Universal Blind Steganalysis of Parity and Substitution Based Image Steganography Algorithm

Chiragkumar B. Patel, Devangkumar U. Shah, Anjulkumari Prajapati

***Abstract*—In the era of digitalization, most of the data is transmitted through the internet. Digital data security is a prominent issue for this century. There are variety of techniques used to protect data during communication. Steganography is an art of hiding the existence of a message between sender and intended beneficiary. Steganography has been used to conceal secret information in various types of files, such as digital images, audio and video. The nature of steganography leaves evident traces within the stego medium. This allows us to expose that a secret communication is taking place using steganalysis techniques. Universal blind steganalysis is prominent technique which is used to detect existence of message in stego medium. It can be considered as a classification problem to detect correct class of tested image. Accurcy of classification and selection of features set for classification are important factor in universal blind steganalysis. In this paper, it is used to detect concealed message in stego image which is generated using newly developed pixel parity and substitution based image steganography algorithm. To evaluate performance classifier accuracy, confusion matrix and ROC curve are used as security assessment matrix. Results proves that performance of pixel swapping and parity based image steganography algorithm is *"stupendous"* in terms of security.**

***Index Terms*—Image Steganography, Universal blind steganalysis, Confusion matrix, ROC, Support Vector Machine, SPAM.**

I. Introduction

THE nature of steganography leaves evident traces within the stego medium. This allows us to expose that a secret communication is taking place using steganalysis techniques. It is also referred to as a warden [1]. In general, there are two types of warden: active and passive. A passive warden only monitors the communication channel to know the presence of hidden messages in media. The warden does not change the content of the media used for communication. Conversely, an active warden may commence distortion to destroy media or interrupt the communication even if there is no indication of secret communication. Currently, most of steganographic methods are premeditated for the passive warden circumstances.

Generally, there are two types of steganalysis: blind and targeted [2]. Targeted steganalysis is intended to attack on one particular steganography algorithm. Although, targeted steganalysis can produce more precise results. Conversely, blind steganalysis is used for detecting different types of steganography algorithm because it can detect a wider class of steganography techniques. Hence, it also known as *universal blind steganalysis*. Although, blind steganalysis is generally less precise; it is used to detect newly developed steganography techniques. Ultimately, blind steganalysis is a unique detection tool if the steganography algorithm is not known or newly developed [3].

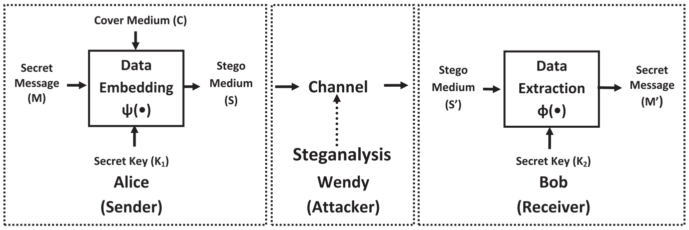


Fig. 1. Steg analysis framework

Steganalysis is often modeled by prisoners problem [3] as illustrated in figure 1. *Alice* and *Bob* are two prisoners and their covert communication is monitored by warden *Wendy*. Using message embedding function $\Psi(\cdot)$ secrete message M suppose to be hidden in cover medium C by Alice with secret key K_1. The resultant stego medium S can describe as a $S = \Psi(C, M, K_1)$. At the receiver side, Bob use message decoding function $\phi(\cdot)$ to decode secrete message from stego medium S' with secret key K_2. The decode message M can well describe by mathematical equation $M' = \phi(S', K_2)$. The steganography system assure that $M' = M$ for error free communication between *Alice* and *Bob*. For public key steganography scheme $K_1 = K_2$ and $K_1 \neq K_2$ for privet key steganography scheme. Once *Wendy* can differentiate stego medium S and cover C, the steganography method is consider broken. Universal blind steganalysis technique use only passive warden scenario for steganalysis. Hence $S = S'$ in steganalysis framework.

II. Related Work

A. Parity and Substitution Based Image Steganography Algorithm

Parity and substitution based image steganography algorithm is basically spatial domain algorithm. Hence, it can provide higher embedding capacity because pixel intensity value is used to embed message. As commonly known, the edges

Mr. Chiragkumar B. Patel is a Ph.D. research scholar of C. U. shah University. In addition to that, He is working as an assistant professor in SAL Engineering and Technical Institute which is affiliated with Gujarat Technological University. Email: chiragkumar@live.com

Dr. Devangkumar U. Shah is working as a Principal in SAL Engineering and Technical Institute which is affiliated with Gujarat Technological University. E-mail: devang.shah@sal.edu.in.

Ms. Anjulkumari Prajapati is graduate student of Gujarat Technological University.

in the image are complex to model, and the edges belonging to each selected area for message embedding are considered as noisy area for embedding. It is seen that embedding in edge pixels of image leads to changes in edges of the stego image. Thus, quality of the image is reduced. To improve quality of image as well as message payload capacity, area for message embedding is identify before message embedding using surrounding pixels intensity value difference. In addition to that, to provide higher security use of cryptography techniques is obvious. Furthermore, random pixel block selection for message embedding using pseudorandom generator and use of stego key boost the security of hidden message.

Message bits are embedded in the cover image using parity and substitution based image steganography algorithm in following steps [4].

1) Read the entire message from text file.
2) Convert message's character stream in binary data stream using *UNICODE* conversion. Let M is a n-bit secret message represented as
$$M = \{m_i | 0 \leq i \leq n, m_i \in \{0, 1\}\} \qquad (1)$$
3) Enter the 16-bit stego key K.
4) Generate cipher binary data stream after encoding binary data stream using *AES* encoder and stego key. M' is cipher data stream generated after encryption
$$M' = \{m'_i | 0 \leq i \leq n, m_i \in \{0, 1\}\} \qquad (2)$$
5) Acquire the gray scale or color image as cover object. Let C is the cover image which has a total $N_C \times M_C \times L_C$ pixels represented as
$$C = \{X_{ijk} | 0 \leq i \leq M_C, 0 \leq j \leq N_C, 0 \leq k \leq L_C\} \qquad (3)$$
where, $X_{ijk} \in \{0, 1, ...255\}$
6) Find out 2×2 non overlapping pixels blocks which are most suitable for message embedding and generate 2×2 pixels block location database D where message bits probably can embed based on their pixel value difference. Set the $Threshold_{min}$ value in such way that, enough 2×2 pixels block location generated to embed secret message.
$$D = \{D_a | 0 \leq a \leq l\} \qquad (4)$$
where, l is a length of message database.
7) Randomly select location D_a in the range $0 < a < l$ from location database D for 8 bit message message embedding.
8) Read 8 bits of cipher binary data $X = x_7x_6x_5x_4x_3x_2x_1x_0$ from cipher stream for data embedding.
9) Select 4 bits of cipher binary data stream $x_3x_2x_1x_0$ and embed using pixel value difference embedding algorithm.
10) Select another 4 bits of cipher binary data stream $x_7x_6x_5x_4$ and changed the least significant bit (LSB) value of pixel $P(x, y)$, $P(x, y + 1)$, $P(x + 1, y)$ and $P(x + 1, y + 1)$ respectively using pixel parity based message encoding algorithm.
11) Repeat step 7, 8, 9, 10 until all cipher binary data stream of message and 16-bit stop flag 1111111111111111 has been embedded.
12) Save image data as a stego image(S) and transmit over channel.

B. Universal blind steganalysis

Universal blind steganalysis can be considered as a classification problem. In general, classification is dividing a set of many possible object into disjoint subsets where is subset forms a class. Usually, the pattern recognition techniques are used to resolve classification problem. Pattern recognize techniques are used to identify complex pattern form given training sample and making intelligent decisions for testing sample.

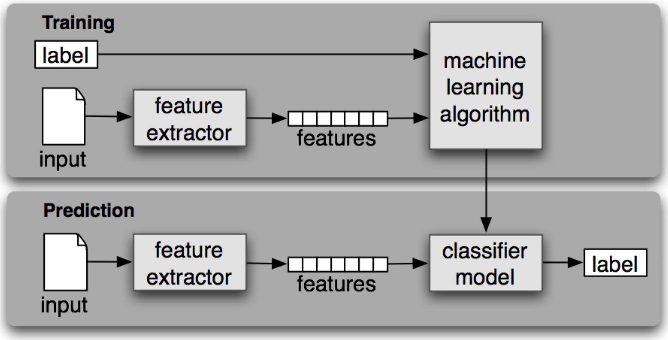


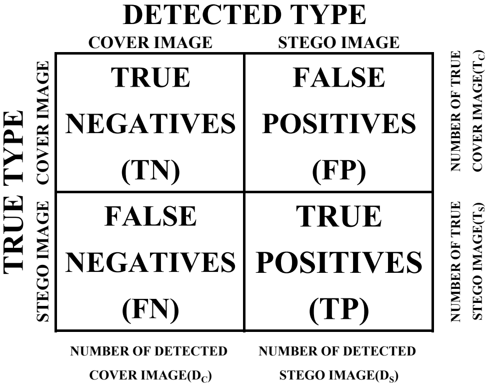
Fig. 2. SVM classifier system

Universal blind steganalysis examines the characteristics of images which is used for training purpose and determine whether this characteristics exhibit as abnormality in test image. As a result of that, classifier should be able to decide the class cover or stego in which the test image belongs. Hence the problem of blind steganalysis can be considered as a binary classification problem. The general framework of the blind steganalysis is shown in figure 2.

In early stage of blind steganalysis research, image quality metrics, moments of image statistic histograms and wavelet decompositions are used as features. More recent features include markov empirical transition matrix, co-occurrence matrix and moment of image statistic from spatial and frequency domains. The features used for blind image steganalysis in this decade available on Binghamton University's website. Among all of available feature extractor source code, SPAM features [5] set is used in this experiment. SPAM feature set is selected because SPAM feature set is highly precise and more appropriate for support vector machine classifier.

Classifiers categories test images into cover or stego image based on their feature vectors. The elementary classification used in universal blind steganalysis is two class supervised learning or machine learning. In machine learning, a set of training samples (image) consist input features and their class labels feed in to classifier for training purpose. Once classified is trained it works as an intelligent machine which can categorize the probably correct class of test (unknown) image based on the given features of test image. For practical steganalysis, main intend to identity the testing medium belongs to stego class or the cover class. When applying image steganalytic method to n image data set of cover image and stego image for detection, There are four possible situations, leftmargin=*

1) Stego medium is correctly detected as stego and it is referred as *True Positive*(TP)
2) Stego medium is incorrectly detected as cover and it is referred as *False Negative*(FN).
3) Cover medium is correctly detected as cover and it is referred as *True Negative*(TN).
4) Cover medium is incorrectly detected as stego and it is referred as *False Positive*(FP).



	COVER IMAGE	STEGO IMAGE	
COVER IMAGE	TRUE NEGATIVES (TN)	FALSE POSITIVES (FP)	NUMBER OF TRUE COVER IMAGE(T_C)
STEGO IMAGE	FALSE NEGATIVES (FN)	TRUE POSITIVES (TP)	NUMBER OF TRUE STEGO IMAGE(T_S)
	NUMBER OF DETECTED COVER IMAGE(D_C)	NUMBER OF DETECTED STEGO IMAGE(D_S)	

Fig. 3. The confusion matrix

The results of test are represented in form of 2×2 matrixas as shown in figure and it is called *Confusion Matrix* as shown in figure 3. Based on confusion matrix some evaluation matrix can be defined as mention below.

$$\text{True Positive Rate}(TPR) = \frac{TP}{TP + FN} \tag{5}$$

$$\text{False Negative Rate}(FNR) = \frac{FN}{FN + TP} \tag{6}$$

$$\text{False Positive Rate}(FPR) = \frac{FP}{FP + TN} \tag{7}$$

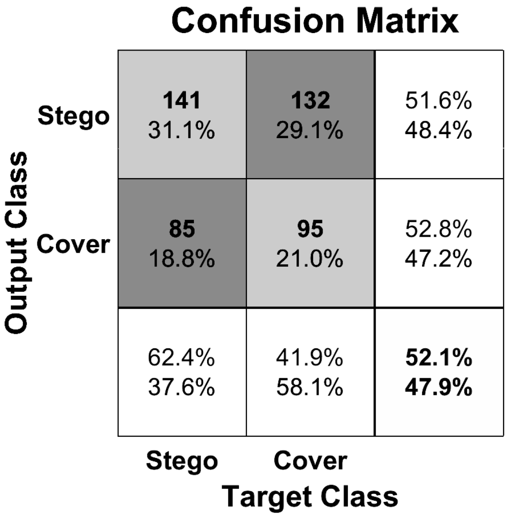
$$\text{True Negative Rate}(TNR) = \frac{TN}{FP + TN} \tag{8}$$

$$\text{Accuracy} = \frac{TP + TN}{TP + FP + FN + TN} \tag{9}$$

$$\text{Precision} = \frac{TP}{TP + FP} \tag{10}$$

III. Performance Evaluation

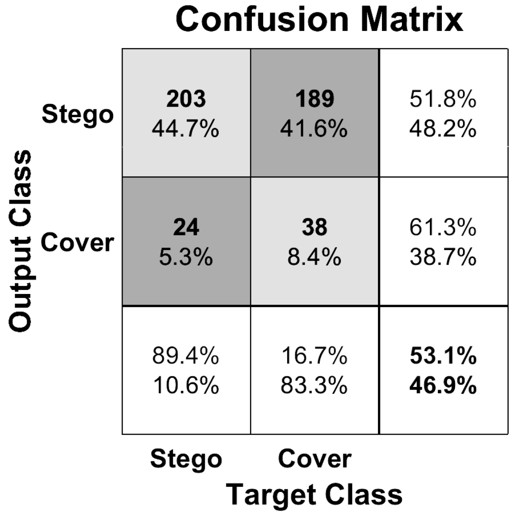
For experimental evaluations of the steganography method, the cover and stego image databases with different sizes are used. The message embedding rate is chosen to generate stego image as per the requirement of the experiment. Prior to the experiments, using MATLAB program the cover images and stego images datasets are divided into training and a testing set of equal sizes, with the same number of cover and stego images. Thus, it is ensured that images in the testing set were not used in any form during the training process or conversely. There is a SVM based classifier use with a gaussian kernel. This classifier must be adjusted to provide optimal results. In addition to that, the SPAM features [5] extraction method is using to extract image features and those features set will be used by SVM classifier. The SVM classifier produces output in form of confusion matrix. For experiment, stego images dataset is generated using 0.01 bpp, 0.05 bpp, 0.1 bpp and 0.2 bpp message embedding rates. Each stego images dataset is jointly used with cover images dataset for classification. Output produced by the SVM classifier for different message embedding rate are shown in figure 4, 5, 6 and 7.



	Stego	Cover	
Stego	**141** 31.1%	**132** 29.1%	51.6% 48.4%
Cover	**85** 18.8%	**95** 21.0%	52.8% 47.2%
	62.4% 37.6%	41.9% 58.1%	**52.1%** **47.9%**

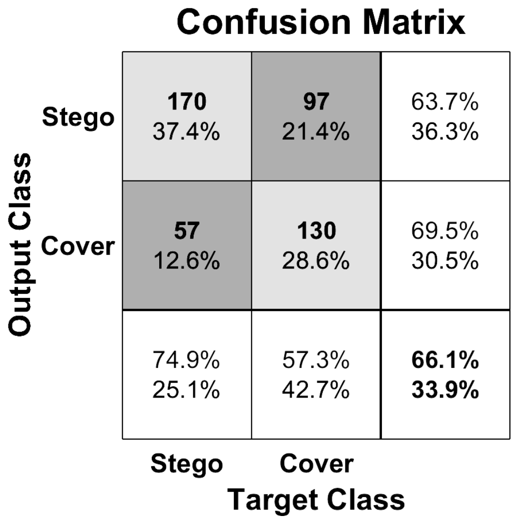
Fig. 4. Confusion matrices of proposed algorithm for 0.01 bpp message payload

It is clearly seen that, true positive rate and true negative rate is smaller for lower message embedding rate. In other words, it can be said that false positive rate and false negative rate is higher for message embedding rate 0.01 bpp and 0.05 bpp. So, SVM classifier does not classify correct class of stego or cover image. Furthermore, classification accuracy is close to 50% or random guessing probability. As message embedding rate is increased true positive rate and true negative rate is increase. Hence, classifier predicated class of image is same to original class of image. Although, in steganography detection system it is highly important not to misclassifying a stego image as a clean image (false positive) as compared to misclassifying a clean image as a stego image (false negative). The consequences of a false positive would be very expensive


Confusion Matrix
Output Class
Target Class

	Stego	Cover	
Stego	**203** 44.7%	**189** 41.6%	51.8% 48.2%
Cover	**24** 5.3%	**38** 8.4%	61.3% 38.7%
	89.4% 10.6%	16.7% 83.3%	**53.1%** **46.9%**

Fig. 5. Confusion matrices of proposed algorithm for 0.05 bpp message payload


Confusion Matrix
Output Class
Target Class

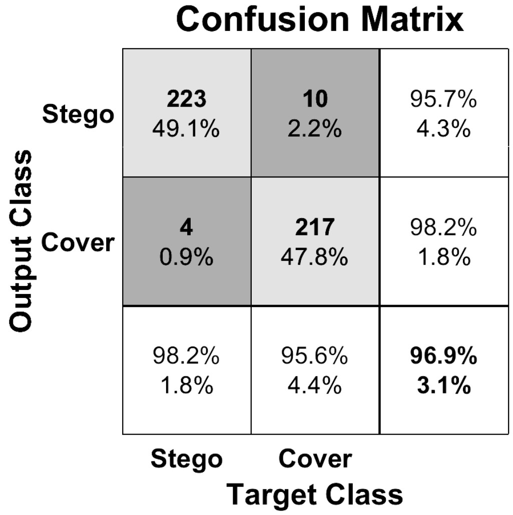
	Stego	Cover	
Stego	**170** 37.4%	**97** 21.4%	63.7% 36.3%
Cover	**57** 12.6%	**130** 28.6%	69.5% 30.5%
	74.9% 25.1%	57.3% 42.7%	**66.1%** **33.9%**

Fig. 6. Confusion matrices of proposed algorithm for 0.1 bpp message payload

from the security point of view. As shown in figure 6, for message embedding rate 0.1 bpp overall classification accuracy is 66.10% which is good indication for proposed steganography algorithm. In other words, when 52000 message bits are embedded in the image SVM classifier does not detect presence of messier bits in image because lower accuracy of the classifier is not consider as good performance of the classifier after training. It is also true that, once message embedding rate is increase over all accuracy of classifier is also increased it can be easily seen in figure 7.

It is worth to say proposed steganography algorithm performing stupendous based on above discussion. Comparison with standard algorithms is required to prove it. Table I present comparison of proposed algorithm with F5 [7], MB1 [3]and Outguess [6]. Here comparison performed with frequency domain algorithm instead of spatial domain algorithm. It is widely known that, spatial domain algorithm provide higher payload capacity and frequency domain algorithm provide higher security against message detection attack. Past results shows that proposed algorithm has a higher payload capacity as compared to other algorithm. Now, it is time to prove that proposed algorithm provided better security against message detection attack. Statics of table I shows that for 0.05 bpp and 0.1 bpp message embedding rates proposed algorithm provide better security compare to F5 [7], MB1 [3] and Outguess [6] algorithm. For 0.5 bpp performance is proposed algorithm is close to MB1 [3]and Outguess [6] algorithm but F5 [7] algorithm provide better security for this case.

Confusion matrix summarizes overall performance of binary classifier in a digit of classification accuracy, which is a good thing. But that is not enough to say any data security algorithm perform better terms of security. For instance, when cover images class is more prevalent than stego images class, the confusion matrix produce biased output. Furthermore, digit of classification accuracy hides information about how the SVM classifier performs on the individual stego image class


Confusion Matrix
Output Class
Target Class

	Stego	Cover	
Stego	**223** 49.1%	**10** 2.2%	95.7% 4.3%
Cover	**4** 0.9%	**217** 47.8%	98.2% 1.8%
	98.2% 1.8%	95.6% 4.4%	**96.9%** **3.1%**

Fig. 7. Confusion matrices of proposed algorithm for 0.2 bpp message payload

TABLE I
SVM CLASSIFIER DETECTION RATE COMPARISON OF PROPOSED ALGORITHM WITH OTHER ALGORITHMS

Rate	Algoritham	TP	FN	FP	TN	Accuracy
0.05 bpp	Outguess [6]	90.1	9.9	12.4	87.6	88.8
	F5 [7]	57.0	43.0	41.4	58.6	57.8
	MB1 [3]	82.0	18.0	20.6	79.4	80.7
	Proposed	55.6	44.4	45.8	54.2	54.9
0.1 bpp	Outguess [6]	96.5	3.5	5.4	94.6	95.5
	F5 [7]	70.2	29.8	31.9	68.1	69.1
	MB1 [3]	93.3	6.7	8.8	91.2	92.2
	Proposed	64.8	35.2	42.8	57.2	61.0
0.2 bpp	Outguess [6]	98.3	1.7	2.80	97.2	97.7
	F5 [7]	88.3	11.7	14.2	85.8	87.0
	MB1 [3]	97.8	2.2	3.3	96.7	97.2
	Proposed	98.2	1.8	4.4	95.6	96.9

and cover image class. This problem is resolved by Receiver Operating Characteristic (ROC) curve.

The steganography security under practical steganalyzers defined as following:

- A Steganography system is said to be γ secure with respect to steganalyzer
  if $\mid TPR - FPR \mid \leq \gamma$, where $0 \leq \gamma \leq 1$.
- A steganography system is said to be perfectly secure with respect to steganalyzer
  if $\gamma = 0$.

It is also represented in graphical form by *Receiver Operating Characteristic*(ROC) performance curve of steganography system. Basically, ROC is curve of True Positive Rate Vs False Positive Rate. Figures 8, 9, 10 and 11 shows ROC curves of stego images when message embedding rate is 0.01 bpp, 0.05 bpp, 0.1 bpp and 0.2 bpp. For this experiment, training and testing images are selected randomly form cover and stego image dataset using MATLAB program and may be in unequal size. Moreover, 70% images are used for training purpose and 30% of images are used for testing purpose.

Any steganography system can get the point on left bottom ($TPR = 0$, $FPR = 0$) and hence classifying everything as negative; similarly any steganography system can get the point on top right ($TPR = 1$, $FPR = 1$) and hence classifying as positive. When $TPR = FPR$, message detection probability is 0.5 and point on diagonal line. The steganography system is said to perfectly secure if we can get point on left top ($TPR = 1$, $FPR = 0$).

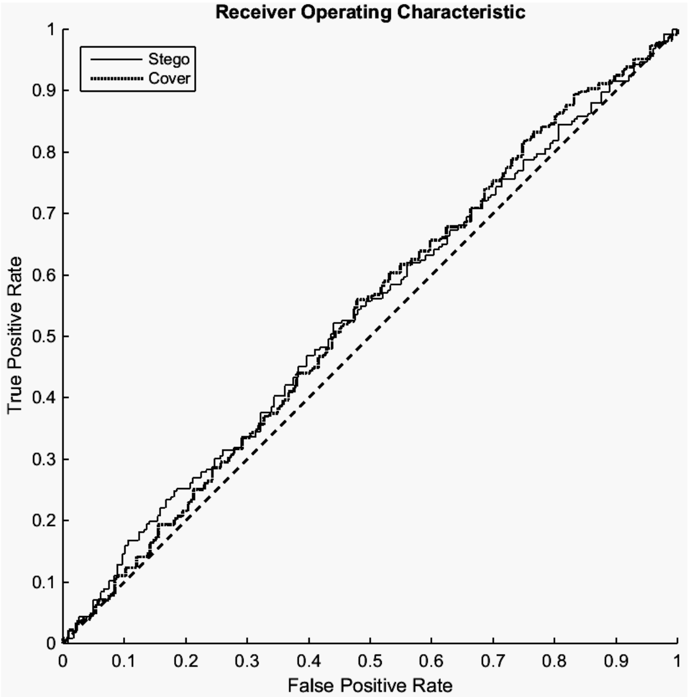


Fig. 8. ROC curves of steganography algorithm for 0.01 bpp message payload

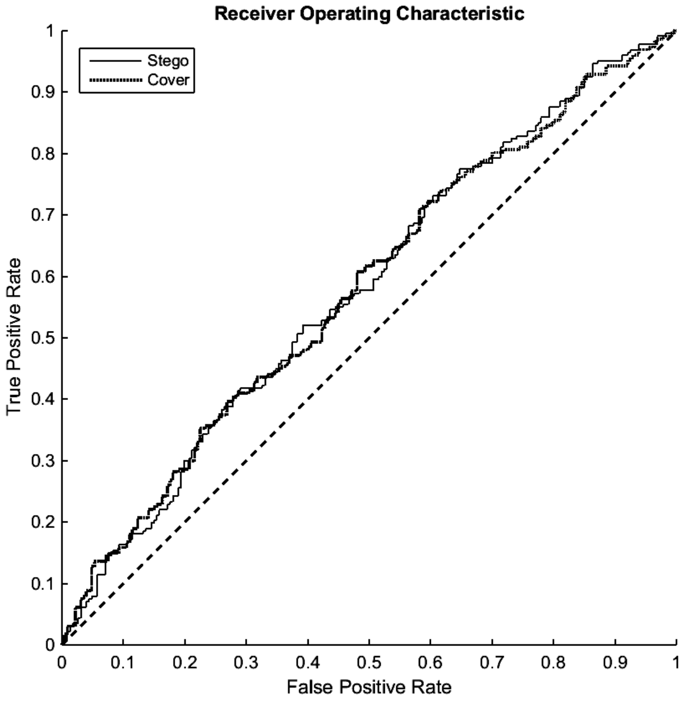


Fig. 9. ROC curves of steganography algorithm for 0.05 bpp message payload

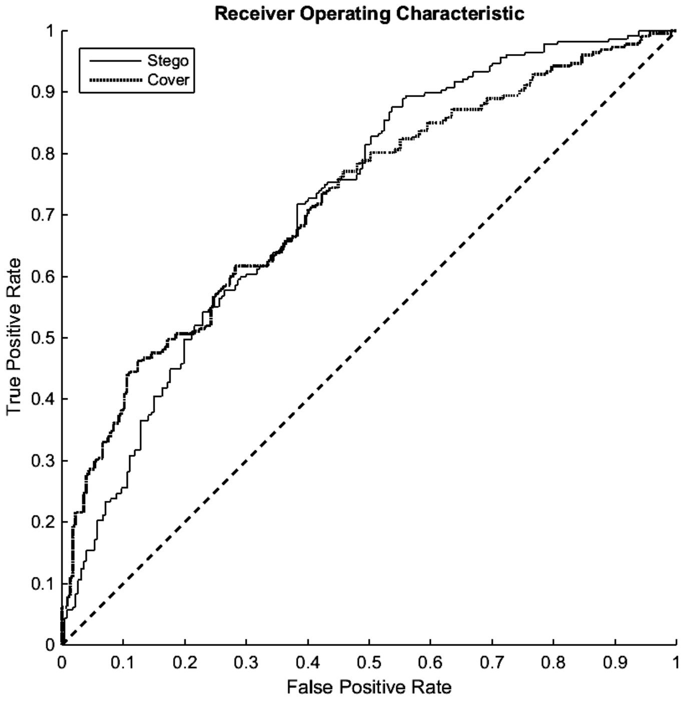


Fig. 10. ROC curves of steganography algorithm for 0.1 bpp message payload

It is clearly seen in figure 8 and 9, at lower message embedding rate ROC curves of stego and cover images class are too close to the ideal characteristic curve which is the indication of random guessing. It indicates that proposed algorithm performance is superior at lower message embedding rate. As message embedding rate is increased ROC curves of stego and cover images moves away from the ideal characteristic curve which is noticeable in figure 10. At higher embedding rate

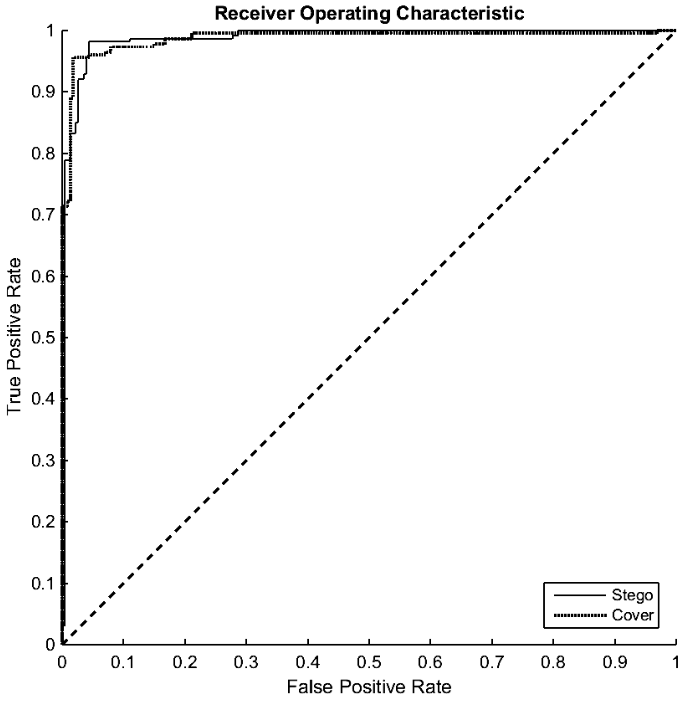


Fig. 11. ROC curves of steganography algorithm for 0.2 bpp message payload

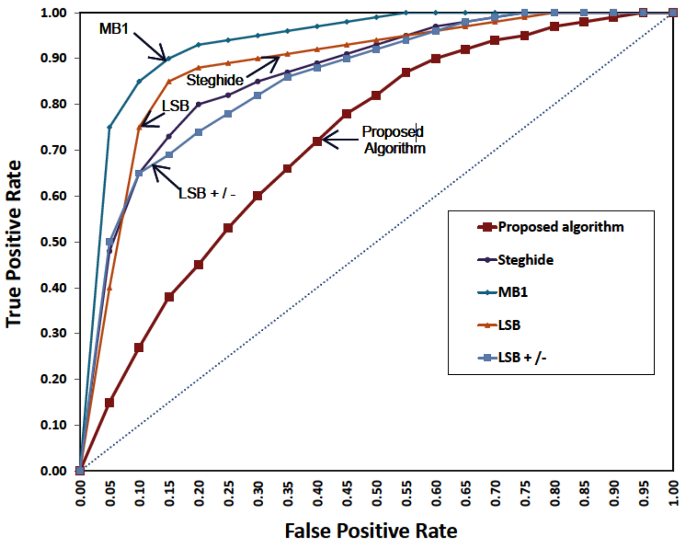


Fig. 12. ROC curves comparison of the proposed algorithm with F5, LSB, LSB ± and MB algorithms

chance of correct classification of class appreciably increased which easily seen when message embedding rate is equal to 0.2 bpp in figure 11.

It also compares ROC of proposed algorithm with F5 [7], LSB [8], LSB± [9] and MB [3] image steganography algorithm curves as shown in figure 12. For ROC curve comparison, message embedding rate 0.1 bpp has been set. Figure 12 undoubtedly indicate proposed algorithm is secured compare to F5 [7], LSB [8], LSB± [9] and MB [3] algorithms because ROC curve of proposed steganography algorithm is near to the ideal characteristic curve in compare to other algorithm. Finally, proposed image steganography algorithm performed stupendous in terms of security against universal blind steganalysis.

IV. Conclusions

Universal blind steganalyzer performance for security analysis plays vital role. For lower message embedding rate, ROC curves of stego and cover images class are too close to the ideal characteristic curve which is the indication of random guessing. Hence, SVM classifier does not classify correct class of stego or cover image because true positive rate and true negative rate are smaller. Once message embedding rate is increased over all accuracy of classifier is also increased and ROC curve moves away from the ideal characteristic curve. ROC curve comparison demonstrates that steganography algorithm is more secure compare to F5, LSB, LSB± and MB algorithms. After the considering above facts, performance of pixel parity and substitution based image steganography algorithm is *"stupendous"* in terms of security.

In future work, Larger image database steganalysis should be performed to verify that hidden message bits in the stego image are not detected by universal image blind steganalysis technique and different feature sets are used in them. Targeted steganalysis frame work should be implemented to test the image steganography algorithm provide better security in compare with other well known image steganography algorithms. Hidden message length estimation is an emerging field in steganalysis world. If image is successfully detected as stego image than applied message length estimation techniques to know the length of hidden message. If algorithm is successfully pass the above mention security test than hardware version should be implemented for higher speed of operation.

References

[1] David Kahn. *The History of Steganography*. First International Workshop, Cambridge, U.K., Lecture Notes in Computer Science. Springer-Verlag Berlin Heidelberg, 1996.

[2] D.B. Rawat. *Security, Privacy, Trust, and Resource Management in Mobile and Wireless Communications*. Advances in Information Security, Privacy, and Ethics:. Information Science Reference, 2013.

[3] N. Johnson, Z. Duric, and S. Jajodia. *Information Hiding: Steganography and Watermarking-Attacks and Countermeasures: Steganography and Watermarking - Attacks and Countermeasures*. Advances in Information Security. Springer US, 2012.

[4] Patel Chiragkumar B and Shah Saurin R. Parity and substitution based novel image steganography algorithm. *International Journal of Computer Applications in Engineering Sciences*, 5(1):10–14, 2015.

[5] T. Pevny, P. Bas, and J. Fridrich. Steganalysis by subtractive pixel adjacency matrix. *Information Forensics and Security, IEEE Transactions on*, 5(2):215–224, June 2010.

[6] Niels Provos. Defending against statistical steganalysis. In *Usenix security symposium*, volume 10, pages 323–336, 2001.

[7] Andreas Westfeld. F5 - a steganographic algorithm. 2137:289–302, 2001.

[8] Xin Liao, Qiao yan Wen, and Jie Zhang. A steganographic method for digital images with four-pixel differencing and modified lsb substitution. *Journal of Visual Communication and Image Representation*, 22(1):1 – 8, 2011.

[9] Parisa Gerami, Subariah Ibrahim, and Morteza Bashardoost. Least significant bit image steganography using particle swarm optimization and optical pixel adjustment. *International Journal of Computer Applications*, 55(2):21 – 25, 2012.