Steganography Capacity: A Steganalysis Perspective

R. Chandramouli[a] and N.D. Memon[b]

[a] Department of E.C.E., Stevens Institute of Technology
[b] Department of Computer Science, Polytechnic University

ABSTRACT

Two fundamental questions in steganography are addressed in this paper, namely, (a) definition of steganography security and (b) definition of steganographic capacity. Since the main goal of steganography is covert communications, we argue that these definitions must be dependent on the type of steganalysis detector employed to break the embedding algorithm. We propose new definitions for security and capacity in the presence of a steganalyst. The intuition and mathematical notions supporting these definitions are described. Some numerical examples are also presented to illustrate the need for this investigation.

Keywords: Steganalysis, capacity, steganography, detection.

1. INTRODUCTION

Steganography deals with hiding messages in a cover signal so that they can be extracted at the receiving side with the help of a secret key. Applications of steganography include covert communications, watermarking and fingerprinting that seem to hold promise for copyright protection, tracing source of illegal copies, etc. There are several issues to be considered when studying steganographic systems. One among the key performance measures used to compare different message embedding algorithms is *steganography capacity*. In a general sense, it is the maximum message size that can be embedded subject to certain constraints. A number of ways to compute the steganography/watermarking capacity using information theory, perceptually based methods, and detection theory have been proposed previously.[1–5] This work extends the recent steganography capacity results in the presence of steganalysis first reported in Chandramouli et. al.[3]

Steganalysis is a relatively new branch of research. While steganography deals with techniques for hiding information, the goal of steganalysis is to detect and/or estimate potentially hidden information from observed data with little or no knowledge about the steganography algorithm and/or its parameters. It is fair to say that steganalysis is both an art and a science. The art of steganalysis plays a major role in the selection of features or characteristics to test for hidden messages while the science helps in designing the tests themselves. While it is possible to design a reasonably good steganalysis technique for a specific steganography algorithm, the long term goal must be to develop a steganalysis framework that can work effectively at least for a class of steganography methods, if not for all. Clearly, this poses a number of mathematical challenges and questions.

One important question in steganography is: what is the trade-off involved in embedding larger message sizes? This question can be answered in several different ways. In Chandramouli et. al.,[3] it is shown that, when the embedding message sizes are larger than a threshold then it becomes easier for a steganalysis algorithm to detect the presence/absence of a hidden message. This defeats the purpose of steganography where the idea is to hide messages in a cover signal such that its very presence can be concealed. We classify steganalysis into two categories:

- **Passive steganalysis**: Detect the presence or absence of a secret message in an observed message or identify the type of embedding algorithm.
- **Active steganalysis**: Estimate/extract some properties of the message or the embedding algorithm. For example, extract a (possibly approximate) version of the secret message from a stego message.

For further information send correspondence to, Email: mouli@stevens-tech.edu

In this paper, we discuss a steganography capacity measure in the presence of passive steganalysis. In order to do this, we have to first define a notion of steganography security. Clearly, this notion has to be statistical and not perceptual. We propose such a definition for this measure and show how to compute the stego capacity based on this. An intuition behind our mathematical formulation is that, stego security and capacity are functions of the steganalysis detector. That is, a good steganalysis detector will be able to detect the presence/absence of an embedded message with high accuracy thus implying that the embedding method is less secure and the maximum embeddable message length is small. Therefore, a decision theoretic formulation is followed in this work.

The paper is organized as follows. Section 2 contains the details of the proposed definition of steganography security and Section 3 deals with the corresponding capacity definition. Concluding remarks are given in Section 4.

2. STEGANALYSIS DETECTOR DEPENDENT STEGANOGRAPHY SECURITY

Notable works in defining steganography security is that of Cachin[5] and Zollner et. al.[6] Cachin defines a steganographyic method to be ϵ-secure ($\epsilon \geq 0$) if the relative entropy between the cover and the stego probability distributions (P_c and P_s, respectively) is at most ϵ, i.e.,

$$D(P_c||P_s) = \int P_c log \frac{P_c}{P_s} \leq \epsilon \tag{1}$$

A stego technique is said to be perfectly secure if $\epsilon = 0$. Existence of perfectly secure algorithms (although impractical) are shown to exist. We observe that there are several shortcomings to this definition. Some of these are listed below.

- While Cachin's definition may work for random bit streams (with no inherent statistical structure), for real-life cover messages such as audio, image, and video, it seems to fail. This is because, real-life cover messages have a rich statistical structure in terms of correlation, higher-order dependence, etc. By exploiting this structure, it is possible to design good steganalysis detectors even if the original probability distribution is preserved (i.e., $\epsilon = 0$) during stego embedding. If we approximate the probability distribution functions using histograms, then, examples such as Jessica et. al.[7] show that it is possible to design good steganalysis detectors even if the histograms of cover and stego are the same!

  Cachin assumes that the cover and stego messages are vectors of independent, identically distributed (i.i.d.) random variables—not true for many real-life cover signals. Perhaps, herein lies the problem. One approach to rectify this is to put the constraint that the the relative entropy computed using the nth order joint probability distributions must be less than, say, ϵ_n. We can then force an embedding technique to preserve the nth order distribution. But, it may then be possible to use $(n+1)$st order statistics for steganalysis. This line of thought clearly poses several interesting questions such as: (a) practicality of preserving nth order joint probability distribution during embedding for medium to large values of n, (b) behaviour of the sequence $\{\epsilon_n\}$, etc. We do not address these issues in this paper. Of course, even if these nth order distributions are preserved, there is no guarantee that embedding induced perceptual distortion will be acceptable. If this distortion is significant, then it is not even necessary to use a statistical detector for steganalysis!

- Consider the following embedding example. Let $P(X = 0) = P(Y = 0) = 1/2$ and the embedding function is the following:

$$Z = X + Y \bmod 2 \tag{2}$$

  We observe that $D(P_Z||P_X) = 0$ but $E(X - Z)^2 = 1$. The non-zero mean squared error value may give away enough information to the steganalysis detector even though $D(.) = 0$ in this case.

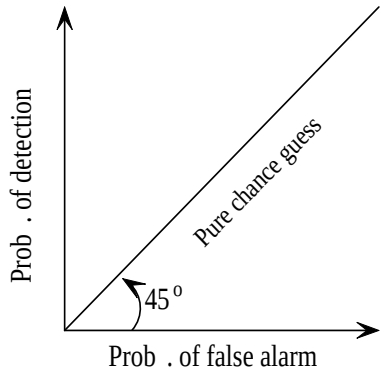

Figure 1. Detector ROC plane.

Given the above arguments, our goal is to investigate an alternative measure for stego security that is perhaps more fundamental to steganalysis. In this regard, we note that the false alarm probability (α=P(detect message present|message absent)) and detection probability (β=P(detect message present| message present) play significant roles. Employing a steganalysis detector optimized for one specific stego algorithm (say, least significant bit embedding), it will be possible to attain low values for α and high values of β for that embedding technique. On the other hand, employing this steganalysis detector to detect some other algorithm will result in higher α and lower β. Therefore, we observe that α and β can possibly assume several different values depending on the problem situation.

The steganalysis detector's receiver operating characteristic (ROC) is a plot of α versus β. Points on the ROC curve represent the achievable performance of the steganalysis detector. The average probability of stegnalysis error is given by,

$$P_e = (1-\beta)P(\text{message embedded}) + \alpha P(\text{message not embedded}). \tag{3}$$

If we assume $P(\text{message embedded}) = P(\text{message not embedded})$ then, from Eq. (3),

$$P_e = \frac{1}{2}\left[(1-\beta)+\alpha\right] \tag{4}$$

Note that, α and β are detector dependent values. For example, for a chosen value of α, β can be maximized by using a Neyman-Pearson hypothesis test[3] or both α and β can be chosen and traded-off with the number of observations by using Wald's sequential probability ratio test. Note that, perceptually based steganalysis techniques also fit into this detector formulation. Observe from Eq. (4) that if $\alpha = \beta$ then $P_e = 1/2$ as shown in Fig. 1. That is, the detector makes purely random guesses when it operates or forced to operate on the 45 degree line in the ROC plane. This means that the detector does not have sufficient information to make an intelligent decision. Therefore, if the embedder forces the detector to operate on the 45 degree ROC curve by choosing proper algorithms or parameters, then we say that the stego message is secure and obtain the following definitions.

DEFINITION 1. *A stego embedding algorithm is said to be $\gamma_{\mathcal{D}}$-secure w.r.t. a steganalysis detector $\mathcal{D}$ if $|\beta_{\mathcal{D}} - \alpha_{\mathcal{D}}| \leq \gamma_{\mathcal{D}}$, where $0 \leq \gamma_{\mathcal{D}} \leq 1$.*

DEFINITION 2. *A stego embedding algorithm is said to be perfectly secure w.r.t. a steganalysis detector $\mathcal{D}$ if $\gamma_{\mathcal{D}} = 0$.*

Clearly, from these definitions, we can think of embedding and steganalysis as a zero sum game where the embedder attempts to minimize $|\beta - \alpha|$ while the steganalyst attempts to maximize it.

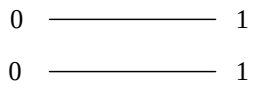

Figure 2. Simple stego channel.

3. γ-SECURITY AND STEGANOGRAPHIC CAPACITY

With the definition of γ-security in place, we are now ready to define steganographic capacity w.r.t. this definition. Before we go into the details, we present a simple example to illustrate why it is useful to define a steganalysis detector dependent steganographic capacity measure. This example also illustrates why Shannon capacity may not be applicable to steganography (even though it is used for watermarking where robustness and not covertness is the main issue). Consider the simple stego channel shown in Fig. 2. The inputs and outputs of the stego channel are shown in this figure. These could be the least significant bits of an image, for example. Let's assume that the steganalysis detector knows that LSB embedding is used, but does not know if the LSB bits are the original message bits or some form of pre-coding has been performed before embedding. Consider the following two steganalysis detectors, $\mathcal{D}_1$ and $\mathcal{D}_2$:

$$\text{"Believe what you see detector"}, \mathcal{D}_1 \quad : \quad \text{decode 0 as 0 and decode 1 as 1} \tag{5}$$

$$\text{"Detect always 1 detector"}, \mathcal{D}_2 \quad : \quad \text{decode 0 as 1 and decode 1 as 1} \tag{6}$$

It is quite obvious that $\alpha_{\mathcal{D}_1} = P(0|0) = 0$ and $\beta_{\mathcal{D}_2} = P(1|1) = 1$, and, $\alpha_{\mathcal{D}_2} = 1$ and $\beta_{\mathcal{D}_2} = 1$. Therefore, $\gamma_{\mathcal{D}_1} = 1$, meaning it is totally insecure w.r.t. $\mathcal{D}_1$, and $\gamma_{\mathcal{D}_2} = 0$ implying that it is perfectly secure w.r.t. $\mathcal{D}_2$. Therefore, we can expect the stego capacity to be zero if $\mathcal{D}_1$ is employed since every embedded message bit can be decoded perfectly by the steganalyst. However, we note that Shannon capacity of the stego channel in Fig. 2 is 1 bit/symbol. This illustrates the fact that Shannon capacity may not be directly applicable to steganography because, by definition, this capacity is the maximum rate for which arbitrary *reliability* can be achieved. However, for steganography *security* is the main concern; reliability and security need not mean the same thing.

Before giving a formal definition of stego capacity in the presence of steganalysis, we present another numerical example. Consider an embedding algorithm that changes the mean value of a Gaussian(0,σ^2) cover signal by $m > 0$ to embed a message bit. Let the length of message symbols be N. Let the steganalysis detector employ the minimum probability of error criterion. Then it is not difficult to show that,

$$\alpha = \frac{1}{2\sqrt{2}}\left[1 - erf\left(\frac{-m\sqrt{N}}{2\sigma}\right)\right] \tag{7}$$

$$\beta = \frac{1}{2\sqrt{2}}\left[1 - erf\left(\frac{m\sqrt{N}}{2\sigma}\right)\right] \tag{8}$$

If we assume that $m^2 << \sigma^2$, i.e., the message-to-noise ratio is small (due to perceptual considerations, etc.) and $\frac{m\sqrt{N}}{2\sigma} << 1$, then, in order to satisfy $|\beta - \alpha| \leq \gamma$, the embedder has to choose,

$$N \leq \frac{2\pi\sigma^2\gamma^2}{m^2}. \tag{9}$$

We observe from this formula that the number of symbols that can be used for embedding and still satisfy the γ-security constraint increases inversely as the message-to-noise-ratio. For other types of steganalysis detectors this rate of increase may different. This means, the embedding *capacity* varies w.r.t. the steganalysis detector and leads us to the following definition.

DEFINITION 3. *Let the number of message carrying symbols be N and let $\alpha_{\mathcal{D}}^{(N)}$ and $\beta_{\mathcal{D}}^{(N)}$ be the corresponding false alarm and detection probabilities for a steganalysis detector $\mathcal{D}$. Then, define the stego capacity as,*

$$N_{\gamma}^{*} = \{\max N \ s.t. \ |\beta_{\mathcal{D}}^{(N)} - \alpha_{\mathcal{D}}^{(N)}| \leq \gamma_{\mathcal{D}}\} \ symbols. \tag{10}$$

4. CONCLUSIONS

We revisit the definitions of stego security and capacity. New definitions for both these quantities are presented from a steganalysis perspective. Arguments are presented to illustrate why some of the current definitions found in the literature may be inadequate. The proposed notions of security and capacity are explained using some numerical examples. Clearly, several questions are now open in this direction of research that we hope to address in the future.

ACKNOWLEDGMENTS

This material is based on research sponsored by Air Force Research Laboratory under agreement number F306020-02-2-0193. The U.S. Government is authorized to reproduce and distribute reprints for Governmental purposed notwithstanding any copyright notation thereon. The views and conclusions contained herein are those of the authors and should not be interpreted as necessarily representing the official policies or endorsements, either expressed or implied, of Air Force Research Laboratory or the U.S. Government.

REFERENCES

1. R. Chandramouli, "Data hiding capacity in the presence of an imperfectly known channel," *SPIE Proceedings of Security and Watermarking of Multimeida Contents II* **4314**, 2001.
2. P. Mouli and M. Mihcak, "The data hiding capacity of image sources," *preprint available at http://www.ifp.uiuc.edu/∼moulin/paper.html* .
3. R. Chandramouli and N. Memon, "Analysis of lsb image steganography techniques," *Proc. ICIP* **3**, pp. 1019–1022, 2001.
4. S. Somasundaram and R. Chandramouli, "Perceptually based waterfilling for watermarking," *Proc. ISCAS* , 2002.
5. C. Cachin, "An information-theoretic model for steganography," *Proc. 2nd International Workshop Information Hiding* **LNCS 1525**, pp. 306–318, 1998.
6. J. Zollner, H. Federrath, H. Klimant, A. Pfitzman, R. Piotraschke, A. Westfeld, G. Wicke, and G. Wolf, "Modeling the security of steganographic systems," *Prof. 2nd Information Hiding Workshop* , pp. 345–355, April 1998.
7. J. Fridrich, D. Soukal, and M. Goljan, "Higher-order statistical steganalysis of palette images," *Proc. SPIE Security and Watermarking of Multimedia Contents V* , 2003.