HUBFIRE - A MULTI-CLASS SVM BASED JPEG STEGANALYSIS USING HBCL STATISTICS AND FR INDEX

Veena H. Bhat[1], Krishna S., P. Deepa Shenoy, Venugopal K. R.
Department of Computer Science and Engineering, University Visvesvaraya College of Engineering, Bangalore, India

L. M. Patnaik
Vice Chancellor, Defence Institute of Advanced Technology, Pune, India

Keywords: Steganography, Steganalysis, Support Vector Machines, Huffman coding.

Abstract: Blind Steganalysis attempts to detect steganographic data without prior knowledge of either the embedding algorithm or the 'cover' image. This paper proposes new features for JPEG blind steganalysis using a combination of Huffman Bit Code Length (HBCL) Statistics and File size to Resolution ratio (FR Index); the Huffman Bit File Index Resolution (HUBFIRE) algorithm proposed uses these functionals to build the classifier using a multi-class Support Vector Machine (SVM). JPEG images spanning a wide range of resolutions are used to create a 'stego-image' database employing three embedding schemes – the advanced Least Significant Bit encoding technique, that embeds in the spatial domain, a transform-domain embedding scheme: JPEG Hide-and-Seek and Model Based Steganography which employs an adaptive embedding technique. This work employs a multi-class SVM over the proposed 'HUBFIRE' algorithm for statistical steganalysis, which is not yet explored by steganalysts. Experiments conducted prove the model's accuracy over a wide range of payloads and embedding schemes.

1 INTRODUCTION

Steganography is the art or practice of concealing a message, image or file within another image or file in such a way that the sender would be able to communicate to the intended recipient covertly. Steganalysis is the science of detecting messages (payload) hidden using steganography; it often deals with scrutinizing the carrier media for anomalies or non-ideal artifacts that are introduced in the process of steganography. While cryptography conceals data by encrypting the message, steganography achieves privacy by hiding the very existence of the message in an innocent looking cover object (Fridrich *et al.*, 2007). According to the terminology as agreed in the First International Workshop on Information Hiding (Pfitzmann, 1996), the embedding of a text or 'payload' in a 'cover image' gives a 'stego-image'.

Transductive Support Vector Machines (SVM), developed by Vladimir Vapnik, a kernel-based learning technique, have their roots in a geometrical interpretation of the classification problem to find a separating hyperplane (classifier) mapped on to a higher dimensional input space by optimization of a convex cost function (Vapnik, 1998 and Kristin, 2000). SVMs based on Structural Risk Minimization principle (Vapnik, 1995), work on maximizing the error margin of the training set, and have been effectively used for classification and regression of real-world data. SVMs have been effectively used in linear and non-linear pattern recognition, regression and classification problems.

[1]Veena H Bhat works as Professor of Information Technology at IBS-Bangalore, India.

2 RELATED WORK

A stego process is defined as a ε-secure process if the Kullback-Leibler divergence ∂ between the probability density functions of the cover document p_{cover} and those of this very same content embedding a message p_{stego} is less than ε:

$$\partial(p_{cover}, p_{stego}) \, 1 \leq \varepsilon. \quad (1)$$

H. Bhat V., S. K., Shenoy P., K. R. V. and M. Patnaik L. (2010).
HUBFIRE - A MULTI-CLASS SVM BASED JPEG STEGANALYSIS USING HBCL STATISTICS AND FR INDEX.
In *Proceedings of the International Conference on Security and Cryptography*, pages 447-452
DOI: 10.5220/0002989004470452

The process is called 'secure' if $\varepsilon = 0$, and in this case the steganography is perfect, creating no statistical differences by embedding of the message (Miche *et al.*, 2009), steganalysis would then be impossible.

Support Vector Machines (SVM) are one of the most effective methods for pattern recognition and multivariate classification (Joachims, 1998; Liang, 2004). A binary SVM identifies whether an instance belongs to a class ensuring a minimum generalisation error with a hyperplane to distinguish the two classes. Multi-class SVM algorithms such as OVA (One Versus All) use winner-takes-all strategy (Boser *et al.*, 1992) and OAO (One Against One) that use max-wins voting method (Liang 2004) are in vogue. Binary SVM are used for Steganalysis as proposed by Lyu and Farid using higher order statistics (Lyu and Farid, 2002). Some of the blind steganalysis techniques developed using multi-class SVM for classification include analysis based on run-length histograms (Dong and Tan, 2008). Quantized DCT features are used as input features to a SVM classifier (Pevny and Fridrich, 2007). Steganalysis using SVM classifier for a huge number of input features, discussed by Xuezeng *et al.*,(2008) employs features extracted from histograms.

3 TESTED EMBEDDING SCHEMES

Steganographic techniques, based on the nature of embedding scheme adopted, are classified into spatial domain, frequency domain and adaptive steganography. The stego-image database is populated, for all these three embedding schemes.

3.1 LSB – Least Significant Bit

Among the several techniques employed for steganography, LSB Steganography is a popular spatial domain technique because of its robustness, fine concealment, high steganographic-embedding capacity and easy realization. In this work, we have employed a scheme that uses both sequential embedding and scattered embedding (Advanced LSB technique) (Chen *et al.*, 2006). To build the stego-image database we have used the open source Matlab code, for this technique by Luigi Rosa.

3.2 JPHS – JPEG Hide and Seek

JPEG Hide-&-Seek (JPHS) is a transform domain tool designed for JPEG files and lossy compression. JPHS uses least significant bit overwriting of the Discrete Cosine Transform (DCT) coefficients used by the JPEG algorithm. Though JPHS fails to preserve the DCT histogram statistics, its statistical detectability is among the lowest based on the results reported in (Tomas, 2007).

The stego-image database is built using the open source Matlab code for this technique, namely 'steganoh', designed by Francisco Echegorri. This program hides a text file in a grayscale image by disordering it into lines and columns using the Ranpermut-Encryption algorithm.

3.3 MBS – Model Based Steganography

Unlike LSB and JPHS embedding techniques, the MBS Technique is an adaptive technique, proposed by Sallee, which tries to model statistical properties of an image and preserve them during embedding process (Sallee, 2005). The embedding operation employs a nonadaptive arithmetic decoder and the coefficients in each histogram bin are modified with respect to embedding rule, while the global histogram and symbol probabilities are preserved. Attacks such as Blockiness (Ullerich and Westfield, 2008) can detect this embedding scheme. Our stego-image database for MBS is built using the open source Matlab code (Sallee, 2005).

Table 1: First Order Histogram statistics for different embedding schemes.

First Order Statistics	Cover	LSB	JPHS	MBS
Mean	4.2948	4.2947	4.3011	4.2951
Variance	0.9936	0.9937	0.9918	0.993
Skewness	-0.8237	-0.8235	-0.8192	-0.8221
Kurtosis	4.2406	4.2397	4.249	4.2364
Energy	0.3438	0.3438	0.3447	0.3438
Entropy	1.2852	1.2853	1.2849	1.2852
File size (kb)	163	164	167	150

The stego-images created as mentioned in sections 3.1, 3.2 and 3.3 and the cover images were checked for first order histogram statistics, table 1 shows some of the observations for images of resolution 2048X1536 for a maximum payload. From table 1, it can be deduced that in the absence of reference to cover images, it would be difficult to differentiate between stego and cover images using

first order histogram statistics for blind steganalysis.

Throughout this paper, for convenience we address these techniques as LSB, JPHS and MBS.

4 HBCL STATISTICS AND FR INDEX

4.1 HBCL - Huffman Bit Code Length Statistics

JPEG, a lossy compression format, employs sequential Huffman Encoding for data compression wherein symbols (DCT coefficients in this case) are taken and encoded with variable length codes that are assigned based on statistical probabilities. A grayscale image employs 2 Huffman tables, 1 each for AC and DC portions. When a JPEG image is embedded with a particular payload, certain non-ideal JPEG artefacts are introduced in the given image, though the enormity of this deviation varies. On artefacts the nature of the DC HBCL statistics of the images in the populated image-database, we have considered HBCL statistics from two to five bits only, as the variation in HBCL from the 6^{th} bit onwards is negligible.

One of the scoring features of Huffman coding algorithm is its 'unique prefix property' that is no code is a prefix to any other code, making the codes assigned to the symbols unique. This fact further supports our choice of the HBCL statistics for our evaluation features to be efficient as the JPEG artefacts introduced by steganography on an image becomes unique and hence can be predicted using a suitable classifier or a prediction model. In our work we extracted these statistics using Huffman decoding for a JPEG image in Matlab.

4.2 FR Index – File size to Resolution Ratio

When a raw image is compressed by JPEG compression, based on the resolution of the image, its quality and compression ratio the resulting JPEG file takes up a particular file size. This indicates that the file size and the resolution of an image and further its quality are interrelated. Thus the ratio of file size of the image in bytes to that of its resolution is found to be unique and in a certain range for a given resolution, this functional is termed 'FR Index' and used as one of the inputs to build the prediction model. Table 4 shows the range of FR Index for the resolutions used in our database.

5 IMPLEMENTATION

5.1 Image Database

For effective performance evaluation of a steganalysis technique, the data set employed in the experiment, which validates the prediction model for its sensitivity and specificity, is very important. This work adopts JPEG grayscale images, as it is harder

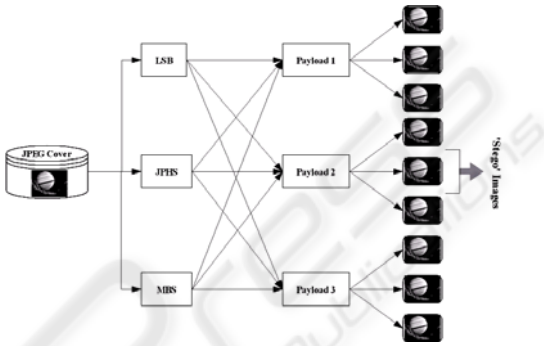

Figure 1: Processing details to create the image database.

to detect hidden data in them compared to color images where steganalysis can utilize dependencies between color channels.

This proposed model uses a dataset of images that would include those images often found on the internet and private domain, at the same taking care to create the entire image dataset within the quality range of 57 to 72, but since one has a freedom in selecting the quantization table when compressing an image using the JPEG algorithm, there is no standard definition of a quality factor. Therefore the quality factor of the images in the dataset is approximated by deploying the publicly available 'JPQ-JPG Quality Estimator' essentially, JPQ estimates the quality factor of the image by comparing its quantization table to the suggested quantization table in the JPEG standard that employs optimized Huffman tables. This quality estimation program assumes the images to be subsampled at 2:1:1 and an approximate recompression error ranging from -0.49 to +0.44, (which can be considered to have a negligible effect on the JPEG features) (Bhat, H, V. *et al.*, 2010).

Among the 10 image sets (Basic Dataset) of 100 images each evaluated, the least resolution is 75X75 and the highest being 2048X1536 which was taken from Jegou's database (Herve, 2008). Table 2 illustrates details of the image database used in our performance evaluation. The Basic Set lists the different resolutions used with 100 images in each set; the Resized Set lists the set of images

correspondingly resized from column 1 to validate against the JPEG Resizing Error. The range of the FR Index for each set is also mentioned.

5.1.1 Cover Images and Stego Images

For each tested method, the cover grayscale JPEG images and several stego grayscale JPEG images embedded with different payloads are prepared. By a cover image, it is understood as an image into which no message is embedded. LSB, JPHS and MBS were the embedding schemes employed with three different payloads in each scheme and for each resolution. Thus for the entire set of 2,000 cover images, 18,000 stego images with different parameters were generated and evaluated.

Table 2: Image dataset details.

Basic Set	FR Index	Resized set	FR Index
75X75	0.12 – 0.59	-	-
130X130	0.09 – 0.45	100X100	0.06 – 0.42
214X214	0.09 – 0.42	200X200	0.08 – 0.34
384X256	0.08 – 0.37	300X300	0.06 – 0.32
481X321	0.06 – 0.31	400X400	0.07 – 0.47
512X512	0.03 – 0.41	500X500	0.04 – 0.52
720x480	0.07 – 0.43	600X600	0.07 – 0.37
800X600	0.03 – 0.31	700X700	0.05 – 0.46
1024X768	0.01 – 0.03	800X800	0.02 – 0.05
2048X 1536	0.004 – 0.03	900X900	0.01 – 0.05
		1KX1K	0.005 – 0.03

In order to check for consistency of the features extracted for the model and to validate this over JPEG Resizing Error, the basic dataset is resized. Table 2 describes the nature of resizing performed and related observations. Thus, a set of 2000 JPEG cover images are obtained on which the consistency of the prediction model is evaluated.

5.1.2 Payloads

Three different payloads for LSB, JPHS and MBS each are tested for the 2,000 images. In case of LSB and MBS embedding schemes, a random data in corresponding to the payload size is embedded, where as in JPHS a text file of the respective size is embedded after encryption. In case of MBS, we used a code that uses optimized Huffman tables and the embedding rate on an average was found to be 1.0431 bits per change (bpc).

For images with resolution 100X100 and less, 340 bytes, 500 bytes and 720 bytes were the three payload sizes used; however our prediction model produced consistent results for payload sizes less than 300 bytes too. For the remaining resolutions up to that of 2048X1536, we use payloads of sizes 1024 bytes (1 KB), 2048 bytes (2KB) and 5120 bytes (5KB). Thus the prediction model is evaluated over 20,000 images (inclusive of cover and stego images).

5.2 Input Functionals used in the Model

In the first phase, to predict steganography, the resolution of the image, file size, quality and HBCL statistics are considered. Though quality does not play a pivotal role in building the prediction model, it helps to build the image database. Resolution and file size of an image are considered because the file size is a direct consequence of the resolution of an image and the JPEG compression performed on the image. Hence, these factors are in a way affected by the JPEG compression (the quantization and the Huffman compression employed).

Table 3: Correlation values for the input attributes selected for steganalysis.

	FR Index	Quality	2 HBCL	3 HBCL	4 HBCL	5 HBCL
FR Index	1					
Quality	0.59	1				
2 HBCL	-0.38	- .05	1			
3HBCL	-0.02	0.02	-0.06	1		
4HBCL	0.20	0.07	-0.26	-0.67	1	
5HBCL	-0.02	-0.07	-0.03	-0.85	0.27	1

HBCL statistics and the reason for choosing them are explained in section 4.1. Correlation between these factors is analyzed to check their interdependency. Table 3 illustrates the correlation between the functionals over the entire image database. '2 HBCL' indicates the number of Huffman code of 2 bit length, '3 HBCL' for number of 3 bit length codes and so on.

5.3 Classifier

Multi-class Support Vector Machine (Burges, C., 1998; Hsu, Chih-Wei., Lin, Chih-Jen., 2002) is used as a classifier with number of classes set to 4 wherein class 1 indicates cover images, class 2; LSB encoding, class 3; JPHS, class 4; MBS scheme. The kernel function opted is Gaussian with suitable λ and cost function with the size of the quadratic programming set to 100. The gamma and cost parameters, (c,γ) for the SVM classification are estimated using grid-search over a multiplicative

grid of 2^{-2} to 2^{+4} over a sampling method of 10 fold cross-validation. This is implemented using an open source Matlab toolbox, SVM-KM.

Let K_C be the entire set of 'cover' images in the database. $K_A \subset K_C$ is a set of ten different resolutions of images; this set K_A is embedded with a payload employing three different steganographic methods as described in section 3 thus generating a 'stego-image' dataset denoted as K_S. The set K_{TRAIN} is used to train the multi-class SVM where $K_{TRAIN} = K_S \cup K_A$. A dataset $K_{TEST} = K_C \cap K_A$ is used to test the trained model in sets of 600 instances with 150 per class such that $K_{TRAIN} \cap K_{TEST} = Null$. One such test is illustrated in the figure 2.

Table 4: HUBFIRE Algorithm.

Input: The test image
Output: The image classified as genuine or stego image.

1. **Data Generation**
   a. Identify the resolution and file size of an image for the sets, K_{TEST} and K_{TRAIN}.
   b. Populate the image database consisting of cover and stego images – LSB, JPHS, MBS with payloads of 1KB, 2KB and 5KB.
2. **Data Preprocessing**
   a. JPEG Huffman decoder is used to extract four HBCL statistical data for each image of the training set, K_{TRAIN}.
   b. Basic properties of the image – filesize, quality, resolution are recorded.
   c. Calculate the FR Index for each image.
3. The functionals created through the data preprocessing step 2 is used to train the Multi-class SVM with number of classes set to 4.
4. The image to be tested is processed against the model trained in step 3.

6 RESULTS

Figure 2 illustrates the confusion matrix of the One-Versus-All SVM algorithm used to classify. The model is highly sensitive to stego-images as against the cover images which increases the false positive response of the system.

The embedding schemes chosen to train the model span over different domains as described in section 3 making the model an efficient blind steganalysis technique. It is observed that the trained model is highly sensitive to MBS in comparison with other embedding schemes.

The model has to be fine tuned to reduce the false positive rate so as to give a reliable detection

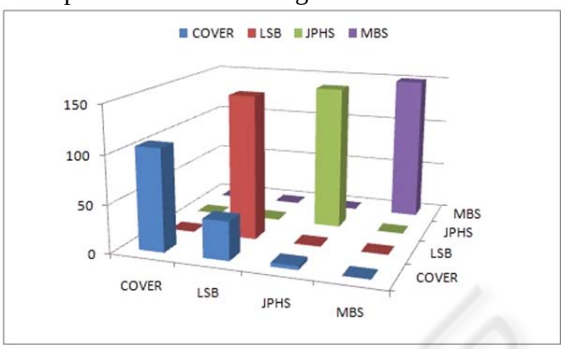

Figure 2: Confusion matrix of HUBFIRE, the multi-class classifier.

even with respect to cover image.

7 PERFORMANCE ANALYSIS

The results of the proposed model are compared with the work of Pevny and Fridrich (2006).

A comparison is illustrated in the table 5. The proposed model is found to have a high false positive rate however it is highly sensitive for classification among the stego-images, thus the model serves as a reliable 'stego-classifier'.

8 FUTURE WORK AND CONCLUSIONS

Our future work will include optimising the multi-class SVM model, by calculating the cost function to reduce the false positive rate. The payloads considered in this work are 1KB, 2KB and 5KB for each of the embedding schemes, however, the model does not analyse the response of the trained model for each of the embedded payload individually. This would reflect upon the sensitivity of the model, with respect to varying payloads. Multi-class SVMs with different kernels of radial basis function and sigmoid function need to be explored.

As explained in section 6, the model is reliable in stego-image detection that is embedded with steganographic schemes that span over several domains. The model gives an efficiency of 100% with respect to detection of MBS.

Table 5: Confusion matrix of the multi-class classifier for the testing set. The leftmost column contains the embedding algorithm. The remaining columns show the results of the classification of HUBFIRE and the model proposed by Pevný and Fridrich (2006).

	HUBFIRE				(Pevný and Fridrich, 2006)			
	Cover	LSB	JPHS	MBS	Cover	LSB	JPHS	MBS
Cover	70.67%	26.67%	2.67%	0.00%	96.45%	0.12%	0.20%	1.44%
LSB	0.00%	99.33%	0.67%	0.00%	0.08%	99.08%	0.53%	0.08%
JPHS	0.00%	0.67%	99.33%	0.00%	0.20%	0.12%	98.32 %	0.56%
MBS	0.00%	0.00%	0.00%	100.00%	1.44%	1.56%	0.72%	94.44%

REFERENCES

Bhat, H, V., Krishna, S., Shenoy, P, D., Venugopal, K, R., Patnaik, L, M., 2010. JPEG Steganalysis using HBCL Statistics and FR Index. *To be published in the Proceedings of Pacific Asia Workshop on Intelligence and Security Informatics*.

Burges, C., 1998. A Tutorial on Support Vector Machines for Pattern Recognition. *Data Mining and Knowledge Discovery*, 2, 121–167.

Boser, B., Guyon, I., Vapnik, V., 1992. A Training Algorithm for Optimal Margin Classifiers. *Proceedings of the Fifth Annual Workshop on Computational Learning Theory*, 5, 144–152.

Chen, Ming.,Zhang, Ru., Niu, Xinxin., Yang, Yixian., 2006. Analysis of Current Steganography Tools: Classification & Features. *Proceedings of the 2006 International Conference on Intelligent Information Hiding and Multimedia Signal Processing*, 384-387.

Cristianini, N., Shawe-Taylor, J., 2000. *An Introduction to Support Vector Machines*, Cambridge University Press.

Herve, Jegou., Matthijs, Douze., Cordelia, Schmid., 2008. Hamming Embedding and Weak Geometry Consistency for Large Scale Image Search. *Proceedings of the Tenth European Conference on Computer Vision*, 304-317.

Hsu, Chih-Wei., Lin, Chih-Jen., 2002. A Comparison of Methods for Multiclass Support Vector Machines. *IEEE Transactions on Neural Networks*, 415-425.

Joachims, T., 1998. Text categorization with support vector machines: Learning with many relevant features. *Proceeding of the Tenth European Conference on Machine Learning (ECML)*, 137–142.

Kristin, P, Bennet., Campbell, Colin., 2000. Support Vector Machines: Hype or Hallelujah ? *SIGKDD Explorations*, 2, 1-13.

Lyu, Siwei., Farid, Hany., 2002. Detecting Hidden Messages Using Higher-Order Statistics and Support Vector Machines. *Proceedings of the Fifth International Workshop on Information Hiding*, 2578, 340-354.

Miche, Yoan., Bas, Patrick., Lendasse, Amaury., Jutten, Christian., Simula, Olli., 2009. Reliable Steganalysis using a Minimum Set of Samples and Features. *EURASIP Journal of Information Security*.

Pevný, Tomas., Fridrich, J., 2006. Multi-class Blind Steganalysis for JPEG Images, *Computer Science and Software Engineering* , 939-942.

Pevný, Tomas., Fridrich, J., 2007. Merging Markov and DCT features for Multi-class JPEG Steganalysis, *Proceedings SPIE - Electronic Imaging, Security, Steganography, and Watermarking of Multimedia Contents*, 03–04.

Pfitzmann, B., 1996. Information Hiding Terminology. *Proceedings of the First International Workshop on Information Hiding*, 347-350.

Sallee, P., Model-based Methods for Steganography and Steganalysis, 2005. *International Journal of Image Graphics*, 167–190.

Ullerich, Christian., Westfeld, Andreas., 2008. Weakness of MB2. *Digital Watermarking*, 127-142.

Vapnik, N, Vladimir., 1995. *Nature of Statistical Learning Theory*, Springer.

Vapnik, N, Vladimir., 1998. *Statistical Learning Theory*, York: Wiley.

Xuezeng, Pan., Li, Zhuo., Jian, Chen., Jiang, Xiaoning., Zeng, Xianting., 2008. A New Blind Steganalysis Method for JPEG Images. *International Conference on Computer Science and Software Engineering.* 939-942.