QUANTITATIVE STEGANALYSIS OF BINARY IMAGES

Ming Jiang, Nasir Memon, Edward Wong

Dept. of Computer and Inf. Science
Polytechnic University
Brooklyn, New York 11201

Xiaolin Wu

Dept. of Electrical & Comp. E.
McMaster University
Hamilton, Ontario, L8G 4K1

ABSTRACT

We propose a quantitative steganalysis method to detect hidden information embedded by flipping pixels along boundaries in binary images. We model steganographic embedding as an additive noise process and use compression rate as a distinguishing statistic that aids in discriminating between stego-images and cover-images. We specifically use the JBIG 2 binary image compression algorithm to derive a quantitative relation between compression rate and embedding rate. Based on this relationship, a practical steganalysis technique is proposed by examining the change of compression rate as embedding rate increases. Experiments conducted show that the proposed technique can reliably detect a steganographic embedding process that flips boundary pixels. Furthermore, it can estimate embedding rate with reasonable accuracy.

1. INTRODUCTION

Steganography is the science of inconspicuously hiding data within data [1]. Although steganography is an old subject, its modern version was first formulated by Simmons as the *prisoners' problem* [2] where Alice and Bob, two prison inmates covertly communicate by embedding a *secret message* M into a *cover-object* C, to obtain the *stego-object* S. The stego-object S is then sent through the public channel. Wendy, the warden, who examines the stego-object is unaware of the embedded message M within S and hence permits the communication to take place.

Steganalysis, in this context, is the art of detecting and sometimes even decoding hidden data within a given medium [3, 4, 5, 6]. The basic idea behind most steganalysis techniques is that they compute image features that are typically not "normal" in given candidate images. Based on these features, a steganalysis technique classifies an image as a stego-image or a cover-image. If a steganalysis technique can determine whether or not a given image contains a secret message with a success rate better than random guessing, the corresponding steganographic system is considered broken.

In the past few years we have seen the development of a number of steganalysis techniques for image data. Perhaps the most successful ones are techniques that can detect the presence of LSB embedding; the most principled and powerful of which were presented in [4, 7, 8]. In [5] and [6] general purpose methods for steganalysis of images that work with a wide variety of embedding techniques are presented. Fridrich et. al. propose some guidelines for practical steganalysis in [3] and a general detection methodology for quantitative steganalysis in [9].

THIS WORK HAS BEEN SUPPORTED BY AFOSR GRANT F30602-03-C-0091

Although there have been many steganalysis techniques proposed for grayscale and color images, the same is not true for binary images. This situation exists despite the fact that many techniques for embedding data in binary images are now known [10]. These techniques include those based on text line, word, or character shifting, boundary modifications, partitioning of image into blocks and selectively flipping image pixels, modification of run-length patterns, or modifications of half-tone images. For a good survey of these techniques the reader is referred to [10].

In this paper we are mainly concerned with examining the steganographic properties of the embedding techniques based on boundary modifications, fixed partitioning of the image into blocks, and modification of run-length patterns. All these techniques share a common characteristic, they embed information by flipping image pixels. By flipping we mean changing a white pixel to black and vice versa. Furthermore, due to perceptibility constraints, these techniques flip pixels only along character or symbol boundaries [10, 11, 12]. Clearly, for most binary images, flipping pixels inside a white or black region would clearly cause noticeable artifacts. We call such techniques as *pixel-flipping embedding* techniques and treat them the same for the purpose of steganalysis. In this paper we focus on the problem of detecting whether a binary image has undergone a data embedding process using a pixel flipping technique.

The rest of this paper is organized as follows. In the next section we present the proposed steganalysis method and in section three we present experimental results. We conclude in section four with a discussion on future work.

2. PROPOSED STEGANALYSIS TECHNIQUE

We adopt the general quantitative steganalysis methodology for digital images proposed by Fridrich et. al. [9]. Based on this approach, our goal is to identify a good distinguishing statistic of an image that predictably changes with the length of the embedded secret message and find the relationship between this statistic and the embedding rate.

In the rest of this section we present our proposed steganalysis technique for binary images which is based on the relationship between compression rate and the data embedding rate. We first argue that the compressed bit rate of a given image increases when data embedding rate increases. We then present an explicit formula for this relationship when using JBIG 2 as the compression algorithm. This allows us to compute the embedding rate based on the observed values of compressed bit rates of binary images. Before we proceed, we would like to acknowledge that our technique is directly inspired by the work in [9, 7, 13] for gray scale im-

0-7803-8554-3/04/$20.00 ©2004 IEEE.

ages. Furthermore, while writing this paper we were made aware of work independently done in [14] which also explores the relationship between compression rates and data embedding rates, albeit for gray scale images and using an entirely different approach,

2.1. Selected Distinguishing Statistics: Compression Bit Rate

Any data embedding technique can essentially be viewed as an additive process. That is, we can write $S = C + M$, where C, S, and M are the original cover signal, the stego signal, and the embedded message respectively. If we assume the bits of M are i.i.d. and independent of C, it is clear that for any embedding process modelled as above, the stego signal has a higher entropy than that of the cover signal [14], that is $H(C) \leq H(S)$. Actually, we can say more. In all information hiding systems where the embedded message is independent of the cover signal, the entropy of the stego signal is a monotonically increasing function of the embedded signal strength, that is

$$H(S_{\alpha_1}) \leq H(S_{\alpha_2}) \tag{1}$$

where $H(S_{\alpha_1})$ and $H(S_{\alpha_2})$ are the entropy of stego signals when the embedded message strength is α_1 and α_2 respectively, and $\alpha_1 \leq \alpha_2$.

Now, the entropy of a signal is difficult to determine as often we do not have an adequate model for the signal. However, we do know that entropy and compression are closely related. A perfect compression technique encodes at a rate equal to the entropy. Hence it is reasonable to use compression bit rate as an estimate for signal entropy and consequentially as a distinguishing statistic for the purpose of steganalysis.

Before we proceed with a description of the proposed steganalysis technique, we make a brief note about notation. Since we are solely focusing on embedding techniques that flip pixels in binary images, we use the terms embedded signal strength and flipping rate interchangeably. We also use the terms embedding rate and relative message length interchangeably. Finally, we denote flipping rate by p and the embedding rate by α where the two are related by $\alpha = 2 * p$.

2.2. Estimation of Flipping Rate by JBIG 2

Since the entropy rate of a binary image increases in the length of embedded message or the flipping rate p, we seek to estimate the embedded message length by examining the lossless compression rate of the test image. If there is an ideal entropy rate lossless binary image codec (i.e., it achieves the entropy rate of the cover image), then this codec can achieve the lossless rate of the stego image. This gives us an expression for the lossless compression rate $R(p)$ of a stego image with flipping rate p as:

$$R(p) = H + ap \log p \tag{2}$$

under the assumption that the embedded message is statistically independent of the cover signal, where H is the entropy of the original image (cover signal). In other words, the embedded message adds extra $ap \log p$ bits to the entropy of the original source, with a being a constant that governs the percentage of the pixels that are amenable for embedding, namely the object boundaries in the case of binary image steganography.

In order to apply (2) to binary image steganalysis, we need a truly universal binary image codec that approaches source entropy. In practice, the best binary image codec we know is the JBIG 2 binary image compression standard [15]. Our goal is to establish a functional relation $R(p)$ between the lossless bit rates of JBIG 2 and the flipping rate p. Unfortunately, JBIG 2 is not strictly universal in an information theoretical sense, and hence (2) is not an accurate estimation scheme. We have to modify (2) in order to obtain a good estimate of p. We propose to use the following functional relationship when using JBIG 2:

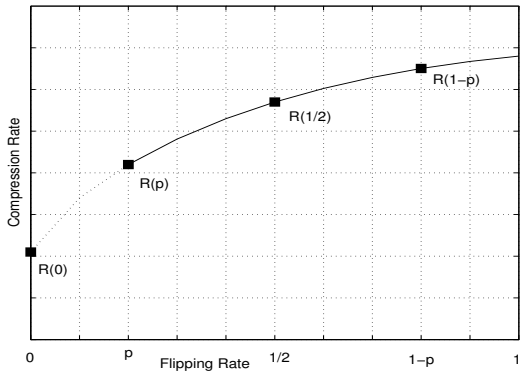

Fig. 1. The function $R(\cdot)$ and the four points on it which can be computed from candidate stego image

$$R(p) = H + ap \log p + bp. \tag{3}$$

To understand the motivation for the above, we recall that JBIG 2 segments a binary image into a set of marks and conditionally saves used marks to build a library [16]. Each entry in the library is stored as bitmaps. The set of marks that together comprise an image can then be represented by storing their spatial locations and also the indices of the corresponding matched entries in the library. Keeping the above JBIG 2 procedure in mind, the second term in the above equation comes from the extra bits needed to represent an entry in the symbol library and the third term from the additional bits needed to represent the symbol library itself.

2.3. Practical Ways to Estimating Message Length

Having established a quantitative relationship between embedding rate and lossless compression rate using JBIG 2, we are now ready to describe our steganalysis procedure. We assume random embedding, i.e., the secret message is randomly spread throughout the entire cover signal. For a given target image, we repeatedly embed messages into it with messages of increasing lengths. Then we examine the corresponding compression rates and use these observed values to arrive at an estimate of the flipping rate p.

Now, for a stego image which already has a message embedded in it, the flipping rate does not change when we embed into the already embedded portion. If the original embedding rate was α_1, and after our embedding, the new embedding rate is α_2, then the combined flipping rate is as follows.

$$p = \alpha/2 = [\alpha_2 + (1 - \alpha_2) * \alpha_1]/2 \tag{4}$$

Using the above observation and in a manner similar to [7], we can compute the values of $R(\cdot)$ at the four points as shown in Figure 1. Their meaning is described as follows:

- *R(p)* is the compressed bit rate of the given image which has an unknown flipping rate of p (possibly zero) that we are trying to estimate.
- *R(1/2)* is the compressed bit rate when the given image is randomly embedded with full capacity as on an average half of the pixels will get flipped.
- *R(1-p)* is the compressed bit rate when all pixels in the image are flipped. Of course this operation restores back the unknown fraction of p pixels already flipped due to message embedding.
- *R(0)* is the compressed bit rate for the cover image. Actually we do not know the exact compression rate of the cover image because we assume we do not have the original cover image. However, we estimate its rate based on noise removal and the idea of soft pattern matching in JBIG 2 algorithm [16]. For a given image, whether it is marked or not, we perform an *opening* morphological operation. We call the image obtained after an opening morphological operation the "reset" image. We claim that the compression rate of the "reset" image is approximately the same as that of the cover image. We conducted extensive experiments to show the validity of our claim. In practice we found the claim to be valid except for the presence of small biases which can be compensated to get a more precise approximation. The estimation accuracy we obtained by our procedure is shown in Figure 2.

Now, given that we can compute the four points above, we can solve four equations to get a, b, H, and p, where p is precisely the flipping rate that we desire to estimate. In the rest of the paper, for simplicity of notation, we rewrite $R(0), R(p), R(1/2)$, and $R(1-p)$ as $R_0, R_p, R_{1/2}$, and R_{1-p} respectively.

3. EXPERIMENTAL RESULTS

To validate our proposed method, experiments were conducted to detect two high-capacity techniques that flip boundary pixels. One was the technique proposed in [11]. The other was the technique proposed in [12]. Both these methods have high embedding capacity.

First, we used the program from the authors of [11] to generate stego images. The test images used contain characters and symbols of different font styles and sizes. They were marked with embedding rates in the range $0 \leq \alpha \leq 1$. We then used 132 test images to examine the estimation accuracy of compression rate of the "reset" images. The histogram of estimate biases are shown in figure 2. The bias is defined as $(\hat{R}_0 - R_0)/R_0$, where $\hat{R}_0$ and R_0 are the estimate and the true compression rate of the cover image, respectively. The x axis is the range of biases, and y axis is the number of images with estimate biases of given values. The results indicate that the estimated rate is fairly close to the true value.

We then used 144 images to investigate the effectiveness of the proposed steganalysis technique. Estimation errors for different flipping rates are shown in Figure 3. Solid line 2 represents the mean of estimation errors. Dotted line 1 and dash-dot line 3 represent the bounds of estimation errors. As can be seen the estimation error gets larger as we approach $\alpha = 0.5$. This is due to the fact that at high embedding rates the three points $R(p), R(1/2)$ and $R(1-p)$ all fall very close to each other and we are not able to estimate the function $R(\cdot)$ accurately. Nevertheless, it should be noted that at these large embedding rates, the detection rate is perfect. That is we are sure there is a large message embedded but an estimation of the length of the message is inaccurate.

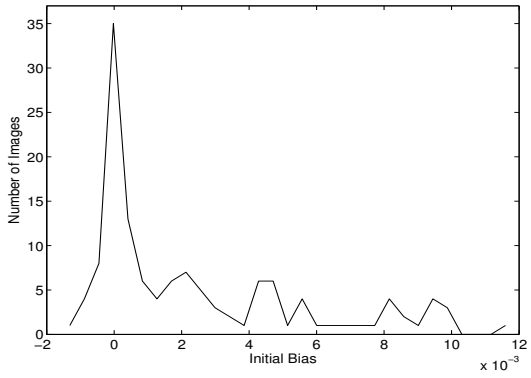

Fig. 2. Estimation error for $R(0)$ for a set of test images. X axis shows error and Y axis the number of images for which we obtained the corresponding estimation error

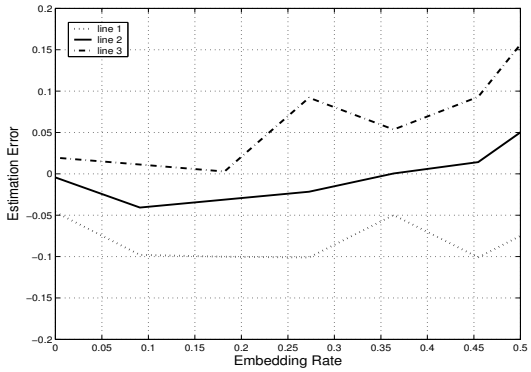

Fig. 3. Estimation Errors of 144 Images for Random Embedding

Next, based on the description in [12], we implemented their data embedding method to generate stego images. We used the same test image set as in the previous experiment. The test images were marked with embedding rates in the range $0 \leq \alpha \leq 1$. Since the embedding method uses shuffling, it essentially provides random embedding. The estimation accuracy of compression rate of the "reset" images is the same as in Figure 2. Again we used 144 images to examine the steganalysis results. The estimation errors are shown in Figure 4. Solid line 2 represents the mean of estimation errors. Dotted line 1 and dash-dot line 3 represent the bounds of estimation errors.

4. CONCLUSION AND FUTURE WORK

We developed a quantitative steganalysis technique for binary images using an additive noise model. Embedded message length estimation was performed according to observed changes in compression rates. The proposed technique is a relatively general one and was designed to detect flipping rate along symbol boundaries. In fact, any binary image steganography technique that employs

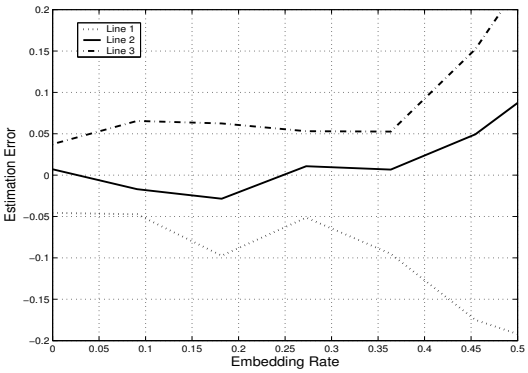

Fig. 4. Estimation Errors of 144 Images for Fixed Partitioning

pixel flipping, has to do so along symbol boundaries to ensure that the imperceptibility condition is satisfied. In this regard, the proposed technique does not consider details of any specific embedding process employed. It simply reports the flipping rate along symbol boundaries. This broadens the scope of its applications.

For random embedding, the techniques provides an accurate estimate of the message length when embedding rate is less than 50 percent. The accuracy of message length estimation partly depends on the estimate of the compression rate of the cover image. For images with single pixel size fonts, estimate of R_0 is not that good according to our observations and hence this gives poor flipping rate estimation results. In fact, we encountered a few such images in our test set. Given this fact, some additional work needs to be done to improve the estimation accuracy of R_0, especially for images which have fonts with single pixel thickness.

Finally, our work can be extended to steganalysis of color/grey images and in this respect, can be considered as an extension of [9].

5. REFERENCES

[1] S. Katzenbeisser and F. Peticolas, *Information Hiding Techniques for Steganography and Digital Watermarking*, Artech House, Boston, London, 2000.

[2] G. J. Simmons, "Prisoner's problem and the subliminal channel," in *Advances in Cryptology: Proceedings of CRYPTO 83*. 1984, pp. 51–67, Plenum Press.

[3] J. Fridrich and M. Goljan, "Practical steganalysis-state of the art," in *Proc. SPIE Photonics Imaging 2002, Security and Watermarking of Multimedia Contents*, Portland, Oregon, USA, April 2002, vol. 4675, pp. 1–13.

[4] A. Westfeld and A Pfitzmann, "Attacks on steganographic systems," in *In Proceedings of Information Hiding, Third International Workshop*, Dresden, Germany, Sept. 1999, pp. 61–75.

[5] Ismail Avcibas, Nasir Memon, and Bulent Sankur, "Steganalysis using image quality metrics," in *Security and Watermarking of Multimedia Contents*, San Jose, CA, Feb. 2001.

[6] S. Lyu and H. Farid, "Detecting hidden messages using higher-order statistics and support vector machines," in *Proc. 5th International Workshop on Information Hiding*, Noordwijkerhout, The Netherlands, 2002.

[7] J. Fridrich, M. Goljan, and R. Du, "Reliable detection of lsb steganography in color and grayscale images," in *Proc. of the ACM Workshop on Multimedia Secuity*, Ottawa, CA, May 2001, pp. 27–30.

[8] Sorina Dumitrescu, Xiaolin Wu, and Zhe Wang, "Detection of lsb steganography via sample pair analysis," *IEEE Transactions on Signal Processing*, vol. 51, no. 7, pp. 1995–2007, July 2003.

[9] J. Fridrich, M. Goljan, D. Hogea, and D. Soukal, "Quantitative steganalysis of digital images: Estimating the secret message length," *ACM Multimedia Systems Journal*, vol. 9, no. 3, pp. 288–302, 2003.

[10] M. Chen, E. K. Wong, N. Memon, and S. Adams, "Recent developments in document image watermarking and data hiding," in *Proc. SPIE Conf on Multimedia Systems and Applications IV*, Denver, CO, Aug 2001.

[11] Q. Mei, E. K. Wong, and N. Memon, "Data hiding in binary text documents," in *SPIE Proc Security and Watermarking of Multimedia Contents III*, San Jose, CA., Jan 2001, vol. 2.

[12] M. Wu, E. Tang, and B. Liu, "Data hiding in digital binary images," in *Proc. IEEE Int'l Conf. on Multimedia and Expo*, New York, Aug 2000.

[13] Jeremiah J. Harmsen and William A. Pearlman, "Steganalysis of additive noise modelable information hiding," in *SPIE Electronic Imaging*, Jan 2003.

[14] Mehmet U. Celik, Gaurav Sharma, and A. Murat Tekalp, "Universal image steganalysis using rate-distortion curves," in *Security and Watermarking of Multimedia Contents*, San Jose, CA, Jan. 2004.

[15] P. Howard and F. Kossentini et. al., "The emerging JBIG 2 standard," *IEEE Trans. on Circuit and Systems for Video Technology*, vol. 8, pp. 838–848, 1998.

[16] Paul G. Howard, "Text image compression using soft pattern matching," *The Computer Journal*, vol. 40, no. 2-3, 1997.