LLRT BASED DETECTION OF LSB HIDING

K. Sullivan, O. Dabeer, U. Madhow, B.S. Manjunath, and S. Chandrasekaran

Dept. of Electrical and Computer Engineering
University of California at Santa Barbara
Santa Barbara CA 93106

ABSTRACT

In this paper we consider a hypothesis testing approach for detection of hiding in the least significant bit (LSB). This steganalysis problem is a composite hypothesis testing problem. We state a regularity condition on the image histogram, which reduces this problem to a simple hypothesis testing problem. We then develop a number of simple practical tests based on the estimation of the optimal log likelihood ratio statistic. We show that our tests significantly outperform Stegdetect, a popular hypothesis test available in the literature. Our approach also leads to good estimates of the hiding rate.

1. INTRODUCTION

Given the proliferation of steganography tools (see, for example, [1, 2, 3, 4]), there is a growing interest in steganalysis tools, which detect hidden data in multimedia. So far, steganalysis research has lagged behind steganography. While there exist promising approaches such as Stegdetect ([5]), Farid's supervised learning framework ([6]), and others ([7, 8, 9, 10]), there is no systematic approach for designing steganalysis tools or analyzing the optimality of such methods. Also, every steganalysis tool has some parameters, which have to be chosen in practice. Ideally, these parameters should be chosen based on the data to attain desired performance. Such data-driven tests are not known in the literature. With these goals in mind, in this paper we start a systematic study of the hypothesis testing approach for steganalysis and develop practical schemes based on the optimal hypothesis tests. We focus our attention on LSB hiding, which is often used in practice (see, for example, [4]). However, in principle our ideas are applicable to steganalysis of any steganography scheme with a statistical description.

The theory of hypothesis testing (see, for example, [11, 12]) provides a natural framework for steganalysis. For example, hypothesis tests for detecting LSB hiding have been developed in [10, 5]. In [10], LSB hiding is (inaccurately) modeled as additive noise for a Gaussian host. In [5], for *actual* LSB hiding in a finite precision host, a test based on the chi-square statistic called Stegdetect is proposed. To the best of our knowledge, Stegdetect appears to be the most promising hypothesis test for steganalysis. Unfortunately, it suffers from the drawback that there is no simple way to choose the test threshold to attain a desired target performance. We show that this problem can not only be alleviated, but we can also significantly improve over Stegdetect by developing simple tests based on optimal hypothesis tests. Furthermore, we also obtain good estimates of the unknown hiding rate.

In Section 2, we describe the basic statistical model for LSB hiding and the host. In Section 3, we formulate the hypothesis testing problem for steganalysis. Since in practice we do not know the hiding rate, this is a *composite* hypothesis testing problem. We state a mild regularity condition on the host probability mass function (PMF), which ensures that the optimal composite hypothesis testing problem reduces to an optimal *simple* hypothesis testing problem. In Section 4, we develop new steganalysis methods based on the optimal simple hypothesis tests and exhibit their superiority over Stegdetect. We also develop estimates of the true hiding rate based on these tests. In Section 5, we state our conclusions.

2. STATISTICAL MODEL FOR LSB HIDING

In this section, we provide a probabilistic description of the host and the LSB hiding mechanism, which is central to the study of statistical steganalysis tools. As a first step, we consider the case of independent and identically distributed (i.i.d.) data samples. This model is commonly used in steganography ([13], [14]). Since the host samples are assumed to be i.i.d., without loss of generality we assume the data to be one dimensional. Suppose the i.i.d. host is $\{h_k\}_{k=1}^N$, where the intensity values h_k are represented by 8 bits, that is, $h_k \in \{0, 1, ..., 255\}$. We use the following model for LSB data hiding with rate R bits per host sample. The hidden data $\{d_k\}_{k=1}^N$ is i.i.d. and,

$$P(d_k = 0) = \frac{R}{2}, \quad P(d_k = 1) = \frac{R}{2},$$
$$P(d_k = \text{NULL}) = (1 - R), \quad 0 < R \leq 1.$$

The hider does not hide in host sample h_k if $d_k = \text{NULL}$, otherwise the hider replaces the LSB of h_k with d_k. With this model for rate R LSB hiding, if the probability mass function (PMF) of h_k is $p(n)$, $n = 0, 1, ..., 255$, then the

This research was supported in part by a grant from ONR #N0014-01-1-0380. Email of corresponding author: sullivak@ece.ucsb.edu

PMF of the data after LSB hiding at rate R is given by,

$$p_R(2l) = \quad 1 - \frac{R}{2} \quad p(2l) + \frac{R}{2} p(2l+1),$$
$$p_R(2l+1) = \frac{R}{2} p(2l) + \quad 1 - \frac{R}{2} \quad p(2l+1).$$

where $l = 0, 1, ..., 127$. For the sake of convenience, we denote the PMF by the 256-dimensional vectors p, p_R, and we write $p_R = Q_R p$, where Q_R is a 256×256 matrix corresponding to the above linear transformation.

The above statistical model can be easily extended to take higher order dependence into consideration. Consider, for example, the joint PMF of neighboring pixels. If we denote this by the 256×256 matrix P, then upon i.i.d. LSB hiding with rate R as described above, the joint PMF is $P_R = Q_R P Q_R$. This idea can be extended to find transformations for any arbitrary order of dependence. In this paper, however, we only consider the case of i.i.d. observations.

3. OPTIMAL COMPOSITE HYPOTHESIS TESTING FOR STEGANALYSIS

In this section, we assume that the host PMF is known to the detector; the lessons learnt are used in the next section to design practical steganalysis schemes which do not assume knowledge of the host PMF.

The theory of hypothesis testing (see, for example, [11, 12]), provides a natural framework for steganalysis. In this approach, the observed data (say an image) is viewed as a realization of a random process. A random process is completely characterized by its probability law and therefore the two hypotheses (presence or absence of hidden data) can be tested based on the probability law of the observed data. An advantage of this approach is that it enables us to study the limits of steganalysis. In this section, for the i.i.d. host and i.i.d. LSB hiding described in Section 2, we study the composite hypothesis testing problem associated with steganalysis.

Suppose we wish to decide between two possibilities: data is hidden at some rate R, where $R_0 \leq R \leq R_1$, or no data is hidden ($R = 0$). The parameters $0 < R_0 \leq R_1 \leq 1$ are specified by the user. We note that R_0 must be strictly positive or else the two hypotheses cannot be distinguished. We use H_R to represent the hypothesis that data is hidden at rate R. The steganalysis problem in this notation is to distinguish between H_0 and $K(R_0, R_1) := \{H_R : R_0 \leq R \leq R_1\}$. The hypothesis that data is hidden is thus *composite* while the hypothesis that nothing is hidden is *simple*. Suppose the observed data is $\{x_j\}_{j=1}^N$, where x_j are i.i.d. and take values in some alphabet $\mathcal{A}$. For grey-scale images, $\mathcal{A} = \{0, 1, ..., 255\}$. Mathematically, a detector δ is characterized by the acceptance region $A \in \mathcal{A}^N$ of hypothesis H_0:

$$\begin{aligned} \delta(x_1, ..., x_N) &= H_0, \text{ if } (x_1, ..., x_N) \in A, \\ &= K(R_0, R_1), \text{ otherwise.} \end{aligned}$$

In the absence of an apriori distribution on R when data is hidden, we use the Neyman-Pearson formulation of the optimal detection problem: for $\alpha > 0$ given, minimize

$$P\,(\text{Miss}) = \sup_{R_0 \leq R \leq R_1} P\,(\delta(x_1, ..., x_N) = H_0 | H_R)$$

over detectors δ which satisfy

$$P\,(\text{False alarm}) = P\,(\delta(x_1, ..., x_N) = K(R_0, R_1) | H_0) \leq \alpha.$$

Suppose that the host PMF satisfies the following 'smoothness' constraint.

$$U(p) := \sum_{k=0}^{127} \quad (p_{2k} + p_{2k+1}) \quad r_k + \frac{1}{r_k} - 2 \qquad < 1, \qquad (1)$$

where $r_k := p_{2k+1}/p_{2k}$. In our related paper [15], we prove that under this regularity condition, the optimal composite hypothesis is solved by the simple hypothesis testing problem: test H_0 versus H_{R_0}. The optimal test for this problem is well-known - it is the log likelihood ratio test (LLRT). Let q denote the empirical PMF (normalized histogram) of the observed data and $D(p\|q)$ is the Kullback-Leibler divergence between the PMFs p and q defined as,

$$D(p\|q) = \sum_{k=0}^{255} p_k \log \quad \frac{p_k}{q_k} \quad .$$

Then the LLRT test declares data to be hidden if

$$D(q\|Q_{R_0}p) - D(q\|p) \leq T(\alpha), \qquad (2)$$

and otherwise declares no data to be hidden. Here $T(\alpha)$ is a real-valued threshold chosen to obtain $P\,(\text{False alarm}) = \alpha$.

To understand the 'smoothness' condition (1), consider the function $f(x) = x + 1/x - 2$. This function has a minimum value zero at $x = 1$ and it monotonically increases as x increases or decreases away from $x = 1$. The condition (1) therefore means that on an average, the ratio p_{2k+1}/p_{2k} is not too large or too small. This assumption would be satisfied for images whose histogram varies smoothly. We have verified that it is true for a digital orthophoto quarter quadrangle (DOQQ) image database with 4000 images.

4. TESTS BASED ON LLRT

Given the discussion in the previous section, we now restrict our attention to the simple hypothesis testing problem: test H_0 versus H_R, $R > 0$. We propose tests based on the estimation of the LLRT statistic and exhibit their superiority over Stegdetect. We also develop estimates of the hiding rate R.

4.1. Estimating the LLRT Statistic

A problem with the optimal LLRT test is that we do not know the host PMF in practice. However, there are two factors that help us to develop good practical tests based on the optimal LLRT.

1. The hiding rate in practice is very low, and therefore, we can estimate the host PMF well; the perturbations introduced by LSB hiding are much smaller compared to the host PMF. We show below that a number of simple estimates of the host PMF based on the assumption that the host PMF is 'smooth' work well.

2. For the optimal LLRT, the threshold that minimizes

$$aP(\text{Miss}) + (1-a)P(\text{False alarm}),\ a \in [0,1]$$

does not depend on the host. In particular, for $a = 0.5$, the optimal threshold $T = 0$. In contrast, for Stegdetect the choice of the threshold depends on a and the host PMF, and there is no known way of making this choice.

With the above motivation, we propose to form an estimate $\hat{p}$ of the host PMF p and then use the following estimated version of the statistic in (2) as an approximate LLRT statistic:

$$S(q) = D(q \| Q_R\hat{p}) - D(q\|\hat{p}).$$

We consider three possible estimates for p, all of which give good results.

1. For natural images the PMF is usually low pass. On the other hand, random LSB hiding introduces high frequency components in the histogram. Hence one simple estimate $\hat{p}$ is to pass the empirical PMF q though a low pass 2-tap FIR filter with taps $(0.5, 0.5)$. We note that normalization will be required after the filtering.

2. Another regularity constraint that we can impose on the host PMF is that local slope is preserved , that is,

$$p_{k+3} - p_k = 3(p_{k+2} - p_{k+1}), k = 0, 4, 8, ..., 252.$$

This regularity constraint can be written as $Ap = 0$ for a suitable 64×256 matrix A. Under this regularity constraint, a natural estimate of p is to project q on to the null space of A. We again need normalization and removal of negative components after this filtering.

3. We also propose a non-linear approach that adapts to the underlying host PMF. We note that LSB hiding only affects the 8^{th} bit plane. Therefore, we impose the regularity constraint that the host PMF is such that we can obtain the host PMF by spline interpolation of the first seven bit planes. The corresponding estimate $\hat{p}$ is obtained by subsampling q, then interpolating using splines, and then normalizing.

We refer to all these tests as the approximate LLRT.

4.2. Simulation Results

In this section we report and discuss a number of simulation results for four thousand images from a DOQQ image set.

In Figure 1 we compare the approximate LLRT test based on the half-half filter for estimating p with Stegdetect. For each point on the curve, the threshold has been fixed over the entire database. At this rate, and other rates we test, the LLRT outperforms Stegdetect. For a fixed host PMF, both these tests perform closely. However, for the database of images we have used, the host PMF varies substantially from image to image. Thus these simulations suggest that Stegdetect is more sensitive to the choice of the threshold than our approximate LLRT test. This is not surprising since we know that to attain a target performance, the choice of the threshold in LLRT does not depend on the host PMF. For example, if we choose $T = 0$ for the approximate LLRT in the case when the hiding rate is 0.05, then we found the operating point to be $P(\text{Miss}) = 0.4043$ and $P(\text{False Alarm}) = 0.3219$. From Figure 1 we can verify that the tangent to the operating curve at this point is of slope approximately 1 as predicted by the theory. The approximate LLRT is therefore closer to the goal of finding a data driven test.

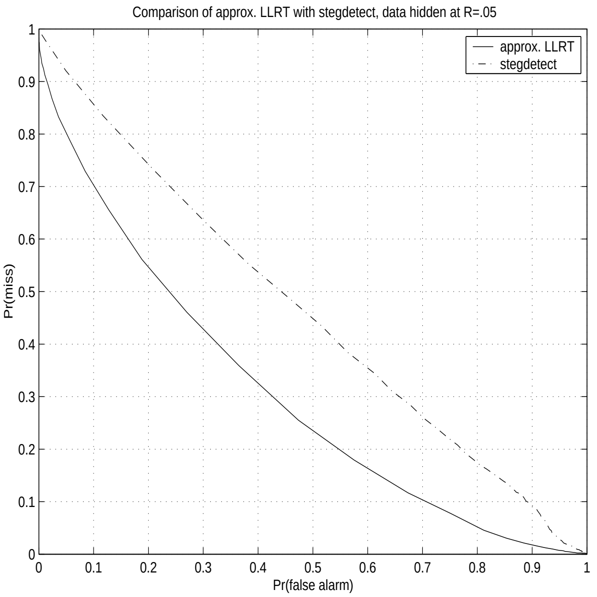

Fig. 1. Approximate LLRT with half-half filter estimate versus Stegdetect: for any threshold choice, our approximate LLRT is superior.

Figure 2 shows that the story remains unchanged if we hide in the LSB of the JPEG coefficients of images (compressed with quality factor 75).

In principle, instead of the simple hypothesis tests as above, we could use the following generalized LLRT (GLLRT) ([12]) type test:

$$\max_{R_h \in (0,1]} \log \frac{p(y|H_0)}{p(y|H_{R_h})} \lesseqgtr T. \qquad (3)$$

This GLLRT performs very close to the (simple) approximate LLRT tests we have developed (which use R_0 instead of R_h). This is not surprising given our result in [15], which states that the optimal composite hypothesis testing problem considered in Section 3 is solved by the simple hypothesis testing problem under the mild constraint (1).

Additionally, we can use the argument R_h that maximizes (3) as an estimate of the actual embedding rate. We find this to work reasonably well in practice, see Figure 3.

Finally, we compare the approximate LLRT scheme based on different estimates of p. The spline estimates of p and the half-half low pass filter estimates perform nearly identically. We have observed that the local slope preserving filter is slightly worse off. This suggests that there might be little to gain from choosing a different host estimate.

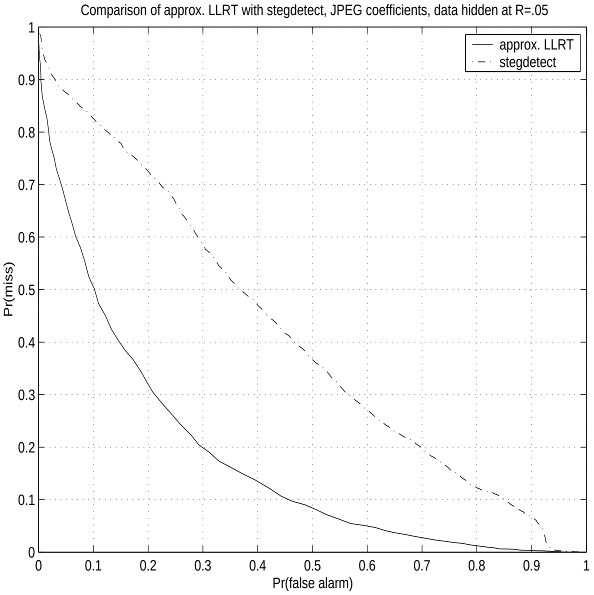

Fig. 2. Hiding in the LSBs of JPEG coefficients: again LRT based method is superior to Stegdetect.

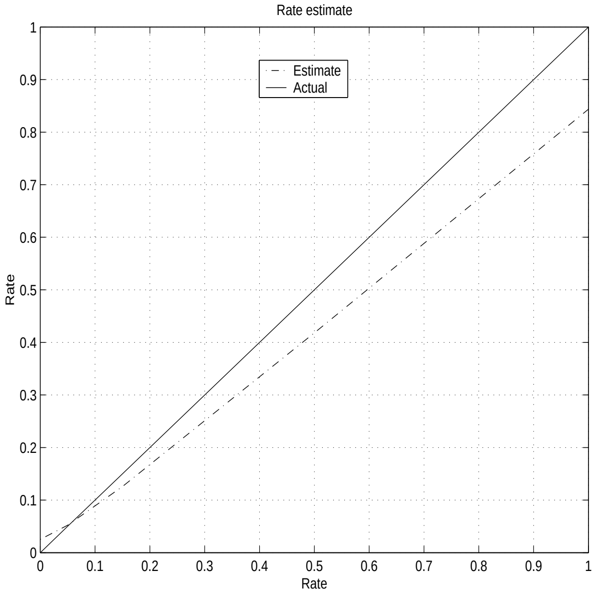

Fig. 3. The rate that maximizes the LRT statistic estimate serves as an estimate of the hiding rate.

5. CONCLUSION

By analyzing the optimal hypothesis test for steganalysis and making justified assumptions about the PMFs of typical real images, we have formulated tools for detecting LSB steganography. Our method performs better than previous hypothesis testing approaches in two ways.

1. They lead to smaller probability of miss for the same probability of false alarm.
2. The choice of the threshold is less sensitive to variations in the host PMF. Thus for typical hiding rates less than 0.1, the choice of threshold $T = 0$ leads to good performance.

Our approach is not limited to LSB hiding alone: these ideas are suitable for any hiding scheme with a good statistical description.

6. REFERENCES

[1] N. Jacobsen, K. Solanki, U. Madhow, B. S. Manjunath a, and S. Chandrasekaran, "Image-adaptive high-volume data hiding based on scalar quantization," in *Proceedings of IEEE Military Communications Conference (MILCOM)*, Anaheim, CA, USA, October 2002.

[2] B. Chen and G.W. Wornell, "Quantization index modulation: A class of provably good methods for digital watermarking and information embedding," *IEEE Trans. Info. Theory*, vol. 47, no. 4, pp. 1423–1443, May 2001.

[3] N. Provos, "Defending against statistical steganalysis," in *In 10th USENIX Security Symposium*, Washington DC, 2001.

[4] http://hacktivismo.com/projects/camerashy/.

[5] N. Provos and P. Honeyman, "Detecting steganographic content on the internet," in *ISOC NDSS'02*, San Diego, CA, 2002.

[6] H. Farid, "Detecting stenographic messages in digital images," Tech. Rep., Dartmouth College, Computer Science, 2001.

[7] J. Fridrich, M. Goljan, and D. Hogea, "Steganalysis of JPEG images: Breaking the F5 algorithm," in *5th International Workshop on Information Hiding*, 2002.

[8] J. Fridrich, M. Goljan, and D. Hogea, "Attacking the Outguess," in *Proceedings of ACM Workshop on Multimedia and Secturity*, Juan-Pins, France, dec 2002.

[9] I. Avcibas, N. Memon, and B. Sankur, "Image steganalysis with binary similarity measures," in *Proceedings of ICIP*, 2002.

[10] R. Chandramouli and N. Memon, "Analysis of LSB based image steganography techniques," in *Proceedings of ICIP*, 2001, pp. 1019–1022.

[11] E. Lehmann, *Testing Statistical Hypothesis*, John Wiley, New York, 1959.

[12] V. Poor, *An introduction to signal detection and estimation*, Springer, NY, 1994.

[13] M.H.M. Costa, "Writing on dirty paper," *IEEE Trans. Info. Theory*, vol. IT-29, no. 3, pp. 439–441, May 1983.

[14] P. Moulin and J.A. O'Sullivan, "Information-theoretic analysis of information hiding," preprint, Dec. 2001.

[15] O. Dabeer, K. Sullivan, U. Madhow, S. Chandrasekaran, and B.S. Manjunath, "Detection of hiding in the least significant bit," 2003, Proceedings of CISS.