# DETERMINING ACHIEVABLE RATES FOR SECURE, ZERO DIVERGENCE, STEGANOGRAPHY

*K. Sullivan, K. Solanki, B. S. Manjunath, U. Madhow, and S. Chandrasekaran*

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

## ABSTRACT

In steganography (the hiding of data into innocuous covers for secret communication) it is difficult to estimate how much data can be hidden while still remaining undetectable. To measure the inherent detectability of steganography, Cachin [1] suggested the $\epsilon$-secure measure, where $\epsilon$ is the Kullback Leibler (K-L) divergence between the cover distribution and the distribution after hiding. At zero divergence, an optimal statistical detector can do no better than guessing; the data is undetectable. The hider's key question then is, what hiding rate can be used while maintaining zero divergence? Though work has been done on the theoretical capacity of steganography, it is often difficult to use these results in practice. We therefore examine the limits of a practical scheme known to allow embedding with zero-divergence. This scheme is independent of the embedding algorithm and therefore can be generically applied to find an achievable secure hiding rate for arbitrary cover distributions.

## 1. INTRODUCTION

Steganography is the application of data hiding for the purpose of secret communication. The steganographer's goal is to embed as much data as possible without the existence of this data being detectable. Intuitively, there is a tradeoff between the amount of data embedded and the risk of detection, however it is difficult to accurately characterize this tradeoff. In preventing detection from a steganalyst, the steganographer has the disadvantage of not knowing the detection method that will be used, and so must assume the steganalyst is using the best possible detector. To measure the capabilities of an optimal statistical detector, Cachin [1] suggested the $\epsilon$-secure measure. Here $\epsilon$ is the Kullback-Leibler (K-L) divergence between the cover distribution and the distribution after hiding. The performance of an optimal statistical test is bound by this divergence, and therefore $\epsilon$ serves as a succinct measure of the inherent detectability of steganography. At zero divergence, an optimal statistical detector can do no better than guessing; the data is undetectable. The hider's key question then is, what hiding rate can be used while maintaining zero divergence?

In [2], Moulin and Wang derive an expression for perfectly secure capacity, the theoretical maximum hiding rate under the constraint of zero divergence. Additionally they provide an example achieving this capacity in the binary-Hamming channel. However it is difficult to extend these results to more complex hiding scenarios.

Rather than deriving the theoretical capacity, we instead seek to derive the achievable hiding rate for a known statistical restoration method capable of hiding with zero K-L divergence [3]. This method is independent of the embedding algorithm (e.g. LSB, spread spectrum, QIM). Therefore this derivation can be generically applied to find a secure hiding rate that is known to be achievable in practice. As an example, we apply this analysis to quantization index modulation (QIM) hiding in randomly generated Gaussian covers, and find 30% of the available coefficients can be used while guaranteeing zero divergence 90% of the time.

## 2. HIDING RATE FOR ZERO K-L DIVERGENCE

We first briefly outline the idea of statistical restoration we use as the basis of our analysis. The basic idea is to hide as usual in some proportion of the symbols available for hiding (e.g. pixels, DCT coefficients) and use the remaining to match the density function to that of the cover, and thus achieve zero Kullback-Leibler divergence. A similar approach was used by Provos to correct histograms [4]. The advantage of our approach is its applicability to continuous data. We earlier presented an application of this approach to reduce K-L divergence [5] and have since extended this method to reduce the divergence to zero. For details and experimental results see [3].

Practically speaking, the steganalyst does not have access to continuous probability density functions (pdf), but instead calculates a histogram approximation. Our data hiding is secure if we match the stego (data containing hidden information) histogram to the cover histogram using a bin size, denoted $w$, the same size as, or smaller than, that employed by the steganalyst. We stress that all values are present and there are no "gaps" in the distribution of values; however, within each bin the data is matched to the bin center. A key assumption is that for small enough $w$, the distribution is uniformly distributed over the bin, a common assumption in source coding [6]. Under this assumption, we can generate uniformly distributed pseudorandom data to cover each bin, and still match the original statistics. Let $f_X(x)$ be the cover pdf and $f_S(s)$ the pdf of the stego data. For $I$ bins centered at $t[i]$, $i \in [1, I]$ with constant width $w$, the expected histogram for data generated from $f_X(x)$ is:

$$P_X^E[i] = \int_{t[i]-w/2}^{t[i]+w/2} f_X(x)dx$$

with a similar derivation of $P_S^E[i]$ from $f_S(s)$. The superscript $E$ denotes that this is the expected histogram, to discriminate it from histograms calculated from random realizations. Let $\lambda \in [0, 1)$ be the ratio of symbols used for hiding. Denoting the cover histogram

This research was supported in part by a grant from ONR #N0014-01-1-0380 and #N00014-05-1-0816.

as $P_X[i]$, and the standard (uncompensated) stego histogram as $P_S[i]$, we have the following constraint: $\lambda \leq \frac{P_X[i]}{P_S[i]}$, [5] which gives us an upper limit on the percentage of symbols we can use for hiding, and from this the rate. Additionally, to prevent decoding problems at the intended receiver (see [5] for details), a worst-case $\lambda$ is chosen: $\lambda^* \triangleq \min_i \frac{P_X[i]}{P_S[i]}$ .

### 2.1. Distribution of Hiding Rate

Our goal is to characterize the rate guaranteeing zero divergence for a given cover distribution and hiding method. In practice, because the data is random, we find a rate that satisfies the zero divergence criteria with a pre-determined probability. To do this, we need to find the distribution of the minimum of the histogram ratio, $\lambda^*$ for a given cover pdf, $f_X(x)$. Our approach is to first find the distribution of the ratio $\frac{P_X[i]}{P_S[i]}$ over all bins, and from this find the distribution of $\lambda^*$.

We note that histograms calculated from real data vary for each realization. In other words, the number of symbols in each bin $i$, $NP_X[i]$, is a random variable. Be analyzing the distribution of these random variables, we can find the distribution of the ratio $\frac{P_X[i]}{P_S[i]}$. Let $V_X[i] = NP_X[i]$ be the number of symbols from $f_X(x)$ falling into bin $i$, then $V_X[i]$ has binomial density function $P_{V_X[i]} = B\{N, P_X^E[i]\}$ [7]. Similarly if $V_S[i]$ is the number of symbols per bin for data from $f_S(s)$, it is distributed as $B\{N, P_S^E[i]\}$. See Fig. 1 for a schematic of finding the distribution of the bins of a histogram.

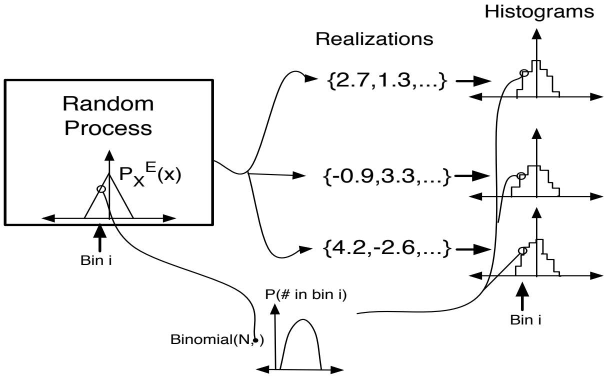

**Fig. 1**. Each realization of a random process has a slightly different histogram. The distribution of the number of elements in each bin is binomially distributing according to the expected value of the bin center (i.e. the integral of the pdf over the bin).

We now define $\Gamma[i] \triangleq \frac{V_X[i]}{V_S[i]} = \frac{P_X[i]}{P_S[i]}$. The cumulative distribution of $\Gamma[i]$, $F_{\Gamma[i]}(\gamma) = P(\Gamma[i] \leq \gamma)$, is given by

$$F_{\Gamma[i]}(\gamma) = \sum_{k=0}^{N} \sum_{l=0}^{\lfloor \gamma k \rfloor} P_{V_S[i]}(k) P_{V_X[i]}(l)$$

Ultimately, we wish to find the distribution of the minimum $\Gamma$ over all bins, giving us a statistical description of $\lambda^*$, our zero-divergence hiding rate. The cumulative distribution of $\lambda^*$ is the distribution of $\min_i \Gamma[i]$ given by

$$F_{\lambda^*}(\gamma) = 1 - \left\{ \prod_i \left[ 1 - F_{\Gamma[i]}(\gamma) \right] \right\}$$

and the density can be found by differentiating. To summarize, given the pdfs of cover and stego, $f_X(x)$ and $f_S(s)$, we can find the distribution of $\lambda^*$: the proportion of symbols we can use to hide in and still achieve zero divergence. Using this, the sender and receiver can choose ahead of time to use a fixed $\lambda$ that guarantees zero-divergence (i.e. $\lambda \leq \lambda^*$) within a desired probability. In Section 2.3 we illustrate this analysis with an example, but first we examine the factors affecting the rate.

### 2.2. General Factors Affecting the Hiding Rate

By examining the derivation of the distribution of $\lambda^*$, we can predict the effect of various parameters on the hiding rate. The key factors effecting the payload are:

1. **Cover and stego pdfs,** $f_X, f_S$: Obviously the “closer” the two pdfs are to one another, the less compensation is required, and the higher the rate. The difference between the pdfs depends on the hiding scheme.
2. **Number of samples,** $N$: The greater the number of samples, the more accurate our estimates of the samples per bin. Therefore it is easier to guarantee a $\lambda$ to be safe with given probability, and so the hiding rate is higher. The number of samples is mostly a function of the size of the image.
3. **Bin width,** $w$**, used for compensation**: Bin width is important to guaranteeing security, but the effect of bin width is not immediately clear. In general the net effect, an increase or decrease in $E\{\lambda^*\}$, depends on the distributions. Fortunately for the steganographer, the steganalyst can not choose an arbitrarily small bin size in order to detect, as the mean integrated square error (MISE) of the detector’s estimate of the pdf is not simply inversely related to bin width [7]. In other words, the steganalyst also faces a challenge in choosing an appropriate bin size.

### 2.3. Maximum Rate of Perfect Restoration QIM

We now apply the analysis to a specific method of embedding: dithered quantization index modulation (QIM), [10]. The basic idea of QIM is to hide the message data into the cover by quantizing the cover with a choice of quantizer determined by the message. The simplest example is so-called odd/even embedding. With this scheme, a continuous valued cover sample is used to embed a single bit. To embed a 0, the cover sample is rounded to the nearest even integer, to embed a 1, round to the nearest odd number. The decoder, with no knowledge of the cover, can decode the message so long as perturbations (from noise or attack) do not change the values by more than 0.5. Since the cover data is quantized, the stego data will only have values at the quantizer outputs. If quantized data is not expected, then steganalysis is trivial: if data is quantized it has hidden data. One solution to this is to dither the quantizer, that is, shift the intervals and outputs by a pseudorandom sequence known by the encoder and decoder. The resulting output no longer “looks” quantized. It is this dithered QIM we examine here. For a given cover pdf $f_X(x)$ we can calculate the expected stego pdf $f_S(s)$ from the cover pdf [8, 9]. Briefly, the cover pdf is convolved with a rectangle function, so the resulting stego pdf is a smoothed version of the original.

In the context of QIM hiding we can more explicitly characterize the factors affecting the amount of data that can be safely embedded. Since we can calculate $f_S$ from $f_X$, of this pair we need only examine the cover pdf. Distributions of typical hiding medium, particularly transform domain coefficients, are sharply peaked, and these peaks tend to become smoothed after hiding. For a particular distribution, $\sigma/\Delta$ is an important parameter charecterizing the detectability of QIM [8]. For large $\sigma/\Delta$, the cover pdf is flat relative to the quantization interval, and less change is caused to the original histogram by hiding, and the expected $\lambda^*$ is large.

Of all the factors, only $w$ and $\Delta$ are in the hands of the steganographer. Decreasing $\Delta$ increases $\sigma/\Delta$, and therefore the safe hiding rate. However, decreasing $\Delta$ also increases the chance of decoding error due to any noise or attacks [10]. Thus if a given robustness is required, $\Delta$ can also be thought of as fixed, leaving only the bin width. For QIM hiding in Gaussians and Laplacians, we found that decreasing the bin size $w$ led to a decrease in $\lambda^*$, suggesting that the steganographer should choose a large $w$. However, as mentioned before, $w$ should be chosen carefully to avoid detection by a steganalyst using a smaller $w$.

We presently examine an idealized QIM scheme, followed by an extension to a practical QIM scheme which prevents decoder errors. As an illustrative example, we provide results derived for hiding in a Gaussian, but note the approach can be used for any $f_X(x)$.

Figure 2 is the density of $\Gamma[i]$, $f_{\Gamma[i]}(\gamma)$ for all $i$ and a range of $\gamma$, for QIM hiding in a zero-mean unit-variance Gaussian. From this density we can see the relationship between $\Gamma$ and bin center. For bins located near zero, $\Gamma[i]$ has a probability concentrated above 1 (though obviously we can not embed in more than 100% of the coefficients). For bins a bit further from the center, the expected value for $\Gamma$ drops. Since the effect of dithered QIM is to smooth the cover pdf this result is not surprising. The smoothing moves probability from the high probability center out towards the tails. Though this result is found for hiding in a Gaussian, we expect this trend from any peaked unimodal distribution, such as the generalized Laplacian and generalized Cauchy distributions often used to model transform coefficients [11]. Near the ends, e.g. $\pm 4$, $\Gamma$ is distributed widely over all possible values. So while it is possible to have a very high $\gamma$ here, it is also possible to be very low; i.e. the variance is very high. The solution we study is to hide only in the high probability region; after hiding, only this region needs to be compensated. This introduces a practical problem, the decoder is not always able to distinguish between embedded coefficients and non-embedded. We address this issue below, but first we examine the ideal case.

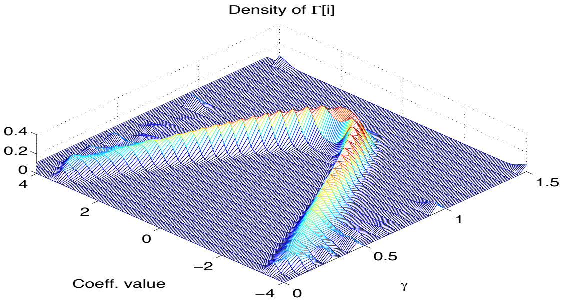

**Fig. 2**. The pdf of $\Gamma$, the ratio limiting our hiding rate, for each bin (coeff. value) $i$. The expected $\Gamma$ drops as one moves away from the center. Additionally, at the extremes, e.g. $\pm 4$, the distribution is not concentrated. In this example, $N$ = 50000, $\sigma/\Delta$ = 0.5, and $w = 0.05$.

Despite the reduction in the number of coefficients we are hiding in, our net rate may be higher due to a higher $\lambda^*$, where $\lambda^*$ is redefined as $\lambda^* \triangleq \min_{i \in \mathcal{H}} \frac{P_X[i]}{P_S[i]}$ where $\mathcal{H}$ is the hiding region, defined as $\mathcal{H} \triangleq [-T, T]$ and $T$ is the hiding threshold. The net hiding rate, no longer simply equivalent to $\lambda^*$, is now $R = \lambda^* G(\mathcal{H})$ where $G(\mathcal{H}) \triangleq \sum_{i \in \mathcal{H}} P_X[i]$. In practice the encoder and decoder can agree on a $\lambda$ which leads to perfect restoration within a predetermined probability, 90% for example. From the distribution of $\lambda^*$, the $\lambda$ guaranteeing perfect restoration with a given probability can be found for each threshold. These 90%-safe $\lambda$s decrease as the threshold is increased, as seen in Fig. 3, along with an example of deriving the 90%-safe $\lambda$ for the threshold of 1.3. The net effect of an increasing $G(T)$ and decreasing safe $\lambda$ is a concave function, as in Fig. 4 from which the maximum rate can be found.

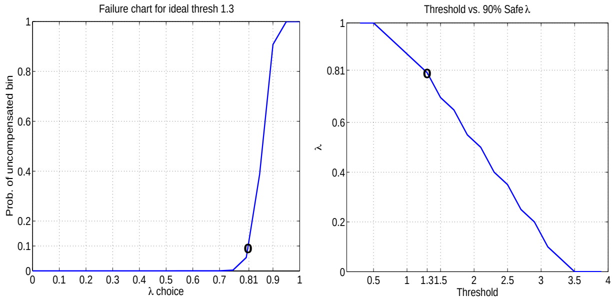

**Fig. 3**. On the left is an example of finding the 90%-safe $\lambda$ for a threshold of 1.3. On the right is safe $\lambda$ for all thresholds, with 1.3 circled.

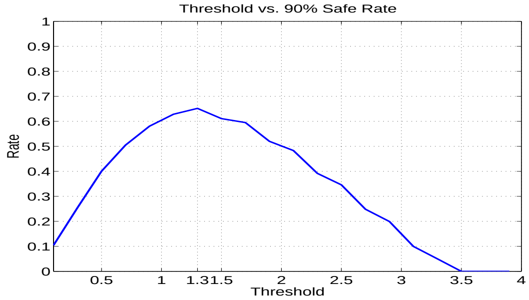

**Fig. 4**. Finding the best rate. By varying the threshold, we can find the best tradeoff between $\lambda$ and the number of coefficients we can hide in.

In Fig. 4 we show the relationship between the chosen threshold and the rate allowing perfect histogram matching in 90% of cases. In this case, the maximum rate is 0.65 bits per coefficient. So, using a threshold of 1.3 and a $\lambda$ of 0.81 (from Fig.3), the hider can successfully send at a rate of 0.65, and the histogram is perfectly restored nine times in ten.

### 2.4. Rate of QIM With Practical Threshold

As noted above, there will inevitably be ambiguity at the decoder with values near the threshold. In the region near the threshold, the decoder does not know if the received value is a coefficient that originally was below the threshold and is now shifted above the threshold after hiding and dithering, or is simply a coefficient that originally was above the threshold and contains no data. Therefore a buffer zone is created near the threshold: if, after hiding, a coefficient would be shifted above the threshold, it is instead skipped over. To prevent creating an abrupt transition in the histogram at the buffer zone, we dither the threshold with the dither sequence [3]. Since the decoder knows the dither sequence, this should not introduce ambiguity. This solution clearly results in a different stego pdf, $f_S(s)$. In the region near the threshold, there is a blending of $f_X(x)$ and a weakened (integrated over a smaller region) version of the standard $f_S(s)$. Beyond the threshold region, the original coefficients pass unchanged and the statistics are unaffected. The cost of this practical fix is a greater divergence between $f_S$ and $f_X$, resulting in a lower overall rate.

As with the ideal threshold case, we can calculate a $\lambda$ guaranteeing perfect restoration a given percentage of the time. Generally the expected $\Gamma$ is increased near the threshold, however it drops quickly after this.

Finally Table 1 shows the 90%-safe rate for various thresholds. Here we would choose a threshold of 1, to achieve a rate of 0.3, about half the rate of the ideal case.

| Threshold vs. Rate | | | |
|---|---|---|---|
| Threshold | 1 | 2 | 3 |
| $G(T)$ | 0.66 | 0.94 | 0.99 |
| 90%-safe $\lambda$ | 0.45 | 0.25 | NA |
| Safe rate | 0.30 | 0.24 | 0 |

**Table 1**. An example of the derivation of maximum 90%-safe rate for practical integer thresholds. Here the best threshold is $T = 1$ with $\lambda = 0.45$. There is no 90%-safe $\lambda$ for $T = 3$, so the rate is effectively zero.

We have compared the derived estimates to Monte Carlo simulations of hiding and found the results to be as expected for different parameters $(n, w, \sigma/\Delta)$. We therefore have an analytical means of prescribing a choice of $\lambda$ and $T$ for maximum hiding rate guaranteeing perfect restoration within a given probability. For experimental results of a practical implementation of the restoration scheme, please see [3].

## 3. CONCLUSION

We have analyzed a hiding scheme designed to avoid detection by eliminating divergence between the statistics of cover and stego. We derive expressions to evaluate the rate guaranteeing secure ($\epsilon = 0$) hiding within a specified probability for practically realizable statistical restoration mechanism. In a specific example, we find for QIM hiding in Gaussian covers, about a third of the coefficients can be used and still achieve zero divergence nine times in ten.

## 4. REFERENCES

[1] C. Cachin, "An information theoretic model for steganography," *Int'l Workshop on Information Hiding, LNCS*, vol. 1525, pp. 306–318, 1998.

[2] P. Moulin and Y. Wang, "New results on steganographic capacity," in *Proceedings of Conference on Information Sciences and Systems (CISS)*, 2004.

[3] K. Solanki, K. Sullivan, U. Madhow, B. S. Manjunath, and S. Chandrasekaran, "Provably secure steganography: Achieving zero K-L divergence using statistical restoration," in *Proceedings of ICIP*, Atlanta, Georgia, USA, Oct 2006.

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

[5] K. Solanki, K. Sullivan, U. Madhow, B. S. Manjunath, and S. Chandrasekaran, "Statistical restoration for robust and secure steganography," in *Proceedings of ICIP*, Genoa, Italy, Sep 2005.

[6] A. Gersho and R.M. Gray, *Vector quantization and signal compression*, Kluwer Academic Publishers, 1992.

[7] D. W. Scott, "On optimal and data-based histograms," *Biometrika*, vol. 66, no. 3, pp. 605–10, 1979.

[8] K. Sullivan, Z. Bi, U. Madhow, S. Chandrasekaran, and B. S. Manjunath, "Steganalysis of quantization index modulation data hiding," in *Proceedings of ICIP*, Singapore, Oct 2004.

[9] M. T. Hogan, N. J. Hurley, G. C. M. Silvestre, F. Balado, and K. M. Whelan, "ML detection of steganography," in *Proc. SPIE Symp. on EIS&T*, San Jose, CA, Jan 2005.

[10] 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.

[11] A. Srivastava, A.B. Lee, E.P. Simoncelli, and S.-C. Zhu, "On advances in statistical modeling of natural images," *Journal of Mathematical Imaging and Vision*, vol. 18, pp. 17–33, 2003.