ISSN (Print) : 0974-6846
Indian Journal of Science and Technology, *Vol 8(24), DOI: 10.17485/ijst/2015/v8i24/79991, September 2015* ISSN (Online) : 0974-5645

A Novel Low-D Feature based Generic Steganalyzer to Detect Low Volume Payloads

S. Arivazhagan*, W. Sylvia Lilly Jebarani, S. T. Veena and M. Shanmugaraj

Department of Electronics and Communication Engineering, Mepco Schlenk Engineering College, Sivakasi - 626005, Tamil Nadu, India; s_arivu@yahoo.com, vivimishi@yahoo.co.in, veena_st@yahoo.com, petshanmugaraj@rediffmail.com

Abstract

Data hiding techniques whenever used to hide mammoth payloads disturb statistical properties of the cover medium thus leaving a characteristic artifact. These artifacts can provide useful information to the watchful eyes of the steganalyst to identify potential carriers. But the probability of detection sharply declines when the amount of data getting embedded is reduced. Intelligent steganographers as a measure of evading significant artifacts hide only minimal amount of data. This work is an effort to differentiate stego images from innocuous cover images especially when they carry very minimal payloads. A novel low dimensional feature set has been used along with an ensemble classifier.

Keywords: Composite Feature Set, Ensemble Classifier, Payload, Steganalysis, Steganography

1. Introduction

Steganalysis, the science of detecting covert communication tries to extract and identify significant artifacts left as a result of the data embedding process. The artifacts serve as a signature and give many vital clues to the steganalyst regarding amount of data embedded, location of embedded data as well as the tool used for the steganographic process. The clues that the steganalyst obtain from the stego images are dependent on the size of the payload. When the amount of data embedded becomes less than 5% of the embedding capacity of the cover image, most of the time it goes undetected and steganalysis miserably fails[1]. Due to abundant availability of images in the World Wide Web and the need to search for hidden content, an automated steganalyzer is necessary. Hence an effort has been made in this approach using a handful of features in training an ensemble steganalyzer.

Generic Steganalysis needs to identify stego images, irrespective of format of the medium, steganographic algorithm used either in spatial or transform domain, data hidden sequentially or randomly, whether busy areas or significant areas of the image are used for hiding and most importantly, whatever may be the size of the payload. Steganographers use this technique of minimal embedding as their covert channel goes unnoticed. But this really makes the task of the steganalyst tough in looking for fine and intricate details or skews in stego images. Steganalysis in the recent past has used many features. Westfeld and Pfitzmann presented both visual attacks, making use of the ability of humans to clearly discern between noise and visual patterns, and statistical attacks which are much easier to automate[2]. Provos and Honeyman presented an automatic detection framework that includes tools to retrieve images from the World Wide Web and automatically detect whether they might contain steganographic content[3]. Fridrich, Goljan and Hogea presented a steganalytic method that can reliably detect messages (and estimate their size) hidden in JPEG images using the steganographic algorithm F5[4].

* *Author for correspondence*

Avcibas, Memon and Sankur used the hypothesis that steganographic schemes leave statistical evidence that can be exploited for detection with the aid of image quality features and multivariate regression analysis[5].

Farid described an approach to detect hidden messages in images that uses a wavelet-like decomposition to build higher-order statistical models of natural images[6]. Fridrich, Goljan and Hogea presented a general methodology for developing attacks on steganographic systems for the JPEG image format[7]. Fridrich introduced a new feature-based steganalytic method for JPEG images and used it as a benchmark for comparing JPEG steganographic algorithms and evaluating their embedding mechanisms[8]. Agaian and Cai presented an universal blind steganalysis method using features derived from color wavelet decomposition[9]. Shi used Moments of Characteristic Functions (MOCF) derived from Wavelet Decomposition and Prediction-Error images[10]. Zou proposed a Markov chain based model and had clearly shown the dependency of detection accuracy on the size of the payload[11]. Lyu and Farid used higher order image statistics to perform generic steganalysis successfully on a very large database but admitted their inadequacy in detecting stego images as the message size becomes smaller[1]. Quach et al. improved blind detection of low volume payloads by using distribution of DCT coefficients of images[12]. Bell and Lee presented a fully automated, blind approach to steganalysis for identification of signatures of steganography software[13]. Zhong et al. presented a specific steganalysis technique against reversible data hiding based on difference expansion method[14]. Zong et al. proposed a blind JPEG steganalytic method based on inter and intra wavelet subband correlations[15]. Cho et al. differentiated a stego image from its cover image inspecting decomposed image blocks of DCT coefficients by performing local steganalysis[16]. Holub et al. used higher order cooccurrence derived from an entire family of noise residuals referred to as rich image representation[17]. Pathak et al. extended the concept of image calibration to cross domains for Steganalysis[18]. To thwart steganalytic attacks, data hiding techniques resort to minimal embedding which leaves much of the cover statistics intact. Low volume steganography grabbed attention only in 2006 and a few works have been reported that employ different feature sets with varying degrees of success. This approach is for designing an efficient steganalyzer which should not let go minimal payloads undetected.

The rest of the paper is organized as follows: The composite feature set used is discussed in detail in Section 2. Architecture of the proposed Steganalyzer is explained in Section 3. Results and Discussion are presented in Section 4 and Section 5 concludes the paper.

2. Feature Extraction

The feature set has features derived from both spatial domain and transform domain as the steganographic software work can work in any of these domains. Changes that have affected the detail part of an image have been captured by deriving features from only the detail subbands of a wavelet decomposed image. As different steganographic algorithms affect different bit planes of an image, features have also been derived from all bit planes of the images. In addition, a performance measure which appraises image encryption process has been used as a feature for Steganalysis as it characterizes the changes that happen during embedding.

2.1 Spatial Domain Features

2.1.1 Net Pixel Change Rate/Number of Changing Pixel Rate (NPCR)

It is a quantifier used in cryptography to evaluate the strength of image encryption algorithms[19]. It is mostly used to evaluate against differential attacks. It is given by the equation

$$N(P,C) = \sum_{i,j} \frac{D(i,j)}{M \times N} \text{ x } 100\ \% \qquad (1)$$

$$where\ D(i,j) = \begin{cases} 0\ if\ C(i,j) = P(i,j) \\ 1\ if\ C(i,j) \neq P(i,j) \end{cases}$$

and P(i, j) and C(i, j) are the plain-image and the corresponding encrypted image of size M x N. A high NPCR means high resistance to differential attacks. And also since this vividly captures the number of pixels changed in an image this can be used as an efficient feature for steganalysis. As Blind Steganalysis cannot have any information regarding the cover image, to derive the NPCR parameter, the cover image needs to be predicted from the stego image at hand. Prediction can always be done by exploiting the fact that the pixels in neighbourhood have high correlation. Thus prediction method of a pixel x is adopted as given by Figure 1.

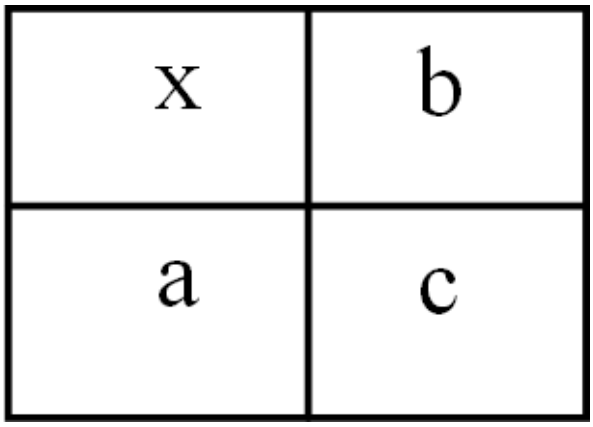

Figure 1. Prediction neighbourhood.

where a is the vertical-down neighbour and b is the horizontal-right neighbor and c is the diagonal-down neighbour. Using these three neighbours the prediction value of x can be brought out as,

$$\hat{x} = \begin{cases} \max(a,b) \text{ if } c \leq \min(a,b) \\ \min(a,b) \text{ if } c \geq \max(a,b) \\ a + b - c \quad \text{otherwise} \end{cases} \qquad (2)$$

This prediction is used as the P. The NPCR value is collected for each image and is a 1D vector. This is a very, very low dimensional feature compared to all those features which aid for Steganalysis existing in literature yet.

2.1.2 Co-Occurrence Features Derived from Different Bit Planes of an Image

Steganographic algorithms differ in their choice of bit plane to embed data. Most algorithms use the least significant bit plane to hide data so that quality of stego image gets maintained. Other algorithms make use of higher bit planes with the help of compensation procedures. To obtain those artifacts, statistical features like mean and variance, shape distribution features like skewness and kurtosis and entropy and co-occurrence[20] features namely correlation, contrast, energy, local homogeneity, cluster shade and cluster prominence are computed from all bit planes.

2.2 Transform Domain Features

2.2.1 Features from Image Detail Subbands (FID)

The main idea in the Discrete Wavelet Transform (DWT) is that a time-scale representation of a digital signal is obtained using digital filtering techniques. The signal is passed through a series of high pass filters to analyze the high frequencies, and it is passed through a series of low pass filters to analyze the low frequencies. The DWT is identical to a hierarchical sub band system where the sub bands are logarithmically spaced in frequency and represent octave-band decomposition. By applying DWT, the image is actually divided, i.e., decomposed into four sub bands and critically sub sampled as shown in Figure 2(a).

The decomposition results in two-dimensional array of coefficients in four bands, each labeled as LL (Low-Low), LH (Low-High), HL (High-Low) and HH (High-High)[1]. The LL band can be decomposed once again in the same manner, to get second level of DWT decomposed sub-bands. The one level, two levels, three levels and four level DWT decomposed sub-bands are shown in Figure 2(a), 2(b), 2(c) and 2(d) respectively.

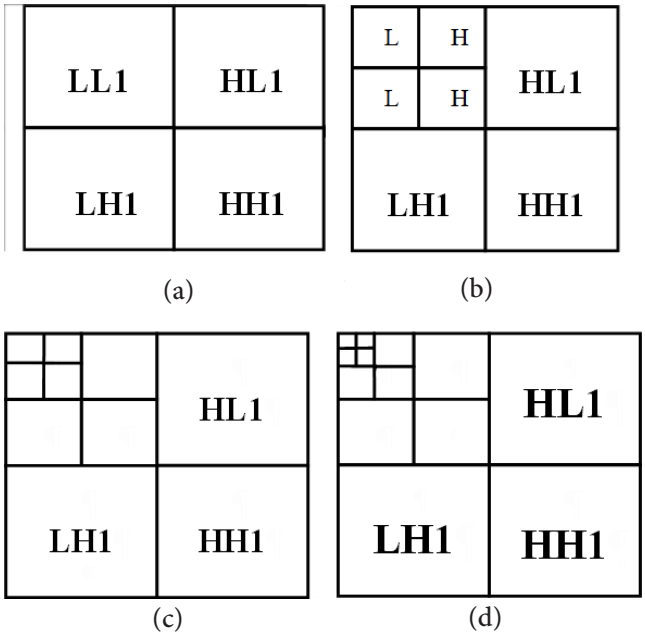

Figure 2. Decomposition of an Image using DWT. (a) One level. (b) Two level. (c) Three level. (d) Four level.

Most natural images have smooth intensity variations, with the fine details are represented as sharp edges in between the smooth variations. Technically, the smooth variations in intensity can be termed as low frequency variations and the sharp variations as high frequency variations. The low frequency components (i.e., smooth variations) constitute the base of image, and the high frequency components (the edge which give the detail) add upon them to refine the image, thereby giving a

detailed image. In general the embedding process done on the edges of the cover images, these embedding distortions are highlighted by using the detailed bands of the image. So that the approximation bands is set to zero and the statistical measures like mean, variance, skewness, kurtosis, entropy and cooccurrence features like correlation, contrast, energy, local homogeneity, shade and prominence are extracted only the detailed bands.

3. Proposed Steganalyzer

Features concentrating on different characteristics of steganographic software are derived and fed to an ensemble classifier. The training phase involves providing sample features from both cover and stego images comprising about 80% of database. Rest of the database, unknown to the steganalyzer is used for testing. Experimentation was carried out using four different approaches for classification. The first two approaches had a classifier built for every class and the classifier will deal with stego images created by one steganographic algorithm whereas the third and fourth approaches learn about all steganographic algorithms by considering samples from each class. Both the techniques have been subjected to two decision rules as indicated in Figures (3) and (4). The steganalyzer as used in the testing phase is shown in Figure 3.

3.1 Ensemble Decision Function

3.1.1 Mode or Majority Rule

Here the output class of the image is decided by the majority voting of decisions made by the group of classifiers on feature yielding maximum result.

$$C = \text{mode}_{j}(\text{argmax}_{i}(D_{j}(I_{fi}))) \quad (3)$$

3.1.2 Max or Maximum Rule

Here the output class of the image is decided by the maximum decision on feature set and maximum by classifier.

$$C = \text{argmax}_{j}(\text{argmax}_{i}(D_{j}(I_{fi}))) \quad (4)$$

where j is the index of classifier, and I is the index of the feature model. Dj is the decision of the jth classifier on ith feature model f on Image I.

4. Results and Discussion

4.1 Data Base

As Steganalysis is targeted over Image data, initially, a collection of images are used to create the image database. From this database, large size images are resized into 512 × 512 to be used as cover images (BMP format) and the images having a dimension less than 256 × 256 are used as secret images. In our system totally 500 Cover images and 500 secret images are used. From the steganographic software collected from internet, five are used namely, Image Protector (IP)[21], Invisible Secrets (IS)[22], Third Eye (TE)[23], S-Tools (ST)[24] and Wb-stego (WB)[25]. Table 1 shows the details of the Steganographic software employed. All the 500 cover images are subjected to five different steganographic algorithms yielding 2500 stego images in total with each image having different and variable size secret images embedded in them. Table 2 also shows the maximum size that can be embedded with different steganographic algorithms.

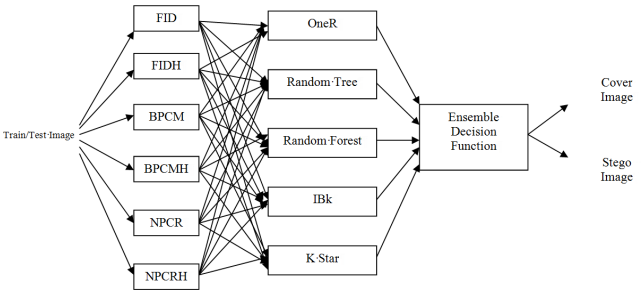

Figure 3. Block diagram for ensemble classifier.

Various Sizes of secret images are embedded in the cover image and the size of secret images as percentage of embedding capacity is shown in the Table 2.

It is important to note from Table 2 that, around 83% of the stego images in the database are embedded with less than 5% of the embedding capacity of the corresponding cover images while around 10% of them are embedded with less than 10% of the embedding capacity. We used 4 different bins based on embedding capacity such as <5%, 5 to 10%, 10 to 25% and 25 to 50%.

The following process was implemented to detect the Steganographic algorithm. From spatial domain as well as the decomposed sub bands after application of DWT, features are extracted and stored in the Features library in the learning phase. The 500 cover images and 2000 stego images (400 stego images for each one of the steganographic software 400 x 5 = 2000) are used for

Table 1. Steganographic software details

Steganographic Software	Carrier file	Secret file	Secret file size	Stego file	Compression
Image Protector (IP)	BMP	Any Format	Max Size 40KB	same as carrier file format	No
Invisible Secrets (IS)	JPEG, PNG, BMP,HTML, WAV	Any Format	Max Size 90KB	same as carrier file format	Yes
Third Eye (TE)	BMP	Any Format	Max Size 75KB	same as carrier file format	No
S-Tools (ST)	BMP & GIF	Any Format	Max Size 97KB	same as carrier file format	No
Wb-Stego (WB)	BMP, TXT, HTML, PDF	Any Format	Max Size 90KB	same as carrier file format	No

Table 2. Secret images size as percentage of embedding capacity

S. No.	Steganographic Algorithm	Max Embedding Capacity	Size of Embedded Secret Images <=5%	5% to 10%	10 to 25%	25 to 50%	Total
1	Image Protector (IP)	40KB	416	52	27	5	500
2	Invisible Secrets (IS)	90 KB	416	52	27	5	500
3	Third Eye (TE)	75 KB	416	51	28	5	500
4	S-Tools (ST)	97 KB	417	51	27	5	500
5	Wb-Stego (WB)	90 KB	416	52	27	5	500
	Total		**2081**	**258**	**136**	**25**	**2500**
	Total in %		**83.24**	**10.32**	**5.44**	**1**	**100**

Table 3. Images used for training and classification

Steganographic Software	No of Images used for Training & Classification							
	<=5%		5-10%		10-25%		25-50%	
	#Train Images	#Test Images	#Train Images	#Test Images	#Train Images	#Test Images	#Train Images	#Test Images
IP	344	72	42	10	13	14	0	5
IS	344	72	42	10	13	14	0	5
TE	344	72	42	9	13	15	0	5
ST	345	72	41	10	13	14	0	5
WB	344	72	42	10	13	14	0	5

training. During classification 500 non-trained stego images are used. When a test image is given as input for the system, it is classified as cover or stego image consulting the Feature library. Randomly chosen 80% of the images are used for training and the remaining unseen images are used for classification as shown in Table 3.

The Detection accuracy obtained for all the four approaches has been given in Table 4. The first three feature sets have been derived from RGB planes and the next three have been derived from HSV domain.

Out of the steganographic tools employed, IS is the easiest to identify. IS leaves a characteristic signature while embedding and its detection has been reported in the Literature by structural means and in contrast, the designed steganalyzer has achieved it statistically. IS is best identified by NPCR feature, TE and ST by BPCM feature, WB by FID feature all derived from HSV domain and IP is best detected by NPCR feature derived from spatial domain. Another achievement by the developed Steganalyzer is that it has achieved a decent detection accuracy with a dimensionality as low as three features even, which is a rare phenomenon in the domain of Steganalysis. The third and fourth approaches which have been designed to improve the generalization of the classifier have also performed well in identifying low volume payloads. The maximum detection accuracy obtained for different algorithms and different approaches has been highlighted in Table 5.

Table 4. Experimentation results - single feature

S. No.	Feature Set (Dimensionality)	Train Tool	Gen- Max	Gen- Maj	Mix - Max	Mix - Maj
1.	FID (99)	IP	56	45.5	52	46.5
		IS	55.5	46		
		TE	55	46.5		
		ST	54.5	44.5		
		WB	56	42		
2.	BPCM (264)	IP	58	51.5	58	52
		IS	91.5	92.5		
		TE	58	53		
		ST	56	45.5		
		WB	57	51.5		
3.	NPCR(3)	IP	63	57.5	55.5	45.5
		IS	57	53		
		TE	55	49		
		ST	50	48.5		
		WB	49	46		
4.	FID – HSV(99)	IP	59	44.5	54	50.5
		IS	66	60.5		
		TE	58	53		
		ST	58	52		
		WB	60	56.5		
5.	BPCM – HSV (264)	IP	58.5	53	61	60.5
		IS	93	89		
		TE	60.5	56.5		
		ST	61.5	52.5		
		WB	56.5	53.5		
6.	NPCR – HSV (3)	IP	54	49	61	55.5
		IS	94.5	94.5		
		TE	57	54.5		
		ST	58.5	52.5		
		WB	49	45		

Table 5. Experimentation results - composite feature

S. No.	Composite Feature Set	Mix - Max	Mix - Maj
1.	BPCM NPCR (267)	57.5	54.5
2.	BPCM FIDH (363)	61.5	56
3.	BPCM BPCMH (198)	61	60
4.	BPCM NPCRH (267)	58.5	51.5
5.	NPCR FIDH (102)	53	51.5
6.	NPCR BPCMH (267)	61	60.5
7.	NPCR NPCRH (6)	61	61
8.	FIDH BPCMH (363)	61	61
9.	FIDH NPCRH (102)	61	55
10.	BPCMH NPCRH (267)	61	61

The features individually contributed for Steganalysis have been combined two at a time to be fed in the designed ensemble steganalyzer and experimentation was carried out for all the four approaches. Composite feature set was not able to enhance the detection accuracy of individual algorithms but able to contribute and marginally improve the detection accuracy in case of the generalization improved version.

Comparison of the developed steganalyzer with the existing methods cannot be done straight away since different authors use different databases, different methods to characterize payloads and also there is a difference in the cover media format as well as steganographic algorithms employed to create stego images. Hence Tables 6 and 7 present the comparison of the proposed method with existing techniques which have specified the embedding capacity in bits per pixel and as a percentage of embedding capacity respectively.

From Table 6, it is evident that the designed Steganalyzer has performed better than all other existing steganalyzers considering the fact that our steganalyzer has worked with raw and uncompressed images. Table 6 also highlights the different definitions available for steganography. Row 5 of Table 6 presents the results obtained by using the SPAM feature on our database provided by author at[26]. Comparable detection accuracy has been obtained in our work with a very low dimensional self built feature set against SPAM which has a dimensionality of 686 features.

Table 6. Comparison with existing works (data embedding rate in bpp)

S. No	Authors	Data Embedding Rates (Bits Per Pixel)	Detection Accuracy in %
		0.01 bpp	39.54
1	Sullivan	0.02 bpp	40.81
		0.05 bpp	44.38
2	Lie and Lin	0.5 bpp	78.78
		0.01 bpp	52.28
3	Zou	0.02 bpp	59.46
		0.05 bpp	75.14
		0.1 bpp	64.36
4	Holub and Fridrich	0.2 bpp	76.03
		0.4 bpp	88.28
5	Pevny, Bas and Fridrich	<=0.05 bpp	63.5
6	Proposed Method	<=0.05 bpp	61.5

Table 7. Comparison with existing works (data embedding rate in %)

S. No	Authors	Data Embedding Rates in %	Detection Accuracy in %
		2%	73
1	Agaian and Cai - BiO9	3%	73
		5%	83.3
		2%	73
2	Agaian and Cai - QMF12	3%	66.7
		5%	86.7
		2%	56.7
3	Agaian and Cai - DB8	3%	56.7
		5%	70
4	Lyu and Farid	5%	1.2
5	Quach et al.	5%	58.46
6	Proposed Method	<=5%	61

Number of altered pixels in %

Table 7 compares the results obtained with similar approaches. Data embedding rate will be generally greater than percentage of altered pixels. In that context, the designed steganalyzer proves superior to state-of-the-art techniques.

5. Conclusion

A generic steganalyzer which effectively detects low volume payloads with a novel, low dimensional feature set has been proposed. The Steganalyzer specially deals with raw and uncompressed image formats and achieves comparable detection accuracy reported in the literature for compressed images. An ensemble classifier has been used along with the novel feature set framed in identifying the otherwise statistically undetectable artifacts. The composite feature set employed enhanced the generalization of the developed Steganalyzer.

6. References

1. Lyu S, Farid H. Steganalysis using Higher-Order Image Statistics. IEEE Transactions on Information Forensics and Security. 2006 Mar; 1(1):111–9.
2. Westfeld A, Pfitzmann A. Attacks on steganographic systems. Dresden, Germany: Proc Third Int Workshop on Information Hiding. 1999.
3. Provos N, Honeyman P. Detecting steganographic content on the internet. Ann Arbor: University Michigan. Tech Rep. 2001.
4. Fridrich J, Goljan M, Hogea D. Steganalysis of JPEG images: Breaking the F5 algorithm. Proc 5th Int Workshop on Information Hiding; Noordwijkerhout, The Netherlands. 2002 Dec.
5. Avcibas I, Memon N, Sankur B. Steganalysis using image quality metrics. IEEE Trans Image Processing. 2002 Feb; 12(2):221–9.
6. Farid H. Detecting hidden messages using higher-order statistical models. Proc International Conference on Image Processing; Rochester, NY. 2002.
7. Fridrich J, Goljan M, Hogea D. New methodology for breaking steganographic techniques for JPEGs. Proc SPIE, Symp Electronic Imaging; Santa Clara, CA. 2003.
8. Fridrich J. Feature-based steganalysis for JPEG images and its implications for future design of steganographic schemes. Proc 6th International Workshop on Information Hiding; Toronto, ON, Canada. 2004.
9. Agaian S, Cai H. Color wavelet based universal blind steganalysis. Proc International TICSP Workshop on Spectral Methods and Multirate Signal Processing; Vienna, Austria. 2004 Sep 11-12. p. 183–9.
10. Shi YQ, et al. Image steganalysis based on moments of

characteristic functions using wavelet decomposition, prediction-error image, and neural network. IEEE International Conference on Multimedia and Expo. 2005 Jul 6–8. DOI:10.1109/ICME.2005.1521412.
11. Zou D, Shi YQ, Su W, Xuan G. Steganalysis based on Markov Model of thresholded prediction-error image. IEEE International Conference on Multimedia and Expo. 2006 Jul 9–12. DOI: 10.1109/ICME.2006.262792.
12. Quach T-T, Perez-Gonzalez F, Heileman GL. Model-based steganalysis using Invariant Features. SPIE Proceedings of Media Forensics and Security. 2009. p. 72540.
13. Bell G, Lee Y-K. A method for automatic identification of signatures of steganography software. IEEE Transactions on Information Forensics and Security. 2010 Jun; 5(2).
14. Zhong S, Liao B, Chen G. Steganalysis against difference expansion based reversible data hiding scheme for 2D vector maps. International Journal of Advancements in computing Technology. 2011 Apr; 3(4).
15. Zong H, Liu F-L, Luo X-Y. Blind image steganalysis based on wavelet coefficient correlation. Digit Investig. 2012 Jun; 9(1):58–68.
16. Cho S, Cha B-H, Gawecki M, Jay Kuo C-C. Block-based image steganalysis: Algorithm and performance evaluation. J Vis Comm Image Represent. 2013 Oct; 24(7):846–56.
17. Holub V, Fridrich J. Random projections of residuals for digital image steganalysis. IEEE Transactions on Information Forensics and Security. 2013 Dec; 8(12):1996–2006.
18. Pathak P, Selvakumar S. Blind image steganalysis of JPEG images using feature extraction through the process of dilation. Digit Investig. 2014 Mar; 11(1):67–77.
19. Kader Mastan JM, Sathishkumar GA, Bhoopathy Bagan K. A color image encryption technique based on a substitution-permutation network. Advances in Computing and Communications. 2011; 193:524–33. DOI:10.1007/978-3-642-22726-4_54.
20. Sun Z, Hui M, Guan C. Steganalysis based on co-occurrence matrix of differential image. Proc International Conference on Intelligent Information Hiding and Multimedia Signal Processing. 2008 Aug 15–17. p. 1097–100. DOI:10.1109/IIH-MSP.2008.176.
21. Available from: http://www.sharesoftware24.com/free-downloads/windows/security-privacy/encryption-tools/info/id-image-protector-2232.html
22. Available from: http://www.invisiblesecrets.com
23. Available from: http://www.securekit.net/index.html
24. Available from: http://www.spychecker.com/download/download-stools.html
25. Available from: http://wbstego.wbailer.com
26. Available from: http://dde.binghamton.edu/download/