A Better Detection of 2LSB Steganography via Standard Deviation of the Extended Pairs of Values

Omed Khalind*, Benjamin Aziz
School of Computing, University of Portsmouth, Portsmouth, Po1 3HE, UK
{omed.khalind, benjamin.aziz}@port.ac.uk, *+44 77 0902 0299

ABSTRACT

This paper proposes a modification to the Extended Pairs of Values (EPoV) method of 2LSB steganalysis in digital still images. In EPoV, the detection and the estimation of the hidden message length were performed in two separate processes as it considered the automated detection. However, the new proposed method uses the standard deviation of the EPoV to measure the amount of distortion in the stego image made by the embedding process using 2LSB replacement, which is directly proportional with the embedding rate. It is shown that it can accurately estimate the length of the hidden message and outperform the other methods of the targeted 2LSB steganalysis in the literature. The proposed method is also more consistent with the steganalysis methods in the literature by giving the amount of difference to the expected clean image. According to the experimental results, based on analysing 3000 never-compressed images, the proposed method is more accurate than the current targeted 2LSB steganalysis methods for low embedding rates.

Keywords: 2LSB steganalysis, 2LSB steganography, Pairs of values analysis, detection, information hiding

1. INTRODUCTION

Steganography is the art and the science of hiding secret data in another media for communication, without raising suspicion by the third party[1]. Usually, Multimedia digital objects are excellent media for steganography, as they have a high degree of redundancy[2]. So, digital images are one of the best and most commonly used digital media for this purpose.

LSB embedding is the most commonly used method of image steganography because; it is easy to implement, has a reasonable capacity, and visually imperceptible. However, it could be easily detected due to the imbalance distortion on the intensity histogram of the image and producing 'Pairs of Values'. Extensions to LSB steganography also received a great attention by steganographers and nowadays there are a number of publicly available steganography tools that could be used for this purpose, for example SilentEye[3]. One of the extensions of the LSB steganography method is 2LSB data embedding which even has a higher capacity than LSB method with more complicated changes on the intensity histogram of the cover image that makes it harder to detect.

During the past decade a number of detection methods have been proposed to detect the extended methods of LSB. Some methods proposed to detect multiple LSB steganography[4-7], which are expected to have lower accuracy than the 2LSB specific steganalysis methods. Other methods are specific to the detection of 2LSB steganography, as explained below.

Based on the quartic equation, Luo et al.[8] proposed a detection method for 2LSB steganography in digital images. The method constructs a finite-state machine based on the sample pairs of the image pixel values, and then builds a quartic equation via the relation of the conversion states to obtain the estimated embedding rate. This method uses a very complex calculation and needs a long time to analyse the given image[9].

Ker[10] also proposed a steganalysis method to detect 2LSB message embedding in digital images by extending the structural analysis of the image. This method uses statistics of many variances to form the equation that estimates the message length. This is also considered as a complex detection method because it needs lots of calculations.

Another method of detecting 2LSB steganography is proposed by Zhang et al.[11] based on the statistical characteristics in the two least significant bits of the pixel values in the image. The detection accuracy can reach 90% only when the embedding rate is not less than 0.2. So, it limits the performance of detection for lower embedding rates.

The last and the most accurate detection method of 2LSB steganography in digital images is proposed by Niu et al.[9]. It can accurately detect and estimate the length of the embedded message by constructing a weighted stego image and using least square equation. The authors have compared their results with the detection method proposed by Ker[10] and showed a better accuracy and faster detection on the same set of images. Hence, this paper considers this method to compare the accuracy of the proposed method.

In this paper we propose a modification to the EPoV analysis of detecting 2LSB replacement in still images. The proposed method uses the standard deviation of the frequency of occurrences in EPoVs to estimate the length of the hidden message, which become a probabilistic classifier rather than being a discrete classifier. The evaluation is done over a set of 3000 never-compressed images[12] after converting them into grey-scale with streams of pseudo random binary values as a secret message; to make it very close to the encrypted version[11]. The results of the proposed method are compared to the results of the method proposed by Niu et al.[9]. It is shown that the proposed method is more accurate in detection for low embedding rates.

This paper is organised as follows; it starts with a brief description of two least significant bits steganography and its features. Then, the proposed method is explained with all supporting information. In section 4, the experimental results are explained and shown in graphs for clarification. Finally, it ends with conclusions about this research work.

2. 2LSB STEGANOGRAPHY IN STILL IMAGES

Embedding in two least significant bits could be divided into two major types[10, 11]; 2LSB and I2LSB. The 2LSB, directly replaces both two least significant bits of the selected pixel values with two bits of the secret message. While I2LSB, independent 2LSB, replaces the two least significant bits of the chosen pixel values independently. For example, it can start with replacing the first-LSBs of all selected pixel values then the second-LSBs separately, or vice versa.

Both methods of two least significant bits steganography, 2LSB and I2LSB, are clearly transferring pixel values into each other in such a way that bounds them in groups of four pixel values, as shown in Figure 1. They also lead to breaking the correlation between 7^{th} and 8^{th} bit-planes in each pixel value, by inserting random stream of binary values. However, this correlation between 7^{th} and 8^{th} bit-planes is not random in clean images, which would be the base for the proposed detection method.

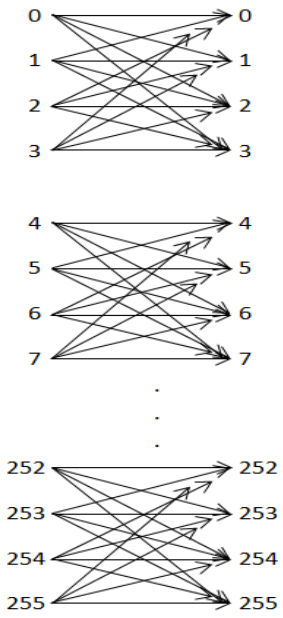

Figure 1: Possible pixel value transitions of 2LSB steganography

3. THE PROPOSED METHOD

The 2LSB steganography results in more complicated changes on the intensity histogram of the pixel values than LSB methods, which makes the detection process harder to perform. It changes the two lower bit-planes (7^{th} and 8^{th}, for 8-bit pixel values) and bounds the transition of pixel values into groups of four values called EPoV[13]. So, analysing extended pairs of values would be a very good choice for detection. Instead of separating the detection from the estimation of the hidden message length, here another method is proposed to measure the amount of change in the stego image by the embedding process.

As mentioned earlier, the 2LSB embedding causes the insertion of random sequence of binary values, it results in a broken correlation in lower bit-planes (7^{th} and 8^{th}), which is not random in clean images. Moreover, it is expected to have more different pairs of bit values in lower two bit-planes (xxxxxx01, xxxxxx10) after embedding with 2LSB steganography, more details could be found in[13]. Hence, they are grouped into same or similar (xxxxxx00, xxxxxx11) and different (xxxxxx01, xxxxxx10) 2LSBs pixel values, as shown in Figure 2.

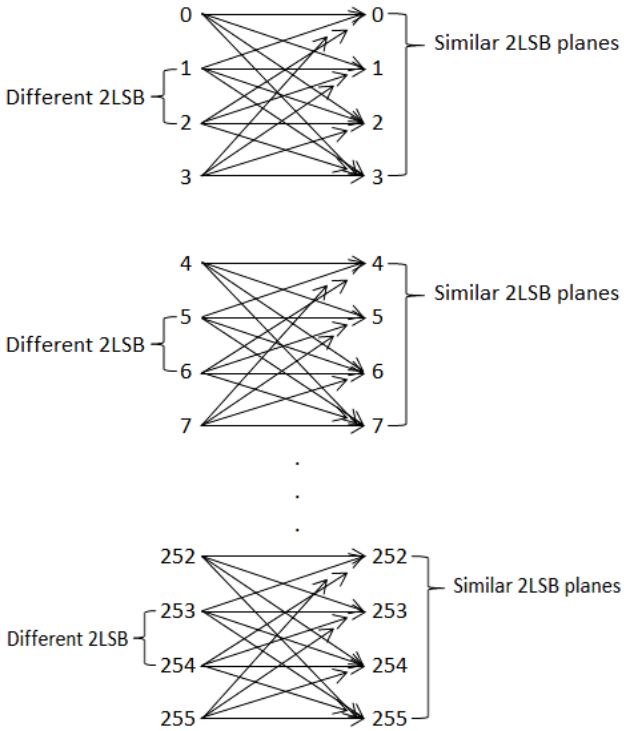

Figure 2: Grouping of pixel values within each EPoV

If k indicates the index of the EPoV, which could range from 0 to 63 for 8-bit pixel values, then the same and different 2LSB pixel values within a certain EPoV would be $(4k, 4k+3)$ and $(4k+1, 4k+2)$ respectively.

The sum of frequency of occurrences within each EPoV stays unchanged before and after embedding process. Hence, taking the arithmetic mean of the frequency of occurrences of both same and different 2LSB pixel value groups in each EPoV would be considered to measure the imbalance between same (00, 11) and different (01, 10) two least significant bits in the image before and after the embedding place has taken place.

According to analysing 3000 never-compressed images, more than 97% of their standard deviation of the set of arithmetic means of frequency occurrences for both groups (same and different) in each EPoVs were very close to the standard deviation of the set of frequency of occurrences for the same 2LSBs in each EPoVs. So, dividing the standard deviation of the arithmetic means of both groups in the EPoVs by the standard deviation of the same 2LSBs group would be very close to 1 in clean images.

Based on this conclusion, to find the amount of changes by the 2LSB embedding, we subtract the expected value of the clean image, which is 1, and the remaining will be the modification rate. According to the experimental results this value will reach up to 1.5 in the corresponding stego image for the embedding rate of 1. So, if we subtract the expected value for clean images we get 0.5 which implies that half of the image is modified. In other words, the total capacity of the image has been used by the embedding process. Hence, the modification rate, after subtracting 1, ranges from 0 to 0.5 which is directly proportional with the embedding rate. For the embedding rate of 1, it gives values very close to 0.5, which means that half of the 2LSBs of the image pixel values are modified. The detection process is shown in Figure 3.

Input: Image I
Output: Double modificationRate
Start
Array $X_{64} = 0, Y_{64} = 0$
For all pixel values P of I
  $indexOfEPoV = P/4 + 1$
  If $2\text{LSB}(P) = 11\ OR\ 2LSB(P) = 00$
    Increment X(indexOfEPoV)
  else
    Increment Y(indexOfEPoV)
  End
  For i=1 to 64
    $Z = round\Big(\big(X(i) + Y(i)\big)/2\Big)$
  End
End
$modificationRate = std(Z)/std(X) - 1$
End

Figure 3: The pseudo code of detection algorithm

Figure 4 and Figure 5, clearly show the differences between the clean and stego versions of the Lenna image. They show the frequency of occurrences for the same 2LSB pixel values (X), the arithmetic mean of same and different 2LSB pixel values (Z) in each EPoVs, and their standard deviation. As could be noted, they are very close for the clean version of the image and different for the stego version with an embedding rate of 1. Moreover, as could be seen in Figure 6, the detection result is very close to Zero for the clean version and 0.5 for the stego version of the Lenna image with the embedding rate of 1, after subtracting the expected value of the clean image which is 1.

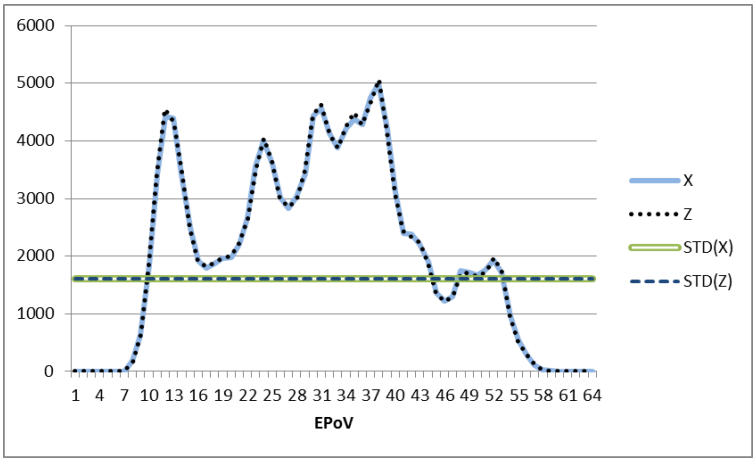

Figure 4: Analysis of Lenna clean image

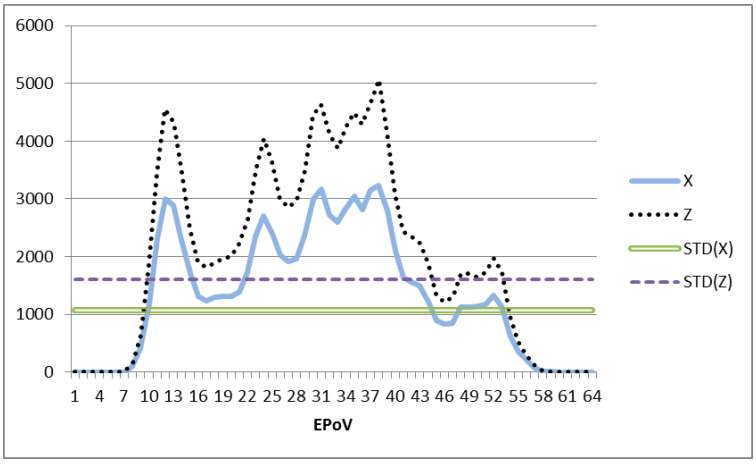

Figure 5: Analysis of Lenna stego image with an embedding rate of 1

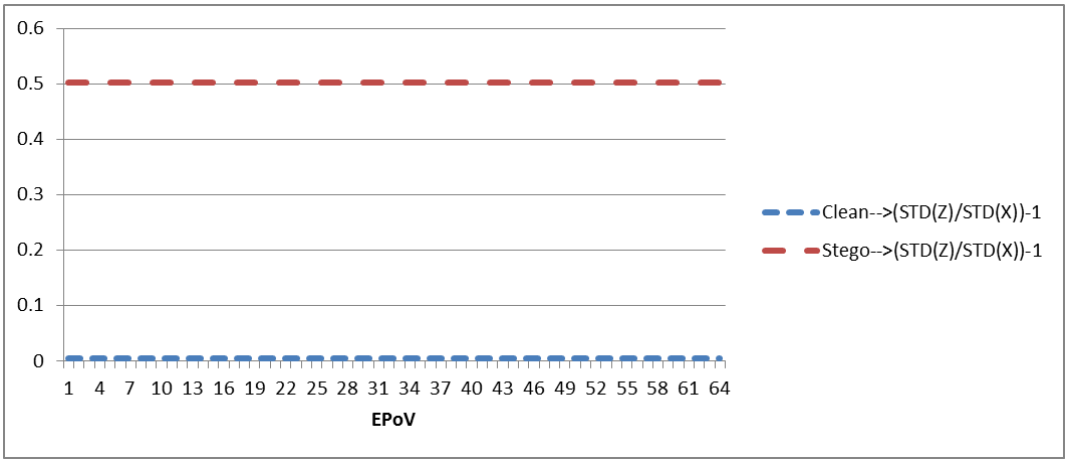

Figure 6: The detection results of the clean and stego version of Lenna image

4. EXPERIMENTAL RESULTS

As a basic evaluation, the three common images among steganographers (Lenna, Pepper, and Baboon) are taken into consideration, shown in Figure 7. The results of the estimated amount of the image that has changed are shown in Tables 1 and 2 for both proposed method and the existing one (WS2[9]).

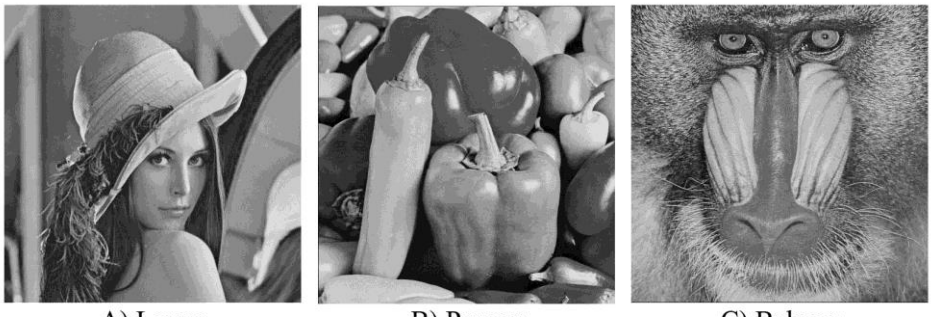

Figure 7: Standard images used in Steganography

Table 1: Detection results of the proposed method

Images	Embedding rate						
	0%	5%	10%	20%	50%	75%	100%
Lenna	0.005	0.031	0.054	0.096	0.205	0.340	0.497
Pepper	0.000	0.020	0.045	0.103	0.246	0.367	0.505
Baboon	0.002	0.015	0.023	0.047	0.168	0.301	0.498

Table 2: Detection results of the WS2

Images	Embedding rate						
	0%	5%	10%	20%	50%	75%	100%
Lenna	0.008	0.024	0.038	0.072	0.174	0.270	0.385
Pepper	0.007	0.028	0.039	0.076	0.181	0.287	0.398
Baboon	0.028	0.035	0.056	0.086	0.189	0.278	0.403

The estimation of the message length (or the modification rate of the image) could also be evaluated by comparing it with a perfect classifier, which practically does not exist. Table 3, shows the average of differences for all embedding rates of the three images with the perfect classifier. As could be noticed, the proposed method is more accurate than the WS2.

Table 3: The difference between the detection methods and the perfect classifier

Detection methods	Average difference
Proposed method	0.018
WS2	0.046

To evaluate the proposed steganalysis method, a set of 3000 never-compressed images[12] are used as cover objects after converting them into grey-scale. For each embedding rate (5%, 10%, 20%, 50%, 100%) the images are loaded with a stream of pseudo random bits as a secret message, to have all the statistical properties of the encrypted version of it[14]. The stego images, then, fed into both the proposed method and the most accurate detection method of the targeted 2LSB steganalysis[9] for comparison. The results are shown in the form of ROC graph for both detection methods in Figures 8 and 9. As could be noticed, the proposed method outperforms the weighted stego method (WS2) for low embedding rates (less than 50%). This is because the weighted stego method relies on the probabilistic model of the cover image, which is expected to not always be very accurate, especially for low embedding rates. However, the proposed method relies on the arithmetic mean of the frequency of occurrences in each EPoV which has the same value for both clean and stego versions of the image with any embedding rate.

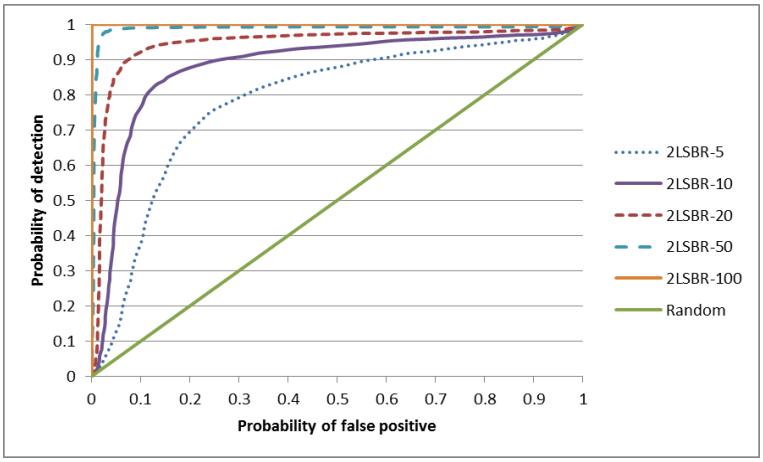

Figure 8: The ROC graph of the proposed method for 3000 images

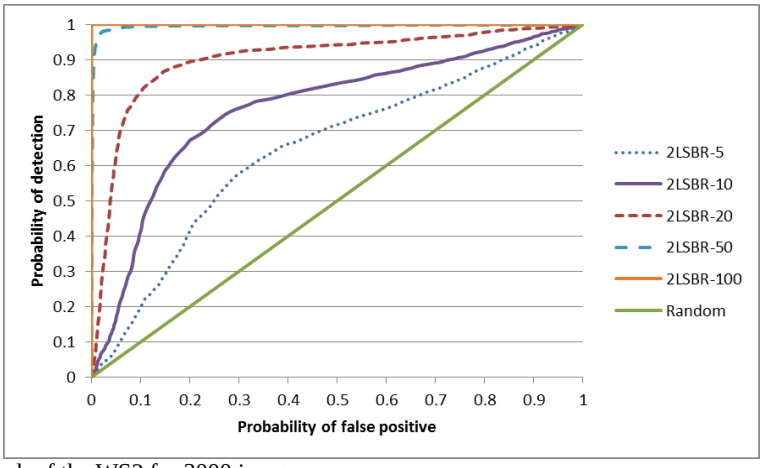

Figure 9: The ROC graph of the WS2 for 3000 images

5. CONCLUSION AND FUTURE WORK

The proposed targeted 2LSB steganalysis method gives a better accuracy in detection for low embedding rates than the existing methods. It also can accurately estimate the length of the hidden message for any embedding rate. Therefore, it can be more useful than the traditional EPoV and improve the current accuracy of the 2LSB steganalysis methods in the literature. Moreover, as it considers the arithmetic mean of the frequency of occurrences in each extended pairs of values, which would be the same before and after embedding for a certain image. Hence, the proposed detection method can maintain its high accuracy for low embedding rates as well. As a future work, the accuracy of the proposed method might be possible to improve using other statistical methods with the EPoV grouping method.

ACKNOWLEDGEMENT

The authors would like to thank the anonymous reviewers for their useful comments and feedback during the review process. The travel expenses of presenting this research paper were funded by the University of Portsmouth, Faculty of Technology Research Capital Investment Fund (RCIF) number 46175.

REFERENCES

[1] R. Chandramouli, and N. D. Memon, "Steganography capacity: a steganalysis perspective," Proc. SPIE 5020, 173-177 (2003).

[2] R. Chandramouli, and N. Memon, "Analysis of LSB based image steganography techniques," Proc. International Conference on Image Processing, 1019-1022 (2001).

[3] A. Chorein, "SilentEye - Steganography is yours," 2008, <http://www.silenteye.org>.

[4] X. Yu, T. Tan, and Y. Wang, "Extended optimization method of LSB steganalysis," Proc. IEEE International Conference on Image Processing, 1102-1105 (2005).

[5] X. Yu, and N. Babaguchi, "Weighted stego-image based steganalysis in multiple least significant bits," Proc. IEEE International Conference on Multimedia and Expo, 265-268 (2008).

[6] X. Luo, F. Liu, C. Yang, S. Lian, and Y. Zeng, "Steganalysis of adaptive image steganography in multiple gray code bit-planes," Multimedia Tools and Applications, 57, 651-667 (2012).

[7] C. Yang, F. Liu, X. Luo, and B. Liu, "Steganalysis frameworks of embedding in multiple least-significant bits," IEEE Transactions on Information Forensics and Security, 3, 662-672 (2008).

[8] X. Luo, Q. Wang, C. Yang, and F. Liu, "Detection of LTSB steganography based on quartic equation," Proc. 8th International conference of Advanced Communication Technology, 1199-1204 (2006).

[9] C. Niu, X. Sun, J. Qin, and Z. Xia, "Steganalysis of two least significant bits embedding based on least square method," Proc. International Colloquium on Computing, Communication, Control, and Management, 124-127 (2009).

[10] A. D. Ker, "Steganalysis of Embedding in Two Least-Significant Bits," IEEE Transactions on Information Forensics and Security, 2, 46-54 (2007).

[11] K. Zhang, H.-Y. Gao, and W.-s. Bao, "Steganalysis Method of Two Least-Significant Bits Steganography," International Conference on Information Technology and Computer Science, ITCS 2009, 350-353 (2009).

[12] "Never-compressed image database," <http://www.shsu.edu/~qxl005/New/Downloads/index.html>.

[13] O. Khalind and B. Aziz, "Detecting 2LSB steganography using extended pairs of values analysis," Proc. SPIE 9120, 912003-12 (2014).

[14] A. Westfeld, and A. Pfitzmann, [Information Hiding], "Attacks on Steganographic Systems," Springer Berlin Heidelberg, 61-76 (2000).