

Data Hiding Capacity in the Presence of an Imperfectly Known Channel

R. Chandramouli

Department of Electrical and Computer Engineering
Stevens Institute of Technology

ABSTRACT

We consider a data hiding channel in this paper that is not perfectly known by the encoder and the decoder. The imperfect knowledge could be due to the channel estimation error, time-varying active adversary etc. A mathematical model for this scenario is proposed. Many important attacks such as scaling, geometrical transformations etc. fall under the proposed mathematical model. Minimal assumptions are made regarding the probability distributions of the data-hiding channel. Lower and upper bounds on the data hiding capacity are derived. It is shown that the popular additive Gaussian noise channel model may not suffice in real-world scenarios; the capacity estimates using the additive Gaussian channel model tend to either over- or under-estimate the capacity under different scenarios. Asymptotic value of the capacity as the signal to noise ratio becomes arbitrarily large is also given. Many existing data hiding capacity estimates are observed to be a special case of the formulas derived in this paper. We also observe that the proposed mathematical model can be applied to real-life applications such as data hiding in image/video. Theoretical results are further explained using numerical values.

Keywords: Data hiding channel, capacity, information theory

1. INTRODUCTION

Data hiding is emerging as an important area of research with the advent of digital technologies. Watermarking, authentication, and steganography are some of the important applications of data hiding techniques to various real-life scenarios.¹ As the applicability of data hiding methods increases it becomes essential to study them more rigorously. A thorough mathematical analysis of general data hiding principles under broad assumptions will prove very useful in their analysis and optimization. Moreover, a general mathematical model will be useful in predicting the performance of a wide range of algorithms. There has been some previous attempts towards this goal.²⁻⁷ Mathematical tools from information theory have provided some insights on the data hiding problem.^{2-4,6,8-12} The number of bits that can be hidden in a given host signal, namely, the *data-hiding capacity* when the host signal undergoes specific kinds of processing/attacks with known probability distributions have been studied in some of these works. The Gaussian probability distribution is a popular model for the data-hiding channel. This model gives rise to closed-form solutions for the data-hiding capacity. A commonality between the majority of the studies on data-hiding is that the type of processing the hidden data undergoes is assumed to be known at the receiver. The processing/attack is usually modeled as additive noise. But, the attack by an active adversary is not guaranteed to be known at the receiver and need not be only additive; *e.g.* scaling and rotation operations are not additive. Therefore, a more general mathematical model is need. Differing from the information theoretic analysis, a decision theoretic method to compute the *watermark length capacity* for a specific noise distribution is introduced in.⁷ The length capacity is the minimum number of signal samples that have to be watermarked so that the watermark can be detected with a given probability of error.

We consider an active adversary in this paper whose strategy could change with time. The time-varying attack strategy of the active adversary means that the estimate of the attack channel by the receiver will always remain imperfect to some extent. On the other hand, if the attack is time-invariant then it can be estimated with arbitrary reliability by using sufficiently large *training data*. In this paper, we propose a channel model that has a random multiplicative and an additive component. This can be thought of as a data-hiding channel with a *fading* component (analogous to the fading in communication theory). We note that most of the attack models studied so far in the literature are subsets of the proposed one. The probability distributions of the random components of the data-hiding

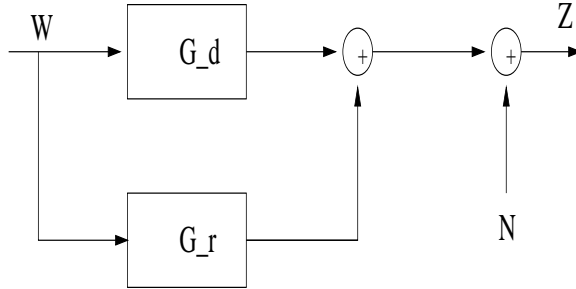


Figure 1. Mathematical model for the imperfectly known data-hiding channel.

channel need not be perfectly known at the receiver. However, it is possible to estimate the deterministic component of the random channel. This gives us partial knowledge about the channel. Our goal is to compute bounds on the data-hiding capacity and study the effect of the imperfect knowledge of the channel on the data hiding capacity. We give both lower and upper bounds for the capacity. Theoretical results are further explained using numerical results. We note that our analysis is valid even for oblivious watermarking schemes where statistical estimators are used to estimate the host signal or some of its parameters. The time-varying attack model can also be interpreted as time-varying channel characteristics when the host signal together with the hidden data is transmitted over randomly fluctuating real-life transmission channels. For example, consider the case where a sender uses an image to hide a message intended for a particular receiver. When this image is sent to the receiver it may undergo compression and also experience additive noise effects. The receiver may not know the compression ratio (or, even the compression algorithm) and the mean and variance of the additive noise. If the receiver does not have access to the original image (which is usually the case) it may have to estimate some of these unknown parameters to extract the hidden data. A typical example is an oblivious watermarking scheme. The parameter estimation process results in various kinds of errors and effects that have significant effects on the data hiding capacity per channel use.

The organization of the paper is as follows. The proposed problem is defined mathematically in Section 2. Theoretical results are derived in this section. These are explained further via numerical analysis in Section 3. The conclusions of this study are given in Section 4.

2. PROBLEM DEFINITION

We attempt to compute and/or bound the data-hiding capacity when the knowledge about the data-hiding channel (or attack) is imperfect and remains to be imperfect. Figure 1 shows the proposed mathematical model for the proposed problem. In the figure, the random variable W stands for the hidden signal (or data), G_d and G_r are the deterministic and the random components of the time-varying attack, *i.e.*, the random gain, G is decomposed as $G = G_d + G_r$ and N is the additive noise component. Formally, the received hidden signal can be written as

$$Z = GW + N \quad (1)$$

$$= G_dW + G_rW + N \quad (2)$$

where, the random variables G, W , and $N \in \Re$ are assumed to be statistically independent of each other. The time-varying nature of the active adversary is characterized by the condition, $\sigma_G^2 = \text{Variance}(G_r) > 0$. This means that there is always some uncertainty in the estimation of the attack. It also quantifies the error in estimating the random gain, G , at the receiver in order to undo the effect of the attack. Further, we make the following reasonable assumptions :

- $E(W^2) \leq \sigma_W^2$
- $N \sim \text{Gaussian}(0, \sigma_N^2)$
- $E(G) = G_d$, $E(G_r) = 0$, and $E(G_r^2) = \sigma_G^2$

Then, the mutual information between Z and W is defined as

$$I(Z; W) = h(Z) - h(Z|W) \quad (3)$$

where $h(\cdot)$ denotes the differential entropy of the random variable. Therefore, if the probability density function (pdf) of Z is $f_Z(z)$ and $W \sim f_W(w)$ then, from,¹³

$$h(Z) = - \int_{\mathfrak{R}} f_Z(z) \ln f_Z(z) dz \quad (4)$$

$$h(Z|W) = \int_{\mathfrak{R}} f_W(w) h(wG_r + N) dw \quad (5)$$

The data hiding capacity is given by

$$C = \sup_{f_W(w): E(W^2) \leq \sigma_w^2} I(Z; W) \quad (6)$$

In real-life scenarios we may not know the probability distribution of G_r making the computation of Eq.(5) non-feasible. That is, we do not know what strategy was used by the active adversary or what kind signal processing operations were performed on the hidden signal. Therefore computing the capacity given by Eq.(6) may not be possible. To overcome this difficulty we attempt to bound the mutual information $I(Z; W)$. The bounds will give us an idea about the data hiding capacity.

We first compute an upper bound for $I(Z; W)$. Let $E(W) = 0$ without loss of generality. Observing that

$$I(Z; W) = I(Z; W|G) - I(W; G|Z) \quad (7)$$

we get

$$I(Z; W) \leq I(Z; W|G) \quad (8)$$

If G is given then Eq.(1) reduces to the well-known data hiding problem in a Gaussian channel as follows :

$$\begin{aligned} I(Z; W|G) &= h(Z|G) - h(Z|W, G) \\ &= E_G(h(Z|G = g)) - h(N) \\ &\leq E_G(1/2 \ln(2\pi e(g^2 \sigma_W^2 + \sigma_N^2))) - 1/2 \ln(2\pi e \sigma_N^2) \end{aligned} \quad (9)$$

$$\leq 1/2 \ln E_G(\ln(2\pi e(g^2 \sigma_W^2 + \sigma_N^2))) - 1/2 \ln(2\pi e \sigma_N^2) \quad (10)$$

$$= 1/2 \ln \left(\frac{G_d^2 \sigma_W^2 + \sigma_G^2 \sigma_W^2 + \sigma_N^2}{\sigma_N^2} \right) \quad (11)$$

$$= 1/2 \ln \left(1 + \frac{G_d^2 \sigma_W^2 + \sigma_G^2 \sigma_W^2}{\sigma_N^2} \right) \quad (12)$$

where Eq.(9) follows from the fact that the Gaussian distribution has the maximum differential entropy ($=1/2 \ln(2\pi \sigma^2)$) for a given variance σ^2 .¹³ Eq.(10) is the result of using Jensen's inequality¹³ which states that, if X is a random variable such that $E(X)$ exists, and if a function $m(\cdot)$ is convex \cap , then

$$E(m(X)) \leq m(E(X)) \quad (13)$$

Since $\ln(\cdot)$ is such a function the result follows. Therefore from Eq.(6) and Eq.(12) we get the following upper bound for the data hiding capacity,

$$\begin{aligned} C &= \sup_{f_W(w): E(W^2) \leq \sigma_w^2} I(Z; W) \\ &\leq 1/2 \ln \left(1 + \frac{G_d^2 \sigma_W^2 + \sigma_G^2 \sigma_W^2}{\sigma_N^2} \right) \end{aligned} \quad (14)$$

To get a lower bound on the data hiding capacity we make one additional assumption, namely, W has a Gaussian distribution. Note that this may not attain channel capacity for any arbitrary distribution of G . A lower bound on $I(Z; W)$ is then readily found by using Theorem 7.4.3 in¹³ by interpreting $G_r W + N$ as additive noise. We restate the theorem here for completeness.

THEOREM 2.1.¹³ *Let a discrete memoryless channel have additive noise of variance σ^2 . Let X and Y denote the input and output of the channel respectively and $E(X^2) \leq \mathcal{E}$. Then $I(X; Y) \geq 1/2 \log \left(1 + \frac{\mathcal{E}}{\sigma^2}\right)$.*

Using this result we have the following lower bound :

$$I(Z; W) \geq 1/2 \ln \left(1 + \frac{G_d^2 \sigma_W^2}{\sigma_G^2 \sigma_W^2 + \sigma_N^2}\right) \quad (15)$$

From Eq. (14) and Eq. (15) we infer the following :

- The mutual information attains the lower bound if $G_r W$ is also Gaussian. Hence, in the worst case, the adversary will attack using a Gaussian distribution or $G_r W$ behaves as additive white Gaussian noise
- The upper bound is attained if the hidden signal behaves as a Gaussian signal with variance $(G_d^2 + \sigma_G^2) \sigma_W^2$.
- We have $I(Z; W|G) - I(Z; W) \leq 1/2 \ln \left(1 + \frac{\sigma_G^2 \sigma_W^2}{\sigma_N^2}\right)$. This is the maximum amount of loss in the mutual information due to the uncertainty at the receiver regarding the attack due to the adversary.
- Clearly,

$$\lim_{\sigma_G^2 \rightarrow 0} I(G; W) = 1/2 \ln \left(1 + \frac{G_d^2 \sigma_W^2}{\sigma_N^2}\right)$$

This is a familiar result—channel capacity of a Gaussian signal in Gaussian noise. We can interpret this as the following. When the uncertainty at the receiver due to the adversary goes to zero, the data hiding capacity equals the Gaussian channel capacity.

- We define signal to noise ratio (SNR) to be equal to σ_W^2 / σ_N^2 . Then, it can be seen that

$$\lim_{SNR \rightarrow \infty} 1/2 \ln \left(1 + \frac{G_d^2 \sigma_W^2}{\sigma_G^2 \sigma_W^2 + \sigma_N^2}\right) = 1/2 \ln \left(1 + \frac{G_d^2}{\sigma_G^2}\right) \quad (16)$$

This is the least amount of data that can be hidden for arbitrarily large SNR.

3. NUMERICAL ANALYSIS

Figure 2 and Figure 3 show the data hiding capacity estimates for various values of the SNR using the upper bound and the lower bound respectively. In the figures AWGN stands for additive white Gaussian noise. The values for σ_G^2 for these two figures have been chosen such that the curves can be differentiated clearly. It is clear from these two figures that a simple additive Gaussian data-hiding channel model could lead to either an over estimate or an under estimate of the capacity depending on the scenario. If the channel gain improves the strength of the hidden signal (or, the effect of estimation error is additional signal power) the data hiding capacity increases. On the other hand, if the randomness due to the channel gain also behaves as additive noise component then the data hiding capacity decreases. From this analysis we observe that a simple additive Gaussian channel model may not suffice for accurate data hiding capacity estimates in real-life scenarios where the signal processing attacks on the host signal may be unknown.

We see in Figure 2 that the data hiding capacity per channel use increase with σ_G^2 . This is because as $\sigma_G^2 \uparrow$ the adversary's attack (or, channel estimation error) results in more signal power leading to improvements in the capacity. On the other hand, Figure 3 shows how the data hiding capacity decreases when the adversary's attack behaves as additive noise with increasing power. Also, we observe from Figure 3 that the minimum number of bits per channel use that can be hidden tends to a finite value as the $SNR \rightarrow \infty$; this is in agreement with the theoretical limiting value given in the previous section.

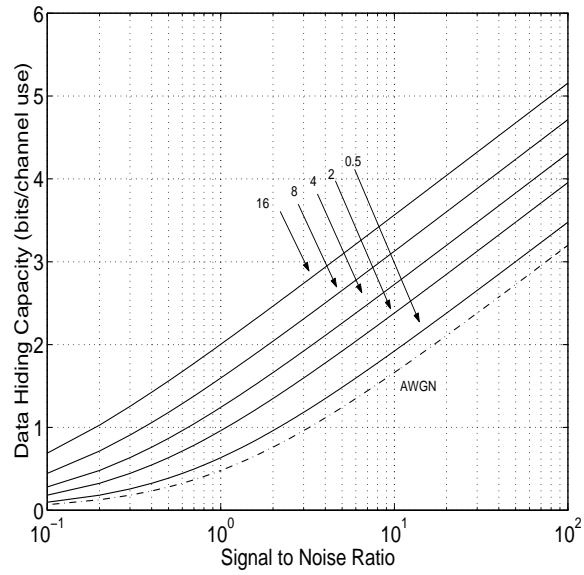


Figure 2. Capacity estimates using the upper bound. $G_d^2 = 1$ and the each curve corresponds to the value of σ_G^2 as shown in the figure.

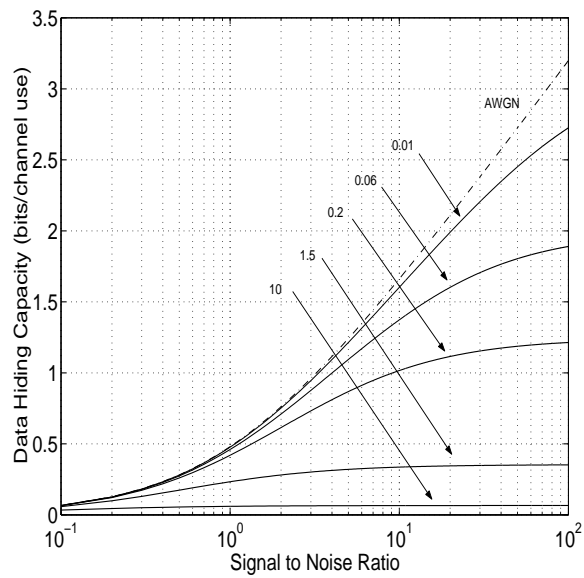


Figure 3. Capacity estimates using the lower bound. $G_d^2 = 1$ and the each curve corresponds to the value of σ_G^2 as shown in the figure.

4. CONCLUSIONS

This study tells us how much data can be hidden when there is a randomly time-varying adversary or channel. A mathematical model that captures this general setting is proposed. Lower and upper bounds on the mutual information between the received signal and the covert signal allows us to compute data hiding capacities under general scenarios. We find that, by not accounting for the error in the estimation of the attack at the receiver we may over or under estimate the data hiding capacity. A simple additive Gaussian channel model is not sufficient, in general. Various existing capacity estimates can be derived as a special case of the proposed work. We plan to extend this work to data hiding in real-life signals such as images that undergo various kinds of image processing operations. Particularly, it is of interest to us to study the effect of estimation errors on the capacity using oblivious watermark detection. It is known that these estimation errors do not have a satisfactory probabilistic description. So, by simply computing their first and second order statistics we may be able to predict the capacity using the theoretical methods discussed in this paper.

REFERENCES

1. F. Petitcolas, R. Anderson, and M. Kuhn, "Information hiding—a survey," *Proc. of the IEEE* **87**, pp. 1062–1078, July 1999.
2. P. Moulin and J. A. O'Sullivan, "Information-theoretic analysis of information hiding," in <http://www.ifp.wiuc.edu/~moulin/paper.html> (preprint), September 1999.
3. P. Moulin and J. A. O'Sullivan, "Information-theoretic analysis of watermarking," in *Proc. of ICASSP*, June 2000.
4. B. Chen and G. Wornell, "An information-theoretic approach to the design of robust digital watermarking system," in *Proc. of ICASSP*, March 1999.
5. M. Swanson, M. Kobayashi, and A. Tewfik, "Multimedia data embedding and watermarking strategies," *Proc. of the IEEE* **86**, pp. 1064–1087, June 1998.
6. N. Merhav, "On random coding error exponents of watermarking systems," *IEEE Trans. on Information Theory* **46**, pp. 420–430, March 2000.
7. R. Chandramouli and N. Memon, "How many pixels to watermark?," in *Proc. of IEEE International Conference on Information Technology: Coding and Computing* (available on-line at <http://www.ece.stevens-tech.edu/~mouli>), pp. 11–15, March 2000.
8. C. Cachin, "An information-theoretic model for steganography," in *Proc. of 2nd Workshop on Information Hiding* (D. Aucsmisht, ed.), *Lecture Notes in Computer Science*, Springer, 1998.
9. D. Kundur, "Implications for high capacity data hiding in presence of lossy compression," in *Proc. of IEEE International Conference on Information Technology: Coding and Computing*, pp. 16–21, March 2000.
10. M. Rankumar and A. Akansu, "Information theoretic bounds for data hiding in compressed images," *IEEE 2nd Workshop on Multimedia Signal Processing*, pp. 267–272, December 1998.
11. J. Ruanaidh, W. Dowling, and F. Bowland, "Watermarking digital images for copyright protection," *IEE Proceedings* **143**, pp. 250–256, August 1996.
12. S. Servetto, C. Podilchuk, and K. Ramachandran, "Capacity issues in digital image watermarking," in *Proc. of Intl. Conference on Image Processing*, pp. 445–449, 1998.
13. R. Gallager, *Information theory and reliable communication*, John Wiley and Sons, 1968.