

Detection of LSB Steganography via Sample Pair Analysis

Sorina Dumitrescu, Xiaolin Wu* and Zhe Wang

Department of Electrical and Computer Engineering

McMaster University

Hamilton, Ontario, Canada L8S 4M2

sorina/xwu@mail.ece.mcmaster.ca/zwang@grads.ece.mcmaster.ca

©2003 IEEE. Personal use of this material is permitted. Permission from IEEE must be obtained for all other uses, in any current or future media, including reprinting/republishing this material for advertising or promotional purposes, creating new collective works, for resale or redistribution to servers or lists, or reuse of any copyrighted component of this work in other works.

Abstract– This paper introduces a new, principled approach to detecting LSB steganography in digital signals such as images and audio. It is shown that the length of hidden message embedded in the least significant bits of signal samples can be estimated with relatively high precision. The new steganalytic approach is based on some statistical measures of sample pairs that are highly sensitive to LSB embedding operations. The resulting detection algorithm is simple and fast. To evaluate the robustness of the proposed steganalytic approach, bounds on estimation errors are developed. Furthermore, the vulnerability of the new approach to possible attacks is also assessed, and counter measures are suggested.

I. INTRODUCTION

Steganography is an art of sending a secret message under the camouflage of a carrier content. The carrier content appears to have totally different but normal (“innocent”) meanings. The goal of steganography is to mask the very presence of communication, making the true message not discernible to the observer. The wide use of the internet as a mass communication means and the proliferation of digital multimedia on the web present unique opportunities for modern steganography. Recent years have seen increased interests and even commercial software in using digital media files, such as images, audio, and video files, as carrier contents of steganography. A popular digital steganography technique is so-called least significant bit embedding (LSB embedding). With the LSB embedding technique, the two parties in communication share a private secret key that creates a random sequence of samples of a digital signal. The secret message, possibly encrypted, is embedded in the least significant bits of those samples of the sequence. This digital steganography technique takes the advantage of random noise present in the acquired media data, such as images, video and audio. Since the magnitude of signal noise is comparable to that of the least significant bits, embedding message bits in the least significant bit plane will not cause any discernible difference from the original visual or audio signals. Earlier works on steganalysis of LSB embedding in grey-scale and color images were reported in [3], [5], [1], [2], and a survey of steganography techniques can be found in [4].

In this paper we present a new robust steganalytic technique for detection of LSB embedding in digital signals. The technique is based on a finite state machine whose states are selected multisets of sample pairs, called *trace multisets*. Some of the trace multisets are equal in

their expected cardinalities, if the sample pairs are drawn from a digitized continuous signal. Random LSB flipping causes transitions between these trace multisets with given probabilities, and consequently alters the statistical relations between the cardinalities of trace multisets. Furthermore, the statistics of sample pairs is highly sensitive to LSB embedding, even when the embedded message length is very short. By analyzing these relations and modeling them with the finite-state machine, we arrive at a simple quadratic function that can estimate the length of embedded message with high precision, under an assumption that is true in reality for continuous signals such as natural images and audio. Furthermore, we can also bound the estimation error in terms of the degree that this assumption deviates from the reality.

The paper is structured as follows. In Section 2 we will study, as the foundation of our new approach of steganalysis, some interesting and useful statistical properties of sample pairs of a continuous signal. Some special multisets of sample pairs, called trace multisets, are introduced. The behavior of trace multisets under LSB embedding operations is modeled by a finite-state machine. Then in Section 3 we use the structure of the finite-state machine to establish quadratic equations for the length of embedded message in terms of the cardinalities of trace multisets. The accuracy of the estimated hidden message length computed by the quadratic equations is analyzed in Section 4. We also discuss how to use trace multisets and how to draw sample pairs from a signal to minimize estimation errors. Section 5 presents our experimental results with a test set of 29 continuous-tone images. Possible attacks to the proposed steganalytic method are examined and counter measures are suggested in Section 6. Section 7 relates the proposed approach of steganalysis to the RS method of [2], and proves some key observations on which the RS method was based. In order not to obscure our main ideas we put necessary but lengthy mathematical developments in four appendices.

II. TRACE MULTISSETS OF SAMPLE PAIRS

In this section, to motivate the proposed approach of steganalysis, let us study the effects of LSB embedding on some selected sets of sample pairs. Assuming that the digital signal is represented by the succession of samples s_1, s_2, \dots, s_N (the index represents the location of a sample in a discrete waveform), a sample pair means a two-tuple (s_i, s_j) , $1 \leq i, j \leq N$. We use sample pairs rather than individual samples as the basic unit in our steganalysis to utilize higher order statistics such as sample correlation. Let \mathcal{P} be a set of sample pairs drawn from

a digitized continuous signal. We will later come back to the issue of how these sample pairs should be chosen to aid steganalysis.

In the following development, however, it is more convenient to treat \mathcal{P} , a set of sample pairs, as a multiset of two tuples (u, v) , where u and v are the values of two samples. In the sequel, unless otherwise explicitly stated, two-tuples (u, v) , or members of \mathcal{P} , always refer to *values* of two different samples drawn from a signal. Denote by D_n the submultiset of \mathcal{P} that consists of sample pairs of the form $(u, u + n)$ or $(u + n, u)$, i.e., the two values differ exactly by n , where n is a fixed integer, $0 \leq n \leq 2^b - 1$, and b is the number of bits to represent each sample value. In order to analyze the effects of LSB embedding on D_n , it is useful to introduce some other submultisets of \mathcal{P} that are closed under the embedding operation, in terms of the pairwise difference of sample values. Since the embedding affects only the LSB, we use the most significant $b - 1$ bits in choosing these closed multisets. For each integer m , $0 \leq m \leq 2^{b-1} - 1$, denote by C_m the submultiset of \mathcal{P} that consists of the sample pairs whose values differ by m in the first $(b - 1)$ bits (i.e., by right shifting one bit and then measuring the difference).

To summarize the above, we introduced the multisets D_n , $0 \leq n \leq 2^b - 1$, to characterize the changes caused by the LSB embedding in the difference between two sample values. We also introduced the multisets C_m , $0 \leq m \leq 2^{b-1} - 1$, which are invariant under the LSB embedding. Note that the multisets D_n form a partition of \mathcal{P} , and the multisets C_m form another partition of \mathcal{P} . It is interesting to investigate the relation between these two partitions. It is clear that D_{2m} is contained in C_m . Indeed, if (u, v) is a pair in D_{2m} (i.e. $|u - v| = 2m$), then both u and v are either even or odd. By right shifting one bit and taking the absolute difference, the value obtained is exactly $|u - v| / 2$, hence $(u, v) \in C_m$. This is not true however for D_{2m+1} . The sample pairs of D_{2m+1} are shared between C_m and C_{m+1} . Specifically, if (u, v) is a pair in D_{2m+1} , then the pair can have one of the following forms: $(2k - 2m - 1, 2k)$, $(2k, 2k - 2m - 1)$, $(2k - 2m, 2k + 1)$ or $(2k + 1, 2k - 2m)$ for some k . The pairs $(2k - 2m - 1, 2k)$ and $(2k, 2k - 2m - 1)$ are in C_{m+1} . This is because by right shifting one bit, the values $2k$ and $2k - 2m - 1$ become k and $k - m - 1$ respectively, which differ by $m + 1$. But the other two forms of pairs, $(2k - 2m, 2k + 1)$ and $(2k + 1, 2k - 2m)$, are in C_m (by right shifting one bit the values $2k + 1$ and $2k - 2m$ become k and $k - m$ respectively, which differ by m).

Since D_{2m+1} is shared between C_m and C_{m+1} we partition D_{2m+1} into two submultisets X_{2m+1} and Y_{2m+1} , where $X_{2m+1} = D_{2m+1} \cap C_{m+1}$ and $Y_{2m+1} = D_{2m+1} \cap C_m$, for $0 \leq m \leq 2^{b-1} - 2$,

and $X_{2^{b-1}} = \emptyset$, $Y_{2^{b-1}} = D_{2^{b-1}}$. Consequently, X_{2m+1} is the submultiset of sample pairs of the form $(2k - 2m - 1, 2k)$ or $(2k, 2k - 2m - 1)$, and Y_{2m+1} the submultiset of sample pairs of the form $(2k - 2m, 2k + 1)$ or $(2k + 1, 2k - 2m)$. A simpler characterization of these two types of submultisets, which reveals both common and distinctive features of them, is the following. Both types contain pairs (u, v) that differ by $2m + 1$ (i.e. $|u - v| = 2m + 1$). Those pairs in which the even component is larger are in X_{2m+1} , whereas those pairs in which the odd component is larger are in Y_{2m+1} . For natural signals, the chance for a sample pair in D_{2m+1} to have a larger or smaller even component is the same, meaning that for any integer m , $0 \leq m \leq 2^{b-1} - 2$,

$$E\{|X_{2m+1}|\} = E\{|Y_{2m+1}|\}. \quad (1)$$

In section 4 we present empirical evidence collected from 29 natural continuous-tone image that validates assumption (1). In that section we will also analyze how the validity of assumption (1) impacts on the precision of our steganalytic method.

In order to analyze the effects of LSB embedding on sample pairs, let us consider all four possible cases of LSB flipping, labelled by four so-called modification patterns π : 00, 01, 10, 11, with 1 indicating which sample(s) of a pair has(have) the LSB reversed, 0 indicating intact sample(s). For each m , $1 \leq m \leq 2^{b-1} - 1$, the submultiset C_m is partitioned into X_{2m-1} , D_{2m} and Y_{2m+1} . It is clear that C_m is closed under the embedding, but X_{2m-1} , D_{2m} and Y_{2m+1} are not. Take an arbitrary sample pair (u, v) of X_{2m-1} . Then $(u, v) = (2k - 2m + 1, 2k)$ or $(u, v) = (2k, 2k - 2m + 1)$. By modifying the sample pair (u, v) with the pattern 10 the sample pair obtained is $(u', v') = (2k - 2m, 2k)$ or $(u', v') = (2k + 1, 2k - 2m + 1)$. Likewise, if (u, v) is modified by the pattern 01, then $(u', v') = (2k - 2m + 1, 2k + 1)$ or $(u', v') = (2k, 2k - 2m)$. These observations illuminate on the usefulness of multisets X_{2m} and Y_{2m} for steganalysis, where X_{2m} is defined as the submultiset of \mathcal{P} consisting of all pairs of the form $(2k - 2m, 2k)$ or $(2k + 1, 2k - 2m + 1)$, and Y_{2m} is defined as the submultiset of \mathcal{P} consisting of all pairs of the form $(2k - 2m + 1, 2k + 1)$ or $(2k, 2k - 2m)$. It is clear that X_{2m} and Y_{2m} form a partition of D_{2m} .

In summary, multiset C_m with $1 \leq m \leq 2^{b-1} - 1$, can be partitioned into four submultisets X_{2m-1} , X_{2m} , Y_{2m} and Y_{2m+1} , called the *trace submultisets* of C_m . And multiset C_m is closed but its four trace submultisets are not under the LSB embedding operations. This phenomenon can be modeled by a finite-state machine as depicted by Fig. 1. The finite-state machine shows

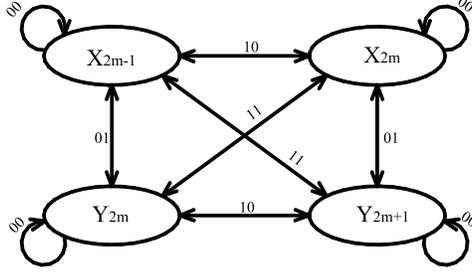


Fig. 1. The finite-state machine whose states are trace multisets of C_m . Note that C_m is closed under LSB steganography but its four subsets are not.

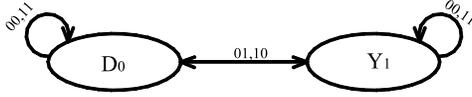


Fig. 2. The finite-state machine associated with C_0 .

how the sample pairs are driven from and to the multisets X_{2m-1} , X_{2m} , Y_{2m} and Y_{2m+1} by different LSB modification patterns. Each arrow drawn from multiset A to multiset B , labeled by a modification pattern, means that any sample pair of A becomes a pair of the multiset B , if modified by the specified pattern. It is straightforward to construct the finite-state machine of Fig. 1 based on the definition of the four trace submultisets of C_m . We have shown that $X_{2m-1} \xrightarrow{10} X_{2m}$ and $X_{2m-1} \xrightarrow{01} Y_{2m}$, and the other transitions can be similarly derived.

The finite-state machine of Fig. 1 does not apply to the multiset C_0 . We need to model the behavior of C_0 under embedding separately. Multiset C_0 is closed under LSB embedding, and can be partitioned into Y_1 and D_0 . The transitions within C_0 are illustrated in Fig. 2.

The significance of the finite-state machines of Fig. 1 and Fig. 2 is that one can statistically measure the cardinalities of the trace multisets before and after the LSB embedding using the probabilities of modification patterns applied to each multiset. Moreover, as we will see in the next section, if the LSB embedding is done randomly in the time domain, then these probabilities are functions of the length of the hidden message.

III. DETECTION OF LSB STEGANOGRAPHY

For each modification pattern $\pi \in \{00, 10, 01, 11\}$ and any submultiset $A \subseteq \mathcal{P}$, denote by $\rho(\pi, A)$ the probability that the sample pairs of A are modified with pattern π as a result of the

embedding. Let p be the length of the embedded message in bits divided by the total number of samples in a multimedia file. Then the fraction of the samples modified by the LSB embedding is $p/2$. Assuming that the message bits of LSB steganography are randomly scattered in the time domain (or image domain if the cover object is an image) we have:

- 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 A and B be submultisets of \mathcal{P} such that $A \subseteq B$. We say that multiset A is unbiased with respect to B if $\rho(\pi, A) = \rho(\pi, B)$ holds for each modification pattern $\pi \in \{00, 10, 01, 11\}$. (When $B = \mathcal{P}$ we simply say that A is unbiased.) If all four trace submultisets of C_m are unbiased, C_m is said to be unbiased. As a convention in sequel, we denote each multiset defined above by A or A' , depending on if the multiset is obtained from the original signal or tampered signal of LSB embedding. The same convention also applies to sample values such that (u, v) and (u', v') are the values of a sample pair before and after LSB embedding. When the message bits LSB steganography are randomly scattered in the time domain, it follows that each C_m is unbiased and one can derive from the finite-state machine of Fig. 1 that

$$|X_{2m-1}|(1-p)^2 = \frac{p^2}{4}|C_m| - \frac{p}{2}(|D'_{2m}| + 2|X'_{2m-1}|) + |X'_{2m-1}|, \quad (2)$$

$$|Y_{2m+1}|(1-p)^2 = \frac{p^2}{4}|C_m| - \frac{p}{2}(|D'_{2m}| + 2|Y'_{2m+1}|) + |Y'_{2m+1}|, \quad (3)$$

where $1 \leq m \leq 2^{b-1} - 1$. And for the special case $m = 0$, we have from Fig. 2 that

$$|Y_1|(1-p)^2 = |C_0|\frac{p^2}{2} - \frac{p}{2}(2|D'_0| + 2|Y'_1|) + |Y'_1|. \quad (4)$$

A proof of (2) and (3) is presented in Appendix A in order not to disrupt the presentation of our main ideas.

From equations (2), (3) and (4) together with the property $E\{|X_{2m+1}|\} = E\{|Y_{2m+1}|\}$, $0 \leq m \leq 2^{b-1} - 2$, we finally obtain the following quadratic equations to estimate the value of p :

$$\frac{(|C_m| - |C_{m+1}|)p^2}{4} - \frac{(|D'_{2m}| - |D'_{2m+2}| + 2|Y'_{2m+1}| - 2|X'_{2m+1}|)p}{2} + |Y'_{2m+1}| - |X'_{2m+1}| = 0, \quad m \geq 1, \quad (5)$$

and

$$\frac{(2|C_0| - |C_1|)p^2}{4} - \frac{(2|D'_0| - |D'_2| + 2|Y'_1| - 2|X'_1|)p}{2} + |Y'_1| - |X'_1| = 0, \quad m = 0. \quad (6)$$

Note that all quantities in (5) and (6) can be obtained from the signal being examined for possible presence of LSB embedding in it. No knowledge of the original signal is required. The smaller root of quadratic equation (5) (or equation (6)) is the estimated value of p , provided that $|C_m| > |C_{m+1}|$ and $|D_{2m}| \geq |D_{2m+2}|$ (or $2|C_0| > |C_1|$ and $2|D_0| \geq |D_2|$). Indeed, the inequalities

$$2|C_0| > |C_1| > |C_2| > \cdots > |C_m| > |C_{m+1}| > \cdots, \quad (7)$$

$$2|D_0| > |D_2| > |D_4| > \cdots > |D_{2m}| > |D_{2m+2}| > \cdots \quad (8)$$

hold under rather relaxed conditions. Let U and V be discrete random variables corresponding to the first and second values of the sample pairs of \mathcal{P} that have joint probability mass function (pmf) $P(u, v)$. Consider the difference between U and V , a new random variable $Z = U - V$. Then the probability mass function of Z , $P_Z(z)$, is a projection of the joint pmf $P(u, v)$ in the direction $(1, 1)$. If the sample pairs of \mathcal{P} are drawn at random, then clearly $P_Z(z)$ has zero mean since $E\{U\} = E\{V\}$.

Also note that if

$$2|D_0| > |D_1| > |D_2| > \cdots > |D_i| > |D_{i+1}| > \cdots, \quad (9)$$

then (7) and (8) follow, based on assumption (1). A sufficient condition for (9) to hold is that $P_Z(z)$ is unimodal and peaks at mean. This condition is satisfied by a large class of joint distributions, including the family of Kotz-type elliptical joint distributions

$$P(u, v) = \alpha(r, s) |\Sigma|^{-1/2} \exp\{-r[((u, v) - (\mu_u, \mu_v))\Sigma^{-1}((u, v) - (\mu_u, \mu_v))']^s\} \quad (10)$$

where r and s are constants, and α is a scaling function in r and s to make $P(u, v)$ a probability function. This family includes the joint Gaussian distribution as a special case. If \mathcal{P} consists of spatially adjacent sample pairs rather than randomly drawn, then $|D_i|$ even has an exponential decay in i (see Fig. 5 for a preview of the distribution in practice).

For proving that the actual value of p equals the smaller of the two real roots of equation (5) it suffices to show that

$$p \leq \frac{(|D'_{2m}| - |D'_{2m+2}| + 2|Y'_{2m+1}| - 2|X'_{2m+1}|)}{|C_m| - |C_{m+1}|}. \quad (11)$$

The right side of the above inequality represents the semisum of the two solutions of equation (5). Relation (11) is equivalent to

$$p \leq \frac{(|C_m| - |C_{m+1}| + |Y'_{2m+1}| - |X'_{2m-1}| + |Y'_{2m+3}| - |X'_{2m+1}|)}{|C_m| - |C_{m+1}|}. \quad (12)$$

Using (40) and $|C_m| > |C_{m+1}|$, (12) becomes

$$(p - 1)(|C_m| - |C_{m+1}|) \leq (1 - p)(|Y_{2m+1}| - |X_{2m-1}| + |Y_{2m+3}| - |X_{2m+1}|) \quad (13)$$

Applying (1), the above inequality reduces to

$$0 \leq (1 - p)(|D_{2m}| - |D_{2m+2}|). \quad (14)$$

IV. ACCURACY OF ESTIMATED HIDDEN MESSAGE LENGTH

In this section we examine the factors that influence the robustness of the steganalytic technique developed above, and suggest ways of improving the accuracy of estimated hidden message length.

Given a chosen multiset \mathcal{P} of sample pairs, the proposed LSB steganalytic technique hinges on assumption (1). The accuracy of the estimated hidden message length \hat{p} made by (5) or (6) primarily depends on the actual difference

$$\epsilon_m = |X_{2m+1}| - |Y_{2m+1}|. \quad (15)$$

An immediate reaction to this observation is to compute the estimate \hat{p} with (5) or (6) for an m value such that $|\epsilon_m|$ is as small as possible. For natural signals that have reasonably smooth waveforms, the smaller the value of m , the smaller the difference $|\epsilon_m|$. In Fig. 3 we plot the value

$$\frac{|\epsilon_m|}{|X_{2m+1}| + |Y_{2m+1}|}$$

averaged over 29 continuous-tone test images as a function of m .

However, a more robust estimate of hidden message length can be obtained by combining trace multisets for a range of m values in which $|\epsilon_m|$ is small. For arbitrary $1 \leq i \leq j \leq 2^{b-1} - 1$,

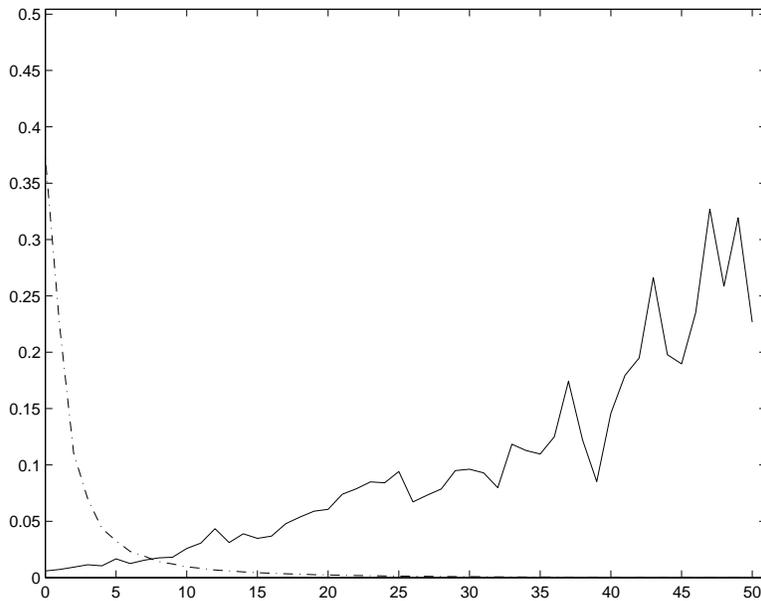


Fig. 3. Solid line: relative error $|\epsilon_m|/(|X_{2m+1}|+|Y_{2m+1}|)$ for $0 \leq m \leq 50$; Dash-dot line: the percentage of $|X_{2m+1}|+|Y_{2m+1}|$ in $\sum_{m=0}^{127} (|X_{2m+1}|+|Y_{2m+1}|)$. Note the exponential decay of $|X_{2m+1}|+|Y_{2m+1}|$.

the finite-state machines of Fig. 1 for C_m , $1 \leq m \leq 2^{b-1} - 1$, can be combined and extended to $\cup_{m=i}^j C_m$ by replacing the trace multisets X_{2m-1} , X_{2m} , Y_{2m} , Y_{2m+1} with $\cup_{m=i}^j X_{2m-1}$, $\cup_{m=i}^j X_{2m}$, $\cup_{m=i}^j Y_{2m}$, $\cup_{m=i}^j Y_{2m+1}$ respectively. We say that the multiset $\cup_{m=i}^j C_m$ is unbiased if the four unions of trace multisets considered above are unbiased. The advantage of combining multiple trace multisets for different m values is that

$$E\{|\cup_{m=i}^j X_{2m+1}|\} = E\{|\cup_{m=i}^j Y_{2m+1}|\} \quad (16)$$

is a more relax condition to satisfy than (1) with respect to individual m . In other words, $|\sum_{m=i}^j \epsilon_m|$ tends to be significantly smaller than $|\epsilon_m|$ for a fixed m , which is a determining factor of the accuracy of the proposed steganalytic approach as we will see shortly. Note that (16) does not require that (1) holds for all m . Instead, (16) only requires that for a sample pair $(u, v) \in \mathcal{P}$ with $|u - v| = 2t + 1$, $i \leq t \leq j$, the even value of u and v has equal probability to be larger or smaller than the odd value of u and v . This is true for natural signals. To corroborate on this assertion we plot in Fig. 4 the relative error term

$$\frac{||\cup_{m=i}^j X_{2m+1}| - |\cup_{m=0}^j Y_{2m+1}||}{|\cup_{m=i}^j X_{2m+1}| + |\cup_{m=0}^j Y_{2m+1}|}$$

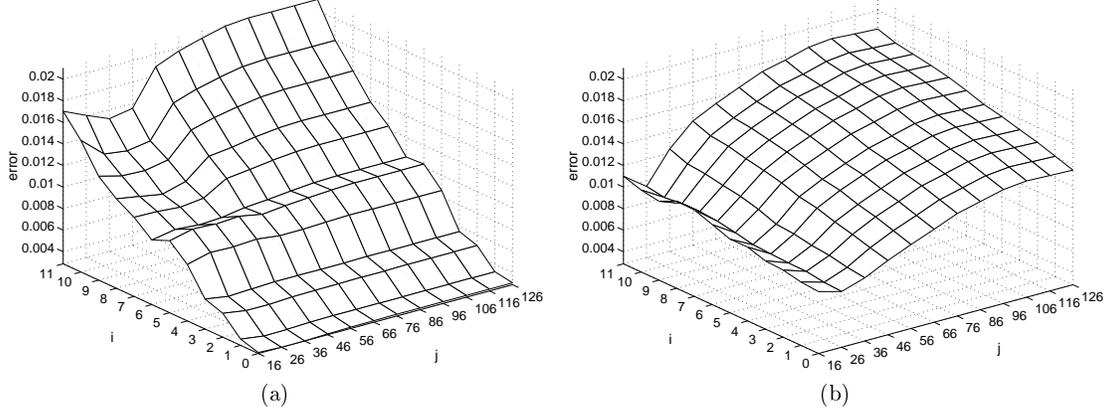


Fig. 4. Relative error of (16): $\frac{||\cup_{m=i}^j X_{2m+1}| - |\cup_{m=i}^j Y_{2m+1}||}{|\cup_{m=i}^j X_{2m+1}| + |\cup_{m=i}^j Y_{2m+1}|}$. a): the case of spatially adjacent sample pairs in \mathcal{P} ; b): the case of randomly chosen sample pairs in \mathcal{P} .

as a function of i and j . The statistics of Fig. 4 is collected from the 29 test images of our experiment (see Section V). The graph shows that the error of (16) is very small for $i = 0$ and appropriate j value. Given a j value the error of (16) increases in i . The error takes on the minimum when $i = 0$ and j is approximately 30. Another important observation is that (16) is far more accurate if \mathcal{P} consists of spatially adjacent sample pairs than if it consists of randomly chosen sample pairs.

As pointed out above, the four unions of trace multisets $\cup_{m=i}^j X_{2m-1}$, $\cup_{m=i}^j Y_{2m+1}$, $\cup_{m=i}^j X_{2m}$, and $\cup_{m=i}^j Y_{2m}$ have the same finite-state machine structure as in Fig. 1. Based on this finite-state machine structure, the statistical relation of (16), and the fact that the multisets $\cup_{m=i}^j C_m$ and $\cup_{m=i+1}^{j+1} C_m$ are unbiased if LSB steganography is done via random embedding, we can derive, in analogous way to Appendix A, the following more robust quadratic equations for estimating p :

$$\begin{aligned} & \frac{p^2}{4} (|C_i| - |C_{j+1}|) - \frac{p}{2} (|D'_{2i}| - |D'_{2j+2}|) + \\ & 2 \sum_{m=i}^j (|Y'_{2m+1}| - |X'_{2m+1}|) + \\ & \sum_{m=i}^j (|Y'_{2m+1}| - |X'_{2m+1}|) = 0, \quad i \geq 1. \end{aligned} \quad (17)$$

In fact, by summing up (5) for consecutive values of m , $1 \leq i \leq m \leq j \leq 2^{b-1} - 2$, we can also arrive at (17).

Similarly, based on (16) and the assumption that the multisets C_0 , $\cup_{m=1}^j C_m$ and $\cup_{m=1}^{j+1} C_m$ are

unbiased for $0 = i \leq j \leq 2^{b-1} - 2$, which is true for random LSB embedding, we have

$$\begin{aligned} & \frac{p^2}{4}(2|C_0| - |C_{j+1}|) - \frac{p}{2}[2|D'_0| - |D'_{2j+2}|] + \\ & 2 \sum_{m=0}^j (|Y'_{2m+1}| - |X'_{2m+1}|) + \\ & \sum_{m=0}^j (|Y'_{2m+1}| - |X'_{2m+1}|) = 0, \quad i = 0. \end{aligned} \quad (18)$$

We can solve either of the two quadratic equations in p , depending on the start index value i , for the smaller root that is the estimated p .

Next we develop a bound on the estimation error of (17) and (18). The error bound is a function of the actual differences

$$\epsilon_m = |X_{2m+1}| - |Y_{2m+1}|, \quad (19)$$

$0 \leq m \leq 2^{b-1} - 2$. For $1 \leq i \leq j \leq 2^{b-1} - 2$, denote

$$e_{ij} = \frac{2 \sum_{m=i}^j \epsilon_m}{|D_{2i}| - |D_{2j+2}|}, \quad (20)$$

and for $0 = i \leq j \leq 2^{b-1} - 2$, denote

$$e_{0j} = \frac{2 \sum_{m=0}^j \epsilon_m}{2|D_0| - |D_{2j+2}|}. \quad (21)$$

Mention that, under some very easy to met assumptions, the denominator of e_{ij} is positive. We can bound the estimation error as below

$$|p - \hat{p}(i, j)| \leq \frac{2|e_{ij}|}{1 - e_{ij}}(1 - p), \quad (22)$$

for all $0 \leq i \leq j \leq 2^{b-1} - 2$, where $\hat{p}(i, j)$ is the estimated value of p obtained by solving (17) (when $i \geq 1$) or (18) (when $i = 0$), provided that $e_{ij} < 1$ and the LSB embedding is done randomly in the time or spatial domain of the signal. The derivation of error bound (22) is given in Appendix B.

To reduce estimation error we want to make $|e_{ij}|$ small. In other words, we would like to reduce $|\sum_{m=i}^j \epsilon_m|$ and increase $|D_{2i}| - |D_{2j+2}|$. Observe from Fig. 4 that $|\sum_{m=i}^j \epsilon_m|$ decreases in general as the difference between i and j increases. But more critically to robust estimation of p , given an i , the larger the distance $j - i$, the larger the difference $|D_{2i}| - |D_{2j+2}|$. This is because $|D_{2i}|$ is a monotonically decreasing function in i (see Fig. 5). Therefore, we should let $i = 0$ and choose a sufficiently large j in (18) to obtain robust estimate of p .

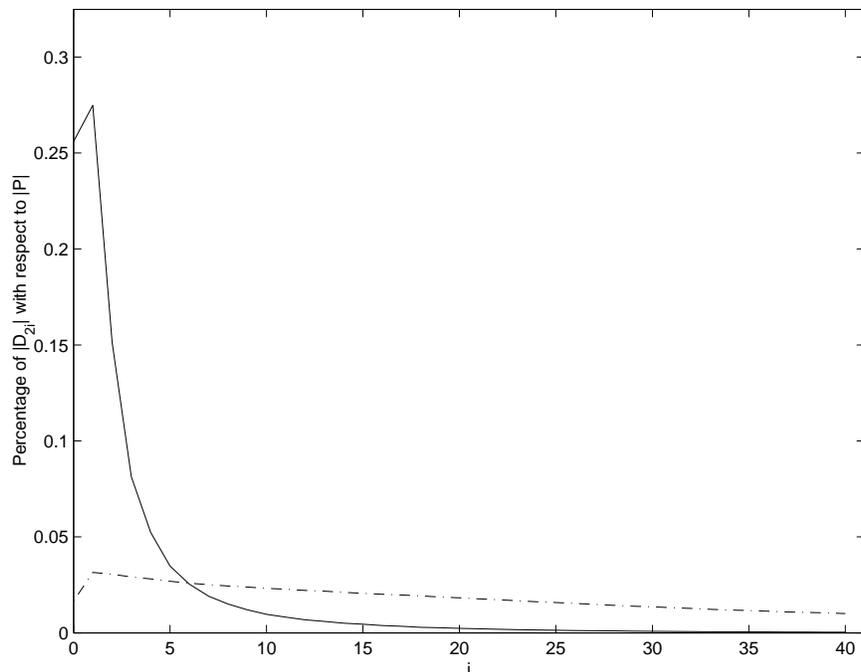


Fig. 5. Probability function $P(|D_{2i}|)$ for spatially adjacent sample pairs (solid line), and for randomly selected sample pairs (dash-dot line). The sample statistics is collected from a set of 29 continuous-tone images.

The estimate accuracy is also affected by the way how the sample pairs of the multiset \mathcal{P} are chosen. Appendix C shows that the more the two values of sample pairs are correlated, the faster $|D_i|$ decreases in i . This means that given i and j , $|D_{2i}| - |D_{2j+2}|$, the denominator of (20), are larger if the sample pairs of \mathcal{P} are drawn from closer positions of a signal waveform. Consequently, for more robust estimate of p the members of multiset \mathcal{P} should be pairs of two spatially adjacent samples (assuming the signal source is Markov). This reasoning is well corroborated in practice. To illustrate this fact we plot in Fig. 5 the probability mass function of $P(|D_{2i}|)$ against all possible i values for two different multisets \mathcal{P} : one consisting of spatially adjacent sample pairs, and the other consisting of randomly selected sample pairs.

The analysis of Appendix C also means that the estimate \hat{p} is more robust if samples of the signal are more highly correlated, and vice versa.

V. EXPERIMENTAL RESULTS

The proposed LSB steganalytic technique is implemented and tested on a set of 29 continuous-tone images of both color and gray-scale types. This test image set includes all original ISO/JPEG

test images such as barb, balloon, goldhill, girl, etc., and those of the kodak set. Twenty-four sample images of our test set are given in Fig. 6. As we can see the test set includes a wide range of natural images, from natural scenery to man-made objects like buildings, and from panoramic views to close-up portraits. This makes the test results to be reported in this section indicative of the performance of the proposed steganalytic technique in reality.

Guided by our estimation error analysis of the proceeding section, in our experiments we form the multiset \mathcal{P} by selecting all pairs of 4-connected pixels. The inclusion of both vertically and horizontally adjacent pixel pairs in \mathcal{P} also accounts for sample correlation in both directions.

The accuracy of the LSB steganography detection technique is evaluated for hidden message lengths $p = 0, 3, 5, 10, 15, 20\%$, where p is measured by the percentage of the number of message bits in the total number of pixels in the test image. In our simulation the embedded message bits are randomly scattered in a test image. Fig. 7(a) plots the distribution of the estimates \hat{p} of different test images for different embedded message lengths p . The vertical difference between an estimate point and the diagonal line is the estimation error $\hat{p} - p$. It is evident from Fig. 7(a) that our LSB steganalytic technique is highly effective, making very good estimate of p . The average error magnitude is only 0.023, and it stays almost the same for different p values.

We define false alarm rate as the probability that the steganalytic technique reports the existence of embedded message when the input signal is truly original, and the missing rate as the probability that a tampered signal evades the steganography detection. If one is to set a threshold of $\hat{p} > 0.018$ to determine whether a hidden message is embedded in the image, then for the test set the false alarm rate when $p = 0$ is 13.79%, and the missing rate is 11.03% when $p = 3\%$. The missing rate quickly drops to 0 if the embedded message length $p > 3\%$.

Our detection algorithm for LSB steganography is compared with the RS method of [2], which is the most accurate and robust LSB steganalytic algorithm in the literature. The two algorithms perform almost identically by the criteria of false alarm and missing rates, and also in terms of average error in estimated hidden message length. These empirical findings should not come as a surprise as we will see in Section VII that the two techniques are essentially based on the same principle and assumptions.



Fig. 6. Sample images of the test set

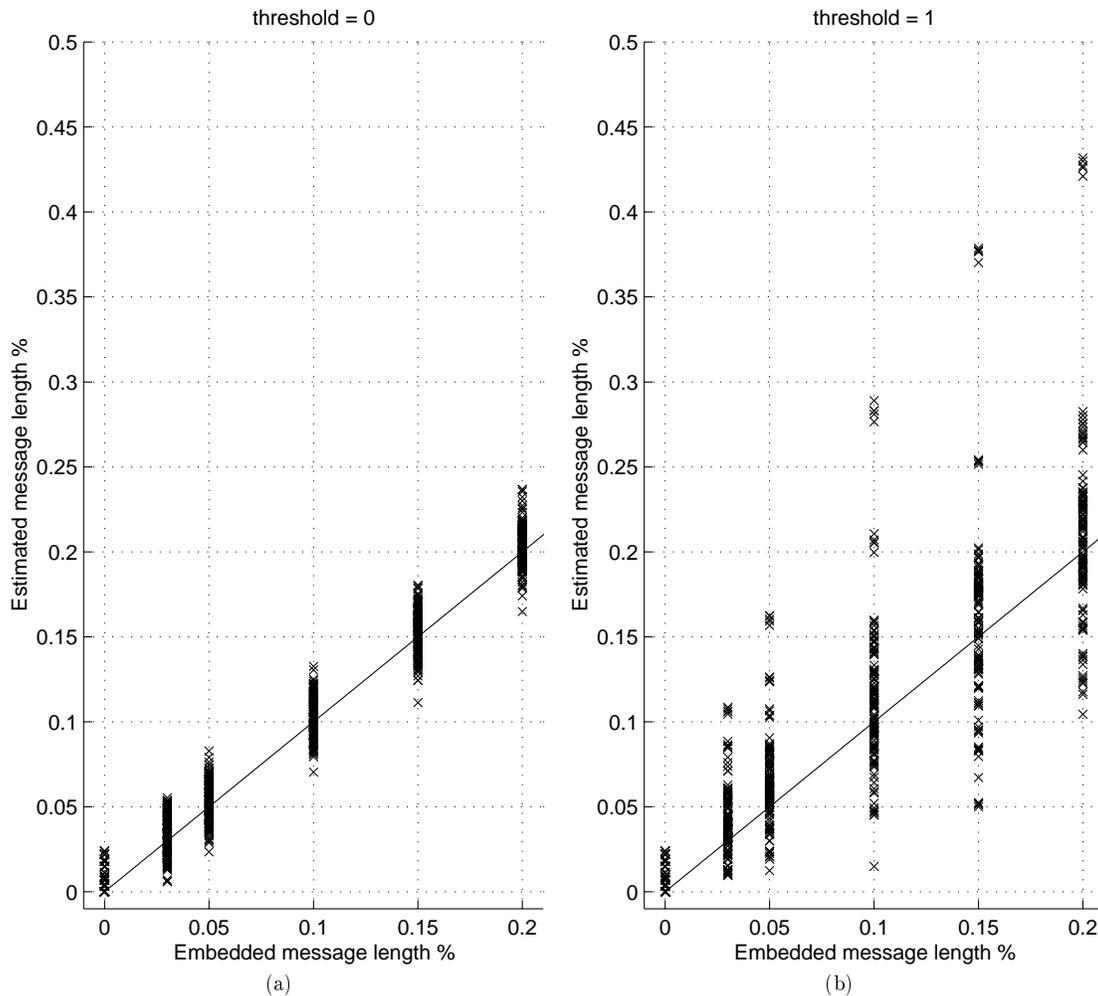


Fig. 7. Performance of the proposed LSB steganalytic technique. a): the case of random LSB embedding; b): the case of selective embedding, with $\tau = 1$.

VI. POSSIBLE ATTACKS AND COUNTER MEASURES

If the message bits are scattered randomly among the least significant bits of all signal samples, then the use of spatially adjacent sample pairs makes the estimate of p more robust. But this choice of \mathcal{P} opens a door for possible attacks on the detection method. An adversary can try to fool the detection method by avoiding hiding message bits at locations where some of adjacent sample pairs have close values. For instance, if the adversary does not embed in adjacent sample pairs that differ by less than 3 in value, then he makes $\rho(\pi, D_0) = 0$, $\pi \in \{01, 10, 11\}$. In other words, the adversary purposefully tricks C_0 to be biased, violating an assumption that ensures

	p=0%	p=3%	p=5%	p=10%	p=15%	p=20%
$\tau = 0$	0.1379	0.1103	0	0	0	0
$\tau = 1$	0.1379	0.0828	0.0069	0.0069	0	0

TABLE I

PROBABILITY OF WRONG DECISION: MISSING RATE WHEN $p > 0$, AND FALSE ALARM RATE WHEN $p = 0$, WITH THE DECISION THRESHOLD SET AT $\hat{p} > 0.018$.

the accuracy of (18). An attack of this type is to only embed message bits among candidate sample positions where all adjacent sample pairs are in C_t such that $t \geq \tau$, where τ is prefixed threshold. In other words, any sample pair (u, v) that is tampered by LSB embedding satisfies $|u - v| \geq 2\tau - 1$, and $|u' - v'| \geq 2\tau - 1$, where (u', v') represents the values of the two samples after LSB embedding. Clearly, this LSB embedding scheme conditioned on C_t such that $t \geq \tau$ can be decoded, because both encoder and decoder can refer to the same C_t , $t \geq \tau$, to decide whether a sample is a candidate for embedding.

The effects of the attack by embedding only at positions where all adjacent sample pairs are in C_t such that $t \geq \tau$ are demonstrated by Fig. 7(b) of threshold $\tau = 1$. By comparing Fig. 7(b) with Fig. 7(a) of threshold $\tau = 0$, we see that the distribution of estimated message lengths \hat{p} has significantly wider spread as τ changes from 0 (random embedding) to 1 (selective embedding).

Table 1 tabulates the false alarm rates when $p = 0$ and the missing rates when $p > 0$ for different p and for $\tau = 0, 1$. The statistics of Table 1 is collected from the set of test images. Our empirical evidence indicates that the proposed LSB steganalytic technique cannot be fooled by selective LSB embedding scheme that avoids embedding in smooth waveforms. As we can see in Table 1 that for $p = 3\%$ the missing rate actually drops from random embedding ($\tau = 0$) to selective embedding ($\tau = 1$), and it only increases very slightly for larger p .

In general, the proposed method is open for attack if the locations of chosen sample pairs in \mathcal{P} are known, and if the algorithm examines a specific close set C_s and the chosen s is also known. Fortunately, to the benefit of steganalysis the detection algorithm can solve (17) for different choices of i and j . In other words, the steganalyst can choose different multisets $\cup_{m=i}^j C_m$ and $\cup_{m=i+1}^{j+1} C_m$ to estimate p . The estimate will be improved as long as $\cup_{m=i}^j C_m$ and $\cup_{m=i+1}^{j+1} C_m$ are unbiased. It is extremely difficult, if not impossible, to select locations of embedded message

bits in such a way that all of C_m , $0 \leq m \leq 2^{b-1} - 1$, become biased. The research on this type of counter measures against attacks is underway.

We conclude this section with an analysis on the capacity of the aforementioned selective steganography. The adversary's objective is to make the multiset

$$C_{<\tau} = \cup_{0 \leq t < \tau} C_t$$

to be void of the message bits. The event that $(u, v) \in \mathcal{P}$ but $(u, v) \notin C_{<\tau}$ has the probability

$$P_\tau = 1 - \sum_{t=0}^{2\tau-2} P(|D_t|) - \frac{P(|D_{2\tau-1}|)}{2}. \quad (23)$$

In the case of steganography in images, if we include all 4-connected sample pairs in \mathcal{P} , then a sample u can be candidate for LSB embedding only if we simultaneously have $(u, n) \notin C_{<\tau}$, $(u, s) \notin C_{<\tau}$, $(u, w) \notin C_{<\tau}$, and $(u, e) \notin C_{<\tau}$, where n , s , w , and e denote the samples to the north, south, west and east of u . In two-dimensional image signals, it is reasonable to assume that (u, n) and (u, s) are mutually dependent but are independent of (u, w) and (u, e) . Then the probability for a sample to be candidate for LSB embedding is

$$\begin{aligned} & P((u, n) \notin C_{<\tau} | (u, s) \notin C_{<\tau}) P((u, s) \notin C_{<\tau}) \cdot \\ & P((u, w) \notin C_{<\tau} | (u, e) \notin C_{<\tau}) P((u, e) \notin C_{<\tau}) \leq P_\tau^2. \end{aligned} \quad (24)$$

From Fig. 5 and Appendix C we observe that P_τ has an exponential decay in τ . Therefore, the capacity of selective steganography diminishes exponentially in threshold τ , and the rate of decay is greater for highly correlated signals.

VII. REMARKS ON THE RS METHOD

Very recently Fridrich *et al.* proposed a steganalytic technique, called the RS method, to detect LSB embedding in continuous-tone images [2]. The RS method was demonstrated in experiments to be very effective. The RS method uses groups of four pixels (2×2 blocks) versus our choice of pixel pairs. Interestingly, our analysis presented in the proceeding sections offers a proof of some key observations underlying the RS method if it is applied to sample pairs.

The RS method partitions an image into N/n disjoint groups of n neighboring pixels, where N is the total number of pixels in the image. In [2] the authors considered the case $n = 4$. A *discrimination function* $f(\cdot)$ that captures the smoothness of a group of pixels is defined $f(G) =$

$f(x_1, x_2, \dots, x_n) = \sum_{i=1}^{n-1} |x_{i+1} - x_i|$, where x_1, x_2, \dots, x_n are the values of the pixels in the group G . In addition, three invertible operations, $F_n(x)$, $n = -1, 0, 1$ on pixel values x , are introduced. $F_1(x)$ is the operation that flips the LSB of a pixel, i.e., $F_1 : 0 \leftrightarrow 1, 2 \leftrightarrow 3, \dots, 254 \leftrightarrow 255$. $F_{-1}(x)$ maps pixel values in the opposite direction to F_1 . Specifically, $F_{-1} : -1 \leftrightarrow 0, 1 \leftrightarrow 2, \dots, 255 \leftrightarrow 256$. $F_0(x)$ is defined as the identity function.

Operations F_1 and F_{-1} are applied to a group of pixels G with a mask M (a n -tuple with components $-1, 0$ or 1), which specifies where and how pixel values are to be modified. For example, if the values of the four pixels of a group G are $39, 38, 40, 41$ and $M = (1, 0, 1, 0)$, then $F_M(G) = (F_1(39), F_0(38), F_1(40), F_0(41))$. Given a mask, operations F_1 and F_{-1} , and the discrimination function f , a pixel group G is classified into one of the three categories:

$$G \in R(M) \Leftrightarrow f(F(G)) > f(G)$$

$$G \in S(M) \Leftrightarrow f(F(G)) < f(G)$$

$$G \in U(M) \Leftrightarrow f(F(G)) = f(G)$$

where $R(M)$, $S(M)$, and $U(M)$ are called *Regular*, *Singular*, and *Unusable Groups*. The RS method is based on the statistical hypothesis that, when no message is embedded in the image, the following equalities hold

$$E\{|S(M)|\} = E\{|S(-M)|\}, \quad (25)$$

$$E\{|R(M)|\} = E\{|R(-M)|\}, \quad (26)$$

where mask $-M$ is obtained by negating all the components of M . For instance, if $M = (1, 0, 1, 0)$ then $-M = (-1, 0, -1, 0)$. Furthermore, the authors of [2] observed that $|R(-M)|$ and $|S(-M)|$ were linear functions of the embedded message length and the two quantities diverge as p increases (remind that p is the number of embedded samples divided by the total number of samples). Also, $|R(M)|$ and $|S(M)|$ are quadratic functions in p , and $|R(M)| = |S(M)|$ when $\frac{p}{2} = 0.5$.

Based on these observations, the RS method estimates the value of p using a quadratic equation, whose coefficients are computed based on the sizes of the regular and singular groups for the masks M and $-M$, for the input image and for the image obtained by flipping the LSB of all pixels.

Next we prove the linear and quadratic functions observed in the experiments of [2] for the case $n = 2$, i.e. each group consist of a pair of pixels, and the mask is $M = (0, 1)$ or $M = (1, 0)$.

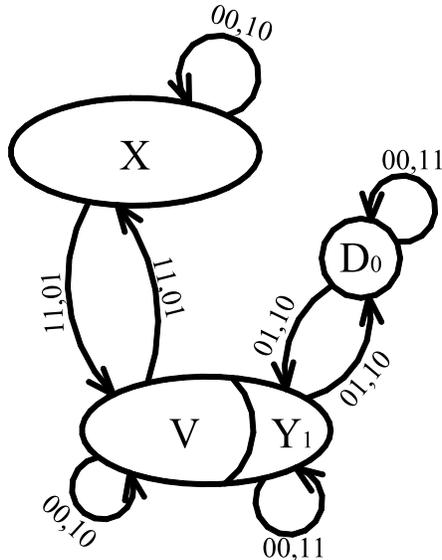


Fig. 8. The finite-state machine to verify the RS method.

In this case $R(M)$, $S(M)$, $R(-M)$ and $S(-M)$ are multisets of pairs of pixels values as defined by this paper. Let us consider mask $M = (0, 1)$ (hence $-M = (0, -1)$). The case for $M = (1, 0)$ is analogous. Define the multisets X and Y :

$$X = \cup_{i=1}^{2^b-1} X_i, \quad (27)$$

$$Y = \cup_{i=1}^{2^b-1} Y_i.$$

In other words, X is the multiset 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 multiset of pairs $(u, v) \in \mathcal{P}$ such that v is even and $u > v$, or v is odd and $u < v$. Then, from the definitions of the discrimination function f and of the flipping functions F_M and F_{-M} in [2], it follows that:

$$R(M) = X \cup D_0, \quad S(M) = Y \quad (28)$$

$$R(-M) = Y \cup D_0, \quad S(-M) = X. \quad (29)$$

Note that the statistical hypotheses (25) and (26) are equivalent to the following assumption:

$$E\{|X|\} = E\{|Y|\}. \quad (30)$$

By the same analysis of transitions under embedding, between the trace multisets in Section 2, we obtain the finite-state machine of Fig. 8, where $V = Y - Y_1$. Let $R'(M)$, $R'(-M)$,

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

The finite state machine described in Fig. 9 together with (28) and (29) leads to the following relations (which are proved in Appendix D):

$$|R'(-M)| = |R(-M)| + \frac{p}{2}|Y_1|, \quad (31)$$

$$|S'(-M)| = |S(-M)| - \frac{p}{2}|Y_1|, \quad (32)$$

$$|R'(M)| = |R(M)| - \frac{p}{2}(2|D_0| - |Y_1|) - \frac{p^2}{2}(|Y_1| - |D_0|), \quad (33)$$

$$|S'(M)| = |S(M)| + \frac{p}{2}(2|D_0| - |Y_1|) + \frac{p^2}{2}(|Y_1| - |D_0|) \quad (34)$$

In order to simplify the derivation of (31) through (34), we replaced the hypotheses $E\{|S(M)|\} = E\{|S(-M)|\}$ and $E\{|R(M)|\} = E\{|R(-M)|\}$ by $|S(M)| = |S(-M)|$ and $|R(M)| = |R(-M)|$. Thus, the equations (31) through (34) should be understood to hold after taking expectations at the both sides of the equations. 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 $\frac{p}{2} = 0.5$. Therefore, our derivations corroborate with the observations on $|R'(-M)|$, $|S'(-M)|$, $|R'(M)|$ and $|S'(M)|$ made by Fridrich *et al.* These observations form the basis of the RS detection technique of [2].

Furthermore, we can obtain the quadratic equation for the estimation of p in a straightforward manner. The equation is

$$0.5|C_0|p^2 + (2|X'| - |\mathcal{P}|)p + |Y'| - |X'| = 0 \quad (35)$$

and its derivation is given in Appendix D.

VIII. CONCLUSION

A new approach is proposed to detect LSB steganography embedded in digital signals, and to estimate the length of the hidden message length. The estimate is remarkably accurate under mild assumptions that are true for continuous signals. The estimation error is analyzed in terms of the degree that the input signals deviate from the assumptions, and error bounds are given. Possible attacks to the proposed steganalytic method are examined and corresponding counter

measures are discussed. Experiments are conducted on a set of continuous-tone images. Empirical observations made in the simulations agree with our analytic results.

Appendix A. Proof of Equations (2) and (3).

First note that the multiset $X'_{2m-1} \cup X'_{2m}$ consists of the sample pairs of $X_{2m-1} \cup X_{2m}$ modified by the patterns 00 or 10, and of the sample pairs of $Y_{2m} \cup Y_{2m+1}$ modified by the patterns 01 or 11. The probability that an arbitrary sample pair of $X_{2m-1} \cup X_{2m}$ is modified by the patterns 00 or 10 equals $(1 - p/2)^2 + p/2(1 - p/2) = 1 - p/2$. Also the probability that an arbitrary sample pair of $Y_{2m} \cup Y_{2m+1}$ is modified by the patterns 01 or 11 equals $p/2(1 - p/2) + (p/2)^2 = p/2$. These observations enable us to express the cardinality of $X'_{2m-1} \cup X'_{2m}$, which equals $|X'_{2m-1}| + |X'_{2m}|$, as follows:

$$|X'_{2m-1}| + |X'_{2m}| = (|X_{2m-1}| + |X_{2m}|)(1 - p/2) + (|Y_{2m}| + |Y_{2m+1}|)p/2. \quad (36)$$

Similarly, the results in Fig. 1 allows us to evaluate the cardinality of the multiset $Y'_{2m} \cup Y'_{2m+1}$ as follows:

$$|Y'_{2m}| + |Y'_{2m+1}| = (|Y_{2m}| + |Y_{2m+1}|)(1 - p/2) + (|X_{2m-1}| + |X_{2m}|)p/2. \quad (37)$$

Subtracting (37) from (36) yields

$$|X'_{2m-1}| - |Y'_{2m+1}| + |X'_{2m}| - |Y'_{2m}| = (|X_{2m-1}| - |Y_{2m+1}| + |X_{2m}| - |Y_{2m}|)(1 - p). \quad (38)$$

Further, by observing that the multiset $X_{2m-1} \cup Y_{2m}$ exchanges sample pairs only with $X_{2m} \cup Y_{2m+1}$, and vice versa, and that the exchanged pairs are only those modified by the patterns 10 or 11, and the fact that an arbitrary sample pair of each of the above mentioned multisets has the probability $p/2$ of being modified by the patterns 10 or 11, we have

$$|X'_{2m-1}| - |Y'_{2m+1}| + |Y'_{2m}| - |X'_{2m}| = (|X_{2m-1}| - |Y_{2m+1}| + |Y_{2m}| - |X_{2m}|)(1 - p). \quad (39)$$

Adding (38) and (39) results in (after making the necessary cancellations and simplification):

$$|X'_{2m-1}| - |Y'_{2m+1}| = (|X_{2m-1}| - |Y_{2m+1}|)(1 - p). \quad (40)$$

The next step is to derive the cardinality of $X'_{2m-1} \cup Y'_{2m+1}$ using the finite-state machine of Fig. 1. Note that $X_{2m-1} \cup Y_{2m+1}$ exchanges pairs only with the multiset D_{2m} (remind that $D_{2m} = X_{2m} \cup Y_{2m}$). It follows that

$$|X'_{2m-1}| + |Y'_{2m+1}| = (|X_{2m-1}| + |Y_{2m+1}|)[(1 - p/2)^2 + (p/2)^2] + |D_{2m}|p(1 - p/2). \quad (41)$$

Since $|D_{2m}| = |C_m| - |X_{2m-1}| - |Y_{2m+1}|$, it follows further that

$$|X'_{2m-1}| + |Y'_{2m+1}| = (|X_{2m-1}| + |Y_{2m+1}|)(1-p)^2 + |C_m|p(1-p/2). \quad (42)$$

By multiplying by $(1-p)$ both sides of (40), and adding the obtained equation to (42), we obtain

$$|X'_{2m-1}|(2-p) + |Y'_{2m+1}|p = 2|X_{2m-1}|(1-p)^2 + |C_m|p(1-p/2). \quad (43)$$

But the multiset C_m is closed under embedding, hence

$$|C_m| = |X'_{2m-1}| + |Y'_{2m+1}| + |D'_{2m}|. \quad (44)$$

Finally, combining (43) and (44) establishes (2). Similarly, by multiplying by $(1-p)$ both sides of (40), then subtracting the obtained equation from (42), and further using (44), equality (3) also follows.

The derivation of equation (4) is similar, and is omitted.

Appendix B. Derivation of Error Bound (22).

We shall prove inequality (22) only for $i \geq 1$. The case $i = 0$ is analogous and omitted.

Let us fix some i and j with $1 \leq i \leq j \leq 2^{b-1} - 2$. For simplicity we shall use the notation \hat{p} instead of $\hat{p}(i, j)$. Hence \hat{p} satisfies the relation

$$\begin{aligned} \frac{\hat{p}^2}{4}(|C_i| - |C_{j+1}|) - \frac{\hat{p}}{2}(|D'_{2i}| - |D'_{2j+2}|) + 2 \sum_{m=i}^j (|Y'_{2m+1}| - |X'_{2m+1}|) + \\ \sum_{m=i}^j (|Y'_{2m+1}| - |X'_{2m+1}|) = 0. \end{aligned} \quad (45)$$

First note that

$$\sum_{m=i}^j (|Y'_{2m+1}| - |X'_{2m+1}|) = \sum_{m=i}^j (|Y'_{2m+1}| - |X'_{2m-1}|) + |X'_{2i-1}| - |X'_{2j+1}|. \quad (46)$$

Using (40) and (15) we further obtain

$$\begin{aligned} \sum_{m=i}^j (|Y'_{2m+1}| - |X'_{2m+1}|) &= (1-p) \sum_{m=i}^j (|Y_{2m+1}| - |X_{2m-1}|) + |X'_{2i-1}| - |X'_{2j+1}| = \\ &= (1-p) \sum_{m=i}^j (-\epsilon_m) + |X'_{2i-1}| - (1-p)|X_{2i-1}| - |X'_{2j+1}| + (1-p)|X_{2j+1}|. \end{aligned} \quad (47)$$

Replacing in (45) it follows that

$$\begin{aligned} & \frac{\hat{p}^2}{4}(|C_i| - |C_{j+1}|) - \frac{\hat{p}}{2}(|D'_{2i}| - |D'_{2j+2}|) + \\ & (1 - \hat{p})[(1 - p) \sum_{m=i}^j (-\epsilon_m) + |X'_{2i-1}| - (1 - p)|X_{2i-1}| - |X'_{2j+1}| + (1 - p)|X_{2j+1}|] = 0. \end{aligned} \quad (48)$$

Rearranging the relation above yields

$$\begin{aligned} & \frac{\hat{p}^2}{4}(|C_i| - |C_{j+1}|) + (1 - \hat{p})(1 - p) \sum_{m=i}^j (-\epsilon_m) + \\ & \frac{-\hat{p}}{2}|D'_{2i}| + (1 - \hat{p})|X'_{2i-1}| - (1 - \hat{p})(1 - p)|X_{2i-1}| - \\ & \left(\frac{-\hat{p}}{2}|D'_{2j+2}| + (1 - \hat{p})|X'_{2j+1}| - (1 - \hat{p})(1 - p)|X_{2j+1}|\right) = 0. \end{aligned} \quad (49)$$

Now let us evaluate the expression

$$\frac{-\hat{p}}{2}|D'_{2m}| + (1 - \hat{p})|X'_{2m-1}| - (1 - \hat{p})(1 - p)|X_{2m-1}| \quad (50)$$

for an arbitrary $m \geq 1$. By replacing $|D'_{2m}|$ by $|C_m| - |X'_{2m-1}| - |Y'_{2m+1}|$ and rearranging the terms, we obtain

$$\begin{aligned} & \frac{-\hat{p}}{2}|D'_{2m}| + (1 - \hat{p})|X'_{2m-1}| - (1 - \hat{p})(1 - p)|X_{2m-1}| = \\ & \frac{-\hat{p}}{2}|C_m| + \frac{\hat{p}}{2}(|Y'_{2m+1}| - |X'_{2m-1}|) + |X'_{2m-1}| - (1 - \hat{p})(1 - p)|X_{2m-1}|. \end{aligned} \quad (51)$$

From (40) and (42) it follows (by adding them and afterwards dividing by 2 the coefficients of the resulted equality) that

$$\begin{aligned} |X'_{2m-1}| = \frac{1}{2}(|X_{2m-1}| - |Y_{2m+1}|)(1 - p) + \frac{1}{2}(|X_{2m-1}| + |Y_{2m+1}|)(1 - p)^2 + \\ |C_m| \frac{p}{2} \left(1 - \frac{p}{2}\right). \end{aligned} \quad (52)$$

By replacing in (51) $|Y'_{2m+1}| - |X'_{2m-1}|$ from (40) and $|X'_{2m-1}|$ from (52), and afterwards using the equality

$$|X_{2m-1}| + |Y_{2m+1}| = |C_m| - |D_{2m}|, \quad (53)$$

we obtain

$$\begin{aligned} & \frac{-\hat{p}}{2}|D'_{2m}| + (1 - \hat{p})|X'_{2m-1}| - (1 - \hat{p})(1 - p)|X_{2m-1}| = \\ & \frac{p}{2} \left(\frac{p}{2} - \hat{p}\right) |C_m| - \frac{1}{2}(1 - p)(\hat{p} - p)|D_{2m}|. \end{aligned} \quad (54)$$

Applying (54) to $m = i$ and $m = j + 1$, and replacing in (49), together with the replacement of $\sum_{m=i}^j (-\epsilon_m)$ by $\frac{-e_{ij}}{2}(|D_{2i}| - |D_{2j+2}|)$ (according to (20)), implies that

$$\begin{aligned} \frac{1}{2}(|C_i| - |C_{j+1}|)(\hat{p} - p)^2 - (1 - p)(|D_{2i}| - |D_{2j+2}|)(1 - e_{ij})(\hat{p} - p) + \\ (1 - p)^2(-e_{ij})(|D_{2i}| - |D_{2j+2}|) = 0. \end{aligned} \quad (55)$$

We shall treat the above relation as a function in variable $(\hat{p} - p)$.

Assertion. Let x_0 be the smallest solution of the equation

$$ax^2 + bx + c = 0, \quad (56)$$

where a, b, c are real numbers with $b^2 - 4ac \geq 0$ and $b < 0$. Then the inequality

$$|x_0| \leq \frac{2|c|}{-b}. \quad (57)$$

Indeed, we have

$$|x_0| = \left| \frac{-b - \sqrt{b^2 - 4ac}}{2a} \right| = \left| \frac{b^2 - (b^2 - 4ac)}{2a(-b + \sqrt{b^2 - 4ac})} \right| \leq \frac{2|c|}{-b}. \quad (58)$$

Assuming that $e_{ij} < 1$ and $|D_{2i}| - |D_{2j+2}| > 0$, we apply Assertion to (55) and subsequently conclude (22).

Appendix C. Sample Correlation and Multiset Cardinality

Suppose that the sample pairs (u, v) of \mathcal{P} are randomly drawn. Since the marginal distributions $P_U(u)$ and $P_V(v)$ of the joint distribution $P(u, v)$, are themselves the distribution of sample values. Hence the random variables U and V have the same mean:

$$E\{U\} = E\{V\}, \quad (59)$$

and the same variance:

$$\text{Var}(U) = \text{Var}(V) = \sigma^2. \quad (60)$$

Now consider the difference $Z = U - V$, a new random variable. Since $Z = (U, V)(1, -1)^t$, it follows that the variance of Z is given by

$$\text{Var}(Z) = (1, -1)\Sigma(1, -1)^t, \quad (61)$$

where Σ is the covariance matrix of the random vector (U, V) . Denoting by $Cov(U, V)$ the covariance between the two random variables U and V , and using (60), it follows that

$$Var(Z) = 2\sigma^2 - 2Cov(U, V). \quad (62)$$

Let ρ denote the correlation between the random variables U and V . Then

$$\rho = \frac{Cov(U, V)}{\sqrt{Var(U)Var(V)}} = \frac{Cov(U, V)}{\sigma^2}. \quad (63)$$

Hence

$$Var(Z) = 2\sigma^2(1 - \rho). \quad (64)$$

This means that the variance of $Z = U - V$ decreases as the correlation between U and V increases. Note that

$$|D_i| = (P_Z(i) + P_Z(-i))|\mathcal{P}|, \quad i \geq 1. \quad (65)$$

Hence we conclude that the probability $P(|D_i|)$ decreases more rapidly in i when the correlation between U and V becomes higher, if $P_Z(z)$ is unimodal and peaks at 0, which is a rather relaxed condition satisfied by many two-dimensional joint distributions of U and V .

Appendix D. Derivations of Key Results of Section VII.

Note from Fig. 1 (in Section 2) that the multiset X exchanges sample pairs only with the multiset V and vice versa, and the exchanged pairs are only those modified by the patterns 01 or 11. Consequently equations (66) and (67) can be derived in the same way as (36) in Appendix A.

$$|X'| = |X|(1 - \frac{p}{2}) + |V|\frac{p}{2} \quad (66)$$

$$|V'| = |V|(1 - \frac{p}{2}) + |X|\frac{p}{2} \quad (67)$$

Assumption (30) implies that

$$E\{|X|\} = E\{|V|\} + E\{|Y_1|\}. \quad (68)$$

Applying (68) in (66) and (67) obtains

$$|X'| = |X| - |Y_1|\frac{p}{2} \quad (69)$$

$$|V'| = |V| + |Y_1|\frac{p}{2} \quad (70)$$

Since $S(-M) = X$ from (29), (69) immediately implies (32). Also, since $R(-M) = Y \cup D_0$ from (29), we have

$$|R'(-M)| = |Y'| + |D'_0| = |V'| + |Y'_1| + |D'_0|. \quad (71)$$

Then, the relation (31) follows from (71), (70), and the obvious equality $|Y'_1| + |D'_0| = |Y_1| + |D_0|$.

From the finite state machine depicted in Fig. 2 (Section 2), we see that the multiset Y_1 exchanges sample pairs only with the multiset D_0 and vice versa, and the exchanged pairs are only those modified by the patterns 01 or 10. These facts yield

$$|Y'_1| = |Y_1|(1 - p + \frac{p^2}{2}) + |D_0|p(1 - \frac{p}{2}), \quad (72)$$

$$|D'_0| = |D_0|(1 - p + \frac{p^2}{2}) + |Y_1|p(1 - \frac{p}{2}). \quad (73)$$

in a similar way to relation (41) in Appendix A. Starting from $|R'(M)| = |X'| + |D'_0|$ of (28) and applying (69) and (73), we arrive at (33). Similarly, the relation $|S'(M)| = |Y'| = |V'| + |Y'_1|$ of (28) together with (70) and (72) leads to (34).

Relations (66), (67) and (68) imply that

$$|X'| - |V'| = (|Y_1|)(1 - p). \quad (74)$$

Further, replacing $|D_0|$ by $|C_0| - |Y_1|$, equation (72) becomes

$$|Y'_1| = |Y_1|(1 - p)^2 + |C_0|p(1 - \frac{p}{2}). \quad (75)$$

The elimination of $|Y_1|$ from (74) and (75) leads to

$$|Y'_1| = (|X'| - |V'|)(1 - p) + |C_0|p(1 - \frac{p}{2}). \quad (76)$$

Since $|X'| + |V'| + |Y'_1| + |D'_0| = |\mathcal{P}|$, relation (35) follows.

REFERENCES

- [1] J. Fridrich, R. Du, M. Long, "Steganalysis of LSB Encoding in Color Images", *Proc. of ICME 2000*, NYC, July 31-Aug. 2, USA.
- [2] J. Fridrich, M. Goljan and R. Du, "Reliable Detection of LSB Steganography in Color and Grayscale Images". *Proc. of ACM Workshop on Multimedia and Security*, Ottawa, Oct. 5, 2001, pp. 27-30.
- [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, 1998.
- [4] 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.
- [5] A. Westfeld, A. Pfitzmann, "Attacks on Steganographic Systems", in Andreas Pfitzmann (Ed.): *Information Hiding*, LNCS 1768, pp. 61-76, Springer-Verlag, 1999.