ROC Curves for Steganalysts

Andreas Westfeld

Technische Universität Dresden
Institut für Systemarchitektur
01062 Dresden
mailto:westfeld@inf.tu-dresden.de

1 Introduction

There are different approaches in the literature for the assessment of steganographic algorithms and steganalytic attacks. In the early papers it was considered sufficient to show the existence of an effect for one or a few examples only. The more the area of steganography evolved, the more diverse became the goals and the harder to measure the improvements. Many branches of science are facing the same problem. More and more elaborate methods are used for assessment. We discuss aspects of the analysis of receiver operating characteristics (ROC) from a steganographer's point of view. ROC curves permit a reliable assessment of steganalytic detectors, independent of their decision threshold.

2 Output of a Detector

There are a number of detectors for steganography. Some return a binary decision (something embedded/nothing embedded). In most cases this decision is based on comparison with a predefined threshold. The reliability can be judged by the detector's error rates. We can distinguish

type I errors, which occur if a message is detected in a pristine carrier medium (false positives), and
type II errors if a steganogram is falsely considered "clean" (false negatives).

In quantitative steganalysis detectors estimate the relative length of the embedded message (embedding rate: 0… nothing embedded, 1… full capacity used). Precision is a further criterion for assessing these quantitative methods [1]. However, a lower error rate goes along with a more precise estimation in most cases.

Our showcase is a slightly idealised attack that returns the estimated embedding rate. We want to assess the reliability of this attack. We have a set of carrier media (e. g., images) in which we embed messages with rate 0.1, yielding a set of steganograms. The detector returns Gaussian distributed values with mean $\mu_1 = 0$ for carriers and $\mu_2 = 0.1$ for steganograms (cf. Fig. 1). Ker's criterion [2] requires at most 5 % type I errors (false positive rate, FPR) for 50 % detection rate (true positive rate, TPR), or $\mathrm{FPR}_{0.5} \leq 0.05$. The attack under investigation just fulfils this criterion: only 5 % of the carrier files cause a detection value ≥ 0.1. The standard deviation for carriers is $\sigma_1 = 0.0608$.

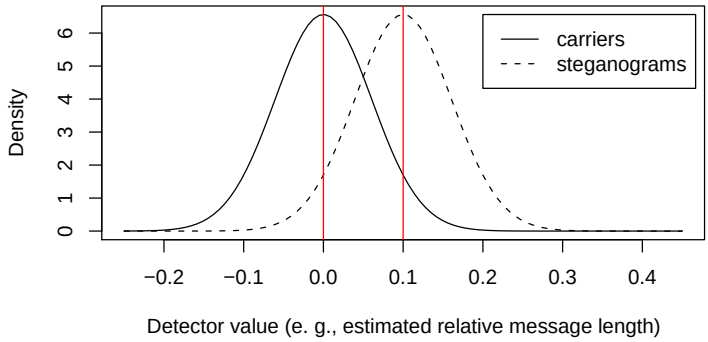

Fig. 1. Density of detector values for carriers (solid) and steganograms (dashed)

A ROC curve is a plot of the true positive rate with respect to the false positive rate. The ROC curve for a perfect detector reaches the point at FPR=0 and TPR=1. For practical results (finite number of points), the ROC curve consists exclusively of vertical and horizontal line segments. Figure 2 shows the ROC curve for a small number (left) and a large number of tested files (right). The detection power is measured by the reliability ρ, which is twice the area under the curve minus one. Note that this measure is independent of any thresholds. Each point of the curve corresponds with a particular threshold. In theory the threshold equal to the median of the detector values for carriers (which is identical to the mean μ_1 in our example) is associated with FPR=0.5 and the threshold equal to the median of the detector values for the steganograms (identical to μ_2 in our example) is associated with TPR=0.5. The deviation from these values decreases with increasing number of observations.

The false positive rate at 50 % detection rate ($\mathrm{FPR}_{0.5} \leq 0.05$) is independent of the standard deviation for steganograms σ_2. In practice, this deviation may vary. The effect of this variation is shown in Fig. 3. The false positive rate is constant for different σ_2 and fixed threshold.

Finally, the curve shape is influenced by the skedasticity of the distribution. Figure 4 shows a detector that returns Cauchy distributed results. This second showcase also just meets Ker's criterion, because the 0.95 quantile of carrier results and the median of the steganogram results coincide.

The Cauchy distribution with its heavy tails results in a ROC curve with a sharper corner (cf. Fig. 5). Note that the curve is convex outside the interval $0 = \mu_1 \leq \mathrm{FPR} \leq \mu_2 = 0.1$ for Cauchy distributed errors. The reliability can be improved by randomising the detection for results below μ_1 and above μ_2, which will replace the convex parts of the curve with straight lines.

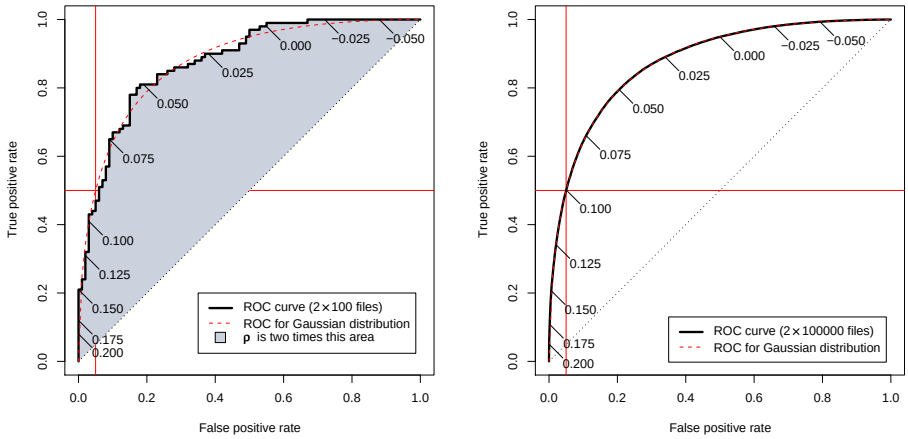

Fig. 2. Receiver operating characteristics for 100 files (left) and 100,000 files (right)

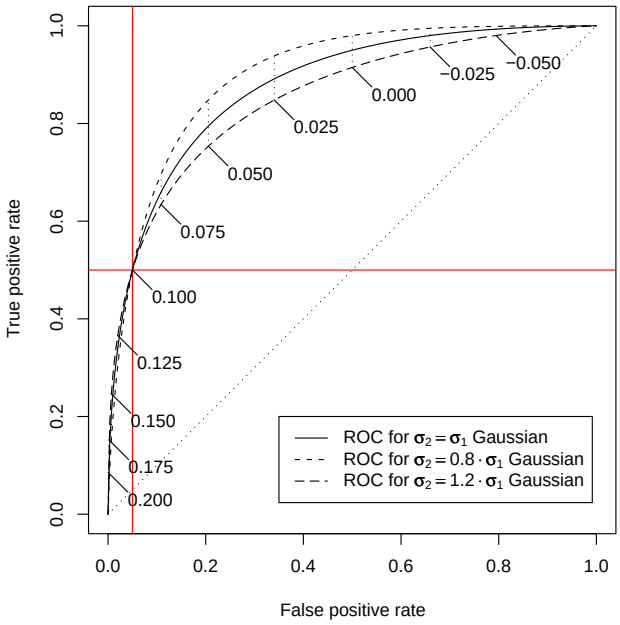

Fig. 3. Varying standard deviations of detection values for steganograms

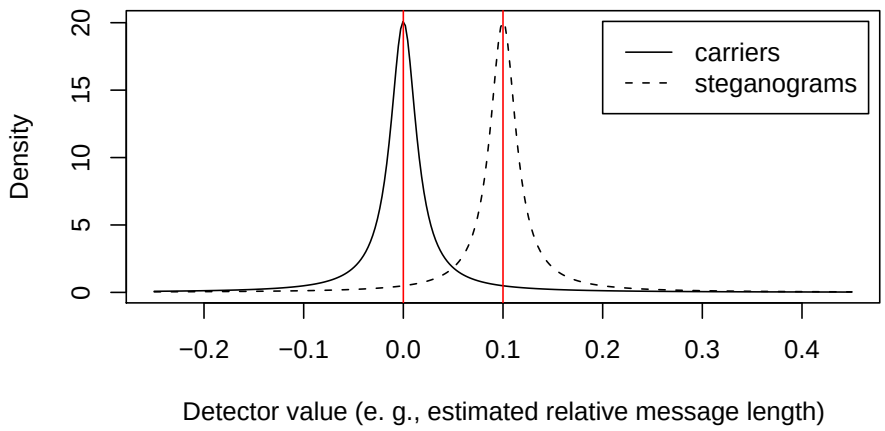

Fig. 4. Density of Cauchy (t_1) distributed detector values for carriers (solid) and steganograms (dashed)

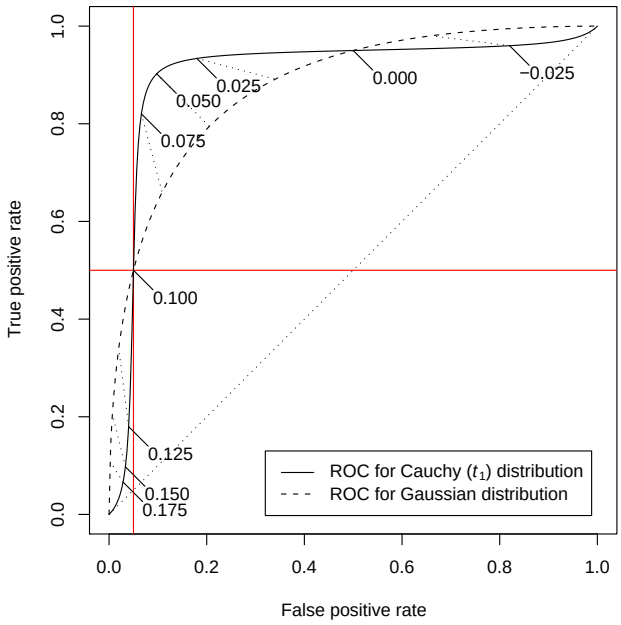

Fig. 5. Correspondences between ROC curves for Cauchy and Gaussian distributed detector results

3 Results

In this section we will compare the stability of five measures from the literature, namely

1. the area under the curve (AUC), converted to the reliability [3],

$$\rho = 2 \cdot \mathrm{AUC} - 1$$

2. the measures by Lyu and Farid [4], who measured the true positive rate at 0 % false positives (TPR_0) and
3. at 1 % false positives ($\mathrm{TPR}_{0.01}$),
4. the equal error rate (EER), which is mostly used in biometrics where the decision threshold is set in order to have approximately equal numbers of false positives and false negatives, as well as
5. the false positive rate at 50 % detection rate ($\mathrm{FPR}_{0.5}$), which is adapted from Ker's criterion [2].

Table 1. Simulated confidence of the resulting reliability measures for 2×1000 files with *Gaussian* distributed results (100,000 fold repetition)

	TPR_0	$\mathrm{TPR}_{0.01}$	$\mathrm{FPR}_{0.5}$	EER	ρ
Median	0.060	0.255	0.050	0.205	0.755
2.5 % quantile	0.007	0.180	0.035	0.188	0.725
97.5 % quantile	0.153	0.334	0.067	0.223	0.784
corresponding lower ρ	—	0.682	0.711*	0.719*	0.725
corresponding upper ρ	—	0.820	0.800*	0.789*	0.784
Rank by confidence	5	4	3	2	1

* The lower ρ of $\mathrm{FPR}_{0.5}$ and EER corresponds to the upper quantile.

Table 1 presents the results of 100,000 simulated ROC curves. Each ROC curve was set up using 1000 Gaussian random numbers for carrier results ($\mu_1 = 0$, $\sigma_1 = 0.0608$), and another 1000 for steganograms ($\mu_2 = 0.1$, $\sigma_2 = \sigma_1$). The parameters have been chosen to cause a median of 0.05 for the $\mathrm{FPR}_{0.5}$ (Ker point). We determined the five measures for each of the ROC curves. The table presents three quantiles: 0.5 (median), 0.025, and 0.975. The latter two are the bounds of a 95 percent confidence interval. Since we know the distribution of the "detection" results, we can express the boundaries of this interval in terms of the reliability ρ. The theoretical reliability for Gaussian detector results is $\rho = 0.7552$ if the curve goes through the point $\mathrm{FPR}_{0.5}$=0.05. For the lower $\mathrm{FPR}_{0.5}$=0.035, which is the 2.5 percent quantile in our simulation, the corresponding curve has a higher reliability $\rho = 0.800$ (cf. "corresponding upper ρ" in Table 1). The 97.5 percent quantile of $\mathrm{FPR}_{0.5}$=0.067 corresponds to the lower $\rho = 0.711$. On this

basis, it is possible to compare the 95 percent confidence intervals. We can score the measures according to their confidence interval. The only exception is TPR_0, which is 0 in theory. The finite number of 2×1000 observations per ROC curve leads to the nonzero TPR in the table. This finding is very dependent on the number of observations. A continuity correction would be needed here, since we rather measure the TPR at $\mathrm{FPR} = \frac{1}{2000}$ instead of 0. However, it is obvious that the TPR_0 is the most volatile measure.

Table 2. Simulated confidence of the resulting reliability measures for 2×1000 files with *Cauchy* distributed results (100,000 fold repetition)

	TPR_0	$\mathrm{TPR}_{0.01}$	$\mathrm{FPR}_{0.5}$	EER	ρ
Median	0.001	0.013	0.050	0.098	0.805
2.5 % quantile	0.000	0.004	0.037	0.085	0.771
97.5 % quantile	0.005	0.031	0.064	0.111	0.837
corresponding lower ρ	—	−0.973	0.754*	0.778*	0.771
corresponding upper ρ	—	0.941	0.854*	0.830*	0.837
Rank by confidence	5	4	3	1	2

* The lower ρ of $\mathrm{FPR}_{0.5}$ and EER corresponds to the upper quantile.

Table 2 presents the results of 100,000 simulated ROC curves. This time, each ROC curve was set up using 1000 Cauchy distributed random numbers for carrier results (location 0, scale 0.0158), and another 1000 for steganograms (location 0.1, scale 0.0158). Again, the parameters have been chosen to cause a median of 0.05 for the $\mathrm{FPR}_{0.5}$ (Ker point). The confidence intervals are larger than in the previous Gaussian case. For the $\mathrm{TPR}_{0.01}$ we found the 2.5 percent quantile below the diagonal, yielding $\rho < 0$. The confidence intervals of the EER and the reliability ρ are still very close to each other. For the Cauchy distribution, the EER is slightly more stable. This shows a dependency between the error distribution and the stability of the measure. In our future work will investigate this finding for particular error distributions of steganographic attacks.

Acknowledgements

Travel to the Information Hiding Workshop was supported in part by the European Commission through the IST Programme under contract IST-2002-507932 ECRYPT. This work was partially supported by the Air Force Office of Scientific Research, Air Force Material Command, USAF, under the research grant number FA8655-06-1-3046.

References

1. Böhme, R.: Assessment of steganalytic methods using multiple regression models. In Barni, M., Herrera-Joancomartí, J., Katzenbeisser, S., Pérez-González, F., eds.: Information Hiding (7th International Workshop). Volume 3727 of LNCS., Berlin Heidelberg, Springer-Verlag (2005) 278–295
2. Ker, A.D.: Improved detection of LSB steganography in grayscale images. In Fridrich, J., ed.: Information Hiding (6th International Workshop). Volume 3200 of LNCS., Berlin Heidelberg, Springer-Verlag (2004) 97–115
3. Fridrich, J.: Feature-based steganalysis for JPEG images and its implications for future design of steganographic schemes. In Fridrich, J., ed.: Information Hiding (6th International Workshop). Volume 3200 of LNCS., Berlin Heidelberg, Springer-Verlag (2004) 67–81
4. Lyu, S., Farid, H.: Detecting hidden messages using higher-order statistics and support vector machines. In Petitcolas, F.A.P., ed.: Information Hiding (5th International Workshop). Volume 2578 of LNCS., Berlin Heidelberg, Springer-Verlag (2003) 340–354