

Channel models for high-capacity information hiding in images

Johann A. Briffa^a, Manohar Das^b

School of Engineering and Computer Science
Oakland University, Rochester MI 48309

ABSTRACT

We consider the scenario of blind information hiding in images as a communications channel, where the channel noise is caused by the embedding and blind extraction method as well as by any lossy compression method utilized to store and transmit the image. We assume that the objectives of the information hiding method are the maximization of payload and its visual and statistical imperceptibility; also, we assume that the warden is passive. For the specific method of Spread Spectrum Image Steganography (SSIS) we show that the channel can be modeled as a Laplacian distribution, and use this to estimate the channel SNR to be expected for any given signal embedding strength by applying the technique to a range of typical images. Finally, we model the effects of various signal extraction methods and lossy compression. This allows a fair comparison with respect to payload capacity. The results shown in this paper are useful for maximizing the channel usage.

Keywords: steganography, information hiding, channel models, capacity

1. INTRODUCTION

Several techniques for blind information hiding in images have been presented in the literature^{1,2}. Most of these are intended for watermarking, where the main requirement is a high robustness to ensure survivability even in the presence of intentional removal attacks; in such a scenario, the perceptibility of the embedded message is of no concern as long as it does not degrade the quality of the cover beyond a certain degree. Also, in watermarking the required amount of embedded data is low; indeed as little as one bit may be sufficient (to confirm ownership or otherwise). On the other hand, comparatively much less work has been done in the related field of steganography where the objective is to hide as much data as possible within the cover; this can be used in applications as diverse as covert data transfer, in-band captioning, and image augmentation. In this scenario, bandwidth efficiency plays a vital role; imperceptibility is also a requirement, at varying degrees depending on the application. In covert communications, for instance, the process should leave the cover statistically similar to typical images in all possible ways. In other applications it may be sufficient to keep the process perceptually invisible. It is clear that with so much diversity in their requirements, the design problem is quite different for information hiding techniques intended for steganography as compared to watermarking.

Most high-payload techniques in the literature are based on the least-significant-bit (lsb) hiding technique³. We consider lsb hiding in the spatial domain unusable because the process is statistically perceptible by examining the image histogram; besides, most such methods assume that the stego-image will not go through a lossy-compression stage. This is a crippling assumption given that photographic images will invariably be compressed using the DCT-based JPEG⁴ or wavelet-based SPIHT⁵ algorithms. Variants of this technique involve lsb hiding in some transform domain (such as DCT coefficients in a JPEG file); these also have been shown to be statistically perceptible in most cases⁶. Two alternatives that were recently proposed go by the name of spread-spectrum techniques - the first⁷ is based on Direct-Sequence Spread-Spectrum (DSSS) as used in communication systems and thus trades off bandwidth for achieving error-free communication, making it less suitable for our purposes; the second⁸ method actually only embeds one bit per pixel (before error-correction) so there is really no bandwidth expansion or spectrum-spreading. However, the system was proven to embed a

a. j.briffa@ieee.org; Telephone: (248) 370-2177; Fax: (248) 370-4633; <http://www.oakland.edu/~jabriffa>

b. das@oakland.edu; Telephone: (248) 370-2237; Fax: (248) 370-4633

Gaussian signal whose statistical properties do not depend on the message⁹; this ensures imperceptibility as long as the embedded signal strength is kept close to typical natural levels of thermal noise. It is this latter system, called Spread-Spectrum Image Steganography (SSIS) by its creators, that we focus on and seek to improve.

A present void in the steganographic literature is the absence of a common frame of reference for comparing different systems; it is usually very hard to judge the relative merits of different embedding schemes. We believe this is mostly due to the tight integration of any error-correction system used, such that one cannot evaluate the embedding method separately from the error-correction used to overcome its deficiencies. We propose to fill this void by presenting a suitable frame of reference - a signal-space representation of the embedded symbols at the transmitter and extracted symbols at the receiver. We present this in Section 2 for the SSIS process. Using this frame of reference we then proceed to model the channel as a Laplacian distribution; we also compare various signal extraction methods, estimate the channel SNR to be expected for any given signal embedding strength, and consider the effect of lossy compression in Section 3. Finally, we draw our conclusions in Section 4.

2. SIGNAL-SPACE REPRESENTATION

Between the embedding of encoded data within an image and its estimation and extraction at the receiving end lies a medium which introduces errors into our estimate; we model this medium as a communications channel. For SSIS, errors are introduced in three ways:

- Quantization of the pixel intensity values in the stego-image (at the transmitting end).
- Estimation errors during signal extraction (at the receiving end).
- Lossy compression during image storage and/or transmission.

2.1 The SSIS modulation process

In SSIS, every pixel in the cover image embeds a single bit before error correction. This is achieved as follows:

- For every pixel, a random number u is generated from a uniform distribution in $[0, 1)$.
- To embed a zero, this number is left as generated; to embed a one, it is transformed using:

$$g(u) = \begin{cases} u + 0.5 & 0.0 \leq u < 0.5 \\ u - 0.5 & 0.5 \leq u < 1.0 \end{cases} \quad (1)$$

- The result is transformed into a Gaussian distribution using the inverse Gaussian cumulative distribution function (cdf), and then scaled to the required embedding strength and embedded.

2.2 Representation of embedded and extracted signals in signal-space

We consider the channel to consist of all the steps between the generation of the embedded signal (including transformation into a Gaussian distribution and scaling) and its estimation at the receiving end. We should thus represent the embedded and extracted signals as points in signal space, such that the communications channel can be treated as a (possibly conditional) probability distribution between the transmitted and received points in signal space. This allows us to easily compute valid soft-information (in the form of *a priori* conditional probabilities) for the received signal.

At the transmitter, the embedded data bit at each pixel is represented by either u or $g(u)$, as defined earlier. For simplicity, at every pixel we map u_0 (the value representing a zero) and u_1 (the value representing a one) to points in signal-space corresponding to BPSK modulation:

$$S_u = \begin{cases} 1 + j0 & u_0 = u \\ -1 + j0 & u_1 = g(u) \end{cases} \quad (2)$$

At the receiving end, whatever estimate is computed for the embedded Gaussian variable will again be transformed into the range $[0, 1)$ using the Gaussian cdf. Thus, for any particular u_0 and u_1 representing a 0 and 1 respectively, there is only a limited range possible for the received value. Since the distance between u_0 and u_1 has the same magnitude for all values of u , we assume a linear relationship between the received value v and its signal-space representation S_v :

$$S_v = 1 - 2 \frac{v - u_0}{u_1 - u_0} + j0 \quad (3)$$

3. RESULTS

In this section we report simulation results on a set of four images with a variety of characteristics. The original images are shown in Figure 1; each image is 8-bit grayscale at a size of 256×256 pixels. The Barbara, Fishingboat, and Lena images are desaturated and subsampled from the original 512×512 color images commonly used in the image processing community. The Lochness image is cropped to 512×512 from a larger color image, then desaturated and subsampled to 256×256 pixels. This image can be considered a typical scenic image, having a large smooth area at high pixel intensity (the lake), another area with low-amplitude detail (the forest in the distance), and areas with high-amplitude detail (foreground trees and castle).



Barbara



Fishingboat



Lena



Lochness

Figure 1. Original Images

In the information-hiding scenario we require some way to estimate the embedded signal at the receiving end. We also want the system to work blindly at the receiving end, i.e. without any knowledge of the original image. Marvel *et al.*⁸ consider this to be a question of estimating the cover image (using some noise-reduction filter) and subtracting this estimate from the stego-image. We prefer to consider it as a direct estimator of the noise component within the stego-image (although this may be performed as a two-step process of noise reduction and subtraction from the stego-image). The reasoning behind our preference is primarily philosophical - it is, after all, the embedded “noise” that we require. However, this change in viewpoint has consequences that reach farther than that. In particular, when computing metrics to quantify the quality of the estimate, we are more concerned with the error relative to the embedded message than the error relative to the original cover image. We also go one step further: our concern is with the system as a communications channel - we thus consider the errors as seen not merely after the “noise” extraction, but also after conversion to the signal-space domain, as described in Section 2.

3.1 Channel Models

We illuminate the distribution of signal extraction errors by embedding a random message (at a strength of -34.2dB relative to peak) into the image of Lena. Note that even at this low embedding strength, we found the noise to be already visible in the stego-image; it is worth noting here that the cover image has an energy of -14.5dB (relative to peak), thus the embedding is actually being performed at -19.7dB relative to the image energy^a. It is this latter metric that can be interpreted as an SNR of 19.7dB in the final stego-image, explaining the visibility of the embedded noise. At the receiving end, we estimate the embedded signal using an Alpha-Trimmed Mean (ATM) filter operating on the 3×3 pixel neighborhood with $\alpha = 1$ (i.e. removing the largest and smallest values before computing the mean). Finally, we normalize the estimated message and use the Gaussian cdf to obtain the corresponding point in signal space to our estimate at each pixel. We assume in this analysis that the receiver knows the embedding strength and that the image is transmitted without lossy compression or any other further processing. Eventually, both assumptions will need to be removed:

- We want to avoid requiring knowledge of the embedding strength because we want to vary the embedding strength at the transmitter depending on the image in question, and the receiver must operate blindly.
- Lossy compression must be allowable because the mere fact that a photographic image is transmitted without lossy compression might arouse suspicion in a covert communications scenario; alternatively, in commercial scenarios (such as in-band captioning) the lack of support for lossy compression would severely limit applicability.

We choose, however, to ignore both these issues for the moment; this allows us to consider the effect of the signal extraction process independently. We plot the distribution of the signal-space error $S_e = S_v - S_u$ in Figure 2; we also plot on the same axes a Laplacian distribution defined by:

$$P(x) = \frac{1}{2\lambda} e^{-|x-\mu|/\lambda} \quad (4)$$

where $\mu = 0$ is the mean value, $\lambda = \frac{\sigma}{\sqrt{2}}$ is the Laplacian parameter, and σ is the standard deviation of S_e . It is clear from

the graph that the Laplacian model represents the error distribution reasonably accurately. For the same simulation, we also plot the error distribution separately for those pixels where we embed a zero and those where we embed a one; this is shown in Figure 3. Note how the distribution differs significantly, particularly in the tail regions - this would indicate that the error distribution is in some way dependent on the embedded data. The reason behind this requires additional mathematical analysis, and is currently being investigated.

a. This is the same as what Marvel refers to as Stego-SNR.

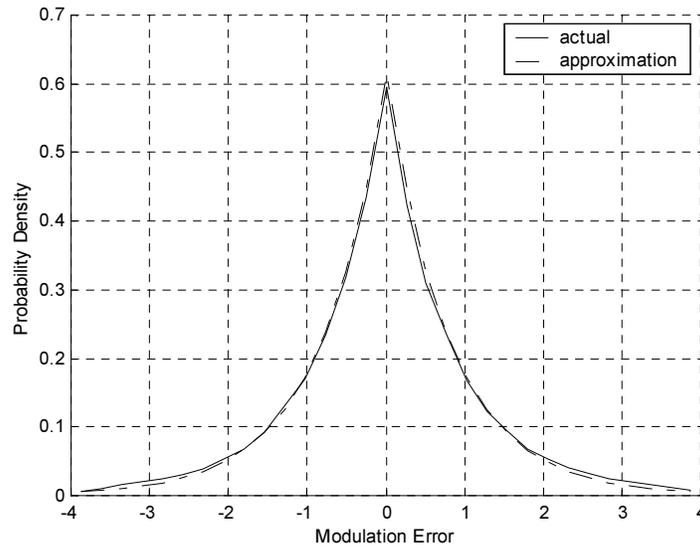


Figure 2. Error Distribution for Lena (Stego-SNR of -19.7dB , ATM filter)

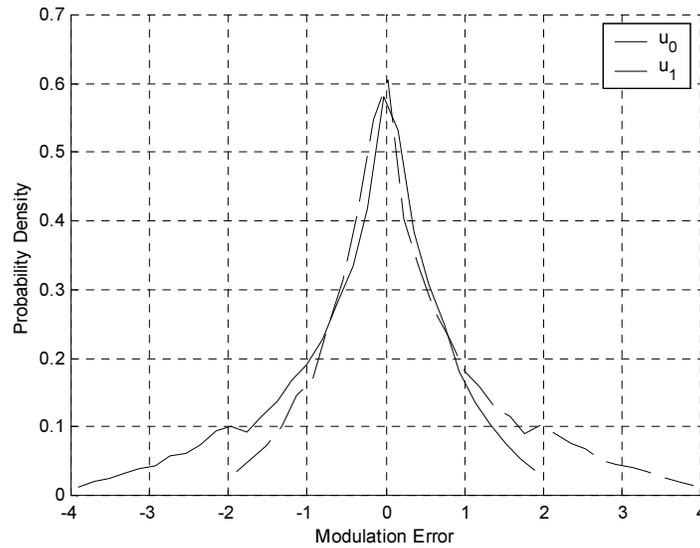


Figure 3. Error Distribution for Lena (Stego-SNR of -19.7dB , ATM filter)

3.2 Comparison of Signal Extraction Methods

We next use a number of different filters to extract an estimate of the embedded signal. We then convert the estimate at each pixel into the equivalent point in signal-space, and use this to compute the hard error rate and the raw signal-to-noise ratio, as listed in Table 1. Our results for the Adaptive Wiener (AW), Alpha-Trimmed Mean (ATM), Mean and Median filters agree with Marvel's results^a; in particular we re-iterate an observation of hers, that while the AW filter results in the lowest error in the mean-squared sense (and thus the best raw SNR), it is outperformed by the ATM filter in terms of hard error rate (though only slightly). The ATM filter was thus considered by Marvel to be the optimal choice, since the hard

error rate is the metric of concern for the hard-decision error-control coding systems she used initially. We believe, however, that when using soft-decision decoding with an appropriate channel model, better performance will be achieved by using a filter with the best overall raw SNR, rather than hard error rate. We also simulate the performance of wavelet shrinkage using the Symmlet-8 wavelet and *VisuShrink* threshold selection, with both hard and soft thresholding¹⁰. Compared to AW and ATM filters, however, the performance of wavelet shrinkage falls far short. The use of better threshold selection criteria, such as those based on Stein’s Unbiased Risk Estimate (SURE) and Donoho’s Hybrid+ scheme, may improve performance.

Table 1. Performance of Signal Extraction Filters

Filter	Hard Error Rate	Raw Signal to Noise Ratio
Adaptive Wiener (AW) ^a	22.8%	-0.62dB
Alpha-Trimmed Mean (ATM)	22.4%	-1.21dB
Mean	22.8%	-1.47dB
Median	28.2%	-1.65dB
Wavelet Shrinkage ^b (Hard)	27.7%	-2.19dB
Wavelet Shrinkage (Soft)	30.7%	-3.13dB

- a. Region size is the 3×3 neighborhood, where applicable.
- b. Wavelet shrinkage filters use the Symmlet-8 mother wavelet and the Donoho *Visu* threshold selection criterion.

3.3 Effect of Signal Embedding Strength

Restricting ourselves to the AW filter as used earlier, we next investigate the effect of embedding strength on performance. We embed a random message at peak strengths ranging from -48dB (where the embedded message has an average amplitude of just one quantization step) to -26dB, in steps of 1dB, into each of the four images in Figure 1. For each case, we estimate the embedded signal using the AW filter and convert to the corresponding signal-space representation; we then compute the hard error rate and raw SNR by comparing to the embedded signal.

We plot the hard error rate against Stego-SNR in Figure 4 - as expected the BER drops as the embedding strength increases. Notably, though, the performance also depends on the cover image itself - invariably, smoother cover images such as *Lochness* result in a lower BER at the same Stego-SNR than more active images such as *Barbara*. The same trend is also visible when we plot the raw channel SNR against Stego-SNR, as in Figure 5. In this case, the distinction between different cover images was even more pronounced, particularly at higher embedding strengths.

A well-known result from information theory is that soft-decision decoding results in better performance than hard-decision decoding; in our scenario, we would expect the achievable capacity of the Laplacian channel to be higher than that of the BSC. We can obtain a first approximation to the Laplacian channel’s capacity using the Gaussian channel capacity formula $C = W \log(1 + S/N)$, where W is the available bandwidth, and S/N is the ratio of signal to noise power. In our case, $W = 1/2$ cycles/pixel, $S = 1$ is the average signal power for our BPSK constellation, and $N = \sigma^2 = 2\lambda^2$ is the noise power. We plot this together with the capacity of the BSC channel for the four images in Figure 6. Note how the capacity of the soft-decision channel is always significantly higher than that of the BSC, for all images.

a. We can, of course, only compare our hard error rate to Marvel’s “Embedded Signal BER”. There is no equivalent in Marvel’s work to our raw SNR.

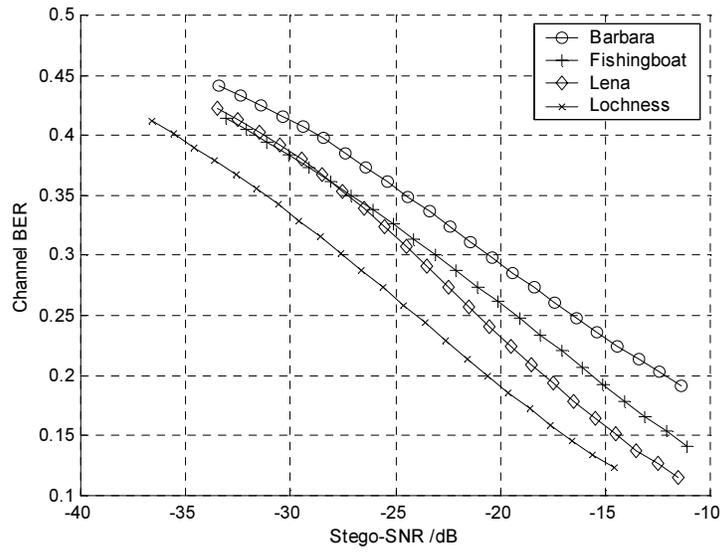


Figure 4. Channel BER against Stego-SNR (Binary Symmetric Channel)

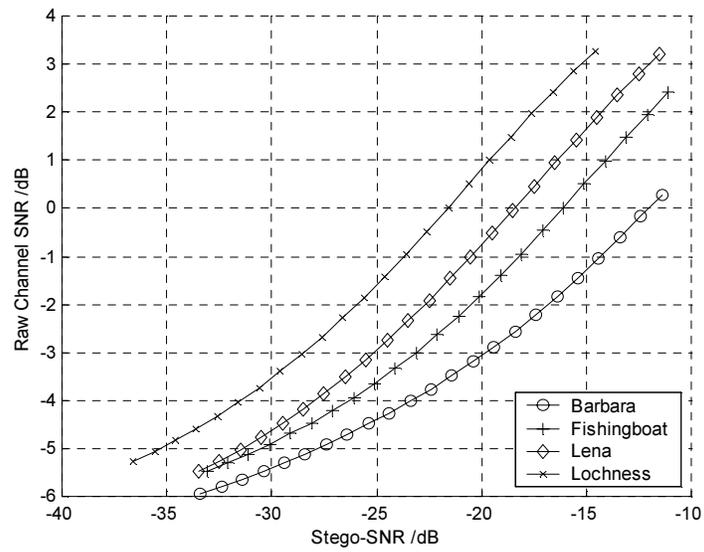


Figure 5. Channel SNR against Stego-SNR (Laplacian Channel)

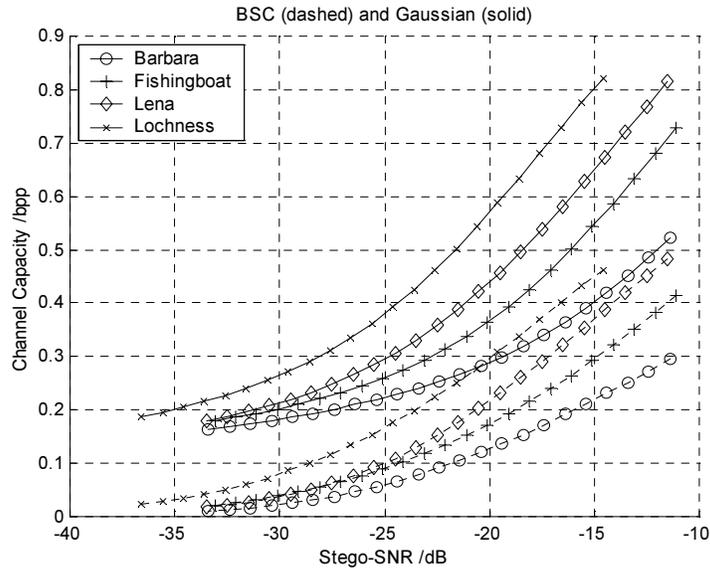


Figure 6. Channel Capacity against Stego-SNR

3.4 Effect of Lossy Compression

Finally, we investigate the effect of JPEG compression on SSIS. We embed a random message at an embedding strength of -34.2dB relative to peak (equivalent to a Stego-SNR of -19.7dB) into Lena, as in Section 3.1. Before decoding, however, we pass the stego-image through a JPEG compression/decompression cycle for quality settings 0-12 (in Adobe Photoshop's TIFF JPEG encoding). We plot the resulting hard error rate against the SNR of the compressed stego-image (which indicates the JPEG quality) in Figure 7. As expected, the error rate increases as the quality of the compressed image is reduced. Note, however, that the increase in error in this case is much sharper than that observed in Figure 4. We also plot the raw channel SNR against the SNR of the compressed stego-image in Figure 8. Again the channel conditions deteriorate as compression fidelity is reduced.

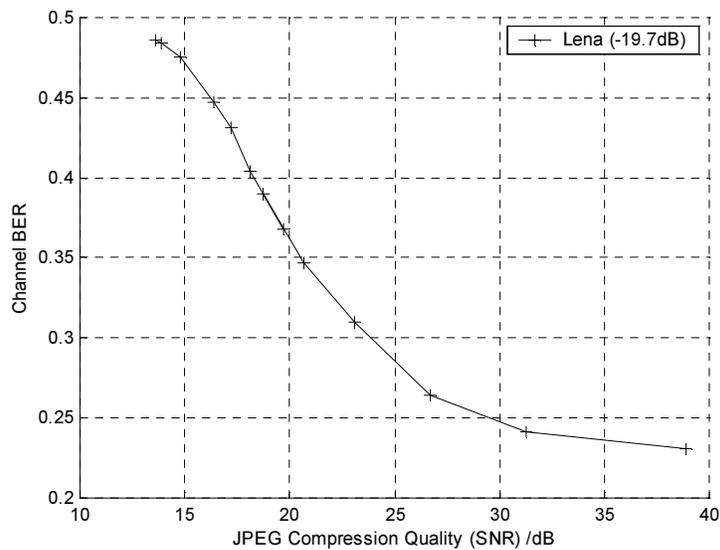


Figure 7. Channel BER against JPEG Compression Quality

To illustrate the difference between the two channel models, we plot once more the capacity of the BSC channel and a Gaussian approximation to the Laplacian channel in Figure 9. Note how the difference between the two channel models is even more pronounced in this case - while the BSC capacity quickly drops to unusable levels, the Gaussian capacity remains within usable limits throughout the range. This would indicate that SSIS is usable at lower quality levels than was thought possible before.

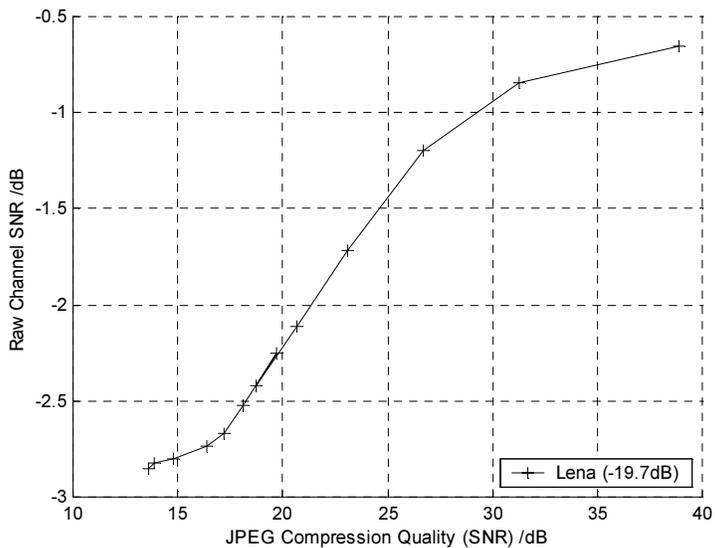


Figure 8. Channel SNR against JPEG Compression Quality

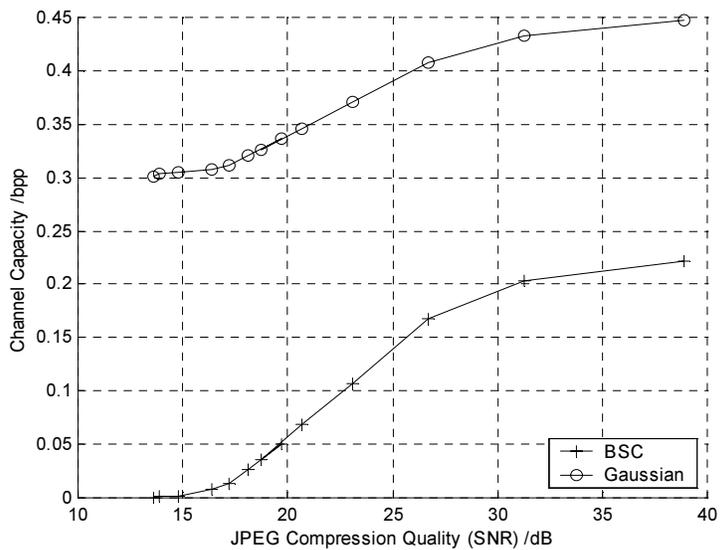


Figure 9. Channel Capacity against JPEG Compression Quality

4. CONCLUSIONS

In this paper we have presented a new frame of reference for comparing high-capacity information hiding schemes, and developed a model for SSIS. We also investigated the performance of various signal extraction filters with respect to this model, as well as the effects of embedding strength and lossy compression. Finally, we took a preliminary look at the impact the new model has on steganographic capacity, particularly noticeable when lossy compression is used. A significant contribution is the decoupling of our model of the channel from any error-correction system employed; indeed, this should allow a more appropriate choice of error correction for any given condition. Several results in this paper necessitate further development:

- A more complete analysis of the dependence of the channel error on the embedded data (as indicated in Section 3.1); this should lead to more advanced channel models and possibly higher achievable capacities.
- Alternative or improved signal extraction filters may lead to better performance.
- A capacity metric appropriate for BPSK modulation with an additive Laplacian noise channel still needs to be developed, and would give tighter estimates on what can be achieved.
- Suitable error-correction codes for this channel (at the required operating SNR) need to be developed.

All the above are currently being actively pursued by the authors.

REFERENCES

1. G. C. Langelaar, I. Setyawan, and R. L. Lagendijk, "Watermarking Digital Image and Video Data. A state-of-the-art overview", *IEEE Signal Processing Magazine*, **17**(5), pp. 20-46, 2000.
2. F. A. P. Petitcolas, R. J. Anderson, and M. G. Kuhn, "Information hiding - a survey", *Proceedings of the IEEE*, **87**(7), pp. 1062-1078, 1999.
3. N. F. Johnson and S. Jajodia, "Exploring Steganography: Seeing the Unseen", *IEEE Computer Magazine*, **31**(2), pp. 26-34, 1998.
4. W. B. Pennebaker and J. L. Mitchell, "JPEG still image data compression standard", Chapman & Hall, New York, 1993.
5. A. Said and W. A. Pearlman, "A new fast and efficient implementation of an image codec based on set partitioning in hierarchical trees", *IEEE Transactions on Circuits and Systems for Video Technology*, **6**(3), pp. 243-250, 1996.
6. N. Provos and P. Honeyman, "Detecting Steganographic Content on the Internet", *ISOC NDSS'02*, San Diego CA, February 2002.
7. M. Kutter and S. Winkler, "A Vision-Based Masking Model for Spread-Spectrum Image Watermarking", *IEEE Transactions on Image Processing*, **11**(1), pp. 16-25, 2002.
8. L. M. Marvel, C. G. Boncelet, and C. T. Retter, "Spread Spectrum Image Steganography", *IEEE Transactions on Image Processing*, **8**(8), pp. 1075-1083, 1999.
9. L. M. Marvel, *Image Steganography for Hidden Communication*, Ph.D. thesis, University of Delaware, 1999.
10. D. L. Donoho, "De-Noising by Soft-Thresholding", *IEEE Transactions on Information Theory*, (**41**)3, pp. 613-627, 1995.