Receiver Operating Characteristic (ROC) Graph to Determine the Most Suitable Pairs Analysis Threshold Value

Emelia Akashah P.A, Savita K. Sugathan

Computer and Information Sciences Dept
Universiti Teknologi PETRONAS
Bandar Sri Iskandar
Perak, MALAYSIA
emeliaakashah@petronas.com.my,
savitasugathan@petronas.com.my

Prof. Anthony TS. Ho

Department of Computing, School of Engineering and
Physical Sciences
University of Surrey
United Kingdom
a.ho@surrey.ac.uk

Abstract—Steganography is the art of hiding the information that is going to be sent from one party to another. Information can be hidden into image, text, audio or video. Steganography allowed communication to happen without other people notice there is transmission of message except the intended party. This paper explains the implementation of Receiver Operating Characteristic (ROC) graph addressing the incorrect classification of images for stegogramme and non-stegogramme classes using Pairs Analysis detection technique. The threshold value to discriminate between the two classes is identified, to reduce the rate of False Negative (FN) .

Keywords—non-stegogramme; pairs analysis; ROC; stegogramme; threshold

I. INTRODUCTION

Steganalysis or the detection of the message in an image is one of the methods to attack the secret communication between two parties. Many researchers conducted a study to break the steganography algorithm.

Fig. 1: Steganalysis system [1]

Figure 1 shows the example of steganalysis system. The source image contains a lot of images and it is not known if they are stegogramme or non-stegogramme.

The research is carried out for Pairs Analysis detection technique developed by Fridrich [1]. The focus will be on greyscale images and Least Significant Bit (LSB) embedding.

Since there is no threshold value to distinguish between stegogramme and non-stegogramme classes, it will lead to the incorrect classification.

This paper will address the limitation of the incorrect classification by reducing the rate of False Negative (FN) in using Pairs Analysis and to set the most appropriate threshold value.

II. RESEARCH WORK

Steganalysis softwares are used to hide message in the carrier images. Pairs Analysis, Chi-squared attack, F5, RS Steganalysis and Outguess become the famous algorithm in attacking the image carrier.

An attack developed by Provos namely Chi-squared is the detection algorithm developed before pairs algorithm which can be applied to any steganographic software [1]. According to Fridrich [1], the detection algorithm works for a fixed set of Pairs of Values (PoVs), or other fixed group of values, are flipped into each other to embed message bits. If we embed the secret message sequentially in the cover image pixels or indices, we will observe an abrupt change in statistical evidence as we encounter the end of the message. This detection algorithm is developed and used for sequential embedding, which means that we embed the message in the sequence order of pixels or indices or coefficients. Chi-squared technique can also be used for random message embedding, but is less effective unless 97% of pixels or coefficients or indices are used for embedding. Westfeld developed an idea to group colours from one pixel or neighbouring pixels and fusing their values using a special hash function. Westfeld claim that

messages as small as 33% of the maximal image capacity can be detected [1].

According to Fridrich [1], an improved chi-squared attack, known as generalized chi-squared is developed by Provos, allowed the random embedding. The uses of sliding window of a fixed size that can be move along the image rather than to increase the window size provides the capability of random message detection. However, Provos did not elaborate or perform the analysis any further on his new proposed technique.

RS Steganalysis is another detection technique which has been introduced before pairs analysis. The estimation of the number of flipped pixels during LSB embedding can be identified, thus it can estimates the length of the message. Pairs analysis is developed as an improvement of RS Steganalysis. It is able to detect the presence or absence of the hidden message in an image, either greyscale image or true colour image. The detection algorithm is applied separately to each colour channel for true colour images . This will increase the value of discrimination function.

Fridrich tested the algorithm on the EzStego steganography software and focusing on gif images [1].

Andrew Kerstated that Pairs and RS Steganalysis attack are threshold-free statistics, which means that the algorithm could detect the presence or absence of the message and try to estimate its length, without having to set the threshold to discriminate between the two classes. He said that the output is only yes for stegogramme or no for non-stegogramme [2].

Pairs analysis algorithm will detects randomly the spread messages in 8-bits images, embedded using LSB flipping of palette indices to a pre-ordered palette [1]. It can detect the message length in the cover image [1]. It is said that this method is reliable and can accurately estimate the secret message length. The advantages of this detection algorithm are, it can be used for many different steganographic systems and also to different image format (jpg, bmp, gif etc).

III. METHODOLOGY OF STUDY

For this research, we choose bmp format natural images to be tested. We generated 0, 10, 20,40,60,80 and 100 percent of message length compared to the image capacity. For 100 percent message length, each bit of the message is corresponding to one pixel of the image.

For the embedding technique, we hide the secret message in random location using the key generated from randperm function in Matlab. This function re-arranges the location of the matrices and the new random location is used as a key for embedding.

Once we got the result for this image database, we tried to perform this algorithm to other steganographic systems. We

gathered all the result from the test and compare them to see how this technique works for different steganography software. We analyzed the result that we got to see whether the detection algorithm is reliable to use for the chosen steganography software or not.

In the first step, we need to extract or split the colours from the image and make the colour pairs. The colour cut is concatenated and put in single stream. The sequence of colour is converted to a binary vector. We assume the image has up to 256 palette colours, $P \leq 256$. The set of colour pairs that will be exchanged during embedding is:

$$Z = \{(c_0, c_1), (c_2, c_3), \dots, (c_{P-2}, c_{P-1})\} \quad (1)$$

For example, let (c_1, c_2) be a colour pair and associate c_1 with a '0' and c_2 with a '1'. The same thing is applied to the rest of the sequence. The next stage is to make shifted pairs and apply the same concept as before. The set of shifted colour pairs is as follows:

$$Z' = \{(c_1, c_2), (c_3, c_4), \dots, (c_{P-1}, c_0)\} \quad (2)$$

After that, we could count the homogenous bit-pairs (eg: the sequence of 11 or 00) for both sequences, E and E' . In this stage, $R(p)$ denote the expected number of homogeneous bit-pairs in Z after flipping the LSB indices, divided by n – the length of Z . The same process is done for Z' where $R'(p)$ is the relative number of homogeneous bit-pairs in Z' [1]. The value that we obtained from those expressions will be used in the quadratic equation to get the result of the unknown message length, q . To get the approximation value of q , we choose the smaller root from the quadratic equation.

It is to prove the relative number of the pairs after embedding a relative message length. In pairs analysis theorem, we should accept one additional assumption where the structure of homogeneous bit-pairs in Z and Z' should be the same if there is no message embedded in the image. There is no reason why they should have different structure [1].

IV. EXPERIMENTAL RESULT

An overlapped distribution graph produced from the experiment performed on the image sources can lead to the incorrect classifications. Figure 2 below is the result of distribution graph for 100 cover images and stegogrammes from our image database. The overlapped area might be or might not be a stegogramme.

Fig. 2: Distribution graph for 100 images (stego & non-stego)

The implementation of FN and FP concept can be applied in distinguishing the image class. To have a threshold value to discriminate between the two classes definitely could help in reducing the mistake to classify the images.

Besides the use of FN and FP in the image classification, the experiment result for the tested images could be true-positive (TP) or true-negative (TN). Such result could be obtained from the non-overlapped area of the graph distribution. Once the threshold has been set and if the output is above the threshold, the test is considered as positive.

$$\text{Cover image} < \text{Threshold value} \leq \text{Stegogramme}$$

Table 1: Classification of cover image and stegogramme

True positive(TP)	Stegogramme which we detect as stegogramme
True negative (TN)	Cover image which we detect as cover image
False positive (FP)	Cover image which we detect as stegogramme
False negative (FN)	Stegogramme which we detect as cover image

Classifying the images to their classes is depending on the accuracy of the threshold value. FP and FN in Table 1 are also referred as Type I and Type II errors.

Table 2 : Type of error

Test result	Actual Condition	Error Type
Stegogramme	Cover image	Type I
Cover image	Stegogramme	Type II

Between the two types of error, Type II or FN is more dangerous if it occurs. We should tolerate with false positives to reduce the number of false negative [7]. Type II error is not to accept something when the condition is true. The observer to detect any secret communication will see it as a normal communication. Type I is just like accusing someone doing something he actually did not commit. FP happens when cover image is assumed as stegogramme [7].

The distribution graph is divided into fractions TP, TN, FP and FN. Axis Y shows the density value of the distributions while axis X shows the value of q , which represents the bitflips. A cut point value is identified on the distribution graph before threshold value can be selected. Figure 3 shows the fraction for TP, TN, FP and FN with cut-off value 0.02. All the cut-off value chosen are belong to the overlapped area.

Fig. 3: Cut-off value =0.02

Fig 4: Overlapped distribution with four cut off value

Figure 4 shows the graph with all the cut-off values that have been chosen. There are four points chosen to be evaluated which are -0.02, 0, 0.02 and 0.05. Among all the values, the experiment will yet to identify which cut-off value will be the most accurate threshold.

Table 3 : Test performed to identify fraction

Test	Fr	N	Fr	N	Total
+ve	TP	A	FP	C	$a + c$
-ve	FN	B	TN	D	$b + d$
Total		$a + b$		$C + d$	

Table 4 : Result of each fraction for each cut-off value

Cut off value = 0.02							
Test	Fr	N		Fr	n		Total
+ve	TP	0.7454		FP	0.4013		1.1467
-ve	FN	0.2546		TN	0.5987		0.8533
Total		1			1		
Cut off value = 0.05							
Test	Fr	N		Fr	n		Total
+ve	TP	0.4013		FP	0.0918		0.4931
-ve	FN	0.5987		TN	0.9082		1.5069
Total		1			1		
Cut off value = 0							
Test	Fr	N		Fr	n		Total
+ve	TP	0.9082		FP	0.5987		1.5069
-ve	FN	0.0918		TN	0.4013		0.4931

Total		1		1	
Cut off value = -0.02					
Test	Fr	N	Fr	n	Total
+ve	TP	0.9772	FP	0.7734	1.7506
-ve	FN	0.0228	TN	0.2266	0.2494
Total		1		1	

V. THRESHOLD VALUE USING ROC

ROC can help us in making decision, for example which threshold value should be chosen to reduce the incorrect classification. To plot ROC curve, only true-positive rate and false-positive rate are needed. True-positive rate also known as sensitivity and false-positive rate is known as specificity. To plot ROC, we take sensitivity value for axis - Y and 1-specificity value for axis-X. ROC curve is sometimes known as sensitivity vs. 1-specificity graph. Each threshold we choose represents one point on the ROC graph. Sensitivity (Se) is a statistical measure of how well a classification test correctly identifies a condition and it is a proportion of true-positives [4]. To perform the test, we require high sensitivity rate. Specificity (Sp) is used to identify negative cases, where the test correctly indicates 'negative' if the image does not contains any hidden message. It represents the proportion of true-negatives of all negative cases [4]. To perform the test, we require high specificity rate. Se and Sp are commonly used to measure the test performance. Errors will occur if we just take the risk and do not consider sensitivity and specificity. Before Se and Sp are calculated, we assume that all of the images that we have, can be allocated in either stegogramme or cover image class without making any error. The frequency of the images falls into false negative, false positive, true positive and true negative area is used to predict the threshold value before we calculate the efficiency of the cut-off value that we choose[5].

Table 5: Result for each cut-off value

Cut-off value = 0.02	
Sensitivity (Se)	0.7454

Specificity (Sp)	0.5987
Cut-off value = 0.05	
Sensitivity (Se)	0.4013
Specificity (Sp)	0.9082
Cut-off value = 0	
Sensitivity (Se)	0.9082
Specificity (Sp)	0.4013
Cut-off value = -0.02	
Sensitivity (Se)	0.9772
Specificity (Sp)	0.2266

There is a diagonal line or known as random guess line that divides the ROC space to determine which plot is the best to be chosen. The plot above the line can be considered as good result, while the plot under the line is the bad result. Below is the figure of ROC curve which we have plot according to the four selected threshold that we choose.

Fig. 5 : ROC plot

From the ROC plot above, we could see there are four points plotted, and they represent each threshold that we choose. There is one point plotted under the random guess line, which lead to the bad result. The other three points are plotted above the guessing line. Because of the point under the diagonal line will give bad result, we will only consider three points above the line. We calculate the accuracy of each point and the result is as follows:

Table 6 : Accuracy percentage for each threshold

Threshold (x)	Accuracy
0.02	65%
0	60%
-0.02	56%

To calculate the accuracy, we divide true-positive rate with false-positive rate. From the table above, we can conclude that the best value to choose as a threshold is 0.02 as it gives the highest accuracy among others. To select the best threshold, we need to consider some factors. We tried to maximize the probability of correct classifications, which are specificity and sensitivity and also to minimize the probability of incorrect classifications, which are false-positive fraction and false-negative fraction.

VI. RESULT FOR TESTED IMAGES

The threshold value that have been chosen using ROC graph will be tested on various numbers of images which are using different type of steganography software to hide the secret message. Steganography softwares such as InfoStego and S-Tools are used to hide secret message in the tested images. The experiment is done to check whether the threshold value that have been chosen using ROC is reliable or vice versa. The secret message is also embedded in the tested images using our simple embedding technique.

S-Tools hide the message bits in the LSB of the image. The message is embedded in random location generated by pseudo-random number generator. The concept of hiding message used by S-Tools is similar with the embedding technique that has been developed to be used in this experiment.

Fig. 6 : Tested images

The pairs analysis algorithm is tested on the database of 10 gif format images. The images that have been chosen are the natural images. For each image, we try to detect the value that we got for cover image to see whether it falls under which class, either stegogramme or non-stegogramme. The algorithm also will be tested whether it can give the right classification for the stego-image. From the experiment result, they can classify them as below:

Table 7 : Detection result for images using our own developed steganography software

Image	Cover Image	Stego Image
1	TN	TP
2	TN	TP
3	TN	TP
4	TN	FN
5	TN	TP
6	TN	TP
7	TN	TP
8	FP	TP
9	FP	FN
10	TN	TP

Table 7 shows that the threshold of 0.02 is quite reliable even though sometimes we will get false positive (FP) or false negative (FN) result. However, it is possible to get FP or FN sometimes because there is some area of the distribution graph which is overlapped. But, this type of error is not always happen. From the result of 10 images, the cover image was wrongly detected as stegogramme twice. It means that eight images are classified in the correct class. For stegogramme, the images are tested with pairs detection to see whether the algorithm can give the correct result. From the table above, eight images are correctly classified in their class while only two images are wrongly assumed as cover image.

Apart from gif format images; other format of image is also used to perform the experiment. Ten bmp formats of natural images are used to be tested.

The same detection process used for gif images is used for bmp images. For each image, we try to detect the value that we got for cover image to see whether it falls under which class, either stegogramme or non-stegogramme. From the test, we will determine whether this algorithm can be used for bmp images or vice versa. From the result that we got for our bmp images, we can classify them as below:

Table 8 : Detection result for images using S-Tools steganography software

Image	Cover Image	Stego Image
1	TN	TP
2	TN	TP
3	TN	TP
4	TN	FN
5	FP	TP
6	FP	TP
7	TN	TP
8	TN	TP
9	TN	FN
10	TN	TP

Table above shows the classification result for bmp images. The same detection process is done for these random bmp images, to classify them in suitable classes. All the stegogramme is produced by S-Tools software and we use text file as secret message which will be hidden in the images.

For cover image, the detection process has correctly classified eight images and detects two cover images as stegogramme.

For stegogramme, there are also eight images classified as true positive and the rest is classified as false negative.

Table 9: Detection result for images using InfoStego steganography software

Image	Cover Image	Stego Image
1	TN	TP
2	TN	TP
3	FP	TP
4	TN	TP
5	TN	TP
6	TN	FN
7	TN	TP
8	TN	TP
9	TN	FN
10	TN	TP

Table above shows the classification result for bmp images. For this random bmp images, the same detection process is used again to classify them in suitable classes. Stegogramme for all the images is produced using InfoStego software. After embedding the secret message inside the image, pairs algorithm is used to detect all the images, including cover images.

For cover image, the detection process has correctly classified nine images and detects one cover image as stegogramme.

For stegogramme, there are eight images is classified as true positive and two are classified as false negative.

VII. CONCLUSION

Steganography is one of the unique ways of communication. Transmitting secret message or file is much easier using this technique. Nowadays, there is lots of steganographic software available, some can be downloaded from the internet and some of them can be purchased. Different software uses different type of image format. Some of them uses image as carrier file and some of them hides message in audio or video format. In our research, we are focusing on hiding message in image file.

From the result that we got, we can conclude that our objective to reduce the rate of FN in using Pairs Analysis is

achieved by choosing 0.02 as the most appropriate threshold value.

For future research, other type of image format can be used to be tested by Pairs Analysis algorithm.

REFERENCES

- [1] J.Fridrich, M.Goljan, and D. Soukal, , "Higher-Order Statistical Steganalysis of Palette Images," University of Binghamton, 2003
- [2] Ker, A., "Quantitative Evaluation of Pairs and RS Steganalysis", Oxford University, 2004
- [3] A. Westfeld, A. Pfitzmann, " Attacks on Steganographic System", Dresden University of Technology, Germany.
- [4] E.Bentley, "RLO:Sensitivity and Specificity", Graduate Entry Medical School, University of Nottingham, 2007.
- [5] G.Vanagas, "Receiver Operating Characteristics Curves and Comparison of Cardiac Surgery Risk Stratification Systems", Interactive Cardio-Thoracic Surgery, 2004
- [6] L.K Westin, "Receiver Operating Characteristics Analysis: Evaluating Discriminance Effects among Decision Support System", Umea University, Sweden.
- [7] David M.Lane, "HyperStat Online Statistics Textbook", Rice University, 2007