Category Attack for LSB Steganalysis of JPEG Images

Kwangsoo Lee[1], Andreas Westfeld[2], and Sangjin Lee[1]

[1] Center for Information Security Technologies (CIST),
Korea University, Seoul, Korea
kslee@cist.korea.ac.kr, sangjin@korea.ac.kr
[2] Technische Universität Dresden,
Institute for System Architecture,
01062 Dresden, Germany
westfeld@inf.tu-dresden.de

Abstract. In this paper, we propose a new method for the detection of LSB embedding in JPEG images. We are motivated by a need to further research the idea of the chi-square attack. The new method simply use the first-order statistics of DCT coefficients, but is more powerful to detect the random embedding in JPEG images. For evaluation, we used versions of Jsteg and Jphide with randomized embedding path to generate stego images in our experiments. In results, the proposed method outperforms the method of Zhang and Ping and is applicable to Jphide. The detection power of both proposed methods is compared to the blind classifier by Fridrich that uses 23 DCT features.

1 Introduction

Steganography aims to hide the existence of secret messages by embedding them into ordinarily looking cover objects, while steganalysis aims to detect stego objects containing hidden messages [1]. Today, in digital era, the digital media such as image and audio are proliferated through the internet so that their transmissions are usual events in our daily life. Besides, many people have thought that digital media contain a lot of redundancies such as natural noises and quantized errors whose changes were expected to make no significant impacts on their perceptual and statistical properties. These have led people to research digital media for steganography, and in particular, digital images have been mostly researched for steganography.

The LSB embedding is a well-known steganographic method that is the way of replacing secret (usually encrypted) message bits with the least significant bits (LSBs) of sample values in digital media. It can be classified into two types according to the way of how to select media samples for carrying message bits. One is the sequential embedding in which message-carrying samples are selected in a fixed order that is publicly known, and the other is the random embedding in which message-carrying samples are randomly selected with a stego key that is shared by communication parties.

Y.Q. Shi and B. Jeon (Eds.): IWDW 2006, LNCS 4283, pp. 35–48, 2006.

There are several attacks on LSB embedding. An earlier approach was proposed by Westfeld and Pfitzmann [2]. Their method, named the chi-square attack, exploits the first-order statistics (histogram) of samples in digital media. The chi-square attack works well for the sequential embedding, but not for the random embedding unless approximately all samples have been used for carrying message bits. Provos and Honeyman [3] presented an extended technique of the chi-square attack and tested it for JPEG based steganography such as Jsteg [11] and Jphide [12]. The extended chi-square attack works better for the random embedding such as OutGuess 0.13b [13] (OutGuess 0.13b can be viewed as a randomized version of the Jsteg algorithm). However, it seems that there still exists a limitation of the detection performance on the extended chi-square attack for the random embedding. Subsequent versions of OutGuess preserve the first-order statistics. Histogram-based attacks will fail to detect OutGuess, however, it is easily detected by comparing to calibrated statistics [4].

In this paper, we propose a new method for detection of LSB embedding in JPEG images. We are motivated by a need to further research the idea of the chi-square attack. The new method simply use the first-order statistics of DCT coefficients, but is more powerful to detect the random embedding in JPEG images. For evaluation, we used versions of Jsteg and Jphide with randomized embedding path to generate stego images in our experiments. In results, the proposed method outperforms another histogram-based detection by Zhang and Ping [7] and is applicable to Jphide. The detection power of both proposed methods is compared to the blind classifier by Fridrich [8] that uses 23 DCT features.

The paper organization is as follows: In the next section, we review the previous histogram-based attacks on LSB embedding. In Section 3, we describe the proposed approach towards an improvement of the idea of the chi-square attack, and some techniques to detect the Jsteg and Jphide embedding. In Section 4, we displays the experimental results of the proposed attack. In Section 5, we evaluate the detection reliability of the proposed attack and provide a fair comparison with previous methods to detect the randomized Jsteg and Jphide embedding. Finally, we conclude this paper in Section 6.

2 Histogram-Based Attacks on LSB Embedding

In this section, we briefly review the histogram-based attacks previously proposed for LSB steganalysis.

2.1 The Original Chi-Square Attack

Westfeld and Pfitzmann [2] proposed a categorical data analysis for detection of LSB embedding in digital images. LSB embedding induces categories of two sample values in which values only differ in the LSBs and so are possibly transformed into each other by LSB embedding operation. We will call them the induced categories throughout this paper, instead of the pairs of values (PoVs) named in their literature. To exemplify the induced categories, let us assume that the digital image is represented by a sequence of samples whose values are

integer numbers and all samples in the digital image are possibly used for carrying message bits. Then induced categories can be represented by the pairs of integer numbers, $(2m, 2m+1)$.

They discovered the fact that if a random message whose bits 0 and 1 are uniformly distributed is embedded in the LSBs of image data, the frequencies of sample values in each of PoVs are likely to be equal. This fact is generally untrue for cover images and was used for their categorical data analysis, named the chi-square attack. The chi-square attack is the way of measuring the degree of similarity between the observed sample distribution and the theoretically expected distribution in the induced categories, by means of a hypothesis test, the χ^2-test.

In order to give a formal description, let h_i denote the observed sample histogram. Then, for the induced categories $(2m, 2m+1)$, the observed distribution $\{o_m\}$ is given by

$$o_m = h_{2m} \ , \tag{1}$$

and the expected distribution $\{e_m\}$ is determined by

$$e_m = \frac{h_{2m} + h_{2m+1}}{2} \ . \tag{2}$$

The difference between the two distributions is measured by the following χ^2 statistics with $\nu - 1$ degrees of freedom,

$$\chi^2 = \sum_{e_m \neq 0} \frac{(o_m - e_m)^2}{e_m} = \frac{1}{2} \sum_{m \in Z} \frac{(h_{2m} - h_{2m+1})^2}{h_{2m} + h_{2m+1}} \ , \tag{3}$$

where ν is the number of different categories. The degree of similarity between the two distributions $\{o_m\}$ and $\{e_m\}$ is then calculated by the complement of the cumulative distribution function (CDF),

$$p = 1 - \int_0^{\chi^2} \frac{t^{(\nu-2)/2} e^{-t/2}}{2^{\nu/2} \Gamma(\nu/2)} dt \tag{4}$$

The p-value p is used for the decision of whether the image contains a secret message hidden with the LSB embedding or not.

To make an allowance for the detection of a sequential embedding, they implemented the χ^2-test on the recurrent samples of progressively increasing sizes, where the samples are selected in the same way of the sequential embedding. If the image contains a sequentially embedded secret message, the chi-square attack will show the result that the p-values are very close to 1 from the start of the test until rapidly fall down to 0 at the end of the hidden message (this can be additionally used to estimate the hidden message length). It is said that the chi-square attack is highly efficient for detecting the sequential embedding. It seems to be generally applicable to any types of digital images, and works very well for Jsteg [11] and Jphide [12] that are the sequential embedding for the JPEG image. However, it can hardly detect straddled messages unless approximately all samples have been used.

2.2 The Extended Chi-Square Attack

Provos and Honeyman [3] extended the chi-square attack by exploiting the sliding window of a fixed size to obtain the sample data, instead of increasing the window size. In this approach, it is important to find the appropriate window size for reliable detection. They implemented the χ^2-test for the shifted categories $(2m-1, 2m)$, $m \in Z$, and find the smallest window size that produces p-values bounded below a certain small threshold. This was formalized and adapted by Fridrich et al. [5]. The extended technique works better for the random embedding such as OutGuess 0.13b [13] that is a random LSB embedding for JPEG images. However, there still exists a limitation of the chi-square attack for the detection of small messages hidden with the random embedding [5].

2.3 Zhang and Ping Method

Another method to target the Jsteg-like algorithm was proposed by Zhang and Ping [7]. They considered the histogram shape of quantized DCT coefficients in JPEG images and assumed that the histogram has symmetry around zero. We will call their method the ZP attack. The following is a brief description of the ZP attack: Let h_i be the histogram of DCT coefficients in a JPEG image, where the indices i denote the values of the DCT coefficients. Let $f_0 = \sum_{i>0} h_{2i} + \sum_{i<0} h_{2i+1}$ and $f_1 = \sum_{i<0} h_{2i} + \sum_{i>0} h_{2i-1}$. In order to determine whether the JPEG image is stego or not, check that $f_1 > f_0$ and then calculate the statistics,

$$\chi^2 = \frac{(f_0 - f_1)^2}{f_0 + f_1} \ . \tag{5}$$

If χ^2 is greater than a certain small threshold, then the image will be determined as the stego image. As an additional information, the method can estimate the length of hidden message as the β value,

$$\beta = \frac{f_1 - f_0}{h_1} \ . \tag{6}$$

We have seen that ZP attack works better than the extended chi-square attack for the randomized Jsteg embedding, however, it does not work for the randomized Jphide embedding.

3 The Category Attack

In this section, we describe our approach to an improved histogram-based attack on LSB embedding. We name it the category attack.

3.1 Main Approach

In our development towards an improvement of the chi-square attack, we make a comparison of the induced categories and the shifted categories, while the authors of the extended chi-square attack used the shifted categories for the

appropriate window size. It is worth to mention that the p-value in Eqn. (4) is not a suitable measurement to discover small changes in sample distribution, that might happen when small messages are hidden with the LSB embedding. The reason is due to the fact that the CDF in Eqn. (4) is not activated unless the χ^2 value in Eqn. (3) decreases below a certain small quantity that depends on the degrees of freedom. In order to achieve an improvement of the chi-square attack, we discard the CDF, and instead use the χ^2 statistics like Eqn. (3) in the comparison of the induced categories and the shifted categories.

3.2 Basic Setting

Without loss of generality, let us assume that the digital image is represented by a sequence of samples whose values are integer numbers. Let X be the random variable of samples in a cover image, and let f_x be the probability distribution of X. Let us define the two statistics χ^2_{ind} and χ^2_{shi} as follows:

$$\chi^2_{\text{ind}} = \frac{1}{2} \sum_{m \in Z} \frac{(f_{2m} - f_{2m+1})^2}{f_{2m} + f_{2m+1}} \text{ , and}$$
$$\chi^2_{\text{shi}} = \frac{1}{2} \sum_{m \in Z} \frac{(f_{2m-1} - f_{2m})^2}{f_{2m-1} + f_{2m}} \text{ .} \tag{7}$$

χ^2_{ind} and χ^2_{shi} will be used as the overall statistics for the degree of difference between sample frequencies in the induced category and in the shifted category respectively. Let X' be the random variable of samples in the stego image which is generated by the LSB embedding with randomized embedding path in the cover image, and let $f'_{x'}$ be the probability distribution of X'. Let $\widetilde{\chi}^2_{\text{ind}}$ and $\widetilde{\chi}^2_{\text{shi}}$ be defined with the stego distribution f' in similar ways of Eqn. (7).

Let ℓ, $0 < \ell < 1$, be the length of hidden message relative to the number of usable samples for carrying message bits. We call ℓ the embedding rate. For example, Jsteg does not modify the quantized DCT coefficients of values 0 and 1, and thus the embedding rate ℓ for Jsteg is the relative length of hidden message in comparison with the number of quantized DCT coefficients unequal to 0 and 1. For Jphide, if we pretend that Jphide does not modify the quantized DCT coefficients of values -1, 0 and 1, the embedding rate ℓ is the relative length of hidden message in comparison with the number of quantized DCT coefficients unequal to -1, 0 and 1. In the subsequent development, for the ease of description, we will assume that all samples in the digital image are possibly used.

3.3 Effect on Induced Categories

Since the random embedding is considered, $\ell/2$ is then the probability that the LSB of a sample could be flipped by LSB embedding. So, we can establish the basic relation between the two distributions f and f' as follows: for $m \in Z$,

$$f'_{2m} = f_{2m} - \frac{\ell}{2}(f_{2m} - f_{2m+1}) \text{ , and}$$
$$f'_{2m+1} = f_{2m+1} + \frac{\ell}{2}(f_{2m} - f_{2m+1}) \text{ .} \tag{8}$$

It is clear that

$$f'_{2m} + f'_{2m+1} = f_{2m} + f_{2m+1} \text{ , and}$$
$$f'_{2m} - f'_{2m+1} = (1-\ell)\,(f_{2m} - f_{2m+1}) \text{ .} \tag{9}$$

This means that, after the LSB embedding, the category frequency (the sum of sample frequencies in the category) of the induced category $(2m, 2m+1)$ is not changed, but the difference between sample frequencies in the category linearly decreases as the embedding rate ℓ increases. It follows that

$$\widetilde{\chi}^2_{\text{ind}} = (1-\ell)^2 \chi^2_{\text{ind}} \text{ ,} \tag{10}$$

and therefore, the LSB embedding causes a decrease in the quantity of χ^2_{ind} for the induced categories:

$$\widetilde{\chi}^2_{\text{ind}} \ll \chi^2_{\text{ind}} \text{ .} \tag{11}$$

3.4 Effect on Shifted Categories

However, the above argument is not true for the shifted categories. From Eqn. (8), we can deduce that

$$f'_{2m-1} + f'_{2m} = f_{2m-1} + f_{2m} + \frac{\ell}{2}(f_{2m-2} - f_{2m-1} - f_{2m} + f_{2m+1}) \text{ ,}$$
$$f'_{2m-1} - f'_{2m} = f_{2m-1} - f_{2m} + \frac{\ell}{2}(f_{2m-2} - f_{2m-1} + f_{2m} - f_{2m+1}) \text{ .} \tag{12}$$

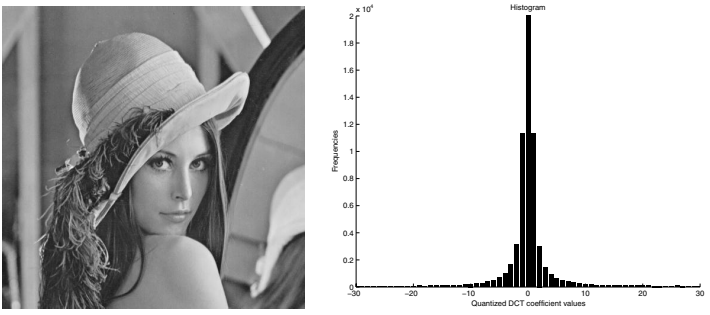
One can see that, for the shifted category $(2m-1, 2m)$, both changes of the category frequency and the difference between sample frequencies in the category are controlled by the frequencies of consecutive four sample values, where the two values are contained in the category and the other two values are externally adjacent to the category. This will lead to a contrast between the induced categories and the shifted categories under the LSB embedding.

In order to analyze the effect of LSB embedding on the difference between sample frequencies in the shifted category, we should define the relation among the frequencies of consecutive four sample values. To exemplify the relation, it is nice to consider the histogram of DCT coefficients in the JPEG image. Fig. 1 shows a part of the histogram of the well-known Lena image transformed in JPEG format with 75 % quality factor (we set the frequency of the coefficient value 0 to $2 \cdot 10^4$). There are a peak at the center of mass, slopes in both sides of the center, and tails at the edges with negligible probabilities. We note that the slopes which are monotonically increasing or decreasing appear in most of small intervals with a significant portion of the distribution, except for the interval containing the value 0 as an internal value.

Let us assume that the histogram is monotonically increasing or decreasing on consecutive four values for a shifted category $(2m-1, 2m)$[1]. That is,

$$f_{2m-2} < f_{2m-1} < f_{2m} < f_{2m+1} \text{ , or}$$
$$f_{2m-2} > f_{2m-1} > f_{2m} > f_{2m+1} \text{ .} \tag{13}$$

[1] It is reasonable when considering JPEG images, but generally untrue for other types of digital images.



(a) Lena (b) Histogram

Fig. 1. Histogram of quantized DCT coefficients for Lena image in JPEG format

From Eqn. (12), we can deduce that

$$f'_{2m-1} + f'_{2m} = f_{2m-1} + f_{2m} \pm \frac{\ell}{2}(|f_{2m-2} - f_{2m-1}| - |f_{2m} - f_{2m+1}|) \ ,$$
$$|f'_{2m-1} - f'_{2m}| = |f_{2m-1} - f_{2m}| + \frac{\ell}{2}(|f_{2m-2} - f_{2m-1}| + |f_{2m} - f_{2m+1}|). \quad (14)$$

After the LSB embedding, the difference between sample frequencies in the shifted category $(2m-1, 2m)$ linearly increases as the embedding rate ℓ increases. The category frequency in the shifted category is also altered, but the change is relatively small in comparison with the change of the frequency difference. Therefore, the LSB embedding causes an increase in the quantity of χ^2_{shi}:

$$\widetilde{\chi}^2_{\text{shi}} \gg \chi^2_{\text{shi}} \ . \quad (15)$$

3.5 Statistical Measurement

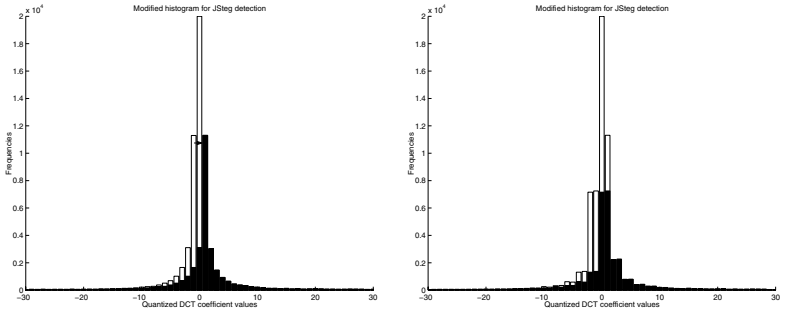
In summary, after the LSB embedding, the quantity of the statistics χ^2_{ind} for induced categories decreases, but the quantity of the statistics χ^2_{shi} for shifted categories increases. This will result in a great difference between the induced categories and the shifted categories under the LSB embedding. For the detection of LSB steganography, we decide to simply use the relative difference of the two statistics defined as follows:

$$R = \frac{\chi^2_{\text{shi}} - \chi^2_{\text{ind}}}{\chi^2_{\text{shi}} + \chi^2_{\text{ind}}} \ . \quad (16)$$

If there are some patterns in the value of R for a certain type of cover histogram, we can use them for LSB steganalysis of the digital image. And we have observed that the R statistics well discriminated between cover images and stego images in JPEG format, where stego images are generated by the Jsteg and Jphide with randomized embedding path.

3.6 Technique for Jsteg Detection

Jsteg [11] can be viewed as the LSB embedding with an exception for usable DCT coefficients; It does not modify the DCT coefficients of the values, 0 and 1. So, for detection of the Jsteg-like algorithm, we ignore them and increase the DCT coefficients of negative values by 2. Fig. 2 displays the modification of the histogram by the preprocessing of the Lena image in JPEG format before and after the Jsteg embedding with full capacity. The white bars are the histogram of all DCT coefficients. We strip two of them that are not used for steganography, namely h_0 and h_1, and regard only the resulting histogram with the black bars. One can see that the modified histogram of the cover image still maintains the symmetry around one value which can deduce that $\chi^2_{\text{ind}} = \chi^2_{\text{shi}}$. This is common for JPEG images and therefore R statistics can be used for the detection.

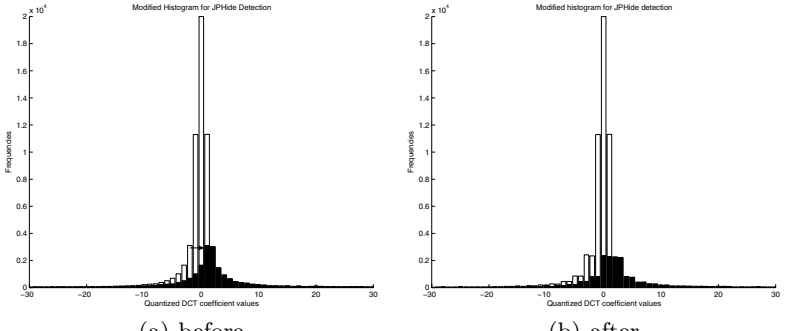


(a) before (b) after

Fig. 2. Histogram of Lena image in JPEG format before and after Jsteg embedding

3.7 Technique for Jphide Detection

Jphide [12] can be viewed as the LSB embedding in a sense that message bits are encoded in the LSBs of the absolute values of DCT coefficients; although



(a) before (b) after

Fig. 3. Histogram of Lena image in JPEG format before and after Jphide embedding

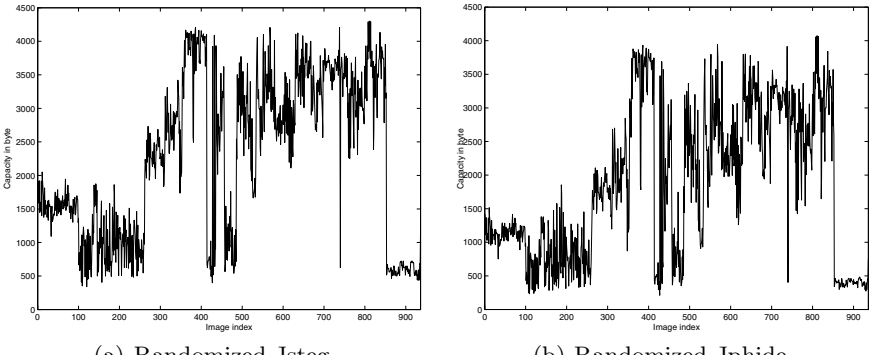
Jphide occasionally modifies the second least significant bits, but these are not frequent and the effect on the statistics is negligible. Jphide also modifies the DCT coefficients of values -1, 0 and 1 in a special way. So, we ignore them and increase the DCT coefficients of negative values by 3. This technique still allows us to detect the Jphide embedding. Fig. 3 displays the modified histogram by the preprocessing of the Lena image in JPEG format before and after the modified Jphide embedding with full capacity. Again, the steganographically unused values, namely h_{-1}, h_0 and h_1, are stripped from the histogram. In this case, however, the modified histogram of the cover image results in $\chi^2_{\text{ind}} > \chi^2_{\text{shi}}$.

4 Experimental Results

4.1 Image Sets

We used 936 JPEG images of the CBIR image database from Washington University [14]. We only used Y channel data of JPEG images in the test. The testing was done for 6 embedding rates, 5 %, 10 %, 20 %, 30 %, 40 % and 50 %, in bits per a usable coefficient (bpc) for each steganography algorithm. We embedded random messages in the LSBs of usable coefficients that are randomly selected from Y channel data. The random message here was newly generated for each stego image. In the simulation, we implemented randomized versions of Jsteg and Jphide algorithms; for the randomized Jphide algorithm, we did not use the DCT coefficients of values -1, 0 and 1, and embedded message bits in the LSBs of absolute values.

Fig. 4 shows message lengths (in bytes) hidden in the stego images for an embedding rate of 10 %. Fig. 4 (a) shows the message lengths for the randomized Jsteg embedding. The maximal capacity is 4302 bytes, and the minimal capacity is 340 bytes, and the average is 2267 bytes. Fig. 4 (b) shows the same for the randomized Jphide embedding. The maximal capacity is 4078 bytes, the minimal capacity is 211 bytes, and the average is 1846 bytes.

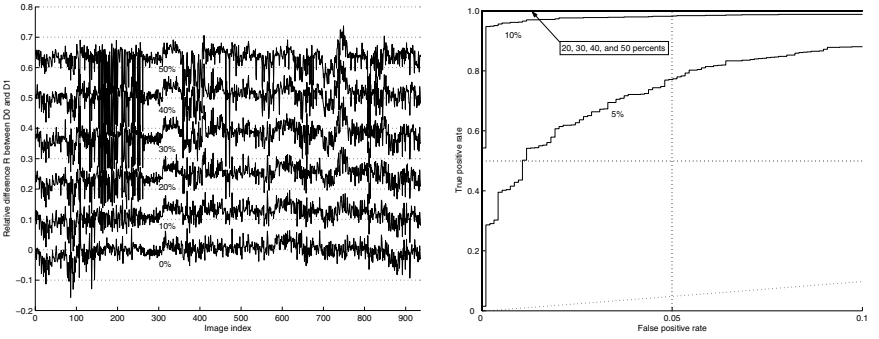


(a) Randomized Jsteg (b) Randomized Jphide

Fig. 4. Embedding capacities for stego images in case of the 10 % embedding rate

4.2 On Randomized Jsteg Algorithm

Fig. 5 displays the results of the proposed attack on the randomized Jsteg algorithm. In (a), one can see that the R statistics is highly sensitive to the low-rate embedding. Even for the 10 % embedding rate, most of stego images seems to be distinguished from the cover images. This well explains the ROC curves for the category attack on the randomized Jsteg algorithm in (b). The ROC curves show that, when the embedding rates are greater than or equal to 20 %, all stego images were perfectly separated from the cover image set.

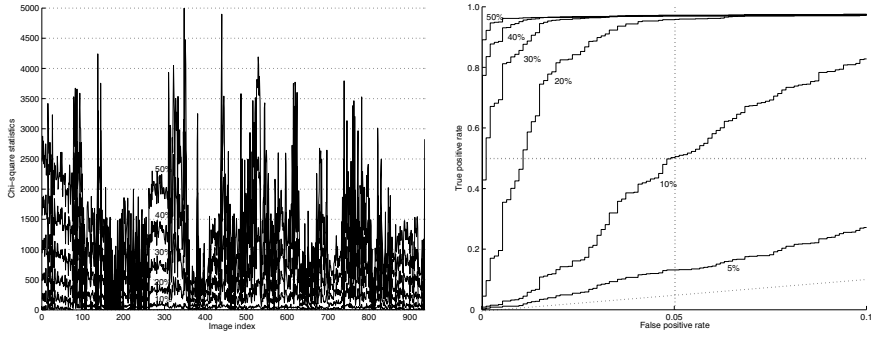


(a) Test Results for R Statistics (b) ROC curves for embedding rates

Fig. 5. Results for the category attack on the randomized Jsteg algorithm

4.3 Vs. ZP Attacks

Fig. 6 displays the results of the ZP attack on the randomized Jsteg algorithm. Here the χ^2 and the relation $f_1 > f_0$ were used. At a glance, the category attack outperforms the ZP attack.

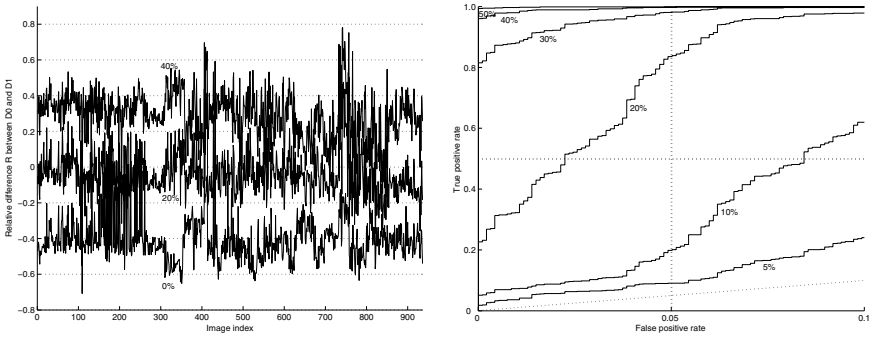


(a) Test Results for χ^2 statistics (b) ROC curves for embedding rates

Fig. 6. Results for the ZP attack on the randomized Jsteg algorithm

4.4 On Randomized Jphide Algorithm

Fig. 7 displays the results of the proposed attack on the randomized Jphide algorithm. The category attack works for the randomized Jphide embedding. However, the initial stats of R statistics for cover images are varied and these will disturb the correct decision for the low-rate embedding.



(a) Test Results for R Statistics (b) ROC curves for embedding rates

Fig. 7. Results for the category attack on the randomized Jphide algorithm

5 Evaluation and Comparison of Detection Reliability

Table 1 and 2 give different quality measures that are used in the literature. Fridrich measures the reliability ρ, defined by twice the area between the ROC curve and the diagonal ($\rho = 1$ means perfect separation, $\rho = 0$ equals random guessing) [8]. Lyu and Farid (LF) measures the true positive rate (TPR) at 1 % false positive rate (FPR) [6]. Ker requires the FPR to be less than 5 % at 50 % TPR [9]. We give the FPR for 50 % TPR in the table. For the LF and Ker values, we give also the separating thresholds. The "Mean" and "Var" columns contain the mean and variance values of the attack results for the respective set steganograms.

The results of the ZP algorithm presented here are based on the β value determined according to Eqn. (6). The original ZP algorithm decides using the χ^2 and the relation $f_1 > f_0$. This relation is also expressed by the sign of β and we found that the degree of negativity can slightly improve the detection reliability.

We implemented the 23 DCT features by Fridrich [8], extracted them from the 12168 files and trained a support vector machine[2] on 2×690 files for the 6 embedding rates and the 2 algorithms. We classified the remaining 2×246. Compared to targeted attacks, the result is rather poor for low embedding rates (this can be also assured at Fig. 8). However, the blind attack is universal and

[2] We use the SVM implementation from the e1071 package of the R software [15].

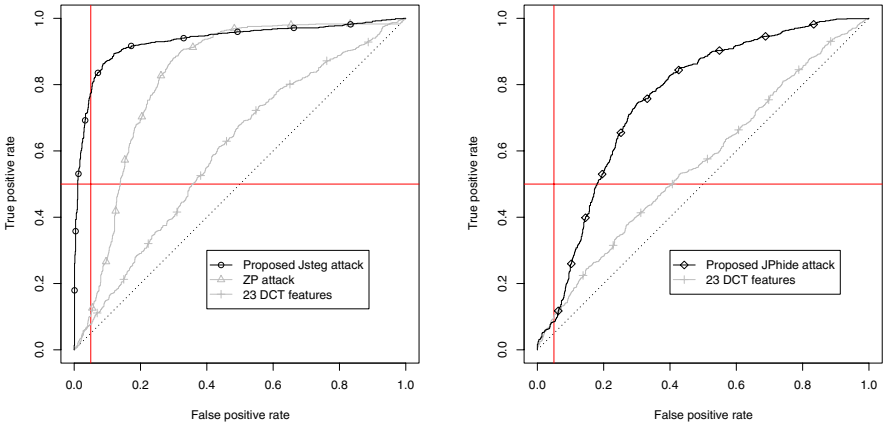
Table 1. Evaluation of randomized Jsteg detection power

Attack	Image set	ρ	Ker	Ker.thr	LF	LF.thr	Mean	Var
CA Jsteg	05	0.8657	0.0107	0.0568	0.4530	0.0600	0.0547	0.0008
	10	0.9897	0.0000	0.1178	0.9605	0.0605	0.1155	0.0009
	20	1.0000	0.0000	0.2466	1.0000	0.0613	0.2432	0.0012
	30	1.0000	0.0000	0.3802	1.0000	0.0613	0.3755	0.0015
	40	1.0000	0.0000	0.5132	1.0000	0.0613	0.5055	0.0016
	50	1.0000	0.0000	0.6361	1.0000	0.0613	0.6268	0.0017
23dctf	05	0.2078	0.3659	0.4011	0.0285	0.8990	0.4644	0.0380
	10	0.5600	0.1057	0.5626	0.0569	0.8997	0.5703	0.0444
	20	0.8469	0.0325	0.7158	0.2398	0.8643	0.7049	0.0602
	30	0.9387	0.0000	0.8074	0.5935	0.7600	0.7731	0.0574
	40	0.9775	0.0000	0.8616	0.7967	0.6505	0.8230	0.0472
	50	0.9935	0.0000	0.8947	0.8699	0.6137	0.8610	0.0381
ZP beta	05	0.6425	0.1389	0.0430	0.0096	2.1429	0.0767	0.6344
	10	0.8414	0.0577	0.0928	0.0096	2.1970	0.1246	0.5871
	20	0.9286	0.0203	0.1934	0.0085	2.1970	0.2204	0.4291
	30	0.9389	0.0171	0.2950	0.0075	2.1970	0.3203	0.3437
	40	0.9419	0.0171	0.3959	0.0075	2.1970	0.4183	0.2356
	50	0.9460	0.0160	0.4959	0.0064	2.1970	0.5140	0.1788

Table 2. Evaluation of randomized Jphide detection power

Attack	Image set	ρ	Ker	Ker.thr	LF	LF.thr	Mean	Var
CA Jphide	05	0.5133	0.1795	-0.3369	0.0331	-0.0051	-0.3176	0.0138
	10	0.7886	0.0812	-0.2396	0.0673	-0.0054	-0.2235	0.0139
	20	0.9392	0.0267	-0.0635	0.2949	-0.0057	-0.0453	0.0164
	30	0.9902	0.0000	0.1209	0.8643	-0.0051	0.1272	0.0178
	40	0.9981	0.0000	0.2983	0.9754	-0.0051	0.2871	0.0193
	50	0.9997	0.0000	0.4560	0.9957	-0.0051	0.4298	0.0204
23dctf	05	0.1393	0.4024	0.6847	0.0244	0.9398	0.6406	0.0405
	10	0.2716	0.2927	0.6204	0.0772	0.8828	0.6006	0.0417
	20	0.5099	0.1504	0.5804	0.1789	0.7938	0.5900	0.0473
	30	0.6661	0.0691	0.6191	0.2602	0.7854	0.6405	0.0494
	40	0.8028	0.0366	0.6742	0.3618	0.7592	0.6827	0.0454
	50	0.8907	0.0081	0.7169	0.5000	0.7169	0.7249	0.0404

detect algorithms like Outguess that preserve first-order statistics. Because the proposed attack decides based on histograms, it is unable to detect algorithms like Outguess.



(a) Jsteg Attacks (b) Jphide Attacks

Fig. 8. ROC curves showing the improved reliability of the category attack on randomized Jsteg and Jphide algorithms. Here, 5 % embedding rate is used for stego images.

6 Conclusion

In this paper, we proposed the category attack for LSB steganalysis of JPEG images. The category attack exploits simply the histogram of DCT coefficients, but is more powerful to detect the randomized Jsteg embedding as well as the randomized Jphide embedding. The proposed method outperformed the Jsteg detection by Zhang and Ping. The detection power of both proposed methods were compared to the blind classifier by Fridrich that uses 23 DCT features.

There seems to exist a relation between the R statistics used in the category attack and the length of hidden messages. The exact formula to estimate the hidden message length will be further researched.

Acknowledgements

This work was supported by grant No. M10640010005-06N4001-00500 from the national R&D Program of MOST and KOSEF.

References

1. S. Katzenbeisser and F.A.P. Petitcolas, *Information Hiding - techniques for steganography and digital watermarking*, Artech House Books, 1999.
2. A. Westfeld and A. Pfitzmann, "Attacks on steganographic systems," in Information Hiding: 3rd International Workshop, IH'99 Dresden, Germany, September 29 – October 1, 1999 A. Pfitzmann, ed., LNCS 1768, pp. 61–76, Springer-Verlag, Berlin Heidelberg, 2000.

3. N. Provos and P. Honeyman, "Detecting Steganographic Content on the Internet," CITI Technical Report 03–11, 2001.
4. J. Fridrich, M. Goljan and D. Hogea, "Attacking the OutGuess," in Proc. of the ACM Workshop on Multimedia and Security 2002, Juan-les-Pins, France, December 6, 2002.
5. J. Fridrich, M. Goljan, and D. Soukal, "Higher-Order Statistical Steganalysis of Palette Images," in Proc. of EI SPIE, Santa Clara, CA, Jan 2003, pp. 178-190.
6. S. Lyu and H. Farid, "Detecting Hidden Messages Using Higher-Oder Statistics and Support Vector Machines," in Information Hiding: 5th International Workshop, IH2002, Noordwijkerhout, The Netherlands, October 7–9, 2002, F.A.P. Petitcolas, ed., LNCS 2578, pp 340–354, Springer–Verlag, Berlin Heidelberg, 2003.
7. T. Zhang and X. Ping, "A Fast and Effective Steganalytic Technique against Jsteg-like Algorithms," in ACM Symposium on Applied Computing, March 9–12, 2003, Florida, USA, 2003.
8. J. Fridrich, "Feature-Based Steganalysis for JPEG Images and its Implications for Future Design of Steganographic Schemes," in Information Hiding: 6th International Workshop, IH2004, Toronto, Canada, May 23–25, 2004, Revised Selected Papers, J. Fridrich, ed., LNCS 3200, pp. 97–115, Springer–Verlag, Berlin Heidelberg, 2004.
9. A. D. Ker, "Improved Detection of LSB Steganography in Grayscale Images," in Information Hiding: 6th International Workshop, IH2004, Toronto, Canada, May 23–25, 2004, Revised Selected Papers, J. Fridrich, ed., LNCS 3200, pp. 97–115, Springer–Verlag, Berlin Heidelberg, 2004.

Internet Sources

10. Steganographic Tool Lists, `http://www.stegoarchive.com`.
11. D. Upham, Jpeg-Jsteg, `http://www.funet.fi/pub/crypt/stegangraphy/jpeg-Jsteg-v4.diff.gz`.
12. Allan Latham, Jphide and JPSeek, `http://linux01.gwdg.de/$\sim$alatham/stego.html`.
13. Niels Provos, OutGuess – Universal Steganography, `http://www.outguess.org`.
14. CBIR Image Database, University of Washington, `http://www.cs.washington.edu/research/imagedatabase/groundtruth`.
15. R: A Language and Environment for Statistical Computing, R Development Core Team, R Foundation for Statistical Computing, Vienna, Austria, 2006, http://www.r-project.org,