2}Q_i^{-1}$ *and the secure total payload is* $M \sim ED^{1/2}\tilde{Q}^{1/2}$, *as* $N \to \infty$ *if the KL distortion bound* D *is fixed.*

(2) Under stronger assumption (A3b) (covers with uniform characteristics) the optimal strategy has $m_i \sim rn_i$ *for a constant* r, *i.e. payload asymptotically proportional to cover size.*

Proof As above, (A0), (A1), (A2b), and (A4a) together give optimization problem:

$$\textbf{Maximize } \sum Ec_i \quad \textbf{s.t.} \quad \sum \phi_i(c_i) \leq D, \tag{1}$$

but we cannot apply Lemma 2 immediately because the ϕ_i are not guaranteed to be convex everywhere.

Using the uniform convergence in (A2b), for any $\epsilon > 0$, there exists $\delta > 0$ and $L > 0$ such that:

$$\text{all } \phi_i(x) \text{ convex on } [0, \delta) \tag{2}$$
$$\text{all } \phi_i(x) > L \text{ on } [\delta, \infty) \tag{3}$$
$$\text{all } \phi_i(x) \in \big((1-\epsilon)Q_i x^2, (1+\epsilon)Q_i x^2\big) \text{ on } [0, \delta) \tag{4}$$
$$\text{all } \phi_i'(x) \in \big((1-\epsilon)2Q_i x, (1+\epsilon)2Q_i x\big) \text{ on } [0, \delta) \tag{5}$$

First, consider the optimization problem restricted to all $c_i \in [0, \delta)$. By Lemma 2 and (2) it has a unique solution with $E/\phi_i'(c_i) = k$, some constant. Using (5) and rearranging,

$$\tfrac{1}{2}Ek^{-1}Q_i^{-1}(1+\epsilon)^{-1} < c_i < \tfrac{1}{2}Ek^{-1}Q_i^{-1}(1-\epsilon)^{-1}. \tag{6}$$

Substituting into the tight constraint $\sum \phi_i(c_i) = D$, and using (4), we have

$$\tfrac{1}{4}E^2k^{-2}\tilde{Q}(1-\epsilon)(1+\epsilon)^{-2} < D < \tfrac{1}{4}E^2k^{-2}\tilde{Q}(1+\epsilon)(1-\epsilon)^{-2}.$$

hence

$$\tfrac{1}{2}ED^{-1/2}\tilde{Q}^{1/2}(1-\epsilon)^{1/2}(1+\epsilon)^{-1} < k < \tfrac{1}{2}ED^{-1/2}\tilde{Q}^{1/2}(1+\epsilon)^{1/2}(1-\epsilon)^{-1}.$$

and, using (6) again,

$$D^{1/2}\tilde{Q}^{-1/2}Q_i^{-1}(1+\epsilon)^{-3/2}(1-\epsilon) < c_i < D^{1/2}\tilde{Q}^{-1/2}Q_i^{-1}(1-\epsilon)^{-3/2}(1+\epsilon).$$

which demonstrates

$$c_i \sim D^{1/2}\tilde{Q}^{-1/2}Q_i^{-1} \tag{7}$$

and therefore $M \sim ED^{1/2}\tilde{Q}^{1/2}$.

We must now verify that (1) cannot have an optimum outside the region of guaranteed convexity $[0,\delta)^N$. By (3), no more than N/L of the c_i can be outside this region, without breaking the distortion constraint. Suppose that some do so: the effect is to reduce the constraint, and force the rest of the c_i into the guaranteed convex region. But finitely many of the c_i can only contribute finitely much to the objective function, and we have shown that, as $N \to \infty$, an unbounded contribution can be achieved by having all c_i inside region of guaranteed convexity. For any $\epsilon > 0$, therefore, there is a sufficiently large N such that all c_i are in $[0,\delta)$ at the optimum.

For part (2), if $Q_i = Q/n_i$ then (7) gives $m_i \sim ED^{1/2}Q^{-1/2}n_i\big[\sum n_i\big]^{-1}$, payload asymptotically proportional to cover size. ■

Finally, we may allow adaptive source coding at the embedder. In this case, the total payload size is superlinear in the number of embedding changes; this alters the objective function.

Theorem 5 *Suppose a local square evidence law and adaptive source coding, making assumptions (A0), (A1), (A2b), (A3a), and (A4b). Then*

(1) The optimal embedding strategy satisfies

$$\tfrac{n_i}{c_i}\log_2\big(\tfrac{n_i}{c_i}\tfrac{q-1}{e}\big) = kQ_in_i \tag{8}$$

where k is a constant. This implies that the secure total payload is $M = \Theta(N^{1/2}\log N)$ as $N \to \infty$ with D fixed.

(2) Under stronger assumption (A3b) (covers with uniform characteristics) the optimal strategy has $m_i \sim rn_i$ for a constant r, i.e. payload asymptotically proportional to cover size.

Proof This time the optimization problem is

$$\textbf{Maximize } \sum \chi_i(c_i) \quad \textbf{s.t.} \quad \sum \phi_i(c_i) \le D.$$

Most of the analysis is similar to that in the previous theorem, and we only sketch the differences. For the same reasons as before, for large enough N all c_i are forced into a region $[0,\delta)$ in which all ϕ_i are convex and χ_i concave, with the former arbitrarily close to $Q_ic_i^2$ and the latter to $c_i\log_2\big(\frac{n_i}{c_i}(q-1)\big)$. Then Lemma 2 applies, with the optimum asymptotically where $\chi_i'(c_i)/\phi_i'(c_i)$ is constant. This simplifies to (8). This equation is difficult to solve analytically (although, of course, the solution can be found numerically if specific values of D, all n_i, and all Q_i are given). However we may still draw a conclusion about the asymptotic growth of the total payload size M, as follows. Recall that n_i and Q_i are

uniformly bounded above, and below away from zero. Write (8) in the form $f\big(\frac{n_i}{c_i}\big) = kQ_in_i$, where $f(x) = x\log_2\big(x\frac{q-1}{e}\big)$; this positive continuous function has strictly positive derivative for $x \geq 1$ so the value of n_i/c_i is bounded above and below away from zero. We may conclude that, for any i and j,

$$0 < a < c_i/c_j < b \tag{9}$$

for some constants a and b independent of N. Now consider the tight distortion bound $\sum \phi_i(c_i) = D$. By prior reasoning, $\phi_i(c_i)$ is arbitrarily close to $Q_ic_i^2$ and together with (9) this forces $c_i = \Theta(N^{-1/2})$. Finally, using $m_i = \chi_i(c_i) \sim c_i\log_2\big(\frac{n_i}{c_i}(q-1)\big)$, we deduce that $m_i = \Theta(N^{-1/2}\log N)$ and hence $M = \Theta(N^{1/2}\log N)$.

The problem is simplified if $Q_i = Q/n_i$, for then (8) becomes $f\big(\frac{n_i}{c_i}\big) = kQ$, a constant, hence $\frac{n_i}{c_i} = l$, a constant. Therefore $m_i = \chi_i(c_i) \sim c_i\log_2\big(\frac{n_i}{c_i}(q-1)\big) = n_il^{-1}\log_2\big(l(q-1)\big)$. Even though the number of embedding changes is no longer proportional to the size of the cover, the optimization problem ensures that the payload embedded in each object remains proportional to cover size. ■

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

Analysis of the Sequential Steganography Problem

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, the optimal strategies all involve N. Therefore these results are not applicable to an endless stream of covers. Although most of our results have phrased capacity asymptotically as $N \to \infty$, in the batch steganography scenario N is fixed.

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: (A0) must mean

$$D_{\mathrm{KL}}\big((X_1,\dots,X_N)\,\|\,(Y_1,\dots,Y_N)\big) \leq D$$

for all N, where $\boldsymbol{X}$ is the stream of covers and $\boldsymbol{Y}$, which depends on the sizes of the embedded data $\boldsymbol{m}$, the stream of stego objects. Since KL divergence is nonnegative, this is equivalent to replacing assumptions (A0) and (A1) with

(A0') The steganographer's distortion bound for the sequential steganalysis problem is $\sum_{i=1}^{\infty} D_{\mathrm{KL}}(X_i\,\|\,Y_i) \leq D$.

It is important to understand where this bound comes from: the steganographer's opponent is a steganalyst who makes a *single* hypothesis test for the presence or absence of payload, based on the objects transmitted up that point, but the steganographer 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 test(s) is still present, so nothing could have been gained by performing the earlier test(s).

We continue to write $M = \sum_1^N m_i$, but now M is a variable which grows with N, and it makes sense to discuss the asymptotic behaviour of M in terms of N. 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. We will illustrate the sequential steganography problem under the most restrictive assumption options (A2a) and (A4a) – an exact square evidence law and no adaptive source coding – and make a relatively easy generalization later. (A3a) will be assumed throughout.

Under (A2a), the distortion bound simplifies to

$$\sum_{i=1}^{\infty} Q_i c_i^2 \leq D \tag{10}$$

and under (A4a), $M = E \sum c_i$. 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, $m_1 = ED^{1/2}Q_1^{-1/2}$, 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 $M = ED^{1/2}\left[\sum_{j=1}^{N} Q_j^{-1}\right]^{1/2}$, exactly as in Theorem 3. 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*:

$$m_i = ED^{1/2}Q_i^{-1/2}2^{-i/2}.$$

This uses half of the distortion budget in the first cover, one quarter in the second, and so on. Unfortunately, the total payload transmitted $M < ED^{1/2}\underline{Q}^{-1/2}\sum_{i=1}^{\infty} 2^{-i/2} = ED^{1/2}\underline{Q}^{-1/2}(\sqrt{2}-1)^{-1}$ 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*:

$$m_i = ED^{1/2}6^{1/2}\pi^{-1}Q_i^{-1/2}i^{-1},$$

meets (10) while $\sum m_i = \infty$. As a function of N, the total payload transmitted M grows without bound, but only asymptotically as fast as $\log N$.

Now the problem becomes clearer. The steganographer must find a sequence (a_i) such that $\sum Q_i a_i^2$ converges, so the distortion bound can be met by $c_i = ka_i$ for a suitable choice of k, but $\sum a_i$ diverges as fast as possible so that the total payload $M = E\sum c_i$ grows as fast as possible. When source coding is permitted, this last quantity changes to $M = \sum \chi_i(c_i)$. But for $\sum Q_i a_i^2$ to converge, the a_i terms must diminish sufficiently fast, and this places a limit on M's growth. It is possible to prove a result which holds under either an exact or a local square law for evidence, and holds in slightly different forms depending on whether adaptive source coding is used.

Theorem 6 (Embedding bound) *Assume (A0') and (A3a).*

(1) Under either (A2a) or (A2b), and fixed source coding (A4a), $M = o(N^{1/2})$.

(2) Under either (A2a) or (A2b), and with adaptive source coding satisfying (A4b), $M = o(N^{1/2}\log N)$.

Proof

Whether (A2a) or (A2b) holds, the distortion bound is $\sum_{i=1}^{\infty}\phi_i(c_i) \le D$ with ϕ_i increasing and $\phi_i(x) \sim Q_i x^2$ uniformly in i. This certainly forces $\phi_i(c_i) \to 0$ as $i \to \infty$, whereby $c_i \to 0$. So there exists j, independent of N, such that $\phi_i(c_i) \ge \frac{1}{2}\underline{Q}c_i^2$ for all $i > j$. Furthermore, $\sum_{i=j+1}^{\infty} c_i^2 < D' = 2D\underline{Q}$.

Write $C = \sum_{i=1}^{N} c_i$. Take any $\epsilon > 0$. Pick k such that $\sum_{i=k+1}^{\infty}\phi_i(c_i) < \epsilon^2/9$. This is independent of N and we may also assume that $j < k < N$.

Write

$$C_1 = \sum_{i=1}^{j} c_i, \quad C_2 = \sum_{i=j+1}^{k} c_i, \quad C_3 = \sum_{i=k+1}^{N} c_i.$$

Recall Cauchy's inequality, that $\sum_{i=1}^{n} a_i \le (\sum_{i=1}^{n} a_i^2)^{1/2} n^{1/2}$. This gives

$$C_2 \le (k-j)^{1/2}\left[\sum_{i=j+1}^{k} c_i^2\right]^{1/2} < k^{1/2}D'^{1/2} \tag{11}$$

and

$$C_3 \le (N-k)^{1/2}\left[\sum_{i=k+1}^{N} c_i^2\right]^{1/2} < N^{1/2}\epsilon/3. \tag{12}$$

Combining (11) and (12) we have

$$CN^{-1/2} = (C_1 + C_2 + C_3)N^{-1/2} < C_1 N^{-1/2} + k^{1/2}D'^{1/2}N^{-1/2} + \epsilon/3 < \epsilon,$$

the final inequality at least if $N > 9C_1^2\epsilon^{-2}$ and $N > 9kD'\epsilon^{-2}$. We have proved that for any $\epsilon > 0$, $C < \epsilon N^{1/2}$ for sufficiently large N.

So for part (1), observe that $M = EC$. We have proved that $C = o(N^{1/2})$ so $M = o(N^{1/2})$.

For part (2), by (A4b) and (A3a) there exists $\delta > 0$ such that $\chi_i(x) \le \psi(x) = 2x\log_2\big(\overline{n}(q-1)/x\big)$ for all i and $x \in [0,\delta)$. For large enough i, $c_i < \delta$ is guaranteed. Since $\psi(x)$ is concave we have

$$M = \sum_{i=1}^{N}\chi_i(c_i) \le \sum_{i=1}^{N}\psi(c_i) \le N\psi\big(C/N\big) = 2C\log_2\big(N\overline{n}(q-1)/C\big).$$

Then $C = o(N^{1/2})$ implies $M = o(N^{1/2}\log N)$. ■

Compare with Theorems 3-5: in the sequential setting, the asymptotic order of growth of M must be *strictly* lower than in the batch setting. Nonetheless, it is possible to come arbitrarily close using the following class of embedding strategies.

Theorem 7 (Zeta embedding) *Suppose (A0'), either (A2a) or (A2b), (A3a), and either (A4a) or (A4b). Let $c_i = ki^{-\nu}$ for constants $k > 0$ and ν.*

(1) If $\nu \leq \frac{1}{2}$ then $\sum_{i=1}^{\infty} \phi_i(c_i)$ diverges whenever $k > 0$, so no distortion bound of the form (A0') can be met.

(2) If $\frac{1}{2} < \nu < 1$ then there exists a $k > 0$ such that $\sum_{i=1}^{\infty} \phi_i(c_i) \leq D$. Then with no adaptive source coding (A4a), $M = \Theta(N^{1-\nu})$, and with adaptive source coding satisfying (A4b), $M = \Theta(N^{1-\nu} \log N)$.

(3) If $\nu = 1$ then there exists a $k > 0$ such that $\sum_{i=1}^{\infty} \phi_i(c_i) \leq D$. Then with no adaptive source coding (A4a), $M = \Theta(\log N)$, and with adaptive source coding satisfying (A4b), $M = \Theta\big((\log N)^2\big)$.

(4) If $\nu > 1$ then $\sum_1^{\infty} m_i$ converges, whether or not adaptive source coding is used, so only a finite amount of information is ever transferred and no secret "channel" has been established.

Proof We use the following elementary facts about infinite series (e.g., Ferrar, 1938): $\sum_{i=1}^{\infty} i^{-p}$ converges if and only if $p > 1$; in the case of divergence the partial sums $s_n = \sum_{i=1}^{n} i^{-p}$ satisfy $s_n \sim n^{1-p}/(1-p)$ for $p < 1$ and $s_n \sim \log N$ for $p = 1$. Similarly, $\sum_{i=1}^{\infty} i^{-p} \log i$ converges if and only if $p > 1$; this time the partial sums satisfy $s_n \sim n^{1-p} \log n/(1-p)$ for $p < 1$ and $s_n \sim \frac{1}{2}(\log n)^2$ for $p = 1$. Also note that when sequences a_n and b_n satisfy $a_n \sim b_n$ then $\sum_{i=1}^{\infty} a_i$ is convergent if and only if $\sum_{i=1}^{\infty} b_i$ is, and when they are divergent the partial sums satisfy $\sum_{i=1}^{n} a_i \sim \sum_{i=1}^{n} b_i$.

(1) If $\nu \leq 0$ then $\phi_i(c_i) \not\to 0$, so $\sum \phi_i(c_i)$ certainly diverges. If $0 < \nu < \frac{1}{2}$ then $\phi_i(c_i) \sim Q_i c_i^2 \geq \underline{Q} k^2 i^{-2\nu}$; by the comparison test $\phi_i(c_i)$ diverges.

(2) Since $c_i \to 0$, for sufficiently large i we have $\phi_i(c_i) \leq 2Q_i c_i^2 \leq 2\overline{Q} c_i^2 = 2\overline{Q} k^2 i^{-2\nu}$. By the comparison test, and because $2\nu > 1$, $\sum \phi_i(c_i) = k^2 S < \infty$. For $k = D^{1/2} S^{-1/2}$, (A0') is met. Under (A4a), $M = E \sum_{i=1}^{N} k i^{-\nu} = \Theta(N^{1-\nu})$, and under (A4b), $M = \sum_{i=1}^{N} \chi_i(c_i)$, which has the same asymptotic order as $\sum_{i=1}^{N} i^{-\nu} \log i = \Theta(N^{1-\nu} \log N)$.

(3) As above, $\phi_i(c_i) \leq 2\overline{Q} k^2 \sum i^{-2}$; the sum is convergent so there exists sufficiently small k to meet (A0'). But in this case, under (A4a) $M = E \sum_{i=1}^{N} k i^{-1} = \Theta(\log N)$, and under (A4b) $M = \sum_{i=1}^{N} \chi_i(c_i)$, which has the same asymptotic order as $\sum_{i=1}^{N} i^{-1} \log i = \Theta\big((\log N)^2\big)$.

(4) $M = \sum_{i=1}^{\infty} m_i = E \sum_{i=1}^{\infty} c_i$ or $\sum_{i=1}^{\infty} \chi_i(c_i)$, according to whether (A4a) or (A4b) is assumed; therefore M has the same asymptotic order as either $\sum_{i=1}^{\infty} i^{-\nu}$ or $\sum_{i=1}^{\infty} i^{-\nu} \log i$, both of which are convergent for $\nu > 1$. Therefore even an infinite number of covers conveys only a finite payload. ■

Harmonic embedding, which we saw earlier, corresponds to $\nu = 1$ and is the worst of the infinite 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 $\sum i^{-1}$ only *just* diverges. By taking ν arbitrarily close to 1/2, we may allow M to grow with rate arbitrarily close to the limits in Theorem 6.

However, there is a penalty for embedding at a rate close to the bound. For simplicity, make the strong assumptions (A2a) and (A4a) and further assume that all Q_i are equal to the constant Q. Then case (2), above, can be refined to:

Theorem 8 *Suppose (A0'), (A2a), $Q_i = Q$, and (A4a). Let $c_i = i^{-\nu} D^{1/2} Q^{-1/2} \zeta(2\nu)^{-1/2}$ for $\frac{1}{2} < \nu < 1$, where ζ is the Riemann zeta function (Abramowitz & Stegun, 1964, Ch. 23). Then (A0') is tight and $M \sim E D^{1/2} Q^{-1/2} N^{1-\nu} (1-\nu)^{-1} \zeta(2\nu)^{-1/2}$.*

Proof Same as for Theorem 7, but keep track of multiplicative constants. Note that $\zeta(s) = \sum_{i=1}^{\infty} i^{-s}$, for $\mathfrak{Re}(s) > 1$, is the definition of the zeta function. ■

By picking $\nu = \frac{1}{2} + \epsilon$ we allow M to grow asymptotically as $K\left(\frac{1}{2} - \epsilon\right)^{-1} \zeta(1 + 2\epsilon)^{-1/2} N^{1/2-\epsilon}$ for a constant K not depending on ϵ. We have a dilemma: the larger the polynomial degree, the smaller the constant multiplier (because $\zeta(1+x) \sim x^{-1}$, as $x \to 0^+$, the multiplicative constant approximates $K(8\epsilon)^{1/2}$). 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.

Discussion

Three other papers deal with the batch steganography problem (Ker, 2006, 2007b, 2007a) 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 $\sqrt{N}$; Theorems 3 and 4 concur with this conclusion. Note that this paper's results are distinct from the other three: Ker (2006) applies to particular steganalysis methods, Ker (2007b) assumes a linear relationship between payload and steganalysis output (but goes further in providing an asymptotically optimal detection strategy), and Ker (2007a) is only for a particular type of evidence pooling behaviour by the steganalyst. We have gone much further, allowing source coding, nonuniform covers, and covering the analytical details so that the growth of the detector's evidence need only be locally square in a suitable sense. It is notable that steganographic capacity, with or without adaptive source coding, remains sublinear. Indeed, source coding only grants an extra logarithmic factor.

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 or 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, batch capacity. In particular, capacity is asymptotically strictly lower in the sequential setting and there is no optimal strategy. However, the zeta embedding class of strategies can provide rates of capacity growth arbitrarily close to the bound, albeit with ever less favourable multiplicative constants.

We should consider carefully the assumptions on which these results rest. Some are unquestionable, for example (A3a). Assumptions such as (A0) and (A1) are essential if we want to measure steganographic security using KL divergence, and there seems to be little alternative. Note that the use of KL divergence 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, but some initial work (Ker, 2008a) shows that complexity of the problem is greatly increased when we grant

the steganalyst less information. And 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. Perhaps future research will shed light on these difficult questions.

The exactly- or locally-square distortion assumptions (A2a) and (A2b) are the cornerstone of this work; some experiments reported in Ker (2007d) seem to confirm that KL divergence is locally square in the number of embedding changes for some real steganalysis methods, but this is not a guarantee that the same applies universally. It would certainly be of significance if a steganalysis method could be found which produces KL divergence growing at a rate *faster* than the square of the number of embedding changes. The assumption about source coding is also not strictly proven: although we know that the upper and lower bounds to capacity (as a function of maximum permitted changes) are concave, there is no guarantee that the function itself is concave. But, in practice, there are codes whose performance approaches the upper bound (Fridrich & Filler, 2007) so any deviation from concavity will be very small.

There are two further assumptions, implicit here, which could be questioned. We measured embedding changes by the *maximum* number possible (over all payloads): this may seem overly pessimistic, since in practice a cover location need not be altered if it coincidentally already contained the correct symbol. However it is reasonable to adopt a pessimistic attitude when measuring the steganographer's risk. Furthermore, it would not materially affect our conclusions if we switched focus to the average number of embedding changes. We also assumed, throughout, that embedding changes and payload sizes can take non-integral values: of course, this is simply untrue and it means that our sequential strategies, in which the payload placed in each object is ever-diminishing, cannot be implemented exactly. In the limit as $N \to \infty$, a fixed total distortion in fact implies that the total number of embedding changes must be *finite*! However it seems that, in practice, such limits are not reached: some numerical computations in earlier work (Ker, 2008b) showed that, for realistic Q-factors and cover sizes, forcing integral embedding changes makes a barely detectable difference to steganographic capacity.

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.

Some analysis of the abstract sequential steganography problem still remains, as we did not optimize the zeta embedding strategies to account for nonuniformity in the covers. Because sequential steganography deals with rates of capacity growth, it is not obvious how an optimization problem can even be constructed. This is a subject for future work.

Acknowledgements

At the time of writing, the author was a Royal Society University Research Fellow. Theorem 6 was proved with the kind assistance of Michael Collins and Roger Heath-Brown.

References

Abramowitz, M., & Stegun, I. (1964). *Handbook of mathematical functions with formulas, graphs, and mathematical tables* (ninth Dover printing ed.). New York: Dover.

Bierbrauer, J. (1998). *On Crandall's problem.* (Unpublished communication available from http://www.ws.binghamton.edu/fridrich/covcodes.pdf)

Bierbrauer, J., & Fridrich, J. (2008). *Constructing good covering codes for applications in steganography.* Berlin: Springer. (To appear in *LNCS Transactions on Data Hiding and Multimedia Security*)

Böhme, R. (2005). Assessment of steganalytic methods using multiple regression models. In M. Barni, J. Herrera-Joancomartí, S. Katzenbeisser, & F. Pérez-González (Eds.), *Information Hiding, 7th International Workshop* (pp. 278–295). Berlin: Springer.

Böhme, R., & Ker, A. (2006). A two-factor error model for quantitative steganalysis. In E. J. Delp III & P. W. Wong (Eds.), *Security, steganography and watermarking of multimedia contents VIII* (Vol. SPIE 6072, pp. 59–74). Bellingham, WA: SPIE.

Boyd, S., & Vandenberghe, L. (2004). *Convex optimization.* Cambridge: Cambridge University Press.

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

Cohen, G. (1983). A nonconstructive upper bound on covering radius. *IEEE Transactions on Information Theory*, *29*(3), 352–353.

Ferrar, W. L. (1938). *A text-book of convergence.* Oxford: Clarendon Press.

Fridrich, J., & Filler, T. (2007). Practical methods for minimizing embedding impact in steganography. In E. J. Delp III & P. W. Wong (Eds.), *Security, steganography and watermarking of multimedia contents IX* (Vol. SPIE 6505, pp. 0201–0215). Bellingham, WA: SPIE.

Fridrich, J., Lisonek, P., & Soukal, D. (2006). On steganographic embedding efficiency. In J. Camenisch, C. Collberg, N. Johnson, & P. Sallee (Eds.), *Information Hiding, 8th International Workshop* (pp. 282–296). Berlin: Springer.

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

Ker, A. (2006). Batch steganography and pooled steganalysis. In J. Camenisch, C. Collberg, N. Johnson, & P. Sallee (Eds.), *Information Hiding, 8th International Workshop* (pp. 265–281). Berlin: Springer.

Ker, A. (2007a). Batch steganography and the threshold game. In E. J. Delp III & P. W. Wong (Eds.), *Security, steganography and watermarking of multimedia contents IX* (Vol. SPIE 6505, pp. 0401–0413). Bellingham, WA: SPIE.

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

Ker, A. (2007c). Derivation of error distribution in least-squares steganalysis. *IEEE Transactions on Information Forensics and Security*, *2*(2), 140–148.

Ker, A. (2007d). The ultimate steganalysis benchmark? In *9th ACM Workshop on Multimedia and Security* (pp. 141–148). New York: ACM Press.

Ker, A. (2008a). *Perturbation hiding and the batch steganography problem.* Berlin: Springer. (To appear in *Information Hiding, 10th International Workshop*)

Ker, A. (2008b). Steganographic strategies for a square distortion function. In E. J. Delp III & P. W. Wong (Eds.), *Security, forensics, steganography and watermarking of multimedia contents X* (Vol. SPIE 6819, pp. 0301–0313). Bellingham, WA: SPIE.

Ker, A., Pevný, T., Kodovský, J., & Fridrich, J. (2008). *The square root law of steganographic capacity.*

Kullback, S. (1968). *Information theory and statistics.* New York: Dover.