ANALYSIS OF LSB BASED IMAGE STEGANOGRAPHY TECHNIQUES

R. Chandramouli

MSyNC Lab
Stevens Institute of Technology
Dept. of Electrical and Computer Engineering
Hoboken, NJ 07030

Nasir Memon

Polytechnic University
Computer Science Department
Brooklyn, NY 11201

ABSTRACT

There have been many techniques for hiding messages in images in such a manner that the alterations made to the image are perceptually indiscernible. However, the question whether they result in images that are statistically indistinguishable from untampered images has not been adequately explored. In this paper we look at some specific image based steganography techniques and show that an observer can indeed distinguish between images carrying a hidden message and images which do not carry a message. We derive a closed form expression of the probability of detection and false alarm in terms of the number of bits that are hidden. This leads us to the notion of steganographic capacity, that is, how many bits can we hide in a message without causing statistically significant modifications? Our results are able to provide an upper bound on the this capacity. Our ongoing work relates to *adaptive* steganographic techniques that take explicit steps to foil the detection mechanisms. In this case we hope to show that the number of bits that can be embedded increases significantly.

1. INTRODUCTION

Steganography is most widely formulated in terms of the prisoner's problem where Alice and Bob are two inmates who wish to communicate in order to hatch an escape plan. However, all communication between them is examined by the warden, Wendy, who will put them in solitary confinement at the slightest suspicion of trouble. Specifically, in the general model for steganography, we have Alice wishing to send a secret message M to Bob. In order to do so she "embeds" M into a *cover-object* C, to obtain the *stego-object* S. The stego-object S is then sent through a public channel. The warden Wendy who is free to examine all messages exchanged between Alice and Bob can be *passive* or *active*. A passive warden simply examines the message and tries to determine if it potentially contains a hidden message. If it appears that it does, then she takes appropriate action else she lets the message through without alteration. An active warden on the other hand can alter messages deliberately, even though she does not see any trace of a hidden message, in order to foil any secret communication that can nevertheless be occurring between Alice and Bob.

Given the above framework, the main goal of steganography is to communicate securely in a completely undetectable manner. That is, Wendy should not be able to reliably distinguish in any sense between cover-objects (objects not containing any secret message) and stego-objects (objects containing a secret message). In this context, *steganalysis* refers to the body of techniques that are designed to distinguish between cover-objects and stego-objects.

Over the past few years, numerous steganography techniques that embed hidden messages in multimedia objects have been proposed. This is largely due to the fact that multimedia objects often have a highly redundant representation which usually permits the addition of significantly large amounts of stego-data by means of simple and subtle modifications that preserve the perceptual content of the underlying cover object. Hence, they have been found to be perfect candidates for use as cover messages. In fact, there are many freely available public domain tools that embed a message into images, audio and video files in different formats. Indeed, if steganography were defined to mean that the perceptual content of the cover object is indistinguishable from the stego-object then such techniques can be deemed to be quite effective and efficient.

However, it should be noted that classical definition of steganography is *statistical* and not *perceptual*. In fact, Cachin [1] has defined a steganography technique to be ϵ-secure if the relative entropy of the probability distribution of cover-objects and stego-objects is less than or equal to ϵ. He calls a steganography technique to be perfectly secure if ϵ is zero. He then demonstrates that there do exist steganographic techniques that are perfectly secure (however, the technique described by him is impractical).

Returning to the issue of image based steganography, as we mentioned before, there have been many techniques for hiding messages in images in such a manner that the al-

0-7803-6725-1/01/$10.00 ©2001 IEEE

terations made to the image are perceptually indiscernible. However, the question whether they result in images that are statistically indistinguishable from untampered images has not been adequately explored. In this paper we look at a specific class of widely used image based steganographic techniques, namely LSB steganography and investigate under what conditions can an observer distinguish between stego-images (images which carry a secret message) and cover-images (image that do not carry a secret message). We derive a closed form expression of the probability of false detection in terms of the number of bits that are hidden. This leads us to the notion of steganographic capacity, that is, how many bits can we hide in an image using LSB based techniques, without causing statistically significant modifications? Our results are able to provide an upper bound on this capacity. Ongoing work is looking at *adaptive* steganographic techniques that take explicit steps to foil the detection mechanisms. In this case we hope to show that the number of bits that can be embedded increases significantly.

Before we proceed, however, we would like to note that although we focus on images in this paper, the techniques we employ and the results we obtain are quite general and apply equally well to other types of multimedia data like audio and video. Also, we would like to note that although our work is in the context of secret communication (steganography) in the presence of a passive warden, the techniques and results are very applicable to the more active areas of watermarking and data hiding, which can be viewed as steganography in the presence of an active warden.

2. LSB BASED STEGANOGRAPHY

In the framework of a passive warden who does not alter images, LSB based steganography is perhaps the most simple and straightforward approach. Here you embed the message into the least significant bit plane of the image. Since this will only effect each pixel by ± 1, if at all, it is generally assumed with good reason that the degradation caused by this embedding process would be perceptually transparent. Hence they are a number of LSB based steganography techniques available in the public domain. See [2] for an excellent survey of such techniques.

LSB based techniques pose a difficult challenge to a steganalyst in the passive warden model as it is difficult to differentiate cover-images from stego-images, given the small changes that have been made. Of course with an active warden, such techniques can be easily defeated by randomizing the LSB. Nevertheless, steganalysis of LSB based techniques is important even for the active warden scenario as the tools developed would be also applicable to steganalysis of this "easier" (from the steganalysts viewpoint) framework.

Now if the message is embedded as is into the LSB of the cover-image, then the resultant structure in the LSB plane of the stego-image would clearly be a giveaway [2]. Hence to maintain a random looking appearance of the LSB, it is suggested that the message be encrypted before it is embedded. In fact a more sophisticated approach would be to randomly insert the encrypted message only into a subset of the pixels in the cover-image [2]. In this case, the natural question that arises is that how many bits can be embedded before the warden, Wendy is able to reliably distinguish between cover-images and stego-images. In the rest of this paper we give for the first time (to the best of our knowledge), a rigorous approach towards answering this question.

3. PROBLEM DEFINITION

In this section we present a mathematical formulation for analysis of LSB based steganographic techniques. LSB based steganographic techniques either change the pixel value by ± 1 or leave them unchanged. This is dependent both on the nature of the hidden bit and the LSB of the corresponding pixel value. Let $I = \{x_i, \ i \in \Omega\}$ where Ω is an index set denote the mean subtracted cover image. The set Ω can be partitioned into three subsets Λ_1, Λ_2, and Λ_3, where, $\Omega = \bigcup_{i=1}^{3} \Lambda_i$ and $\Lambda_i \bigcap \Lambda_j = \phi$ for $i \neq j$. Then, the pixel values in a LSB based stego-image, $I_s = \{y_i, \ i \in \Omega\}$ can be represented as

$$y_i = \begin{cases} x_i + 1 & \text{if } i \in \Lambda_1 \\ x_i - 1 & \text{if } i \in \Lambda_2 \\ x_i & \text{if } i \in \Lambda_3 \end{cases} \quad (1)$$

The goal of a steganalyst is to estimate if I has hidden data. In this case, this goal can be translated to finding if Λ_1 and Λ_2 are non-empty, if so, what are the elements of these two sets? If this can be done, then it will be possible to detect the presence or absence of hidden data within the cover image, I. Ofcourse, during this process, a steganalyst will incur errors due to statistical uncertainties and partial or no *a priori* knowledge of the data hiding key. The magnitudes of these errors depend on the cardinality of the Λ_i's, $i = 1, 2, 3$. This gives rise to the following question: how many bits can be hidden in I such that a steganalyst cannot detect their presence with a desired probability? The answer to this question will provide a measure of the **steganographic capacity** of I. We attempt to shed some light on this important question in the rest of this paper.

We make the following simplistic but realistic assumptions in order to compute the steganographic capacity. x_i, $i \in \Omega$ is Gaussian distributed with zero mean and variance σ^2 (*i.e.*, $x_i \sim N(0, \sigma^2)$). The steganalysis process can then be formulated as the following multiple hypothesis testing problem for each $i \in \Omega$:

$$H_j : y_i = x_i + d_i, \ \ j = 1, 2, 3 \quad (2)$$

where $d_i = -1, 0$ or 1. This means $H_1 : y_i \sim N(1, \sigma^2)$, $H_2 : y_i \sim N(-1, \sigma^2)$, and $H_3 : y_i \sim N(0, \sigma^2)$. We note that under H_3 there is no statistical difference between the stego-image and the cover image. This leaves the steganalyst with detecting H_1 and/or H_2. If either of this hypothesis is detected it means the image contains hidden data. We note that we can safely ignore the case where all the data bits are equal to the LSB's of the corresponding pixels of the cover image match so that no modification is done to that LSB. This can be explained as follows. Let us suppose the probability of a data bit equal to 1 is $0 < p_d < 1$ and the probability of a LSB (denoted by l_i) being 1 is equal to $0 < p_l < 1$. Assuming the hidden bits and the LSB's are independent of each other, the joint probability,

$$P(d_1 = 1, l_1 = 1, \ldots, d_{|\Omega|}, l_{|\Omega|}) = (p_d p_l)^{|\Omega|} \tag{3}$$

tends to zero as $|\Omega|$ becomes arbitrarily large. Here, $|.|$ denotes the cardinality of a set.

The first step in the steganalysis process is to identify which one of the three hypotheses is true for each pixel. We use the minimum probability of error criterion [3] as the cost function for this process. The minimum probability of error detector can be shown to be a maximum *a posteriori* probability (MAP) detector [3], namely, the true hypothesis is given by,

$$H = \arg\max_j P(H_j) P(y_i | H_j) \tag{4}$$

Since y_i is Gaussian the MAP detector becomes

$$H = \arg\max_j P(H_j) exp \frac{-(y_i - d_j)^2}{2\sigma^2} \tag{5}$$

where $d_j = 1, -1$, or 0 corresponding to H_1, H_2, or H_3. This gives the steganalyst an estimate of the pixel locations that have been modified by data hiding (namely, the pixel positions where H_1 and H_2 are detected). Of course, during this process errors are made. Let us denote these error probabilities by $p_{kj} = P(\text{decide } H_j | H_k \text{ true})$, $j, k = 1, 2, 3$. The values of p_{jk} will depend on the variance of the image and techniques for estimating them are given in our earlier work in [4].

4. STEGANOGRAPHIC CAPACITY COMPUTATION

We now proceed to design a test for the presence of hidden message in the image. Towards this goal, once the first pass of steganalysis is over the second pass is begun. Here, a second detector combines the output decisions of the first pass. The output of this detector will tell us if there is any hidden data at all (with a certain probability). Let u_i denote the decision of the first detector for pixel i and $M = |\Omega|$. Therefore, $u_i = H_j$, $j = 1, 2$ or 3. Using a minimum probability of error criterion we observe that hidden data is detected if,

$$P(H_1) \prod_{i=1}^{M} P(u_i | H_1) \begin{cases} > P(H_2) \prod_{i=1}^{M} P(u_i | H_2) & \text{and} \\ > P(H_3) \prod_{i=1}^{M} P(u_i | H_3) \end{cases} \tag{6}$$

or

$$P(H_2) \prod_{i=1}^{M} P(u_i | H_2) \begin{cases} > P(H_1) \prod_{i=1}^{M} P(u_i | H_1) & \text{and} \\ > P(H_3) \prod_{i=1}^{M} P(u_i | H_3) \end{cases} \tag{7}$$

Both these cases will have the same probability of error due to symmetry. So, we consider only the first case. We make another simplification. Only the detection of H_1 versus H_3 is considered because they are statistically *closer* than H_1 and H_2. So, the multiple hypothesis problem has been simplified to binary hypothesis testing. We now have,

$$\frac{\prod_{i=1}^{M} P(u_i | H_1)}{\prod_{i=1}^{M} P(u_i | H_3)} \begin{cases} > \frac{P(H_3)}{P(H_1)} & \text{decide Hidden Data} \\ \text{else} & \text{decide No Hidden Data} \end{cases} \tag{8}$$

which gives

$$\prod_{S_1} \frac{P(u_i = 1 | H_1)}{P(u_i = 1 | H_3)} \prod_{S_2} \frac{P(u_i = -1 | H_1)}{P(u_i = -1 | H_3)} \prod_{S_3} \frac{P(u_i = 0 | H_1)}{P(u_i = 0 | H_3)} \tag{9}$$
$$\begin{cases} > \frac{P(H_3)}{P(H_1)} & \text{decide Hidden Data} \\ \text{else} & \text{decide No Hidden Data.} \end{cases}$$

This in turn implies,

$$\prod_{S_1} \frac{p_{11}}{p_{31}} \prod_{S_2} \frac{p_{12}}{p_{32}} \prod_{S_3} \frac{p_{13}}{p_{33}} \begin{cases} > \frac{P(H_3)}{P(H_1)} & \text{decide Hidden Data} \\ \text{else} & \text{decide No Hidden Data} \end{cases} \tag{10}$$

Here, S_1, S_2, and S_3 denote the set of pixels where 1, -1, and 0 is detected, respectively. If $P_d = P(\text{decide } H_1 | H_1 \text{ true})$ is the probability of correct detection and $P_f = P(\text{decide } H_1 | H_3 \text{ true})$ denotes the false alarm probability of the steganalyst then we see from Eq. (10) that these quantities are functions of $|S_1|$ and $|S_2|$, the number of hidden bits. there are 2^{3^M} possible detection rules the second detector can employ. This includes the optimal detector also. Sometimes, computing the parameters of the global detection rule may be highly computationally intensive. Therefore, we sacrifice optimality for tractability. In this spirit, suppose the second detector uses a J-out-of-M detection rule (*i.e.*, if J or more out of M decisions favor Hidden Data the steganalyst decides Hidden Data) then

$$P_d = \sum_{k=J}^{M} \sum_{r=0}^{M-k} \frac{M!}{k! r! (n-k-r)!} p_{11}^k p_{12}^r p_{13}^{M-k-r} \tag{11}$$

$$P_f = \sum_{k=J}^{M} \sum_{r=0}^{M-k} \frac{M!}{k! r! (n-k-r)!} p_{31}^k p_{32}^r p_{33}^{M-k-r} \tag{12}$$

Therefore, if we want the steganalyst to only achieve a given value of P_d and P_f then the number of bits that can be hidden reliably under this constraint can be computed using Eq. (12) and solving for J. This then gives us a notion of steganographic capacity. This means that it may not be possible to hide more than J bits in an M pixel image without being detected, given a probability of false detection.

5. NUMERICAL RESULTS

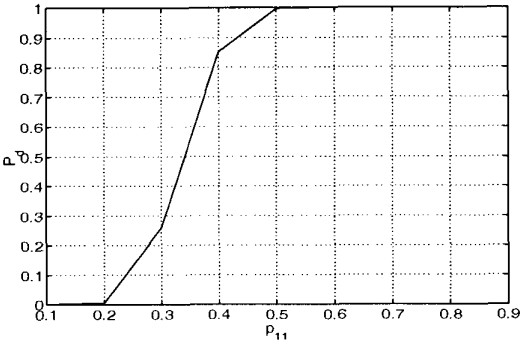

Fig. 1. Performance of steganalysis in terms of p_{11} versus P_d when $J = \lceil (M+1)/2 \rceil$.

We now illustrate the limitations of LSB based image steganography techniques and steganographic capacity computation (as defined previously) with a numerical example when $M = 64$, $p_{12} = 0.05$, $p_{13} = 1 - (p_{11} + p_{12})$, $p_{31} = p_{13}$, $p_{32} = p_{13}$, and $p_{33} = 1 - (p_{31} + p_{32})$. Note that M=64 is not a restrictive assumption because steganalysis can be performed on a block by block basis and we can arrive at 64 blocks for the second stage of detection. Figure 1 shows the performance of the proposed steganalysis technique in terms of p_{11} versus P_d when the second detector uses the MAJORITY decoder ($J = \lceil (M+1)/2 \rceil$). Even though we can design optimal detectors for the second stage of steganalys, heuristic techniques such as MAJORITY logic also produce acceptable performance and are easy to analyze. We see from the figure that as the steganalyst improves the local detection performance (p_{11} increases), the second stage detector is able to detect the presence of LSB data with a very high probability. This means, the steganographic capacity is actually less than 64 bits (one bit/pixel when M=64). In fact, in this numerical example, P_d=0.5 when p_{11}=0.33. This means that the steganalyst is forced to make a random guess regarding the presence of hidden data only when p_{11}=0.33. So, in this case, 44 bits can be reliably hidden (assuming half the number of hidden data do not require replacing a LSB) if the detector in the first stage is forced to achieve p_{11}=0.33. Note that p_{jk}'s depend on the image properties such as the standard deviation of its pixels (σ). So, the image property and the strategy of the steganalyst play a role in determining the data hiding capacity for LSB based schemes. Also, the capacity expression above is an upper bound, since there may be other tests that can be devised to detect the presence of hidden data. For example, in [5], it is shown that classifiers based on image quality metrics are able to quite accurately distinguish between watermarked and unwatermarked images.

6. CONCLUSION

Previous work on steganographic (or watermarking) capacity were based on information and communication theoretic considerations, which are also very valid when considering an active warden who will manipulate the image. In this paper we have taken a novel and rigorous approach at arriving at the steganographic capacity of LSB based image data hiding techniques. We define the capacity in terms of detectibility. That is how many bits can we embed before the warden can start reliably differentiating between stego-objects and cover-objects. Our current formulation is in the framework of a passive warden. In future work we will address an active warden who injects noise in the cover-object.

7. REFERENCES

[1] C. Cachin, "An information-theoretic model for steganography," *Proc. 2nd Information Hiding Workshop*, vol. 1525, pp. 306–318, 1998.

[2] N.F. Johnson, Z. Duric, and S. Jajodia, "Information hiding: Steganography and watermarking - attacks and countermeasures," *Kluwer Academic Publishers*, 2000.

[3] H. Poor, *An introduction to signal detection and estimation*, Springer Verlag, 1994.

[4] R. Chandramouli and N.D. Memon, "A distributed detection framework for watermark analysis," *Proc. ACM Multimedia and Security Workshop*, 2000.

[5] N. D. Memon, I. Avcibas, and B. Sankur, "Steganalysis techniques based on image quality metrics," *Proc. SPIE Security and Watermarking of Multimedia Contents III*, vol. 4314, 2001.