Steganalysis on Images using SVM with Selected Hybrid Features of t-test Feature Selection Algorithm

R. Uma Rani
Department of Computer Science
Sri Sarada College for Women
Salem, Tamilnadu, India
umainweb@gmail.com

S. Deepa
Department of Computer Science
Government Arts College
Dharmapuri, Tamilnadu, India
sdeepamca@yahoo.co.in

***Abstract*- Steganography techniques are classified into many categories based on embedding method used. In spatial domain steganography techniques, image pixels values are converted into binary values and some of the bits are changed for hiding secret data. This work attempts to detect the stego images created by Wavelet Obtained Weights (WOW) algorithm by Steganalysis on Images using statistical attack. It is based on the classification of selected Hybrid image feature sets using Support Vector Machine (SVM) and feature selection algorithm through t-test (SVM-HT) with the combined features of Chen Features, Subtractive Pixel Adjacency Mode (SPAM) Features and Cartesian-calibrated Pev [Ccpev] Features. It uses the first 1000 principal features for training and testing. The proposed approach produced results prove low probability of error rate.**

Keywords- Steganalysis, Steganography, SVM-HT, t-test

1. INTRODUCTION

In spatial domain scheme, the secret messages are embedded directly. It embeds information in the intensity of the pixels. The task is to find out some areas or data that can be modified without having any significant effects on this cover file. Therefore, a secret message can be embedded by replacing the redundant or insignificant parts of a cover file with secret message bits, without adding any significant noise to this cover file. The Least Significant Bit (LSB) substitution is the most commonly used spatial domain technique. In LSB substitution technique the least significant bit of each pixel of the cover is replaced by the secret message bits. The image in which secret message is hidden is called as the stego-image.

1.1. Applications of Steganography

1.1.1. Secret Communication

By using Steganography two parties can communicate secretly without anyone knowing about the communication. Cryptography can only encode the message but its presence is not hidden and thus draws unwanted attention. Steganography, thus, on the other hand, hides the existence of message in some cover media. Steganography provides us with

a) Potential capability to hide the existence of confidential data.

b) Hardness of detecting the hidden (i.e., embedded) data.

c) Strengthening of the secrecy of encrypted data.

1.1.2. Copyright Protection

This is basically related to watermarking i.e., a secret message is embedded in the image which serves as the watermark and thus identify it as an intellectual property which belongs to a particular owner.

1.1.3. Feature Tagging

Features such as captions, annotations, name of the individuals in a photo or location in a map can be embedded inside an image. Copying the stego image also copies all of the embedded features and only parties who possess the decoding stego key will be able to extract and view the features.

1.1.4. Digital Watermarking

This is one of the most important applications of Steganography. It basically embeds a digital watermark inside an image. Digital watermarks may be used to verify the authenticity or integrity of the carrier

signal or to show the identity of its owners. It is prominently used for tracing copyright infringements and for banknote authentication.

1.1.5. Other Applications

Steganography is widely used in areas such as Military, Banking and Market Applications to provide secure communication between the parties. In Industries, Steganography is widely used as a mechanism to prevent piracy. It is also used in biometrics for providing secure and robust biometrics system.

1.2. Drawback of Steganography

Steganography at a large scale can also be used by terrorists, who hide their secret messages in innocent, cover sources to spread terrorism across the country. Rumours were spread about terrorists using steganography when the two articles titled "Terrorist instructions hidden online" and "Terror groups hide behind Web encryption" were published in newspaper. Other media worldwide cited these rumours many times, especially after the terrorist attack of 9/11, without ever showing proof.

2. LITERATURE REVIEW

J. Fridrich et al. (T. Filler, 2010), proposed Wavelet Obtained Weights (WOW) Algorithm that defines additive steganographic distortion in the spatial domain. Additive in the sense the distortion does not consider the effects of individual embedding changes influencing each other. A bank of directional filters is employed to obtain the directional residuals, which assess the content around each pixel along multiple directions. The changes in the residuals caused by embedding are measured and aggregated forcing the embedding cost of a pixel to be high where the content is predictable in at least one direction (smooth areas and along edges) and low where the local content is unpredictable in any direction (e.g., in textured or noisy areas). If the residual values are large for a pixel in all directions, it means that the local content at that pixel is not smooth in any direction. So the embedding prefers changing large values of directional residuals, where the textures and edges are, and preserve the small values, where the content is predictable.

Chen features are Markov features presented by (C.Chen, 2008) and (Y.Q. Shi, 2006), utilizing both intra block and inter block correlations among JPEG coefficients. These are DCT domain features proposed by those authors. It computes transition probability matrix for each difference JPEG 2-D array to utilize the intrablock correlation, and "averaged" transition probability matrices for those difference mode 2-D arrays to utilize the interblock correlation. All the elements of these matrices are used as features for steganalysis. This algorithm gives 486 features that can be used for steganalysis.

SPAM is a method for detection of steganographic methods that embed in the spatial domain by adding a low-amplitude independent stego signal proposed by (Tomas Pevny P. &., 2010). It focuses on evaluation of detection of Least Significant Bit (LSB) matching. First, arguments are provided for modeling the differences between adjacent pixels using first-order and second-order Markov chains. Subsets of sample transition probability matrices are then used as features for a steganalyzer implemented by support vector machines. The steganalyzer is constructed as follows. A filter suppressing the image content and exposing the stego noise is applied. Dependences between neighboring pixels of the filtered image (noise residuals) are modeled as a higher-order Markov chain. The sample transition probability matrix of a higher-order Markov model is then used as a feature vector for steganalysis. The major contribution of the work is the use of higher-order Markov chains, exploiting of symmetry in natural images to reduce the dimensionality of the extracted features. This algorithm extracts spatial domain SPAM features of dimensionality 686 that can be used for steganalysis. It also detects steganography in the transform domain.

Tomas Pevny et al. (Tomas Pevny J., 2007) also devised a Cartesian-calibrated Pev [CcPev] feature set model. It contains 548 features. The feature set was obtained by merging and modifying two previously proposed feature sets with complementary performance (the DCT feature set that captures inter-block dependencies among DCT coefficients and Markov features which capture intra-block dependencies). Although calibration was originally introduced for the JPEG domain, there were attempts to use this powerful concept in the spatial domain as well. In fact, the image obtained using the predictor in WS steganalysis can also be considered as a reference image even though it was not formulated within the framework of calibration.

Calibration was credited with increasing the features sensitivity to embedding while decreasing their image-to-image variations. Indeed, when the pay-load is small, the best estimate of the cover image features are the features derived from the stego image itself.

3. THE ATTACK ON STEGANOGRAPHY

Attacks can be of several types for example, some attacks merely detect the presence of hidden data, some try to detect and extract the hidden data, some just try to destroy the hidden data by finding the existence without trying to extract hidden data and some try to replace hidden data with other data by finding the exact location where the data is hidden.

3.1. Different Approaches of Steganalysis

3.1.1. Visual attacks

By analyzing the images visually, like considering the bit images and try to find the difference visually in these single bit images.

3.1.2. Structural attacks

The format of data file often changes as the data to be hidden is embedded, identifying these characteristic structural changes can detect the existence of image, for example in palette based steganography the palette of image is changed before embedding data to reduce the number of colours so that the adjacent pixel colour difference should be very less. This shows that groups of pixels in a palette have the same colour which is not the case in normal images.

3.1.3. Statistical attacks

In these type of attacks the statistical analyses of the images by some mathematical formulas is done and the detection of hidden data is done based on these statistical results. Generally, the hidden message is more random than the original data of the image thus finding the formulae to know the randomness reveals the existence of data.

4. IMPLEMENTATION OF THE SVM-HT STEGO IMAGE DETECTION

4.1. Feature Selection Algorithm

Many factors affect the success of machine learning on a given task. The representation and quality of the instance data is first and foremost. If there is much irrelevant and redundant information present or noisy and unreliable data, then knowledge discovery during the training phase is more difficult. In real-world data, the representation of data often uses too many features, but only a few of them may be related to the target concept. There may be redundancy, where certain features are correlated so that is not necessary to include all of them in modelling; and interdependence, where two or more features between them convey important information that is obscure if any of them is included on its own.

Generally, features are characterized as:

a) Relevant: These are features which have an influence on the output and their role cannot be assumed by the rest.
b) Irrelevant: Irrelevant features are defined as those features not having any influence on the output, and whose values are generated at random for each example.
c) Redundant: A redundancy exists whenever a feature can take the role of another (perhaps the simplest way to model redundancy).

4.2. Feature Selection using t-test

The proposed algorithm uses t-test as a statistical attack on WOW stego images. The t-test is a parametric test that compares the location parameter of two independent data samples. It is used to determine if two sets of data are significantly different from each other. The formula for the t-test is provided below:

$$\frac{\mu_E - \mu_C}{\sqrt{\frac{\text{var}_E}{N_E} + \frac{\text{var}_C}{N_C}}} \quad (1)$$

In equation (1), μ_E and μ_C are the sample means of the data sets E & C, var_E and var_C are the variances of the data sets and N is the sample size. The above equation gives the t-statistic.

The appropriate degree of freedom is given by:

$$\frac{\left(\frac{\text{var}_E}{N_E} - \frac{\text{var}_C}{N_C}\right)^2}{\frac{\left(\frac{\text{var}_E}{N_E}\right)^2}{N_E - 1} + \frac{\left(\frac{\text{var}_C}{N_C}\right)^2}{N_C - 1}} \quad (2)$$

4.3. The Proposed SVM with Hybrid features of t-test (SVM-HT)

The proposed method attempt to attack WOW stego images because, WOW algorithm is claiming it as a strong stego algorithm. Further, it provided a mid range performance during previous evaluation on modern spatial domain steganography algorithms [2016].

4.3.1. Steps of SVM-HT classification Method

a) Input : WOW Stego Images and Non Stego Images
b) Extract Chen-486, Spam-686 and Ccpev-548 Features of Non-Stego Images and Stego Images at Different BitsPerPixel (0.2 bpp, 0.4 bpp, 0.6 bpp, 0.8bpp)
c) It results in 3 set of features for Non stego Images and 4 set of features with stego images at 4 level of hiding for every feature extraction method.
d) For SVM-HT classification, combine the chen-486, Spam-686 and Ccpev-548 features of the non-stego image (from step b) and the chen-486, Spam-686 and Ccpev-548 features of stego images at 4 level of hiding
e) Reduce the dimension of data (1720 features) using t-test feature selection algorithms and only select the first 1000 principal features from the combined feature dataset.
f) For k=1 to 10
g) Train the SVM neural network with randomly selected 70% of data mentioned in step e
h) classify the remaining 30% of data using the trained SVM network of step g
i) Performance(k) = Estimate the Performance()
j) End
k) Find average performance from Performance(k)

5. THE RESULTS OF STEGANALYSIS AND DISCUSSION

5.1. Image Database

The Images used for this evaluation were originally taken from the BOWS Image Dataset. BOWS (Break Our Watermarking System) was a Contest organized within the activity of the Watermarking Virtual Laboratory (Wavila) of the European Network of Excellence ECRYPT. In fact, the original dataset contains 10,000 images. But the proposed system uses a subset of cover images from BOWS database that were previously used in another work named “Gibbs Construction in Steganography” (T. Filler, 2010). This system uses around 500 images to evaluate the performance of the proposed steganalysis model. The cover images feature sets are extracted using three different feature extraction algorithms and stego images feature sets extracted using three different feature extraction algorithms at 4 different level of hiding such as 0.2 bpp, 0.4 bpp, 0.6 bpp and 0.8bpp.

5.2. Performance of the Classifier or Stego Detection System

Table 1 shows the numerical output of the performance of the classifier in terms of different metrics.

The previous work "steganalysis on images based on the classification of image feature sets using SVM classifier [2016]" evaluated the performance of three state of the art Feature Extractors for Steganalysis using Support Vector Machine (SVM) namely SVM-chen, SVM-ccpev, SVM-spam. The proposed SVM-HT is compared with those models.

Table : 1 Performance of SVM-HT (First 1000 Feature provided by t-test Algorithm)

Iteration	Precision	F_Score	Sensitivity	Specificity	Accuracy	Error Rate
1	100.00	100.00	100.00	100.00	100.00	0.00
2	100.00	100.00	100.00	100.00	100.00	0.00
3	100.00	100.00	100.00	100.00	100.00	0.00
4	97.06	98.51	100.00	90.91	97.67	2.33
5	100.00	100.00	100.00	100.00	100.00	0.00
6	97.37	98.67	100.00	85.71	97.67	2.33
7	97.14	98.55	100.00	90.00	97.67	2.33
8	94.59	97.22	100.00	77.78	95.35	4.65
9	97.37	98.67	100.00	85.71	97.67	2.33
10	100.00	100.00	100.00	100.00	100.00	0.00
AVG	**98.35**	**99.16**	**100.00**	**93.01**	**98.60**	**1.40**

The following graph shows the performance of the stego image classifier or stego image detection system in terms of Error Rate. As shown in the Figure 1, the proposed SVM-HT provided excellent performance than other three proposed models.

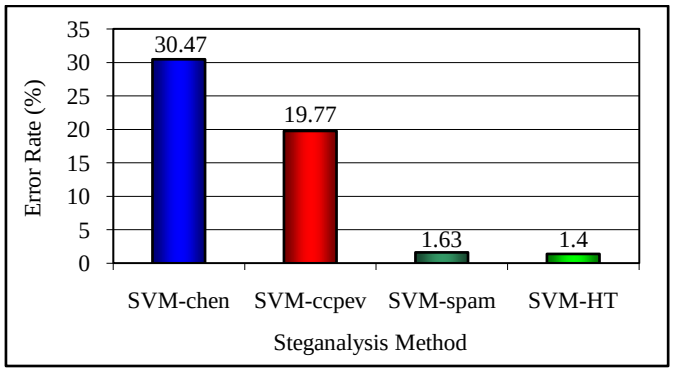

Figure 1. Performance in Terms of Error Rate

The following graph shows the performance of the stego image detection system in terms of Accuracy. As shown in the Figure 2, the proposed SVM-HT model provided excellent performance than other three proposed models.

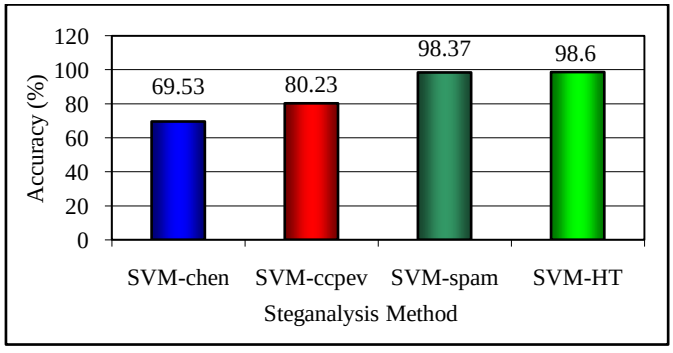

Figure 2. Performance in Terms of Accuracy

The following graph shows the performance of the stego image detection system in terms of Sensitivity. As shown in Figure 3, the proposed SVM-HT model provided excellent performance than other

three proposed models. Here high value of sensitivity case of SVM-HT signifies that the system was able to classify all the non-stego images correctly with 100% accuracy.

The following graph shows the performance of the stego image detection system in terms of Specificity. As shown in the Figure 4, the proposed SVM-HT model provided excellent performance than other three proposed models. Here high value of specificity in the case of SVM-HT signifies that the system was able to classify all the-stego images correctly with high accuracy.

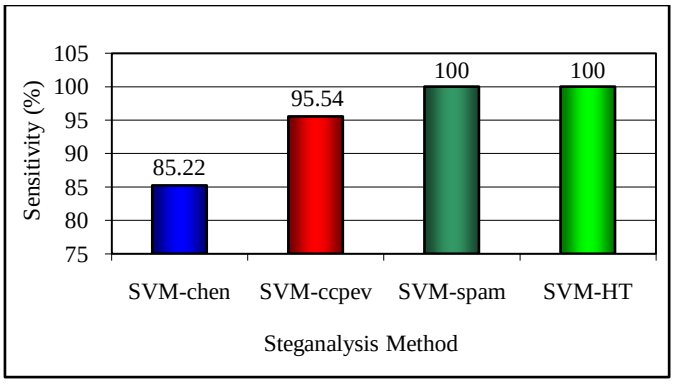

Figure 3. Performance in Terms of Sensitivity

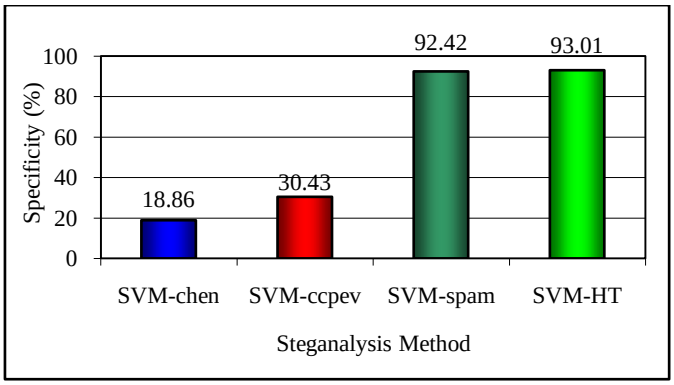

Figure 4. Performance in Terms of Specificity

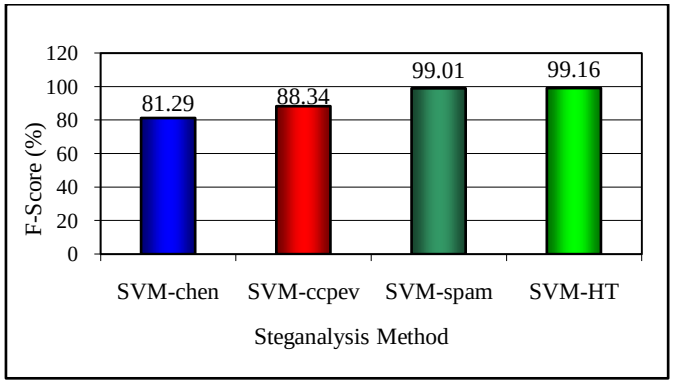

Figure 5. Performance in Terms of F-Score

The above graph shows the performance of the stego image detection system in terms of F-Score. As shown in the Figure 5, the proposed SVM-HT model provided excellent performance than other three proposed models. Here high value of F-Score in the case of SVM-HT signifies that the system was able to classify all the stego images as well as non-stego images with high accuracy.

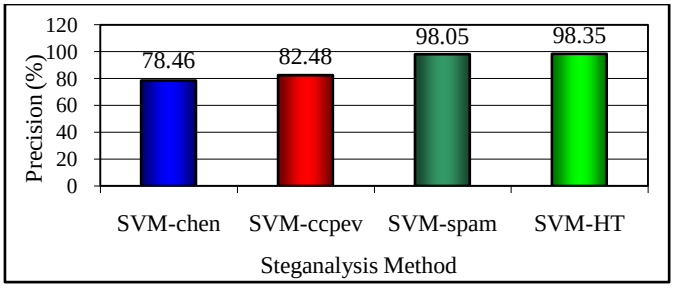

Figure 6. Performance in Terms of Precision

The above graph shows the performance of the stego image detection system in terms of Precision. As shown in the Figure 6, the proposed SVM-HT model provided excellent performance than other three proposed models. Here high value of Precision in the case of SVM-HT signifies that the system was able to classify all the stego images with high accuracy.

5.3. Comparison of Performance with Previous Methods

In the following table and graph, the results of the compared algorithms (1) Ensemble classifier, (2) FLD classifier, (3) Ridge Regression, (4) LSMR Optimization and (5) LASSO were taken from the paper "Is Ensemble Classifier Needed for Steganalysis in High-Dimensional Feature Spaces? (Remi Cogranne, 2015)". Table 2 shows the performance of proposed methods and previous methods in terms of probability of error.

Table : 2 Performance in Terms of Probability of Error

Sl. No.	Steganalysis Method	Probability of Error
1	Ensemble classifier	0.3196
2	FLD classifier	0.3289
3	Ridge Regression	0.3402
4	LSMR Optimization	0.3267
5	LASSO	0.3694
6	SVM-chen	0.3047
7	SVM-spam	0.0163
8	SVM-ccpev	0.1977
9	SVM-HT	0.0140

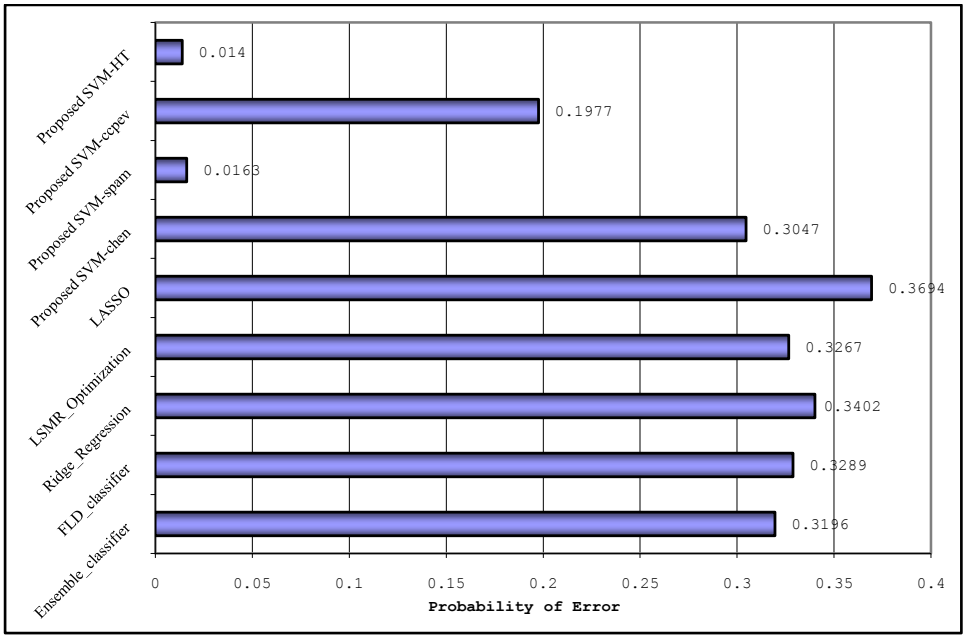

Figure 7. Performance in Terms of Probability of Error

The above graph shows the performance of proposed methods and previous methods in terms of probability of error. As shown in this graph, SVM-spam performed better than all the previous methods. But the performance of SVM-HT was very good and it provided very lower probability of error. The improvement in performance of the proposed model is based on four important aspects.

a) SVM neural network based classifier.
b) The mixed class stego image features with different bpp hiding for training the SVM neural network.
c) The use of combined extracted features from three state of the art feature extraction algorithms.
d) The use of t-test feature selection algorithm provided significant features.

Application: The algorithm can be applied in suspicious stego image communication by terrorists.

6. CONCLUSION

In this paper, a new Steganalysis method based on the selected hybrid features of t-test feature selection algorithm using support vector machine classifier is proposed. The proposed method successfully implements a Matlab framework for doing steganalysis on WOW based stego images. This paper uses the t-test feature selection algorithm and selected 1000 significant features from the combined feature set and used it for the training and testing of SVM classifier. The proposed approach is examined on most popular and publicly available BOWS image dataset. The proposed steganalysis method SVM with Hybrid features of t-test (SVM-HT) performed better than all the earlier methods. The proposed approach (SVM-HT) produced better performance in decrease in Error Rate (1.40) and increase in Accuracy (98.60), Specificity (93.01), F_Score (99.16) and Precision (98.35) of detectability of stego images. Future work is to design the Steganalysis method applied in HD images and movable edited pictures.

REFERENCES

C. Chen, (2008). JPEG image steganalysis utilizing both intrablock and interblock correlations. International Symposium on Fircuits and Systems. USA: IEEE.

Remi Cogranne, V. S. (2015). Is Ensemble Classifier Needed for Steganalysis in High-Dimensional Feature Spaces. Internatinal Workshop on Information Forensics and Security. USA: IEEE.

S. Deepa, (2016). An Evaluation on Modern Spatial Domain Steganography Slgorithms. International Journal of Computer Science Engineering and Information Technology Research, 43-52.

S. Deepa, (2016). Steganalysis on Images based on the classificationof Image Feature Sets using SVM Classifier. International Journal of Computer Science and Engineering, 15-24.

T. Filler, J. (2010). Gibbs Construction in Steganography. Transactions on Information Forensiccs and Security . USA: IEEE, 705-720.

Tomas Pevny, (2007). Merging Markov and DCT features for multiclass JPEG Steganalysis. Security, Steganography, Watermarking of Multimedia Contents IX. USA: SPIE.

Tomas Pevny, (2010). Steganalysis by Subtractive Pixel Adjacency Matrix. Transactions on Information Forensics and Security. USA: IEEE, 215-224.

V. Holub, J. (2012). Designing steganographic distortion using directional filters. WIFS. Tenerife, Spain: IEEE.

Y.Q. Shi, (2006). A Markov process based approach to effective attacking JPEG steganography. Information hiding, 8th International Workshop. Verlag, New York: Springer, 249-264.