Batch Steganography and the Threshold Game

Andrew D. Ker

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

ABSTRACT

In Batch Steganography we assume that a Steganographer has to choose how to allocate a fixed amount of data between a large number of covers. Given the existence of a steganalysis method for individual objects (satisfying certain assumptions) we assume that a Warden attempts to detect the payload by pooling the evidence from all the objects. This paper works out the details of a particular method for the Warden, which counts the number of objects of which the detection statistic surpasses a certain threshold. This natural pooling method leads to a game between the Warden and Steganographer, and there are different varieties depending on whether the moves are sequential or simultaneous. The solutions are intriguing, suggesting that the Steganographer should always concentrate the payload in as few covers as possible, or exactly the reverse, but never adopt an intermediate strategy. Furthermore, the Warden's optimal strategies are instructive for the benchmarking of quantitative steganalysis methods. Experimental results show that some steganography and steganalysis methods' empirical performance accords with this theory.

Keywords: Batch Steganography, Steganalysis, Game Theory

1. INTRODUCTION

Traditional steganalysis aims to tell innocent cover objects from payload-carrying stego objects, with the focus on discriminating individual objects. Similarly, steganography literature concentrates on efficient embedding of data into single covers. In Ref. 1 we asked how this can be extended to groups of objects, formulating and motivating the competing aims of *batch steganography* and *pooled steganalysis*. The topic is a wide one and Ref. 1 leaves many of the general questions unanswered.

In this work we again assume that a Steganographer has to choose how to spread a fixed amount of data between a number of covers. Given the existence of a steganalysis method for individual objects (satisfying certain assumptions) we assume that the Warden attempts to detect steganography by pooling the evidence from all the objects. Ref. 1 has shown that the Warden can detect substantially lower embedding rates if evidence from multiple objects is combined, but that this requires selecting the right method for pooling the evidence, which depends heavily on the way the Steganographer has spread their payload.

This paper works out the details of just one particular option for the Warden, of counting the number of objects for which the detection statistic surpasses a certain threshold. This leads to a game between the Warden and Steganographer, one which is surprisingly awkward to analyze, but the solutions seem interesting. We emphasise that this work only looks at one type of Warden's behaviour (in which they throw away a lot of information) and we do also need some assumptions about steganalysis which are not universally true. General answers to the batch steganography problem are likely to be very difficult, and this paper only makes a start by considering a special case.

The structure of the paper is as follows. In the remainder of this section we state the problem we are addressing, the assumptions required, and the nature of the *Threshold Game* played between Steganographer and Warden. In Sect. 2 we derive optimal strategies for the case when one player is forced to move first. In Sect. 3 we consider (briefly) mixed strategies when the players move simultaneously or without knowledge of their opponent's move. In Sect. 4 we perform some experiments on genuine steganalyzers, to show that the theoretical results have application in practice, and we conclude in Sect. 5.

Further author information: E-mail: adk@comlab.ox.ac.uk, Telephone: +44 1865 283530, Fax: +44 1865 276790

1.1. Batch Steganography and Pooled Steganalysis

Suppose that a Steganographer is going to transmit N (cover or stego) objects each with the same capacity C bits. They aim to insert a total payload of BNC bits, where $B \ll 1$ is the proportionate *bandwidth*, by embedding Cp into each of Nr of the objects and leaving the other $N(1-r)$ alone. Therefore they require $rp = B$, but can vary r and p with one degree of freedom within this constraint and the limitations $r \leq 1$ (no more than N objects can be used, which forces $p \geq B$) and $p \leq 1$ (capacity constraint of the individual objects, which forces $r \geq B$). They have the freedom to spread the steganography as thinly as possible amongst all covers ($p = 1$) or to concentrate in as few as possible ($r = 1$) or somewhere in between. This is the problem of *batch steganography* as described in Ref. 1, and the choice of r and p subject to the constraints is called the Steganographer's *embedding strategy*.

The Warden's task is *pooled steganalysis*: given the N objects treated by the Steganographer, to formulate reliable hypothesis tests for whether any payload has been transmitted, i.e. to decide between:

$$\begin{array}{ll} H_0: & r = 0 \\ H_1: & p, r > 0. \end{array} \tag{1}$$

We are not, for now, concerned with estimation of r and p, or in identification of the individual stego and cover objects. We will assume that the Warden has already developed a steganalysis method for individual objects and combines the outputs of steganalysis of all N objects into a single binary decision for this test; the method they use for this combination is called their *pooling strategy*.

This is a highly relevant problem as it mirrors, for example, the task of a computer forensics expert presented with a hard disk containing many images, only some of which might have been used for steganography. The expert is asked to determine whether some payload exists in some of the images. If there are very many, performing steganalysis on each image singly will lead to many false positives.

We must decide how to benchmark the pooled steganalysis hypothesis test, and we choose the *false positive rate when the false negative rate is 50%*. Although unusual this metric is justified in Ref. 1 and is perhaps most persuasive when it is presented as the median p-value which the Steganographer will present to the Warden, when embedding. More practically, this metric allows analysis of the strategies to become tractable. It is difficult in any case to formulate a one-dimensional metric for performance of a binary classifier (since balancing false positives and false negatives is heavily application-dependent), and obvious possibilities (which, for example, assign a loss function to false positives and negatives) seem very difficult to prove anything about. Finally, we will consider the asymptotic behaviour as $N \to \infty$; this also allows the discrete problem to be attacked with continuous mathematics.

1.2. Assumptions

Analysis of the problem requires a number of assumptions about the nature of the steganalysis which the Warden applies to individual objects. We state the assumptions here, and postpone comment on them until Subsect. 5.1.

We will assume that the Warden uses a *quantitative* steganalysis method for individual objects: an estimator for the size of payload in an individual stego object as a proportion of the maximum, giving results in or close to the region $[0, 1]$. Such estimators are common in the literature, at least for detection of bit replacement in digital images.[2–6] It also seems likely that many other steganalysis methods could be adapted to become quantitative.

Following Ref. 1 we also make the *shift hypothesis*, which states that the quantitative estimator is subject to additive error independent of the true embedding rate and of other errors. That is, if payload of size Cp_i is embedded in object i, the Warden's steganalysis outputs $X_i = p_i + \epsilon_i$, with all ϵ_i independent, identically distributed, continuous random variables. As noted in Ref. 1, and in one case checked here in Sect. 4, this is a reasonable approximation for the behaviour of some real-world steganalysis methods.

Let us write Ψ for the distribution function of the errors ϵ_i, and ψ for the density function. We will need some mild assumptions about the shape of the errors: the density ψ must be "bell-shaped" in a number of ways, which we enumerate thus:

(A1) Ψ is at least three times differentiable,
(A2) ψ is an even function, i.e. $\psi(x) = \psi(-x)$ for all x,
(A3) ψ has only one turning point (it must be a maximum at 0 by (A2)), i.e. $\psi'(x) < 0$ for all $x > 0$,
(A4) $\frac{\psi^2(x)}{\Psi(x)(1-\Psi(x))}$ has only one local maximum (it must at 0 by (A2)).

Given assumptions (A1)-(A3) we can also describe a function derived from Ψ, which will have important connections to the Threshold Game and which we will name $\overline{\Psi}$.

DEFINITION 1.1. *Each function in the family (parameterised by $t > 0$) $\upsilon_t(p) : [0, \infty] \to \mathbb{R}$, defined by*

$$\upsilon_t(p) = p\psi(t-p) - \Psi(t) + \Psi(t-p),$$

has a unique zero in the region $(0, \infty)$. We define $\overline{\Psi}(t)$ to be this root, and extend $\overline{\Psi}$ continuously to $[0, \infty]$ by setting $\overline{\Psi}(0) = 0$.

We must show that this is well-defined, i.e. that $\upsilon_t(p)$ does indeed have a unique zero in $(0, \infty)$. First, if $p \geq 2t$, $\upsilon_t(p) = \int_{t-p}^{t} \psi(t-p) - \psi(x)\,\mathrm{d}x$; by (A2) and (A3) the integrand is negative in this range and therefore so is $\upsilon_t(p)$. On the other hand, if $p \leq t$ then the same integrand is positive and therefore so is $\upsilon_t(p)$. Finally, $\upsilon_t'(p) = -p\psi'(t-p)$ which by (A2) and (A3) is negative for $p > t$. Since everything in sight is continuous by (A1), we have proved that υ_t has a unique root, which is in the range $(t, 2t)$. This also demonstrates that $\overline{\Psi}(0) = 0$ is the correct continuous extension.

Finally, then, we are able to make one more assumption about Ψ, expressed in terms of Ψ, ψ, and $\overline{\Psi}$:

(A5) $\frac{\psi(x-\overline{\Psi}(x))}{\sqrt{\Psi(x)(1-\Psi(x))}}$ has at most one turning point, which is a local maximum, in $[0, \infty)$

1.3. The Threshold Game

In Ref. 1 we considered three different ways for the Warden to pool their steganalysis evidence. We did not consider any case where the pooling strategy is parametric, because it leads to a game. In what we call the Threshold Game, the Warden chooses some threshold t and then counts how many of the N objects in the batch have estimate exceeding t. The reliability of this statistic, as a detector of steganography, depends on the distribution of the steganalysis estimator, t, and p in nontrivial ways: thus the choice of t and p requires analysis.

This pooling strategy is very simple and, it seems, rather inefficient because the magnitudes of the steganalysis estimates over the threshold are discarded. However, when authors show ROC curves for quantitative steganalysis methods, they are implicitly assuming that some threshold is being set and therefore – if there are multiple objects – that the Threshold Game, or something very like it, is being played.

To state the game precisely, then, we assume that the null distribution function of the quantitative steganalysis estimator Ψ is known to both players, as is N*. The Warden chooses the threshold parameter $t \in \mathbb{R}$; the Steganographer chooses how much of each cover is used, when a cover is used at all, $p \in [B, 1]$ which therefore determines the proportion of covers used $r = B/p$. Since we consider behaviour as $N \to \infty$ we can assume that r and p are continuous parameters. The game is zero-sum and, given the shift hypothesis and our selected metric, the payoff is given by:

THEOREM 1.2. *Asymptotically as $N \to \infty$, the median p-value of the observed number of steganalysis estimates greater than t is*

$$m(t, p) = 1 - \Phi\left(B\sqrt{N} \frac{\Psi(t) - \Psi(t-p)}{p\sqrt{\Psi(t)(1-\Psi(t))}} \right)$$

where Φ is the Gaussian distribution function.

Proof. Let us write $X_i = p_i + \epsilon_i$ for the Warden's steganalysis estimate of the proportionate payload p_i in object i. Suppose first the null hypothesis, that no payload is embedded. Then $\Pr[X_i > t] = \Pr[\epsilon_i > t] = 1 - \Psi(t)$. Using

*Later we shall consider whether the Warden should have knowledge of B, and the significance of the lack of such knowledge.

the Gaussian approximation to the binomial (valid for large enough N) we have that the number of estimates Y exceeding t has asymptotic distribution

$$H_0 : \; Y \dot{\sim} \mathrm{N}\Big(N\big(1-\Psi(t)\big), N\Psi(t)\big(1-\Psi(t)\big)\Big).$$

Now suppose the alternative hypothesis. Considering the Nr stego objects, for each of which $\Pr[X_i > t] = \Pr[\epsilon_i > t-p] = 1-\Psi(t-p)$, and the $N(1-r)$ cover objects in the batch separately, and using one of the many variants of the Central Limit Theorem which apply to non-identically distributed components (e.g. the *Berry-Esséen Theorem*, §XVI.5 of Ref. 7), we derive

$$H_1 : \; Y \dot{\sim} \mathrm{N}\Big(Nr\big(1-\Psi(t-p)\big) + N(1-r)\big(1-\Psi(t)\big), \sigma^2\Big)$$

with σ^2 some finite variance which need not concern us. The median of Y under H_1 is therefore asymptotically $N\big(1-\Psi(t)\big) + Nr\big(\Psi(t)-\Psi(t-p)\big)$. Comparing this value to the distribution of Y under H_0, the median p-value is seen to be

$$1 - \Phi\left(r\sqrt{N}\,\frac{\Psi(t)-\Psi(t-p)}{\sqrt{\Psi(t)(1-\Psi(t))}}\right).$$

Using $r = B/p$ gives the required result. □

Note that the performance depends on the quantity $B\sqrt{N}$; this indicates that the size of payload, proportional to BN, can increase only with the square root of N – not linearly – if risk of detection is not to increase. The same phenomenon was noted in Ref. 1 and is shown to be the asymptotically optimal behaviour for pooling strategies in Ref. 8. Therefore the Threshold Game, despite having some inefficiencies in discarding steganalysis information, at least has the right order of growth of detection power with respect to N.

In the Threshold Game the Steganographer wants to present the least evidence to the Warden, that is to maximize the p-value $m(t,p)$; the Warden wants to minimize the same quantity. It will be easier to consider

$$\chi_p(t) = \tfrac{\Psi(t)-\Psi(t-p)}{p\sqrt{\Psi(t)(1-\Psi(t))}}$$

in which $m(t,p)$ is monotone decreasing (i.e. the Warden wants to maximize $\chi_p(t)$ and the Steganographer to minimize it). We extend the functions continuously to χ_p for $p = 0$, so

$$\chi_0(t) = \tfrac{\psi(t)}{\sqrt{\Psi(t)(1-\Psi(t))}},$$

which simplifies the analysis. The results we prove will initially relax the constraint $p \in [B,1]$ to $p \in [0,1]$; as long as B is small then this will be shown to cause very little difference in the results.

The exact shape of the family $\chi_p(t)$, and therefore the optimal strategies for the players, depends on the error distribution Ψ. To illustrate the shape of $\chi_p(t)$ let us take a particular example: a Student t-distribution

$$\psi(x) = \frac{\Gamma(\frac{\nu+1}{2})}{\lambda\sqrt{\nu\pi}\,\Gamma(\frac{\nu}{2})}\left(1+\frac{x^2}{\lambda^2\nu}\right)^{-\frac{\nu+1}{2}}$$

with the degrees of freedom parameter $\nu = 2$ and the scale factor $\lambda = 0.05$ (this is a good model[1] for a slightly magnified error distribution of a genuine steganalysis estimator[3]). We compute the surface $\chi_p(t)$, as a function of $p \in [0,1]$ and t (we only include the interesting range of t), and display it in Fig. 1. It is rather clearer to display some of the functions χ_p separately, which we show as a conventional graph in the same figure.

This establishes the parameters and payoffs for the Threshold Game. However we have not specified which player makes the first move, or whether the moves are simultaneous. These options give rise to games with different solutions, strategies for each player, which we examine in the next two sections.

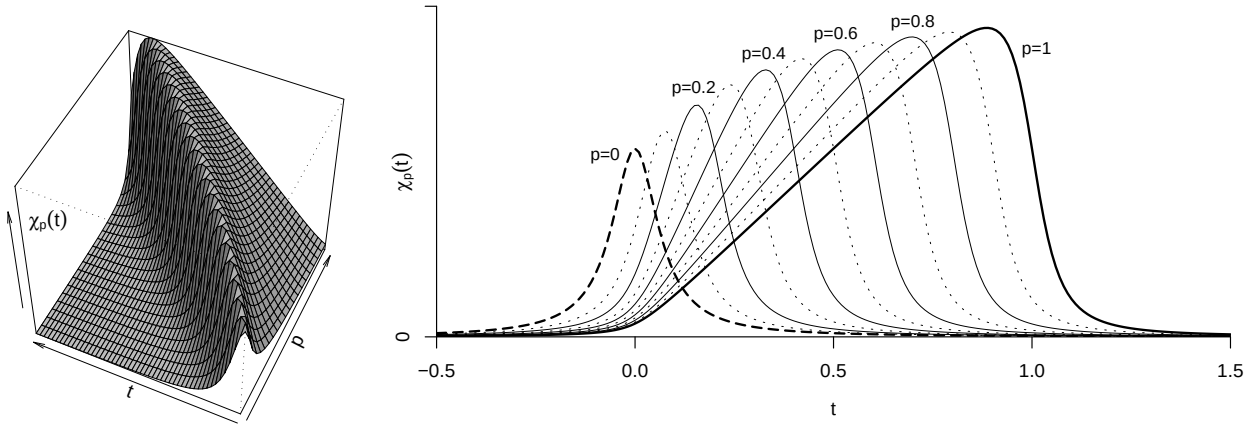

Figure 1. The function family χ_p and the surface they generate, arising from one particular error distribution Ψ. The Steganographer wants to choose p to minimize, and the Warden to choose t to maximize, this value.

2. MINIMAX AND MAXIMIN SOLUTIONS IN PURE STRATEGIES

Take the case when the Warden chooses t first. More precisely, we suppose that t is known (or can be estimated by some trial communications) by the Steganographer before the choice of p is made. The Warden's *maximin strategy* is to choose the value of t which can least be exploited by the Steganographer, i.e. to find

$$\operatorname*{argmax}_t \min_p \chi_p(t).$$

This is the Warden's best option if the Steganographer is allowed to adapt their choice p to t.

Consider the Steganographer's response to a choice of t by the Warden. In Fig. 1 it appears that the two curves χ_0 and χ_1 form the lower envelope, making $p = 0$ or $p = 1$ the only possibilities for the Steganographer's best strategy. But of course we have displayed the family χ_p only for one particular steganalysis distribution Ψ. In fact the same is true for all distributions Ψ satisfying the assumptions of Subsect. 1.2:

THEOREM 2.1.

(a) Consider the equation

$$\Psi(t) - \Psi(t-1) = \psi(t). \tag{2}$$

Given our assumptions it has a unique solution $\alpha \in (0, \frac{1}{2})$. Then,

(b) If $t < \alpha$, $\chi_1(t) < \chi_a(t)$ for $a < 1$.
If $t = \alpha$, $\chi_0(t) = \chi_1(t) < \chi_a(t)$ for $a \in (0, 1)$.
If $t > \alpha$, $\chi_0(t) < \chi_a(t)$ for $a > 0$.

(c) For $t < \alpha$, $\chi_1(t) < \chi_1(\alpha)$. For $t > \alpha$, $\chi_0(t) < \chi_0(\alpha)$.

Proof. (a) If we define $\nu(t) = \Psi(t) - \Psi(t-1) - \psi(t) = \int_{t-1}^{t} \psi(x) - \psi(t)\,\mathrm{d}x$, then (by (A2) and (A3)) for $t \leq 0$ the integrand is negative, for $t \geq \frac{1}{2}$ the integrand is positive. Then consider the derivative $\nu'(t) = \psi(t) - \psi(t-1) - \psi'(t)$; we have $\psi(t) > \psi(t-1)$ for $t \in (0, \frac{1}{2})$ and $\psi'(t) < 0$ for all $t > 0$, so $\nu'(t)$ is positive in the relevant region and we can deduce that ν has exactly one root, which is in the region $(0, \frac{1}{2})$.

(b) We compute the partial derivative

$$\frac{\partial \chi_p(t)}{\partial p} = \frac{p\psi(t-p) - \big(\Psi(t) - \Psi(t-p)\big)}{p^2\sqrt{\Psi(t)(1-\Psi(t))}} \tag{3}$$

and note that its sign depends only on the numerator, which is the function $\upsilon_t(p)$ defined in Subsect. 1.2. Therefore for fixed t, as a function of p, $\chi_p(t)$ has just one turning point when $p = \overline{\Psi}(t)$ and furthermore (because υ'_t is negative) this is a local maximum. We deduce that the lowest value of $\chi_p(t)$, as p varies, can only occur at the extremes $p = 0$ or $p = 1$. Therefore the lower envelope of the curves χ_p is traced out by some combination of χ_0 and χ_1.

Now consider the equation $\chi_0(t) = \chi_1(t)$; this immediately reduces to (2) which we know has only one root α. Moreover, for $t < \alpha$ we have $\nu(t) < 0$ so that $\chi_0(t) > \chi_1(t)$; therefore it is χ_1 which is the lower envelope here. Conversely, if $t > \alpha$ we have $\nu(t) > 0$ so it is χ_0 which is the lower envelope in the other part of the curve.

(c) We check that χ_1 is increasing on $(-\infty, \alpha)$ and χ_0 decreasing on (α, ∞). The former comes from computing

$$\chi'_1(t) = \big(\Psi(t)(1-\Psi(t))\big)^{-3/2}\Big(\big(\psi(t)-\psi(t-1)\big)\Psi(t)\big(1-\Psi(t)\big) + \big(\Psi(t)-\Psi(t-1)\big)\psi(t)\big(\Psi(t)-\tfrac{1}{2}\big)\Big);$$

the first factor is positive, and both of the terms inside the other factor are positive as long as $t < \frac{1}{2}$. For the latter, we only need note that (A4) says that χ_0^2, and therefore, χ_0 is decreasing on $(0, \infty)$. This suffices to establish the required result. □

To summarise, in this version of the game where the Warden first chooses t, the Steganographer (who wants to minimize $\chi_p(t)$) should always choose $p = 0$ or $p = 1$, the former when $t > \alpha$ and the latter when $t < \alpha$. At the critical point $t = \alpha$ the Steganographer is indifferent between $p = 0$ or $p = 1$, but intermediate values of p are inferior. The Warden must therefore choose $t = \alpha$, so that the Steganographer cannot exploit their knowledge of t to reduce $\chi_p(t)$.

Finally, we consider whether the constraint $p \geq B$ makes a significant difference to the outcome. The results of Theorem 2.1 can be adapted (in some cases with rather increased complexity) to give a similar conclusion for any $B < \frac{1}{2}$. So the only difference is that the critical point α becomes the root of the slightly different equation $\chi_B(t) = \chi_1(t)$. We check, in Sect. 4, that in practice the difference caused is slight, as long as B is small.

We now consider the arguably unrealistic situation in which the Steganographer chooses p first, with the Warden able to choose t with full knowledge of p. Practically, this situation seems quite impossible for the Warden, who would have to know the Steganographer's choice of p without also knowing whether the Steganographer is transmitting steganography. It is hard to think of a scenario in which this could happen. We include the analysis only for completeness, and because *Kerkhoffs' Principle* states that the enemy should be allowed to know the (crypto) system. Actually, we believe that Kerkhoffs' Principle should probably not apply to steganography, or rather it might apply in reverse (the Steganographer could possibly know the steganalysis method used by the Warden).

When the Steganographer must move first, their *minimax* strategy is to find

$$\operatorname*{argmin}_p \max_t \chi_p(t).$$

Consider again Fig. 1; this time the choice of curve is made by the Steganographer, and so the Warden will naturally choose t to find the peak of $\chi_p(t)$ (sadly this seems to have no concise form). In this particular figure it appears that the curve with the lowest peak – recall that the Steganographer wants to minimize $\chi_p(t)$ – is χ_0 but this is not the case for every steganalysis distribution Ψ. To demonstrate this we display the equivalent of Fig. 1 arising from slightly different steganalysis error distributions in Fig. 2. Observe that the curve with the lowest peak could be at either end of the range of p.

Generally, given our assumptions on Ψ, we can say:

THEOREM 2.2.

(a) For each p the function $\chi_p(t)$ has a maximum when

$$\big(\Psi(t)-\Psi(t-p)\big)\big(1-\Psi(t)\big)\Psi(t) = \big(\psi(t)-\psi(t-p)\big)\big(\tfrac{1}{2}-\Psi(t)\big)\psi(t). \tag{4}$$

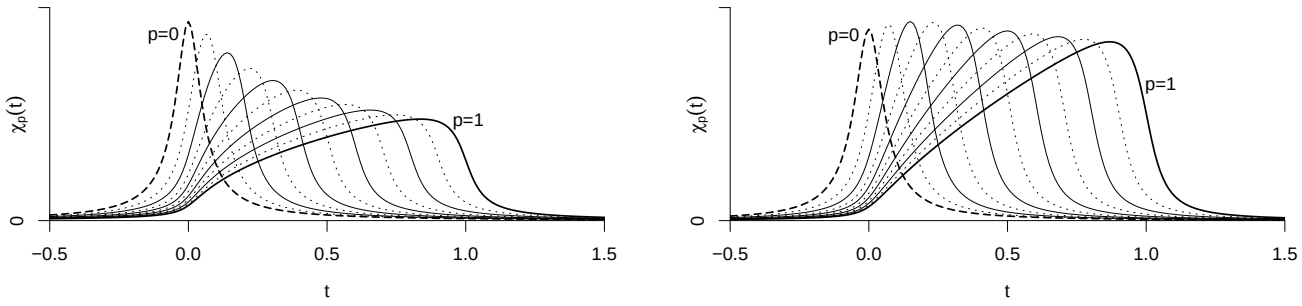

Figure 2. The function families χ_p, arising from two other Student t-distributions. As before, $\lambda = 0.05$, but here we take $\nu = 1$ (*left*) and $\nu = 1.5$ (*right*), to demonstrate the possible shapes of the upper envelope.

(b) The value of the upper envelope of $\chi_p(t)$ has a single maximum, and therefore the lowest peak is to be found either at $p = 0$ or $p = 1$.

Proof. (a) is immediate, once we compute

$$\chi_p'(t) = \big(\Psi(t)(1-\Psi(t))\big)^{-3/2} p^{-1}\Big(\big(\psi(t)-\psi(t-p)\big)\Psi(t)\big(1-\Psi(t)\big) + \big(\Psi(t)-\Psi(t-p)\big)\psi(t)\big(\Psi(t)-\tfrac{1}{2}\big)\Big).$$

(b) The upper envelope of the family $\chi_p(t)$, for fixed t, occurs when $\frac{\partial \chi_p(t)}{\partial p} = 0$. We have already computed this derivative, at (3), and we have already shown that it has a single zero (a maximum of $\chi_p(t)$ with respect to p) which must be when

$$p\psi(t-p) - (\Psi(t) - \Psi(t-p)) = 0.$$

This is exactly when $p = \overline{\Psi}(t)$ (see Def. 1.1). Therefore the upper envelope of $\chi_p(t)$ is the function

$$\frac{\psi\big(t-\overline{\Psi}(t)\big)}{\sqrt{\Psi(t)(1-\Psi(t))}}$$

which, by (A5), has one local maximum as its only turning point and therefore must take a minimum at an extremum of its domain, 0 or 1. □

We conclude that the Steganographer's minimax strategy is either $p = 0$ or $p = 1$, and they can choose which by solving (4) at each and seeing which gives rise to the lower value of χ_p. The Warden chooses the corresponding value of t. Again, the shape of the result is not affected if we force $p \geq B$.

Note how the extreme strategies for the Steganographer – $p = 0$ (or in practice $p = B$) corresponding to spreading the payload across all covers, and $p = 1$ where the payload is concentrated in as few covers as possible – occur in both these situations. These options were prevalent in Ref. 1, and they seem to be equally important in the Threshold Game.

3. EQUILIBRIUM IN MIXED STRATEGIES

When the moves are made simultaneously, or more precisely when each player moves without knowledge of the other's move (the most likely situation for steganography and steganalysis) the appropriate concept is *equilibrium*.[9] A Nash equilibrium for a game is a strategy for each player such that neither is disadvantaged by revealing their strategy in advance. It follows that a Nash equilibrium (for a zero sum game) is both a minimax and a maximin strategy. The previous section shows that both Warden and Steganographer can improve their position if they adapt their choice of t or p to their opponent's behaviour if known, so there is no choice of p and t which is in equilibrium.

However in general Nash equilibria do exist if the players are allowed *mixed strategies*[10]: strategies in which their action is partly random. In the case of the Threshold Game a mixed strategy for the Steganographer is one where the choice of p is made according to some probability distribution *before* the game begins, and similarly a mixed strategy for the Warden is a distribution from which to select t.

The introduction of mixed strategies to the Threshold Game is rather problematic because of our choice of payoff function as the median p-value for the Warden's hypothesis test (1). In the presence of mixing, this median is *not* a linear combination of the values for the pure strategies. For example if just the Steganographer uses a mixed strategy, elementary calculations give

LEMMA 3.1. *The median p-value m, if the Warden chooses a pure strategy t and the Steganographer mixes p according to the distribution P, satisfies*

$$\int \Phi\left(\frac{\Phi^{-1}(m) - B\sqrt{N}\frac{\Psi(t)-\Psi(t-p)}{p\sqrt{\Psi(t)(1-\Psi(t))}}}{\sqrt{1+\frac{B}{p}\left(\frac{\Psi(t-p)}{\Psi(t)}\frac{1-\Psi(t-p)}{1-\Psi(t)}-1\right)}}\right)\mathrm{d}P = \frac{1}{2}. \tag{5}$$

This presents enormous difficulties, even in showing the existence of any equilibrium, let alone computing one, and we do not propose a genuine solution here. The best we can do is to take the payoff function $m(t,p)$ as an entity in itself and imagine that the payoff of a mixture *is defined* as a linear function of this quantity. Since this problem setup does not truly reflect the Threshold Game, we will not include much detail in this section.

If we accept this definition of mixed payoff, then, we have:

THEOREM 3.2. *The Threshold Game has a Nash equilibrium when $t = \alpha$ (a pure strategy) and p is chosen randomly according to the distribution*

$$\Pr[p=0] = \frac{\chi_1'(\alpha)}{\chi_1'(\alpha)-\chi_0'(\alpha)} \qquad \Pr[p=1] = \frac{\chi_0'(\alpha)}{\chi_0'(\alpha)-\chi_1'(\alpha)} \tag{6}$$

Proof. According to Theorem 2.1, the only case in which the Steganographer has other than a single optimal strategy p, i.e. the only time they will want to mix against a pure strategy t, is when $t = \alpha$ and their mixture involves only the extreme strategies $p = 0$ and $p = 1$ (again we ignore the small distinction between $p = 0$ and the genuine minimum $p = B$). If they choose $\Pr[p = 1] = q$ and $\Pr[p = 0] = 1 - q$, the payoff (which the Steganographer aims to *maximize*) becomes

$$m(t,p,q) = q\big(1-\Phi(\chi_1(t))\big) + (1-q)\big(1-\Phi(\chi_0(t))\big)$$

and the maximum is found where the derivative with respect to t is zero. It is routine to check that this gives the required value for q. □

Again, the conclusion is not altered if we restrict $p \geq B$, although the mixture between $p = B$ and $p = 1$ will be a little different from the above. Analysis of mixtures of t remains difficult. We believe, but have not been able to prove for general distributions Ψ:

CONJECTURE 3.3. *This is the only Nash equilibrium for the Threshold Game (perhaps in the presence of some additional constraints on Ψ).*

4. EXPERIMENTAL RESULTS

We conduct a number of separate experiments to verify that the results in this paper, while they require assumptions about steganalysis which will not necessarily hold *exactly*, do apply to the practice of batch steganography. We will focus on a single type of steganography (simple LSB replacement) for which we have lots of experimental data available, and for which most of the leading detectors are indeed quantitative payload estimators.

First, we check that the conclusions of Theorem 2.1 are valid. For this experiment we test the Sample Pairs Analysis (SPA) detector of Ref. 3, which is a quantitative estimator. Using a set of 20000 cover images, all sized

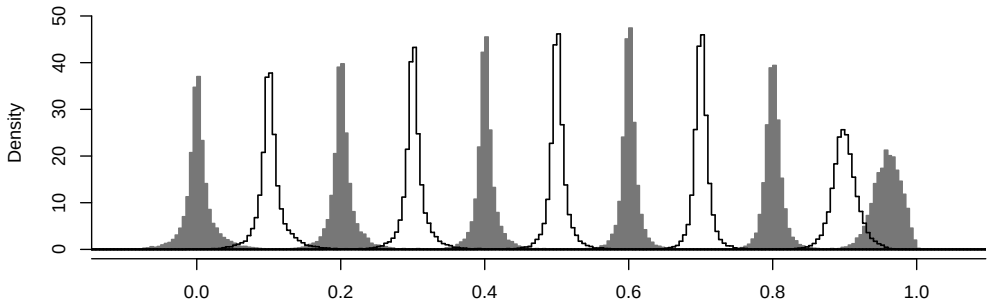

Figure 3. Histograms of the output of the SPA quantitative estimator, computed over a library of 20000 cover images, for true proportionate payloads $p = 0, 0.1, \ldots, 1$.

640×416 so that their individual capacity $C = 266240$ bits (originally stored as colour JPEGs at quality factor 58), we created a detailed profile of the steganalysis estimator distributions by embedding random messages of proportionate lengths $0, 0.001, 0.002, \ldots, 1$ in each image. See Fig. 3 for some of the results; this figure shows that the quantitative estimator does *approximately* satisfy the shift hypothesis except for near-maximal embedding, and has a roughly symmetrical null distribution. That our theoretical results accord with empirical results using this detector will demonstrate that small departures from the assumptions do not invalidate the conclusions.

Using the empirical cdf of the detector response when no payload was embedded (smoothed by a Gaussian kernel) as an estimate for the true distribution, we find numerically the solution to (2), with the result that in this case $\alpha \approx 0.0585$.

We then set a particular batch steganography scenario: $N = 10000$ cover images, and a random payload of 13000Kb, equivalent to proportional bandwidth $B = 0.005$. We repeated the following experiment for a large number of values of $t \in [-0.5, 1.5]$ and $p \in [B, 1]$: simulate batch steganography by picking Nr images embedded with proportionate payload p (with $rp = B$) and a further $N(1-r)$ cover images; count how many of the total batch of N have steganalysis estimate exceeding the threshold t; repeat 5000 times, and compute the median p-value by comparing with the number of times N plain covers exceed the same threshold. This computation-intensive procedure avoids any theoretical modelling, although we must expect that results obtained by Monte Carlo methods have a certain amount of noise in them.

For each t we computed the optimal value of p (the optimal value from the Steganographer's point of view is that which gives the highest median p-value, i.e. gives the least evidence to the Warden), assuming the situation where the former is known to the Steganographer when choosing the latter, and the consequent median p-value. The results of this experiment, focusing on the interesting range of t, are displayed in Fig. 4. We see that the conclusions of Theorem 2.1 are valid: the Steganographer's best strategy switches abruptly from fully-spread embedding $p = 1$ to fully-concentrated embedding $p = B$ once t crosses a certain point, and that this critical value of t leads to the best payoff from the Warden's perspective. The location of this saddlepoint is $t = 0.057$ (accurate to 0.001) which is very close indeed to the theoretically-predicted value α.

Second, we aim to verify the equilibrium found in Sect. 3. Numerical determination of equilibria when there are mixtures of continuous strategies is difficult so we content ourselves with a fine-grained discrete approximation, using again the SPA detector profile computed empirically from 20000 cover images. This time we will imagine that $B\sqrt{N} = 1$, so that the payoff is simply $\Phi(-\chi_p(t))$; this was computed with p varying in steps of 0.001 and t in steps of 0.0001 between -0.5 and 1.5 to create a discrete approximation to the continuous game.

Again we apply numerical methods to the empirical distribution functions, giving $\alpha \approx 0.0585$ (which predicts the equilibrium pure strategy for t) and $\frac{\chi_0'(\alpha)}{\chi_0'(\alpha)-\chi_1'(\alpha)} \approx 0.399$ (which predicts the proportion of $p = 1$ in the mixture $p \in \{0, 1\}$). Standard linear programming techniques[11] are used to find an equilibrium when potentially both p and t are chosen by mixed strategies. The equilibrium found indicates a pure strategy for the Warden of $t \approx 0.0575$ and the Steganographer mixing between only the two options $p = 0$ and $p = 1$, with probabilities 0.444 and 0.556 respectively. The predicted value for t matches to within 0.001 the observed pure equilibrium

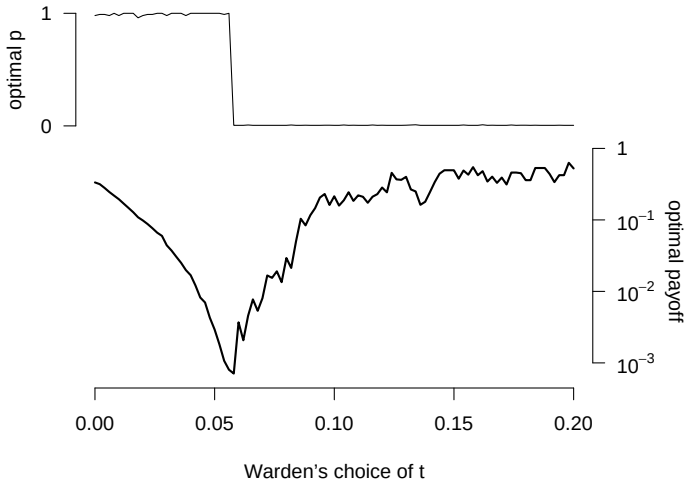

Figure 4. Results of Monte Carlo simulations of the Threshold Game. For each t, the Steganographer's optimal value of p is determined (*above*) and the corresponding median p-value for the Warden's hypothesis test (*below*).

strategy for t; the prediction of (6) is fairly close to the observed mixture between $p = 0$ and $p = 1$. Once again the theoretical results, even though based on assumptions only fulfilled approximately by real-world steganalysis, are in accordance with experimental data.

Our third and fourth sets of experiments still concern LSB replacement embedding but are performed on a wider range of steganalysis methods. We apply the results about the Threshold Game to benchmark the methods themselves, and finally check that the approximation of the constraint $p \geq B$ by $p \geq 0$ does not make a substantial difference to the outcome.

The benchmarking application is a useful byproduct of this work. Most authors display ROC curves for their detectors, but of course it is difficult to compare the performance of two detectors when their ROC curves cross (as they often do). Furthermore, one has to compare ROC curves from a variety of different payload sizes. The results of Sects. 2 and 3 provide a partial answer: the root of (2) determines the point on the ROC curve corresponding to optimal use of the individual steganalysis method, at least when the detector is used for pooled steganalysis with the Warden playing the Threshold Game and either the moves are simultaneous or the Warden moves first. Moreover, if the constraint $p \geq B$ is relaxed to $p \geq 0$ then the players' behaviour is independent of the embedding rate.

We are thus able to find a simple linear benchmark to compare some of the leading quantitative steganalysis methods for the LSB replacement embedding algorithm.[2–6] For some large sets of cover images we compute an empirical density function for the steganalysis estimator when no data is embedded, and deduce a numerical solution for (2). In Tab. 1 we show some of these results, displaying the optimal value t for the Warden's strategy in the Threshold Game, and the corresponding false positive rate which this induces on the detection in individual images. Most important is the value of χ_1 at α. Recall that the median p-value for the hypothesis test (1) is $\Phi(-B\sqrt{N}\chi_1(\alpha))$; therefore the value of χ_1 at α is an overall summary for the performance of the detector, at least as it applies to pooled steganalysis with the Warden playing the Threshold Game. Tab. 1 demonstrates the differences between performance on never-compressed and previously-compressed cover images, and also the superiority of the more recent detectors[†]. It is also interesting that the optimal use of these detectors is where their individual-image false positive rates are around 1-10%, which is often the region tested in the literature.

Finally, we examine the effect of our simplified assumption, in Sect. 2, that the range of p was $[0, 1]$ when in fact it is $[B, 1]$ (this is a constraint in the Batch Steganography problem). If the Warden were somehow to

[†]We note that the validity of these results do rely on the shift hypothesis, plus some other properties of the distribution functions, which may not be exactly satisfied by real steganalysis methods. The Quadruples detector[6] matches the assumptions least accurately.

Table 1. Empirically determined values of α, and the corresponding per-image false positive rate and payoff $\chi_1(\alpha)$, for some LSB replacement steganalysis estimators. The experiments have been repeated with two sets of 3000 covers: one of never-compressed grayscale images and the other of previously JPEG-compressed (quality factor 90) colour images.

	Never-compressed grayscale covers			Previously compressed colour covers		
Estimator	Optimal Threshold α	False Positive Rate (%)	$\chi_1(\alpha)$	Optimal Threshold α	False Positive Rate (%)	$\chi_1(\alpha)$
RS[2]	0.0509	2.74	5.95	0.1846	10.95	2.85
SPA[3]	0.0527	2.08	6.85	0.1672	9.36	3.11
SPA/LSM[4]	0.0594	2.35	6.44	0.1276	2.82	5.88
Triples[5]	0.0452	2.15	6.74	0.0612	2.08	6.85
Quadruples[6]	0.0281	1.89	7.31	0.0660	2.41	6.35

know the Steganographer's desired bandwidth then this would cause a slight change in their behaviour: they should select t as the root β of $\chi_B(t) = \chi_0(t)$, rather than α the root of $\chi_0(t) = \chi_1(t)$, because they know that their opponent is constrained by $p \geq B$. Considering Fig. 1 we see that β will be a little higher than α, and the Warden's corresponding payoff a little better than if they had used $t = \alpha$. Similarly, the Steganographer should choose between $p = 0$ and $p = B$ according to whether t is above or below β, not α. The latter presents no problems: the Steganographer knows his own required bandwidth (payload). But it seems very unlikely that a Warden could know the potential payload embedded by their opponent, so we quantify the loss caused to the Warden by this information asymmetry.

Let us suppose that B is in fact 0.01, and consider whether the root α of $\chi_0(t) = \chi_1(t)$ differs significantly from the true optimum β, the root of $\chi_B(t) = \chi_1(t)$. We compute such values β for the same steganalysis methods as in the previous experiment, and also the corresponding payoffs $\chi_1(\beta)$, and display the results in Tab. 2 (using this time only the never-compressed set of 3000 cover images). We see that the optimal thresholds change only a small amount, and that the Warden loses around 5-10% on their payoff (this equates to requiring 5-10% larger payloads before detection meets a given reliability level). This loss varies, of course, with B but note that $B = 0.01$ is quite a *large* bandwidth in the batch setting: as long as N is at least a few thousand, such a size of payload can be detected with very high reliability.

5. CONCLUSIONS

Before drawing conclusions, we return to the assumptions in Subsect. 1.2.

5.1. Commentary on Assumptions

We have seen, in Sect. 4, that the accuracy of the results obtained in Sects. 2 and 3 is not highly sensitive to the accuracy of the assumptions. The experiments were performed on genuine quantitative estimators, which do

Table 2. Comparison of numerically-computed roots α and β of the equations $\chi_0(t) = \chi_1(t)$ and $\chi_B(t) = \chi_1(t)$ with $B = 0.01$, along with their associated payoffs $\chi_1(\alpha)$ and $\chi_1(\beta)$, for a range of steganalysis methods.

Estimator	α	β	$\chi_1(\alpha)$	$\chi_1(\beta)$
RS	0.0509	0.0550	5.95	6.14
SPA	0.0527	0.0579	6.85	7.41
SPA/LSM	0.0594	0.0645	6.44	7.15
Triples	0.0452	0.0505	6.74	7.35
Quadruples	0.0281	0.0332	7.31	8.29

not exactly satisfy either the shift hypothesis or the assumptions on the nature of steganalysis error (e.g. their null distributions are not perfectly symmetric). It is worthwhile to consider whether the assumptions are at all reasonable.

First we should note that the Threshold Game itself is by no means the only possible framework for the interaction between the Steganographer's embedding strategy and the Warden's pooling strategy. Other pooling methods, which do not throw away so much information, are likely to be superior. Nonetheless, authors often display ROC curves for the classification of covers and stego objects by quantitative steganalysis methods, in which case it is assumed that a simple threshold is set.

Second, the shift hypothesis is a crucial component of this work. Thankfully, it seems that very many steganalysis methods are, or could be adapted to be, quantitative. Certainly we might expect that any steganalysis which first extracts some feature vector, prior to classification, will have a response which is approximately linear in the level of stego noise; unless some source coding method is used (e.g. *matrix embedding*[12]) this should in turn be proportional to payload size. However we should note that the shift hypothesis is not reasonable in the presence of source coding which adapts to the relative payload size, and other work will be necessary to examine the Batch Steganography problem if such embedding methods are permitted.

Finally, what of the assumptions (A1)-(A5) on the nature of steganalysis error? We could reasonably expect (A1)-(A3) to be true of any sensible error distribution (although (A1) rules out the Laplace distribution, this light tailed distribution has not been observed in steganalysis). Similarly, (A4) seems a rather gentle assumption; indeed we have not noticed any error distribution for which it fails, leading us to suspect that (A4) in fact follows from (A1)-(A3), perhaps with some additional weak conditions. (A5) is not in a very natural form, and can be rather difficult to establish even for well-known distributions. Thankfully both (A4) and (A5) have the property that they are *scale-free*, i.e. that if true for some random variable X they remain true for λX, for all $\lambda \in \mathbb{R}$. Hence we can at least check numerically that (A5) is true for common distributions including Gaussian, Cauchy and the rest of the Student t-family, Logistic, and so on.

5.2. Conclusions

The Threshold Game is a natural application of threshold-based steganalysis to multiple objects. It becomes one of "cat and mouse" with the Warden wanting to adapt their choice of t best to detect the Steganographer, and the Steganographer choosing p to evade detection. Solutions to versions of the game, with sequential or simultaneous moves, indicate that the Steganographer should either spread the payload as thinly as possible or concentrate as much as possible, or a mixture of these, but never adopt an intermediate strategy. In the case of some quantitative detectors for LSB replacement, empirically-determined optimal strategies have been shown to coincide closely with the theoretical predictions. This work also suggests new ways to create simple benchmarks for quantitative steganalysis methods' ability to classify cover objects and stego objects.

We emphasise, however, that the Threshold Game represents only a special case of the Batch Steganography problem, and indeed that our metric for detector performance (median p-value) is not automatically the correct one. Analysis of more sophisticated pooling strategies for the Warden, or alternative measures of performance, are likely to lead to more difficult mathematical challenges.

ACKNOWLEDGMENTS

The author is a Royal Society University Research Fellow.

REFERENCES

1. A. Ker, "Batch steganography and pooled steganalysis," To appear in *Proc. 8th Information Hiding Workshop*, 2006.
2. J. Fridrich, M. Goljan, and R. Du, "Reliable detection of LSB steganography in color and grayscale images," *Proc. ACM Workshop on Multimedia and Security*, pp. 27–30, 2001.
3. S. Dumitrescu, X. Wu, and Z. Wang, "Detection of LSB steganography via sample pair analysis," in *Proc. 5th Information Hiding Workshop*, *Springer LNCS* **2578**, pp. 355–372, 2002.

4. P. Lu, X. Luo, Q. Tang, and L. Shen, "An improved sample pairs method for detection of LSB embedding," in *Proc. 6th Information Hiding Workshop*, *Springer LNCS* **3200**, pp. 116–127, 2004.
5. A. Ker, "A general framework for the structural steganalysis of LSB replacement," in *Proc. 7th Information Hiding Workshop*, *Springer LNCS* **3727**, pp. 296–311, 2005.
6. A. Ker, "Fourth-order structural steganalysis and analysis of cover assumptions," in *Security, Steganography and Watermarking of Multimedia Contents VIII*, E. J. Delp III and P. W. Wong, eds., *Proc. SPIE* **6072**, pp. 25–38, 2006.
7. W. Feller, *An Introduction to Probability Theory and its Applications, Volume II*, Wiley, second ed., 1971.
8. A. Ker, "A capacity result for batch steganography." To appear in *IEEE Signal Processing Letters*, 2007.
9. A. Dixit and S. Skeath, *Games of Strategy*, W. W. Norton and Company, 1999.
10. O. Morgenstern and J. von Neumann, *The Theory of Games and Economic Behavior*, Princeton University Press, 1947.
11. G. Dantzig, *Linear Programming and Extensions*, Princeton University Press, 1963.
12. J. Fridrich and D. Soukal, "Matrix embedding for large payloads," in *Security, Steganography and Watermarking of Multimedia Contents VIII*, E. J. Delp III and P. W. Wong, eds., *Proc. SPIE* **6072**, 2006.