A Capacity Result for Batch Steganography

Andrew D. Ker, *Member, IEEE*

***Abstract*—The problems of batch steganography and pooled steganalysis, proposed in [1], generalize the problems of hiding and detecting hidden data to multiple covers. It was conjectured that, given covers of uniform capacity and a quantitative steganalysis method satisfying certain assumptions, "secure" steganographic capacity is proportional only to the square root of the number of covers. We now prove that, with respect to a natural definition of secure capacity, and in a suitably asymptotic sense, this conjecture is true. This is in sharp contrast to capacity results for noisy channels.**

***Index Terms*—Channel capacity, communication systems, information hiding, steganography.**

I. Introduction

STEGANALYSIS aims to distinguish innocent cover objects from payload-carrying stego objects, and it is clear that larger payloads are more easily detectable. Determining the maximum payload for which risk of detection is acceptable is a fundamental problem in steganography and steganalysis; at its heart is how quickly our ability to distinguish covers from non-covers grows with the distortion of the object, and viewed this way it seems inextricably linked with properties of the cover medium itself. This is perhaps why there has been little literature able to provide realistic measures of secure steganographic capacity: a theoretical result in [2] has not yielded information on practical steganography, and an application in [3] has been proved (see, e.g., [4]) to make an assumption (independence of pixels) which invalidate its conclusions as regards specific capacity limits.

The usual focus is on embedding in, and scanning of, individual objects. In [1], we asked how these can be extended to groups of objects, formulating and motivating the competing aims of batch steganography and pooled steganalysis. Apart from presenting an interesting general problem of steganalysis in multiple objects, this allows us to separate media-specific questions (our assumption of which is encapsulated by the shift hypothesis, see below) from the rest of the problem (performing multiple tests for anomaly). It is a framework which allows us to use tools of statistics to answer for the first time the following capacity question: how does secure capacity increase, as the number of cover objects grows?

In this work we will assume that a Steganographer already has a method to embed data in individual objects, and that a Warden already possess a quantitative detector for that form of steganography. A quantitative detector is an estimator for the size of payload in an individual object as a proportion of the maximum. At least for certain types of steganography, such detectors are common; see [5] for some examples.

We further assume that the estimation error distribution does not depend on the true value being estimated. In [1] we called this the shift hypothesis, that the distribution of detector response over plain covers is shifted, by the true embedding rate, when steganography is performed. We will define ψ as the density function of the error, and Ψ as the corresponding distribution function. That is, if X is the random variable representing the estimate of proportionate data embedded in a cover, when in fact proportion p has been embedded, then

$$\Pr[X < x] = \Psi(x - p). \tag{1}$$

(This property does hold, approximately, for some quantitative steganalysis methods in the literature, e.g., [5]. In Section III we argue that this hypothesis, and even the demand that the Warden's detector be quantitative, can be replaced by a much weaker assumption.) We expect that Ψ is continuous and ψ has infinite support, plus other regularity conditions.

The problem of batch steganography, which we generalize from [1], is as follows. Given N different cover objects each with the same capacity C bits,[1] the Steganographer wants to spread a total payload of M bits amongst them. We say that a Steganographer's embedding strategy is a choice of $p_1, \ldots, p_N$, where $0 \leq p_i \leq 1$ is the proportionate use of cover number i, i.e., Cp_i is the amount of payload embedded in cover number i. Possibly some p_i are zero, so that the transmitted objects may be a mixture of covers and stego objects. The Steganographer's constraint is $C \sum p_i = M$; within it they want to choose p_i to make detection difficult.

The Warden's dual problem is pooled steganalysis: given steganalysis payload estimates $X_1, \ldots, X_n$ for the N objects treated by the Steganographer, they aim to detect whether there is any payload. That is, to perform the hypothesis test

$$\begin{aligned} H_0 &: \quad \text{all } p_i = 0 \\ H_1 &: \quad \text{some } p_i > 0 \end{aligned} \tag{2}$$

with best reliability. Any such test is called a pooling strategy. We do not assume that the Warden wants to estimate M or the individual p_i, in this work. Note that the covers are considered uniform, not only in their capacity but also to prevent the Steganographer embedding adaptively. The Warden's steganalysis errors, for the N objects, are assumed independent.

II. Batch Steganographic Capacity

In [1], we investigated a number of strategies for Warden and Steganographer, and noted some commonalities amongst

Manuscript received August 11, 2006; revised September 20, 2006. This work was supported by the Royal Society. The associate editor coordinating the review of this manuscript and approving it for publication was Dr. Min Wu.

The author is with the Oxford University Computing Laboratory, Oxford OX1 3QD, U.K. (e-mail: adk@comlab.ox.ac.uk).

Digital Object Identifier 10.1109/LSP.2006.891319

[1]The assumption that the covers have uniform capacity is important to this proof, but we expect that it can be discarded in future work. In [1] we also imposed the constraint that all cover objects used for embedding had to contain the same amount of payload: the "uniform embedding" assumption. We relax this latter condition in this paper.

the results. In particular we found that if N is increased then the Steganographer must not raise M in proportion, else their risk of detection increases. But the investigation was limited to a particular metric for risk, considered a restricted form of batch steganography, and did not prove a general result. We conjectured that "secure steganographic capacity," whatever that means precisely, might be proportional to the square root of the total cover size. We will prove that this is correct in a fairly general sense.

Definition of "secure steganographic capacity" is not nearly as simple as, say, channel capacity. This is because the notion of "secure" is clearly application-dependent: what is an acceptable level of risk in one situation might be considered completely insecure in another. Motivated by Cachin's work applying information theory to steganographic systems, we propose the following definition for "secure."

Definition 1: Fix any $0 < \alpha^* < 1$ and $0 < \beta^* < 1-\alpha^*$. Suppose that an active Steganographer performs batch steganography with some choice of $p_1, \ldots, p_N$ which embeds total message length M into N cover objects, and that an inactive Steganographer transmits N cover objects.

We say that the Steganographer is at risk (w.r.t. (α^*, β^*)) if the Warden has a hypothesis test for (2), determining whether the Steganographer is active or inactive, with type-I error (false positive) probability $\alpha < \alpha^*$ and type-II error (false negative) probability $\beta < \beta^*$.

This definition allows us to capture any level of risk for the Steganographer, *excluding* the cases of $\alpha^* = 0$ or $\beta^* = 1$. That is, the Steganographer is not allowed to demand zero chance of detection or zero rate of false positives by the Warden (either is an impossible requirement, if ψ has infinite support).

We are now in a position to state the result which bounds secure capacity above and below, regardless of α^* and β^*. Note, however, that we are only able to determine capacity a) asymptotically and b) without regard to the multiplicative constant of this limit.

Theorem 1 (Batch Steganographic Capacity): Suppose the shift hypothesis (1) and that the error density ψ has infinite support, is at least twice continuously differentiable, and has $(\log \psi)''$ bounded below. Fix any α^* and β^* as above.

1) There is a pooling strategy for the Warden such that, no matter what the Steganographer's embedding strategy, if $M/\sqrt{N} \to \infty$ as $N \to \infty$ then for sufficiently large N the Steganographer is at risk.
2) There is an embedding strategy for the Steganographer such that, no matter what the Warden's pooling strategy, if $M/\sqrt{N} \to 0$ as $N \to \infty$ then for sufficiently large N the Steganographer is not at risk.

The two halves of this result are proved next. Morally speaking, the theorem tells us that M can safely increase no faster than $\sqrt{N}$, a result in sharp contrast to those of coding and noisy channels where information transmitted is always proportional to the number of symbols sent.

A. Proof of the Batch Steganographic Capacity Theorem (1)

We exhibit a pooling strategy for the Warden with the required characteristics. It is a modification of the simplest pooling strategy investigated in [1] and certainly is not an "optimal" strategy in practice. However it suffices to prove the asymptotic bound on steganographic capacity. We will have to work a little harder than in [1], for rigour and because we allow non-uniform embedding by the Steganographer.

The Warden will count the number of positive estimates

$$P = |\{X_i | X_i > 0\}|$$

and make the diagnosis of an active Steganographer if P exceeds a critical value P^*. We will define a P^* such that, for sufficiently large N, the type-I and type-II errors are bounded by α^* and β^*.

Write $P = \sum_{i=1}^{N} Y_i$ where the Y_i are independent random variables, indicators for each event $X_i > 0$. Under H_0, $Y_i \sim \mathrm{Ber}(\underline{q})$,[2] where $\underline{q} = 1 - \Psi(0)$. We will set

$$P^* = N\underline{q} - \Phi^{-1}\left(\frac{\alpha^*}{2}\right)\sqrt{N\underline{q}(1-\underline{q})}.$$

Under H_1, $Y_i \sim \mathrm{Ber}(q_i)$, where $q_i = 1 - \Psi(-p_i)$. It will be convenient to write $\bar{q} = 1 - \Psi(-1)$; because Ψ must be monotone increasing, and ψ is assumed to have infinite support

$$0 < \underline{q} \le q_i \le \bar{q} < 1 \quad \text{for all } i. \tag{3}$$

In the case of H_1 we have a sum of not-identically distributed random variables so we cannot use the central limit theorem as in [1]. We will use the following corollary of the Berry–Esséen Theorem (see, e.g., [6, §XVI.5]), which not only gives asymptotic behaviour but also bounds the difference from the asymptotic distribution function.

Lemma 1: Suppose that $Y_i \sim \mathrm{Ber}(q_i)$ are a sequence of independent Bernoulli random variables, with $\sup\{q_1, q_2, \ldots\} < 1$ and $\inf\{q_1, q_2, \ldots\} > 0$. Then there is some constant C_1 such that

$$\left|\Pr\left[\sum_i^N Y_i \le y\right] - \Phi\left(\frac{y - \sum q_i}{\sqrt{\sum q_i(1-q_i)}}\right)\right| < \frac{C_1}{\sqrt{N}}$$

for all N.

Note that the hypotheses about q_i are satisfied in this case, by (3). Applying Lemma 1 with all $q_i = \underline{q}$, we have

$$\alpha = \mathrm{P}\left[\sum Y_i > P^* | H_0\right] < 1 - \Phi\left(\frac{P^* - N\underline{q}}{\sqrt{N\underline{q}(1-\underline{q})}}\right) + \frac{C_1}{\sqrt{N}}$$
$$= \frac{\alpha^*}{2} + \frac{C_1}{\sqrt{N}}$$

which, for sufficiently large N, is less than α^*. Next

$$\beta = \mathrm{P}\left[\sum Y_i \le P^* | H_1\right] < \Phi\left(\frac{P^* - \sum q_i}{\sqrt{\sum q_i(1-q_i)}}\right) + \frac{C_1}{\sqrt{N}}$$
$$\le \Phi\left(\frac{\sum(\underline{q} - q_i)}{\sqrt{N\bar{q}(1-\underline{q})}} - \Phi^{-1}\left(\frac{\alpha^*}{2}\right)\sqrt{\frac{1-\underline{q}}{1-\bar{q}}}\right) + \frac{C_1}{\sqrt{N}}$$

because $\sum \underline{q} - q_i \le 0$ and $-\Phi^{-1}(\alpha^*/2) > 0$, and using (3). Now consider $\underline{q} - q_i = \Psi(-p_i) - \Psi(0)$ and apply the Mean Value Theorem to Ψ on $[-p_i, 0]$. We deduce that $\underline{q} - q_i = -p_i\psi(\hat{p})$,

[2]Throughout, $\mathrm{Ber}(q)$ denotes a Bernoulli random variable, which takes value 1 with probability q and value 0 with probability $1-q$, and Φ represents the Gaussian distribution function.

for some $\hat{p} \in (-p_i, 0)$. Since ψ is continuous and positive on [−1,0], it is bounded below (say by $A > 0$) and therefore

$$\sum \underline{q} - q_i < -A \sum p_i = -\frac{AM}{C}.$$

We conclude that

$$\beta < \Phi\left(C_2 - C_3 \frac{M}{\sqrt{N}}\right) + \frac{C_1}{\sqrt{N}} \tag{4}$$

where C_2 and C_3 are positive constants. Given $(M/\sqrt{N}) \to \infty$ as $N \to \infty$, $\beta \to 0$ so β must be less than β^* (for any positive β^*) for sufficiently large N.

B. Proof of the Batch Steganographic Capacity Theorem (2)

The Steganographer's embedding strategy is simply to split the payload equally between all covers. Let us write $p = M/NC$, so $p_i = p$ for each i.

In a very similar way to Cachin [2] we will give a bound on the ability of the Warden to perform hypothesis test (2) using the concept of Kullback–Leibler (KL) divergence. We include only the briefest sketch of the connection between KL divergence and hypothesis testing; for a more detailed exposition, see [7]. For two continuous random variables X, Y with density functions f, g the KL divergence (or relative entropy) is defined as

$$D_{\mathrm{KL}}(X\|Y) = -\int f(x) \log \frac{g(x)}{f(x)} \mathrm{d}x.$$

This measure satisfies the information processing theorem (see, e.g., [7]) which states that $D_{\mathrm{KL}}(h(X)\|h(Y)) \leq D_{\mathrm{KL}}(X\|Y)$ for any function h. This is usually quoted as "processing cannot increase divergence." Because hypothesis testing is an example of processing (with a binary output), this implies that if a test has type-I error α and type-II error β

$$d(\alpha, \beta) \leq D_{\mathrm{KL}}(X_0\|X_1) \tag{5}$$

where $d(\alpha, \beta)$ is the KL divergence between two Bernoulli distributions with probabilities α and $1-\beta$, which is known to be $\alpha \log(\alpha/(1-\beta)) + (1-\alpha)\log((1-\alpha)/\beta)$. This will suffice to bound the Warden's ability to perform (2), no matter what pooling strategy they use.

First we consider the case of a single cover object. Suppose that $(\log \psi)'' > -B$. Let X_0 be an instance of the Warden's estimator when no data is embedded in a cover, and X_1 an instance of the estimator when proportion p data is embedded; under the shift hypothesis the respective density functions are $\psi(x)$ and $\psi(x-p)$. Expanding $\log \psi$ about x using Taylor's theorem (with Lagrange remainder) we have

$$\begin{aligned} D_{\mathrm{KL}}(X_0\|X_1) &= -\int \psi(x)[\log \psi(x-p) - \log \psi(x)]\mathrm{d}x \\ &= -\int \psi(x)\left[-p\frac{\psi'(x)}{\psi(x)} + \frac{p^2}{2}(\log \psi)''(y(x))\right]\mathrm{d}x \\ &\qquad \text{where } y(x) \in (x-p, x) \\ &< p\int \psi'(x)\mathrm{d}x + \frac{Bp^2}{2}\int \psi(x)\mathrm{d}x \\ &= \frac{Bp^2}{2} \end{aligned}$$

because any continuously differentiable density function satisfies $\int \psi = 1$ and $\int \psi' = 0$.

Now consider embedding in multiple objects, according to the embedding strategy $p_i = p$ for each i. Let $\boldsymbol{X_0}$ be the random vector of all X_i under H_0, and $\boldsymbol{X_1}$ the corresponding vector under H_1. Since all X_i are independent and, with this Steganographer's embedding strategy, identically distributed

$$D_{\mathrm{KL}}(\boldsymbol{X_0}\|\boldsymbol{X_1}) = N D_{\mathrm{KL}}(X_0\|X_1) < \frac{NBp^2}{2} = \frac{BM^2}{2NC^2}.$$

Therefore, if $M/\sqrt{N} \to 0$ as $N \to \infty$, $D_{\mathrm{KL}}(\boldsymbol{X_0}\|\boldsymbol{X_1}) \to 0$ which implies that, for the hypothesis test (2), $d(\alpha, \beta) \to 0$. This forces $\alpha \to 1 - \beta$. If $\beta < \beta^*$ then, for sufficiently large N, $\alpha > 1 - \beta^* > \alpha^*$. Therefore, for sufficiently large N, the Steganographer is not at risk.

III. Discussion

The hypotheses of the capacity theorem are not too onerous and, we believe, realistic. Note that we do not require that the error has zero mean, a condition needed in [1] and noted to be surprisingly cumbersome. The lower bound on $(\log \psi)''$ is convenient to check, and such a bound exists for many common distributions with infinite support including Gaussian, Cauchy, Logistic, and the Student t-family.

The strongest hypothesis is the shift hypothesis. We believe that this is not a necessary condition, indeed it is not even necessary for the steganalysis statistic to be quantitative. Fundamentally we need that, in the limit as the payload per cover tends to zero, the steganalysis response is linear in the number of embedding changes (the number of locations at which the cover is altered). This is most likely to be true for a wide range of steganalysis methods because it is reasonable to assume that the effect of changing, say, two pixels (or samples, or bytes) of a cover will cause about twice the difference in a detector as changing one pixel. We do not attempt a formal proof at this stage, but it does seem likely that the capacity theorem can be extended to steganalysis methods satisfying this quasi-linear property. This suggests that capacity should be defined in terms of embedding change rate, instead of payload length (with sophisticated source coding, these quantities are not necessarily in fixed proportion [8]).

Does it make sense for a quantitative estimator to have infinite support? Clearly the true relative message length must be in the range [0,1], but it is the case that many quantitative steganalysis estimators do sometimes produce answers outside this range. Consider that the estimate derives from some abstract model which specifies properties of cover objects and that no cover will meet the model exactly. Inevitably, some covers are such that a small embedded message moves them closer to the model—these covers are exactly those which produce a negative estimate. Further, suppose that a quantitative estimator were to obey our shift assumption except that its value is clamped in the range [0,1]. It is easy to check that the capacity theorem remains valid.

A key assumption is that the steganalysis errors for the N different covers are independent. We believe that this is realistic (as long as the embedding method is not flawed in having predictable location of content within each cover), since most steganalysis error is due to deviations of the covers from their theoretical model. For different covers, such deviations

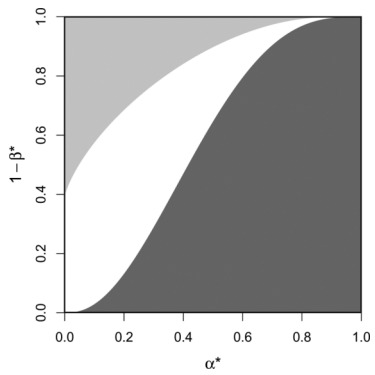

Fig. 1. Bounds given by (6) and (7), if the steganalysis error is standard Gaussian and $\gamma = 1$. The y-axis shows $1-\beta^*$ to correspond with the usual presentation of receiver operating characteristic curves. The light-shaded area indicates (α^*, β^*) for which the steganographer is guaranteed to be not at risk; the dark-shaded area indicates those values for which the steganographer is at risk, regardless of their embedding strategy (the loose inequalities used in Section II-A are evident in the weak bound here). Other values are undetermined.

are likely to be independent. Certainly for presently-known steganalysis methods which work on individual objects, errors between covers can often be shown experimentally to be uncorrelated. The setting of batch steganography allows this independence assumption, and it is a key reason why it is possible to develop capacity results when corresponding results for single covers are much harder.

The capacity theorem tells us that a payload which grows faster than $\sqrt{N}$ will be subject to an unacceptable risk and one which grows slower than $\sqrt{N}$ will eventually have acceptable risk. What if the payload grows exactly as $\sqrt{N}$? To be precise let us normalize by the individual object capacity C and write $(M/C\sqrt{N}) \to \gamma$, for a positive constant γ. Both halves of the capacity theorem proof tell us something here: depending on γ, for certain α^* and β^* the steganographer is at risk for sufficiently large N, and for certain other α^* and β^* the steganographer is not at risk for sufficiently large N. If we pass to the limit we can tighten (4) to

$$\beta^* \le \Phi\left(-\Phi^{-1}(\alpha^*)\sqrt{\frac{1-\underline{q}}{1-\bar{q}}} - \frac{A\gamma}{\sqrt{\bar{q}(1-\underline{q})}}\right) \tag{6}$$

and (5) gives

$$d(\alpha^*, \beta^*) \le \frac{B\gamma^2}{2}. \tag{7}$$

It turns out that these two bounds on detection reliability are a long way apart. As a simple example, suppose that the steganalysis response is standard Gaussian (zero mean and unit variance), so that $A = (2\pi e)^{-1/2}$, $\underline{q} = 1/2$, $\bar{q} = \Phi(1)$, and $B = 1$. We set $\gamma = 1$ and then use (6) to determine a region in (α^*, β^*) such that the Steganographer is at risk, and (7) to determine a region such that the Steganographer is not at risk. We plot these two regions in Fig. 1, and note the substantial gap between them.

The gap is much larger for more realistic steganalysis distributions. As shown in [1], a practical model of an error distribution for one quantitative steganalysis is a Student t-distribution with 2 degrees of freedom and a scale factor of 0.01, the density function of which is $\psi(x) = 10^2.(2 + 10^4 x^2)^{-3/2}$. This distribution gives $A = 0.9997\cdots \times 10^{-4}$, $\underline{q} = 1/2$, $\bar{q} = 0.99995\cdots$ and $B = 15000$.

Let us pick $(\alpha^*, \beta^*) = (0.05, 0.5)$ (the same definition of security has been used in various steganalysis literature including [5]) and compute the lower and upper bounds on γ implied by (6) and (7). The former gives

$$\gamma \le \frac{\sqrt{\bar{q}(1-\underline{q})}}{A}\left(-\Phi^{-1}(\beta^*) - \Phi^{-1}(\alpha^*)\sqrt{\frac{1-\underline{q}}{1-\bar{q}}}\right) = 1.163\cdots \times 10^6$$

and the latter

$$\gamma \ge \sqrt{\frac{2d(\alpha^*, \beta^*)}{B}} = 0.008121\cdots$$

IV. Conclusion

The theorem we have presented here is of theoretical importance—it is the first to show explicitly how capacity is influenced by the number the covers—but not very helpful to the Steganographer in practice because the bounds on γ can be so many orders of magnitude apart. Tightening of the bounds is for further research: we note that it is the extravagant inequalities in Section II-A which are primarily responsible for the gap. Pooled steganalysis is particularly difficult because we cannot realistically assume that the Warden knows the values of $p_1, \ldots, p_N$, so there is no UMP test; the results in [1] illustrate vividly how the performance of a pooling strategy can be heavily dependent on the embedding strategy.

Another problem with applying this work is that we have assumed that N is known to both parties, before any embedding takes place. In practice a Steganographer probably has an increasing number of covers and wants to know how much they can safely embed at each point. This sequential setting is a little different from the batch setting considered here, and further work will be needed to study it.

References

[1] A. Ker, "Batch steganography and pooled steganalysis," in *Proc. 8th Inform. Hiding Workshop*, to be published.
[2] C. Cachin, "An information-theoretic model for steganography," *Inform. Comput.*, vol. 192, no. 1, pp. 41–56, 2004.
[3] P. Sallee, "Model-based methods for steganography and steganalysis," *Int. J. Image Graph.*, vol. 5, no. 1, pp. 167–190, 2005.
[4] Y. Shi, C. Chen, and W. Chen, "A Markov process based approach to effective attacking JPEG steganography," in *Proc. 8th Inform. Hiding Workshop*, to be published.
[5] A. Ker, "A general framework for the structural steganalysis of LSB replacement," in *Proc. 7th Inform. Hiding Workshop*, 2005, vol. 3727, Springer LNCS, pp. 296–311.
[6] W. Feller, *An Introduction to Probability Theory and its Applications*, 2nd ed. New York: Wiley, 1971, vol. II.
[7] R. Blahut, *Principles and Practice of Information Theory*. Reading, MA: Addison-Wesley, 1987.
[8] J. Fridrich and D. Soukal, "Matrix embedding for large payloads," *IEEE Trans. Inform. Forens. Sec.*, vol. 1, no. 3, pp. 390–395, Sep. 2006.