On Steganalysis of Random LSB Embedding in Continuous-tone Images

Sorina Dumitrescu[1], Xiaolin Wu[2], Nasir Memon[3]
Dept. of Computer Science, Univ. of Western Ontario, London, ON, Canada[1]
Dept. of Computer & Info. Science, Polytechnic Univ. Brooklyn, NY[2,3]
sorina@csd.uwo.ca/xwu@poly.edu/memon@poly.edu

Abstract– In this paper we present an LSB steganalysis technique that can detect the existence of hidden messages that are randomly embedded in the least significant bits of natural continuous-tone images. The technique is inspired by recent work of Fridrich *et al.* [1] and just like [1], it can also precisely measure the length of the embedded message, even when the hidden message is very short relative to the image size. The key to our success is the formation of some subsets of pixels whose cardinalities change with LSB embedding, and such changes can be precisely quantified under the assumption that the embedded bits are randomly scattered. Interestingly, our study on steganalysis of LSB embedding sheds light on the recent work of Fridrich *et al.* [1] on the detection of LSB embedding, and offers an analytical proof of an observation made in [1].

1 Introduction

Steganography refers to the science of "invisible" communication. Unlike cryptography, where the goal is to secure communications from an eavesdropper, steganographic techniques strive to hide the very presence of the message itself from an observer. In the general model for steganography, we have Alice wishing to send a secret message m to Bob. In order to do so, she "embeds" m into a cover-object c, to obtain the stego-object s. The stego-object s is then sent through the public channel. In *private key steganography* Alice and Bob share a secret key which is used to embed the message. The secret key is usually a password that can be used to seed a pseudorandom generator to select locations in the cover-object for embedding the secret message (possibly encrypted). Wendy has no knowledge about the secret key that Alice and share, although she might be aware of the algorithm that they could be employing for embedding messages.

The main goal of steganography is to communicate securely in a completely undetectable manner. That is, Wendy should not be able to distinguish in any sense between cover-objects (objects not containing any secret message) and stego-objects (objects containing a secret message). In this context, "steganalysis" refers to the body of techniques that aid Wendy in distinguishing between cover-objects and stego-objects. It should be noted that Wendy has to make this distinction without any knowledge of the secret key which Alice and Bob may be sharing and often also without any knowledge of the specific algorithm that they might be using for embedding the secret message.

Given the proliferation of digital images, and given the high degree of redundancy present in a digital image (despite compression), there has been an increased interest in using digital images as cover-objects for the purpose of steganography. The simplest of such techniques essentially embed the message in a subset of the LSB (least significant bit) plane of the image, possibly after encryption. For a good survey of steganography techniques, the reader is referred to [2]. In this paper we focus our attention on LSB embedding.

The development of techniques for image steganography and the wide-spread availability of tools for the same let to an increased interest in steganalysis techniques for image data. The last few years have seen many new and powerful steganalysis techniques reported in the literature. Perhaps some of the earliest work in this regard was reported by Johnson and Jajodia [3]. A more principled approach to LSB steganalysis was presented in [5] by Westfeld and Pfitzmann [4]. Another steganalysis tools on similar lines but with higher detection rate even for short messages was proposed by Fridrich, Du and Long [5]. A more sophisticated technique that provides remarkable detection accuracy even for short messages was presented by Fridrich et. al. in [6] and described more in detail in Section 3 below. Avcibas, Memon and Sankur [7] present a general purpose technique for steganalysis of images that works with a wide variety of embedding techniques including but not limited to LSB embedding. Chandramouli and Memon [8] do a theoretical analysis of LSB steganography and derive a closed form expression of upper bounds for steganographic capacity

for LSB encoding in general.

In this paper we present a steganalysis technique for LSB embedding. The technique we present was inspired by the work by Fridrich et. al. in [6] and indeed can be considered equivalent to the one in [4], although it is arrived at using a different approach. In fact, our new approach offers analytical explanation of an observation made in [6] and sheds light on why [6] achieved such remarkable accuracy and efficacy.

2 Primary Sets and Their Sensitivity to LSB Embedding

The proposed technique to detect LSB steganography in continuous-tone natural images is based on an important statistical identity related to some sets of pixels. But this identity is very sensitive to LSB embedding, and the change in the identity can quantify the length of the embedded message. Consider the partition of an image into pairs of horizontally adjacent pixels. Let $\mathcal{P}$ be the set of all these pixel pairs. Define the subsets X and Y of $\mathcal{P}$ as follows:

- X is the set of pairs $(u, v) \in \mathcal{P}$ such that v is even and $u < v$, or v is odd and $u > v$.
- Y is the set of pairs $(u, v) \in \mathcal{P}$ such that v is even and $u > v$, or v is odd and $u < v$.

The significance of X and Y in steganalysis of LSB embedding comes from the following assumption.

Assumption 1. Statistically we have

$$|X| = |Y|. \tag{1}$$

The assumption is true for natural images. This is because natural images are isotropic in terms of the gradient of intensity function. In other words the gradient in any direction has equal probability to be positive and negative.

Now let Z be the subset of pairs $(u, v) \in \mathcal{P}$ such that $u = v$. Furthermore, partition set Y into two subsets W and V, with W being the set of pairs in $\mathcal{P}$ of the form $(2k, 2k+1)$ or $(2k+1, 2k)$, and $V = Y - W$. Then $\mathcal{P} = X \cup W \cup V \cup Z$. In the sequel we call the sets X, V, W and Z primary sets.

Let us analyze now what happens when a message is embedded in the last bit plane of pixel values. The embedding modifies the values of some pixels by switching the LSB. Hence the relative ranking of some affected pixel pairs in $\mathcal{P}$ will be changed. Given a pixel pair (u, v), there are four situations:

a) both values u and v remain unmodified;

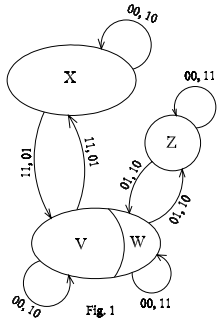

Figure 1: State transition diagram for sets X, V, W, Z under LSB flipping.

b) only u is modified;

c) only v is modified;

d) both u and v are modified.

If we are in situation "a" (b, c, d) we say that the modification pattern due to LSB embedding is 00 (10, 01, 11, respectively). The embedding process leads to change of membership of some pixel pairs between the primary sets, as illustrated by Figure 1. In Figure 1, each arrow drawn from a set A to a set B, labelled by a modification pattern, means that any pixel pair of A becomes a pixel pair of the set B, if modified by the specified pattern.

For each modification pattern $\pi \in \{00, 10, 01, 11\}$ and any subset $A \subseteq \mathcal{P}$, denote by $\rho(\pi, A)$ the probability that the pixel pairs of A are modified with pattern π. The following assumption about a statistical property of LSB steganography is crucial for the success of the steganalytic method to be proposed.

Assumption 2. For each modification pattern $\pi \in \{00, 10, 01, 11\}$ and each primary set $A \in \{X, V, W, Z\}$,

$$\rho(\pi, A) = \rho(\pi, \mathcal{P}).$$

Assumption 2 means that the message bits of LSB steganography are randomly scattered in the image space, independent of image features. Consequently,

i) $\rho(00, \mathcal{P}) = (1 - p/2)^2$;

ii) $\rho(01, \mathcal{P}) = \rho(10, \mathcal{P}) = p/2(1 - p/2)$;

iii) $\rho(11, \mathcal{P}) = (p/2)^2$.

Let p be the length of the embedded message in bits divided by the total number of pixels. Then the fraction of the pixels modified by the LSB embedding is

$p/2$. Assumption 1 and the set migration relationship of Figure 1 allow us to express the cardinalities of the primary sets after the embedding as functions of p and the cardinalities before the embedding. For each $A \in \{X, Y, V, W, Z\}$, denote by A' the set defined in the same way as the set A but considering the pixel values after the embedding. Then we obtain the following equations:

$$|X'| = |X| \cdot (1 - p/2) + |V| \cdot p/2 \quad (2)$$
$$|V'| = |V| \cdot (1 - p/2) + |X| \cdot p/2 \quad (3)$$
$$|W'| = |W|(1 - p + p^2/2) + |Z|p(1 - p/2). \quad (4)$$

Relations (2) and (3) imply that

$$|X'| - |V'| = (|X| - |V|)(1 - p). \quad (5)$$

Since statistically $|X| = |Y|$ we have $|X| = |V| + |W|$, and further

$$|X'| - |V'| = (|W|)(1 - p). \quad (6)$$

Observe from Figure 1 that the embedding process does not modify the set $W \cup Z$, and let $\gamma = |W| + |Z| = |W'| + |Z'|$. Replacing $|Z|$ by $\gamma - |W|$, equation (4) becomes

$$|W'| = |W| \cdot (1 - p)^2 + \gamma \cdot p(1 - p/2). \quad (7)$$

The elimination of $|W|$ from (6) and (7) leads to

$$|W'| = (|X'| - |V'|) \cdot (1 - p) + \gamma \cdot p(1 - p/2). \quad (8)$$

Since $|X'| + |V'| + |W'| + |Z'| = |\mathcal{P}|$, relation (8) is equivalent to

$$0.5\gamma p^2 + (2|X'| - |\mathcal{P}|)p + |Y'| - |X'| = 0. \quad (9)$$

Finally, we arrive at a simple steganalytic method to estimate the length of the embedded message p, based on the values $|X'|, |Y'|$ and $\gamma = |W' \cup Z'|$. The actual value of p is the smaller solution of the quadratic equation (9), provided that $\gamma \neq 0$. Note that X', Y', W', Z' are all defined and can be measured for the image being examined for possible steganography.

If $\gamma = 0$, then (9) becomes an identity regardless of p, and consequently our steganalytic method will fail. Indeed, $\gamma = 0$ implies $|X'| = |X| = |Y| = |Y'| = |\mathcal{P}|/2$. Since $\gamma = |W| + |Z|$, one can see from the definitions of the primary sets that γ is the number of pixel pairs in $\mathcal{P}$ that differ only in the least significant bit. For natural images the probability of zero γ is very small.

3 Relationship with Fridrich *et. al*'s Method

Very recently Fridrich *et al.* proposed a steganalysis technique to detect LSB embedding in continuous-tone images [1]. In this section we describe their technique in some detail and show the relationship between their technique and the approach described in the previous section. Although the approaches taken by the two techniques are quite different, interestingly, it turns out that they are equivalent in some sense as we show below.

The LSB embedding detection method of [1] partitions an image into N/n disjoint groups of n adjacent pixels, where N is the total number of pixels in the image. In [1] the authors considered the case $n = 4$. For simplicity we analyze their method for the case $n = 2$. We adopt the same terminology of [1]. They first define a *discrimination function* $f(\cdot)$ constructed such that it captures the smoothness of a particular group of pixels as follows:

$$f(G) = f(x_1, x_2, ..., x_n) = \sum_{i=1}^{n-1} |x_{i+1} - x_i|$$

In addition, three invertible operations, $F_n(x), n = -1, 0, 1$ on pixel values x, are defined as follows:

- $F_1(x)$ is defined as the operation that performs flipping of the LSB of an image pixel. It effectively maps pixel values as follows:

$$F_1 : 0 \leftrightarrow 1,\ 2 \leftrightarrow 3,\ ...,\ 254 \leftrightarrow 255$$

- $F_{-1}(x)$ maps pixel values in the opposite direction as compared to F_1. Specifically the mapping F_{-1} is defined as follows:

$$F_{-1} : -1 \leftrightarrow 0,\ 1 \leftrightarrow 2,\ ...,\ 255 \leftrightarrow 256$$

- $F_0(x)$ is defined as the identity function.

The specific operations F_1 or F_{-1} can be performed on a group of pixels G by using a mask M, which defines a way to apply the operations on a group of pixels. For example, if $G = (39, 38, 40, 41)$ and $M = (-1, 0, -1, 0)$, then $F_M(G) = (F_{-1}(39), F_0(38), F_{-1}(40), F_0(41))$.

Finally, using the operations F_1, F_{-1} and the discrimination function f, three types of pixels groups are defined as follows:

$$G \in R \Leftrightarrow f(F_1(G)) > f(G)$$
$$G \in S \Leftrightarrow f(F_1(G)) < f(G)$$
$$G \in U \Leftrightarrow f(F_1(G)) = f(G)$$

where R represents a *Regular Group*, S is a *Singular Group*, and U is an *Unusable Group*. Utilizing the definitions above the steganalysis technique consists of the following simple steps:

1. Calculate the number of Regular Groups (R_M) and Singular Groups(S_M) using a positive mask M and a negative mask $-M$. A negative mask is the positive mask with its operations negated, meaning: If $M = (-1, 0, -1, 0)$ then $-M = (1, 0, 1, 0)$.

2. Invert the LSB's of all of the image pixels and repeat step 1.

3. Based on empirical observations, estimate the length of the hidden message from the counts of obtained in steps 1 and 2.

For our analysis of Fridrich's method and its interpretation in the framework of section 2, for brevity, we assume the mask defined in [1] is now a binary word of 2 bits, i.e., $M = (0, 1)$ and $-M = (0, -1)$. Let $R(M)$ denote the set of regular groups with respect to the mask M, and $S(M)$ denote the set of singular groups with respect to mask M. Let $R(-M)$ and $S(-M)$ be the corresponding sets with respect to mask $-M$.

Then, from the definitions of the discrimination function f and of the flipping functions F_M and F_{-M} in [1], it follows that:

$$\begin{aligned} R(M) &= X \cup Z, \qquad S(M) = Y \\ R(-M) &= Y \cup Z, \qquad S(-M) = X. \end{aligned} \tag{10}$$

Note that there are no unusable groups for any of the masks M and $-M$.

Assumption 1 in the previous section is equivalent to the statistical hypothesis of [1] that, for a natural image, the following equalities hold:

$$|S(M)| = |S(-M)|, \quad |R(M)| = |R(-M)|. \tag{11}$$

The random LSB embedding changes regular and singular subsets $R'(M), R'(-M), S'(M), S'(-M)$. Each of these subsets is defined in the same way as the corresponding set without the prime sign, just that now we consider the pixel values after the LSB embedding.

The analysis of the previous section together with (11) lead to the following relations:

$$\begin{aligned} |R'(-M)| &= |R(-M)| + p/2|W|, \\ |S'(-M)| &= |S(-M)| - p/2|W|, \\ |R'(M)| &= |R(M)| - p/2(2|Z| - |W|) \\ &\quad -p^2/2(|W| - |Z|), \\ |S'(M)| &= |S(M)| + p/2(2|Z| - |W|) \\ &\quad +p^2/2(|W| - |Z|) \end{aligned} \tag{12}$$

The first two equations state that $|R'(-M)|$ and $|S'(-M)|$ are linear functions in p, and they diverge as p increases. The next two equations show that $|R'(M)|$ and $|S'(M)|$ are quadratic functions in p, and also $|R'(M)| = |S'(M)|$ when $p/2 = 0.5$. Therefore, we can verify, using a totally different approach, the same observations on $|R'(-M)|$, $|S'(-M)|$, $|R'(M)|$ and $|S'(M)|$ made by Fridrich *et al.* These observations form the basis of the LSB embedding detection technique of [1].

References

[1] J. Fridrich, M. Goljan and R. Dui, "Reliable Detection of LSB steganography in Color and Grayscale Images," *Proc. of the ACM Workshop on Multimedia and Security*, Ottawa, CA, October 5, 2001, pp. 27-30.

[2] N. F. Johnson, S. Katzenbeisser, "A Survey of steganographic techniques", in S. Katzenbeisser and F. Petitcolas (Eds.): Information Hiding, pp. 43-78. Artech House, Norwood, MA, 2000.

[3] [4] N. F. Johnson, S. Jajodia, "Steganalysis of Images created using current steganography software", in David Aucsmith (Ed.): Information Hiding, LNCS 1525, pp. 32-47. Springer-Verlag Berlin Heidelberg 1998.

[4] A. Westfeld, A. Pfitzmann, "Attacks on Steganographic Systems", in Andreas Pfitzmann (Ed.): Information Hiding, LNCS 1768, pp. 61-76, Springer-Verlag Berlin Heidelberg, 1999.

[5] J. Fridrich, R. Du, M. Long, "Steganalysis of LSB Encoding in Color Images", Proceedings of ICME 2000, New York City, July 31-August 2, New York, USA.

[6] J. Fridrich, M. Goljan and R. Du, "Reliable Detection of LSB Steganography in Color and Grayscale Images". Proc. of the ACM Workshop on Multimedia and Security, Ottawa, CA, October 5, 2001, pp. 27-30.

[7] I. Avcibas, N. Memon and B. Sankur, "Steganalysis Using Image Quality Metrics", Security and Watermarking of Multimedia Contents, SPIE, San Jose, CA, February 2001.

[8] R. Chandramouli and N. Memon, "Analysis of LSB Based Image Steganography Techniques." Proceedings of the International Conference on Image Processing, Thessaloniki, Greece, October 2001.