A Sequential Method for Online Steganalysis

Rémi Cogranne
Troyes University of Technology - ICD - LM2S - ROSAS - UMR 6281 CNRS - Troyes, France
remi.cogranne@utt.fr

Copyright ©2015 IEEE. Personal use of this material is permitted. However, permission to use this material for any other purposes must be obtained from the IEEE by sending a request to pubs-permissions@ieee.org
Accepted version. Final to be published online on ieeexplore.ieee.org within WIFS proceedings.

***Abstract*—This paper studies online detection of hidden information in digital images. By online, it is meant that we inspect a flow of images that are transmitted sequentially. This has crucial consequences on the detection of hidden data. By contrast to the usual detection settings, the delay before detection has to be considered in the definition of the correct detection probability, or power function. Similarly, the false alarm probability is considered with respect to a number of inspected cover images. In this paper a new sequential detection method is proposed with the goal to maximize the detection accuracy for a prescribed detection delay. The study of proposed sequential test allows the establishing of detection power as a function of detection delay and also permits us to bound the probability of false alarm for a given number of cover images.**
Numerical results on real image using both the well-known WS type of detector and the recent ensemble classifier show the relevance of the proposed approach and the accuracy of the theoretical findings.

***Index Terms*—Hypothesis testing theory, Sequential detection, optimal detection, Pool steganalysis, Steganography.**

I. Introduction

STEGANOGRAPHY is often referred to as the science of covert communication. On the one hand, the goal of steganography is to hide secret data within a digital objet to create an innocuous-like stego-object. On the other hand, the goal of steganalysis is to inspect digital object in order to detect the presence of hidden data. Steganography is much more mature with digital images as cover files, this paper thus focuses on digital images though the proposed methodology is general and can be applied for a wide range of steganalysis.

Steganography has been evolving at a fast rate since two decades. Modern steganographic methods are content adaptive, that is, the embedding changes are concentrated in areas of the cover image where they are expected to be the most difficult to detect. On the other hand, steganalysis methods have also been remarkably developed to detect with better and better accuracy modern steganographic schemes. Four approaches have mainly been used in steganalysis:

1) Historically, first detectors belong to the class of *structural detectors* that focus on the detection of Least Significant Bit (LSB) replacement using pixels' correlation, see [2] for an overview.
2) Then, the well-known Weighted-Stego image (WS) algorithms has been introduced in [3] and deeply studied in [4]. While it originally aims at estimating the payload, or hidden message length, it has been shown in [5], [6] that the WS detector is closely related to the optimal Likelihood Ratio Test (LRT) under the assumption that pixels' noise is Gaussian.
3) This re-opened the field of optimal detectors, as referred to in [7], that uses a statistical model of cover and steganographic images and exploit hypothesis testing theory to design a test that is optimal for a given criterion. The first detectors from this category used simplistic image model [8] and were latter improved and enlarge for LSB matching detection [9] and JPEG images [10], [11].
4) Last, most of the modern steganalysis methods are built by extracting a set of features that reveal the presence of hidden data and on supervised learning to train a classifier to distinguish between cover and stego images. The firsts feature-based steganalysis methods used the Fisher Linear Discriminant (FLD) analysis and dozens of features. Soon, more accurate learning algorithms have been used such as the popular Support Vector Machine (SVM) and the number of extracted features grew to more than 30 000 for recent rich models [12]. This improvement of feature dimension has been enabled with the ensemble classifier [13] as an efficient alternative learning methods for such high dimensionality.

While the "optimal detectors" can not achieve the detection accuracy of feature-based approaches, they have the indisputable advantage to provide a statistical test with analytic expression of its statistical properties. On the opposite, the statistical properties of features-based steganalysis methods are only known empirically. Note that, it has recently been proposed to formulate the ensemble classifier within the framework of hypothesis testing [14], [15]. This prior work models the projection of ensemble base learners as a multivariate Gaussian which allows the replacing of majority vote decision rule by an optimal LRT and to theoretically study the

Copyright (c) 2015 IEEE. Personal use of this material is permitted. However, permission to use this material for any other purposes must be obtained from the IEEE by sending a request to pubs-permissions@ieee.org
Accepted version. Final to be published online on ieeexplore.ieee.org within WIFS proceedings.

performance of the ensemble classifier.

In this paper, the advantage of steganalysis methods with known statistical properties is used to study their application in a sequential context, when images are transmitted and inspected sequentially, one after the other.

A. Contribution and Organization of this Paper

As noted in [7], the moving of steganalysis "from laboratories to the real life" implies several challenging problems among which the problem of pooled steganalysis received little attention. One of few notable prior work that studies a similar problem [17] proposed to gather images from different users and to identify those who transmit steganographic image, without prior reference, or cover dataset.

By contrast, in this paper, it is proposed to study the problem of online steganalysis, when users are sending images sequentially, one after the other, and when a cover dataset reference is possibly available. The problem is cast within the framework of optimal detection, that is by using hypothesis testing methods, we propose a sequential statistical test for which the probability of false alarm and the detection accuracy are known. It is important to note that, the definition of false alarm and correct detection has to be reformulated to match the problem of sequential steganalysis of several images.

The main contributions of this paper are the following:

1) First the problem of sequential steganalysis for inspection of a flow of images is stated. A general framework is proposed regardless of the strategy used for spreading the hidden data within several images.
2) Then a sequential detection method is proposed and its statistical performance is analytical established. While usual sequential methods focus on mean detection and mean time to false alarm, the proposed approach focuses on maximizing the probability of correct detection under the constraint of a maximal detection delay.
3) Last, and not least, numerical results on real images show the sharpness of the theoretical results.

The present paper is organized as follows. Section II cast the problem of sequential analysis from a flow of transmitted images within the framework of hypothesis testing theory. This first section also presents the definition of detection accuracy used in sequential analysis and recalls the usual methods for sequential detection. Then Section III presents the proposed sequential test and studies its statistical properties, that is false alarm probability for a fixed analysis run length and detection power as a function of detection delay. Section IV presents numerical results that show the relevance of the proposed methodology for different type of steganalysis method and for both spatial and JPEG domain steganographic methods. Finally, Section V summarizes the present work and concludes the paper.

II. Online Steganalysis: Problem Statement

In this paper we do not make any assumption on the steganalysis method used by the warden, as well on the strategy used by steganographer for spreading hidden information over several images. The general framework for representing the warden considers that each image, after have been analyzed, is reduced to a scalar x_i. Let us suppose that the warden has observed N images, then has a sequence $\{x_n\}_{n=1}^N$. In the present paper, the sequential detection problem is modeled as follows: after having sent an unknown number $\nu-1$ of cover images, the steganographer starts embedding hidden message within the following images, with, by Kerckhoff's principle, a known embedding algorithm. Hence the sequence $x_1,\dots,x_N$ can be modeled as:

$$\begin{cases} x_n \sim \mathcal{P}_0 \,,\, \forall n < \nu, \\ x_n \sim \mathcal{P}_{\theta_n} \,,\, \forall n \geq \nu \,,\, \theta_n \geq 0. \end{cases} \tag{1}$$

Here, in Equation (1), it is assumed that before the ν-th image, the steganalysis results x_n are i.i.d (independent and identically distributed) with distribution $\mathcal{P}_0$. On the opposite, the embedding of hidden information, that starts at image number ν changes the distribution of warden analysis results' $x_n \sim \mathcal{P}_{\theta_n}$ where θ_n represents a parameter of the distribution that, without loss of generality, is increased by the use of steganography. Note that it is possible that the steganographer does not use all the images after ν; without loss of generality, it is assumed that $\forall n \geq \nu \,,\, \theta_n \geq 0$ with the inequality being an equality if no data is hidden within an image.

Finally, it should be emphasized that the present paper focus on the case in which the statistical distribution of x_n is established, before and after embedding starts. However, the proposed can be extended when x_n is a decision result that, for instance, takes value in $\{0;1\}$. In such a case x_n would be drawn from a Bernouilli trial with a, possibly unknown, probability p that x_n is 1. In fact in both cases, the problem of the warden can be modeled as a sequential detection of a change of the distribution parameter θ_n.

A. Criterion of Optimality

When dealing with sequential detection, as presented in Equation (1), the goal is to detect the image number at which embedding starts ν as quickly as possible. This problem is known in the literature as the quickest changepoint detection. Formally, a sequential test is a mapping $\delta_N : \mathbb{R}^N \to \{0;1\}$ which, based on the observations $x_1,\dots,x_N$ decide between the two following hypotheses:

$$\mathcal{H}_0 : x_n \sim \mathcal{P}_0 \,,\, \forall n \in \{1,\dots,N\}, \tag{2}$$

$$\mathcal{H}_n : x_n \sim \begin{cases} \mathcal{P}_0 \,,\, \forall n < \nu, \\ \mathcal{P}_{\theta_n} \,,\, \forall n \geq \nu \,,\, \theta_n \geq 0 \end{cases} . \tag{3}$$

The stopping time T of a sequential changepoint detection is the smallest image index n for which $\delta_n(x_1,\dots,x_n) = 1$. When embedding is detected this can be erroneous if $T < \nu$ in which case a false alarm is raised. On the opposite, a flow of steganographic images is correctly detected made when $T \geq \nu$. With the quickest detection of change in mind, it is naturally desirable to minimize the detection delay $T-\nu$ which is referred to as the detection delay.

Since the detection is made sequentially, it does not make

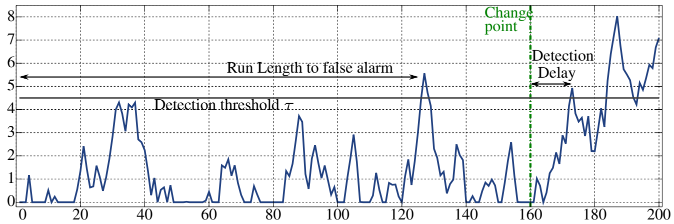


Fig. 1: Illustrative example of a changepoint detection criteria of accuracy with simulated data.

any sense to measure the probability of false alarm for each sequence $x_1, \ldots, x_n$. Instead, the most usual criterion in changepoint detection is the average run length (ARL) to false alarm, that is, the expected number of cover images tested and leading to the false-alarm (erroneous decision that embedding starts):

$$ARL = \mathbb{E}\left[T | T < \nu\right]. \tag{4}$$

Because the ARL does not provide any guarantee on false alarm probability, another criterion that may be used is the the probability of false alarm over a run length, or a number of inspected cover images L, denoted $\alpha_0^{(L)}$ and defines for the i.i.d case by:

$$\alpha_0^{(L)} = \mathbb{P}\left[T \leq L | L < \nu\right] \tag{5}$$

Similarly, the power of a test, defined as the probability of correctly detecting the use of a steganographic algorithm for a "snapshot" test, has very few meaning in sequential detection. Instead it is usually considered that the average detection delay (ADD) should be considered:

$$ADD = \mathbb{E}[T - \nu | T \geq \nu]. \tag{6}$$

The ADD represents the number of images, after the embedding startpoint, that would be required in average to correctly detect steganography.

However, similar to the ARL, the average number does not provide any guarantee on the detection accuracy. Hence, when it is preferable to reliably detect the use of steganography with a fixed detection delay, that is with at most M images after embedding start, the probability of detection before M images has to be used:

$$\beta^{(M)} = \mathbb{P}\left[\nu \leq T \leq \nu + M\right]. \tag{7}$$

The notions of detection delay and run length are illustrated in Figure 1.

B. Changepoint detection: Brief Review

In the literature of sequential analysis, two approaches which are briefly presented below are generally used, see [19] for a detailed review. First, when the changepoint is $\nu = 0$ and it is desired to decide between two hypotheses with as few samples a possible, the Sequential Probability Ratio Test (SPRT) is generally used because it has been shown to be optimal with respect to minimizing average time to detection and error probabilities. The SPRT is associated with the following decision rule:

$$\delta_N^{\text{sprt}} = \begin{cases} 0 & \text{if } \lambda_N = \Sigma_{n=1}^N \Lambda(x_n) < \tau_0, \\ 1 & \text{if } \lambda_N = \Sigma_{n=1}^N \Lambda(x_n) > \tau_1, \end{cases} \tag{8}$$

$$\text{with} \quad \Lambda(x_n) = \frac{f_0(x_n)}{f_\theta(x_n)}. \tag{9}$$

Here f_0 and f_θ are the probability distribution functions associated with distributions $\mathcal{P}_0$ and $\mathcal{P}_{\theta_n}$ and Λ is the Likelihood Ratio (LR) between the two distributions. Note that for the sake of clarity it is assumed here that θ is fixed and known after embedding starts. The two thresholds τ_0 and τ_1 are set to respect constraints on error probabilities. When the SPRT cannot decide between the two hypotheses, this means that the inspection should continue with one more image.

For the problem of changepoint detection, with unknown ν, the most widely used method is the cumulative sum (CUSUM) initially proposed in [18]. The CUSUM method is based on the following decision rule:

$$\delta_N^{\text{cusum}} = \begin{cases} 0 & \text{if } S_N = \max\left\{0; S_{N-1} + \Lambda(x_N)\right\} \leq \tau, \\ 1 & \text{if } S_N = \max\left\{0; S_{N-1} + \Lambda(x_N)\right\} > \tau, \end{cases} \tag{10}$$

with initialization $S_0 = 0$. The CUSUM is thus essentially a cumulative sum of LR, as the SPRT, which is reset to 0 when its value is negative. This can be explained by the fact that the LR has a negative expectation under $\mathcal{H}_0$ that inspected images are cover and hence will decrease in average. Resetting the CUSUM to 0 avoid the negative drift of the SPRT and hence allows a much quicker detection.

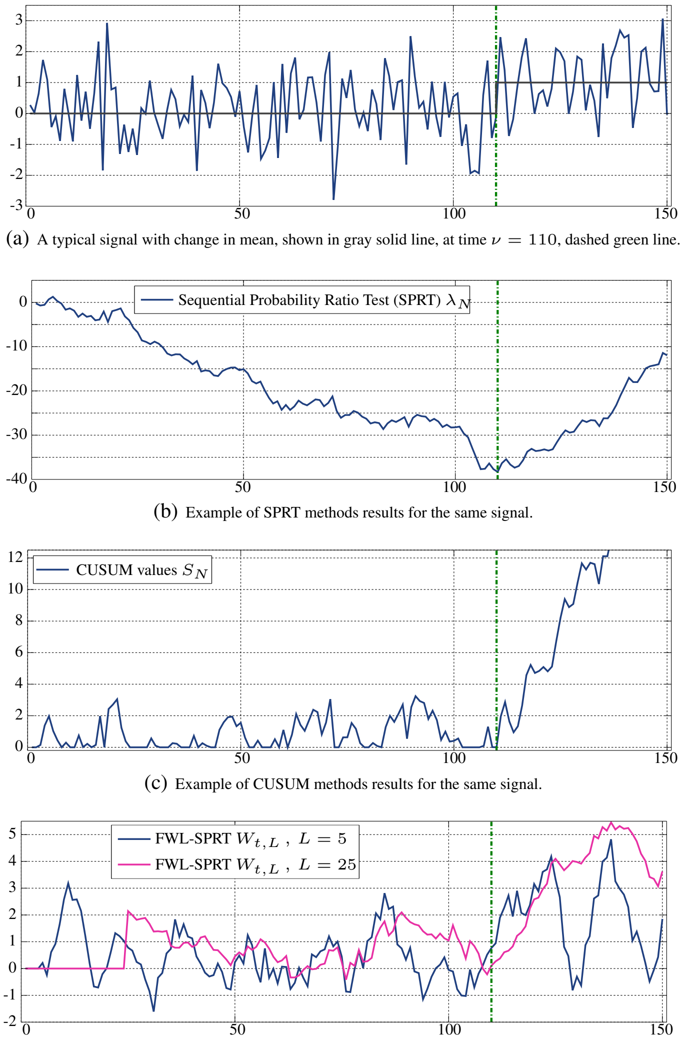
The popularity of the CUSUM may be explained by the fact that its optimality, with respect to minimizing worst case average run length to false-alarm and detection delay, has been proved in several cases (see [19] for a details).

However, the criterion of minimizing false-alarm probability for a given run length has seldom been studied and the changepoint detection method is rarely provided with established statistical properties in term of probability of errors. Such a study has only recently been proposed in the context of transient change detection [20], [21] and it has motivated the present study.

An illustrative example of results from the SPRT and the CUSUM are given in Figures 2.Obviously the SPRT is slowly decreasing before the embedding starts, because it aims at detecting if from the beginning the inspected flow of images is made of only cover or of only stego images. On the opposite the CUSUM, that aims at reacting as quickly as possible after the embedding start, remains around zero until changepoint.

III. Proposed Changepoint Detection Method for Online Steganalysis

The present paper lies within the field of optimal detection and, hence, focuses on guaranteeing a false-alarm probability for a given run-length, or number N of inspected cover images and aims at designing a statistical changepoint detection methods for which the corresponding power function for a



(a) A typical signal with change in mean, shown in gray solid line, at time $\nu = 110$, dashed green line.

(b) Example of SPRT methods results for the same signal.

(c) Example of CUSUM methods results for the same signal.

(d) Example of sliding / Window limited SPRT methods results for the same signal.

Fig. 2: Comparison between the proposed LR test and the majority vote decision rule for spatial domain steganalysis.

fixed detection delay is established.

Let us also recall that the proposed method is used in this paper with the WS-type statistical test [5] and with the ensemble classifier with LRT decision rule [14], [15], though it can be extended to wider range of steganalysis methods. The two steganalysis methods on which the proposed sequential method relies share a common model for the detection statistics. In fact, in both case, when inspecting a cover image their results has been shown to follow a Gaussian distribution with zero mean and unit variance. On the opposite, when inspecting steganographic images, their results follow a Gaussian distribution still with unit variance while the mean $\theta > 0$ is increased. However, a fundamental difference, that is discussed in numerical results, is that the mean θ depends, for the WS, on the payload R but also on pixels' variance. On the opposite, because ensemble classifier is trained on a whole dataset of cover and stego images, the test [14], [15] is designed such that all stego image share the same expectation. We also note that because both tests used are optimal whatever the payload, or more precisely Uniformly Most Powerful, the payload has small influence.

Hence, each steganalysis result x_n can be modeled as:

$$x_n \sim \begin{cases} \mathcal{N}(0,1) & \text{if inspected image is a cover ,} \\ \mathcal{N}(\theta_n,1) & \text{if inspected image is a stego .} \end{cases} \quad (11)$$

And the goal is to detect as quickly as possible the embedding starting image ν at which the expectation change from 0 to $\theta_n > 0$. In this paper it is proposed to target a fixed maximal detection delay M, after which the detection is considered as "too late". Hence, the proposed sequential method has a fixed window length (FWL) M that coincide with the maximal detection delay. The proposed FWL-SPRT test is based on the following decision rule:

$$\delta_N^{\text{fwl}} = \begin{cases} 0 & \text{if } W_{N,M} = \Sigma_{n=N-M+1}^{N} x_n \leq \tau, \\ 1 & \text{if } W_{N,M} = \Sigma_{n=N-M+1}^{N} x_n > \tau. \end{cases} \quad (12)$$

We note that here for used steganalysis methods [5] and [14], [15], the x_n's correspond to values of likelihood ratios. This allows a transparent comparison with the CUSUM and the SPRT. First, the proposed methodology essentially consists in a SPRT which is only applied on the M lasts steganalysis results, as opposed to the SPRT which is computed using all results. More precisely, from Equations (12) and (8) it is obvious that $W_{N,M} = \lambda_N - \lambda_{N-M}$. For this reason the proposed method is referred to as a "Fixed Window Length SPRT" (FWL-SPRT). This process, similar to the reset to 0 for the CUSUM, naturally avoids the drift of the SPRT when the flow of inspected images begins with covers. Besides, the CUSUM may also be re-written as $S_N = \lambda_N - \min_{1 \leq n \leq N} (\lambda_n)$, see details in [19, Chap.8.2], that is it uses an adaptive threshold based on the minimal SPRT value instead of M lasts.

A. Statistical Properties of the Proposed Method

Let us start by noting that the FWL-SPRT $W_{\nu+M-1,M}$, which is last one for which the steganalysis detection is possible under maximal delay M constraint, is given by:

$$W_{\nu+M-1,M} = \Sigma_{n=\nu}^{\nu+M-1} x_n$$

where, from Equation (11) $x_n \sim \mathcal{N}(\theta_n, 1)$. Hence, from the independence of images, it is immediate that:

$$W_{\nu+M-1,M} \sim \mathcal{N}(\|\boldsymbol{\theta}_\nu\|_1, M), \quad (13)$$

$$\text{with } \|\boldsymbol{\theta}_\nu\|_1 = \sum_{n=\nu}^{\nu+M-1} |\theta_n| = \sum_{n=\nu}^{\nu+M-1} \theta_n,$$

the last equality being due to the fact that $\theta_n \geq 0$ for all $n \in \{0, \ldots, T\}$.

It should be noted that the distribution of $W_{\nu+M-1,M}$ as given in Equation (13) is made without any assumption on the strategy used to spread the payload within several images. Hence, it especially includes two specific cases which are embedding with constant payload after embedding start ν and the case in which the embedding is made in only a selected subset of images, and cover images are still sent, for instance in order to try avoiding detection. In this later case, the θ_n will be 0 for cover images.

The distribution of $W_{N,M}$ for cover images is more difficult to establish. In fact, the main difficulty is that $W_{N,M}$ and $W_{N+1,M}$, for instances, are computed using $M-1$ same images. Hence the correlation between $W_{N,M}$ and $W_{N+1,M}$ will be important. Because of this, it is hardly possible to establish the false-alarm probability over a run length of size L. To avoid the difficulty it is proposed to bound the probability $\alpha_0^{(L)}$. Without loss of generality let us assume that $L=\ell M$ with $\ell \in \mathbb{N}$. It is proposed in this paper to bound $\alpha_0^{(L)}$ by above by noting that $\forall n \in \{1,\dots,N-M+1\}$:

$$\mathbb{P}[\max\{W_{n,M},\dots,W_{n+M-1,M}\} > \tau] \geq \mathbb{P}[W_{N,M} > \tau].$$

On the opposite, since the correlation between FWL-SPRT values $W_{N,M},\dots,W_{N+M-1,M}$ is positive is follows that $\forall t \in \{1,\dots,T\}$ [20]:

$$\mathbb{P}[\max\{W_{n,M},\dots,W_{n+M-1,M}\} > \tau] \leq \sum_{t=n}^{n+M-1} \mathbb{P}[W_{N,M} > \tau].$$

Those observations allow us to bound both the power for a maximal detection delay $\beta^{(M)}$ and the probability $\alpha_0^{(L)}$ of false alarm for a run of fixed length L.

Proposition 1. *For any decision threshold $\tau \in \mathbb{R}$ the false alarm probability over $L=\ell M$ inspected cover image is bounded by:*

$$1-\left[\Phi\left(\frac{\tau}{\sqrt{M}}\right)\right]^{\ell} \leq \alpha_0^{(L)} \leq 1-\left[\Phi\left(\frac{\tau}{\sqrt{M}}\right)\right]^{L}, \qquad (14)$$

with Φ and Φ^{-1} the standard normal cumulative distribution function and its inverse.
For any decision threshold τ, the power of the proposed FWL-SPRT is bounded:

$$\beta^{(M)} \geq 1-\Phi\left(\frac{\tau-\|\boldsymbol{\theta}\|_1}{\sqrt{M}}\right) \qquad (15)$$

Since the study of sequential detection method is very difficult, the previous Proposition 1 only offer bounds on the statistical performances of the proposed FWL-SPRT. However, this allows to set a threshold for guaranteeing a false alarm probability for L cover images and to guarantee a minimal detection accuracy.

IV. Numerical Results

The numerical results presented in this paper has been obtained from BOSS database version 1.01 [22], made of $10,000$ grayscale images of size. The proposed method relies on steganalysis detector for which the statistical distribution of the results is known. Hence, the WS-type statistical test proposed in [5] from "optimal detectors" has been used at first.To enlarge the application of the proposed detection method to feature-based approaches, the recently proposed approach for establishing the statistical properties of the well-known ensemble classifier has been used. In this paper we only present, for spatial domain, results obtained using the

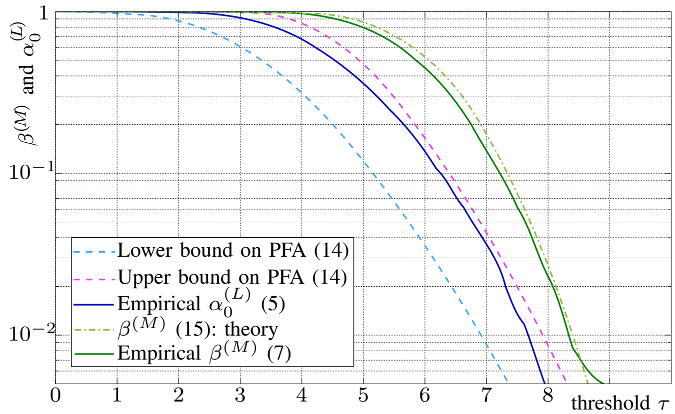


Fig. 3: Comparison between empirical and theoretical probability of false alarm for run length $L=100$ and power for $M=5$ with the WS-like statistical test [5] for LSB replacement embedding at payload $R=0.025$.

improved version of the Spatial Rich Model (SRM) [12], called maxSRMd2 [23], that takes into account the selection channel. The embedding method used is the recent embedding scheme [24], [25], that is based on statistical detectability of LSB matching under a multivariate Gaussian model, with payload 0.2 bit per pixel (bpp). For the JPEG domain, the presented results were obtained using the recent Phase aware projection model (PHARM) feature set [26] and Uniform Embedding Distortion (UED) [27] steganographic algorithm at payload 0.2 bit per non zero-coefficients.
It should be noted that several other embedding schemes and feature sets have been tested, for both spatial and JPEG domain, and similar trends has been observed.

First, because the main goal of this paper is to design a changepoint detection method for steganalysis with known statistical properties, Figure 3 show a comparison between empirical and theoretical results for WS-type test [5]. This figure shows the false-alarm probability over a run length, $\alpha_0^{(L)}$ (5), with $L=100$ and the power for a maximal detection delay $\beta^{(M)}$ (7), with $M=5$. The empirical results obtained for LSB replacement and WS-type algorithm show the relevance of the theoretical bounds. However, we note that this type of algorithm is applied for each image independently, hence the statistical properties of individual results is reliable. On the opposite, Figure 4 show a comparison between empirical and theoretical bounds on false-alarm probability over a run length $L=5$ and $L=100$ for the ensemble classifier with LRT decision rule and spatial domain embedding [24], [25] with maxSRMd2 feature set [23]. While the bounds are rather tight, it should be noted that for large $L=100$ and large threshold value τ its accuracy becomes doubtful. In fact, since this type of steganalysis is trained on a wide range of images, the statistical properties of individual results are not as much reliable; training is based on properties of the whole training set. Hence, it is more likely to find images that are outliers with respect to the average image from training set.

For the JPEG domain steganalysis, Figure 5 presents a comparison between the theoretical power for a maximal

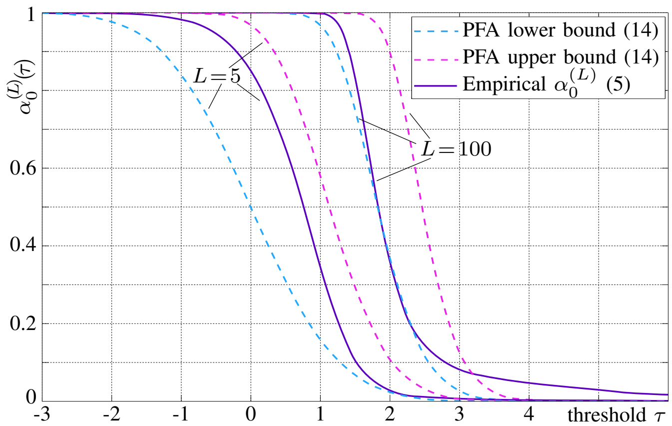


Fig. 4: Comparison between empirical and theoretical probability of false alarm for run length $L = 5$ and $L = 100$ for the ensemble classifier with LRT decision rule [14], [15]. The feature set used in maxSRMd2 [23] and the embedding scheme is [24], [25].

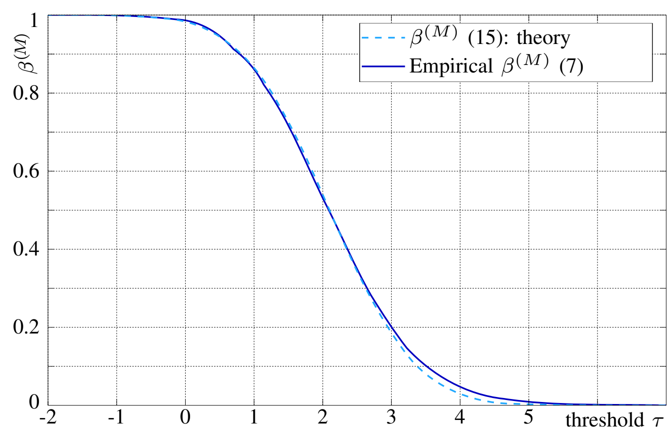


Fig. 5: Comparison between empirical and theoretical power for a maximal detection delay $\beta^{(M)}$, with $M = 5$, as a function of the detection threshold. The ensemble classifier with LRT decision rule [14], [15] is used with with PHARM feature set [26] and UED embedding scheme is [27].

detection delay $\beta^{(M)}$ as function of decision threshold τ. For this figure, PHARM feature set [26] and UED embedding scheme [27] are used. It should be noted that, in this case, the model for spreading the payload over different images is that after embedding start at image ν, 3 randomly selected images are used in the 10 next images (hence 7 are left as cover). Roughly speaking this strategy correspond to case in which the steganographer keep sending cover images in order to prevent detection. This Figure show the accuracy of the theoretical bound of the power for a maximal detection delay $\beta^{(M)}$, with $M = 5$. We also note here that the empirical results does not match perfectly the theoretical bound for large threshold because feature based steganalysis is trained for the whole dataset and hence can hardly deal with outliers images.

Finally, it is worth comparing the performance of the proposed FWL-SPRT methodology with the CUSUM algorithm. We recall here that the proposed approach rely on minimizing

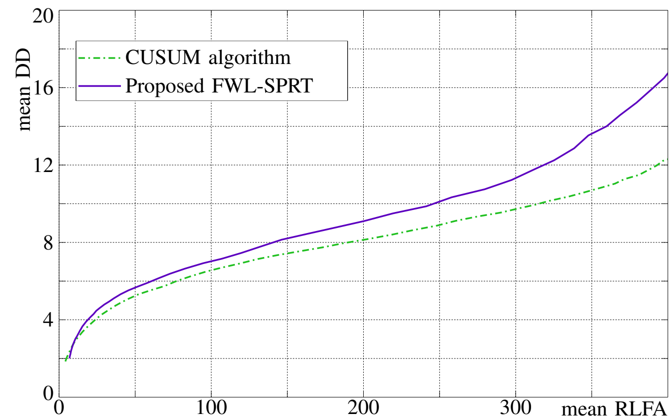


Fig. 6: Comparison between the proposed FWL-SPRT and the CUMSUM algorithm in term of average detection delay as function of average run length to false-alarm. Results obtained with ensemble classifier with LRT decision rule [14], [15] using maxSRMd2 feature set [23] and embedding scheme [24], [25].

probability of false alarm for a given run length while the CUSUM is optimal to minimize the average run length to false-alarm. However, it is proposed in Figure 6 to compare the CUMSUM and the proposed FWL-SPRT through the average detection delay as a function of average run length to false-alarm. While the CUMSUM has been shown to be optimal for such criterion in several case, we note that the proposed method only performs slightly worse. The results presented in Figure 6 are also obtained with the ensemble classifier with LRT decision rule using spatial domain embedding [24], [25] and maxSRMd2 features set [23].

V. Conclusion

This paper studies the problem of sequential steganalysis, when a flow of images has to be inspected, within the framework of "optimal detectors" relying on hypothesis testing. To the best of our knowledge, the problem of inspecting several images one by one with a goal to detect steganographic embedding startpoint as quickly as possible has never been studied. This paper presents the performance criterion that has to be considered for sequential steganalysis and briefly review usual sequential methods. Then a methodology is proposed with the aims of detecting when steganographic embedding start with a maximal detection delay. Bound are provided for the performance of the proposed method and their sharpness is verified on digital images using WS-type statistical test and ensemble classifier as an example of feature based steganalysis method.

Future works will study the application of the proposed method with steganalysis detector that output a binary results and study the optimal strategy to spread the payload with respect to the proposed test.

Acknowledgements

The work of Rémi Cogranne was funded in part by the STEG-DETECT program for scholar mobility and by IDENT research grant both from Conseil Régional de Champagne-Ardenne.

REFERENCES

[1] T. Filler, J. Judas, and J. Fridrich, "Minimizing additive distortion in steganography using syndrome-trellis codes," *Information Forensics and Security, IEEE Transactions on*, vol. 6, no. 3, pp. 920–935, Sept 2011.

[2] A. Ker, "A general framework for structural steganalysis of LSB replacement," in *Information Hiding*, ser. LNCS, vol. 3727, 2005, pp. 296–311.

[3] J. Fridrich and M. Goljan, "On estimation of secret message length in LSB steganography in spatial domain," *IS&T/SPIE Electronic Imaging conf.*, vol. 5306, Security, Steganography, and Watermarking of Multimedia Contents, 2004, pp. 23–34.

[4] A. D. Ker and R. Böhme, "Revisiting weighted stego-image steganalysis," in *Security, Forensics, Steganography, and Watermarking of Multimedia Contents X, Proc. SPIE 6819*, 2008, pp. 501–517.

[5] R. Cogranne, C. Zitzmann, L. Fillatre, I. Nikiforov, F. Retraint and P. Cornu, "A cover image model for reliable steganalysis," in *Information Hiding*, ser. LNCS vol. 6958, 2011, pp. 178–192.

[6] C. Zitzmann, R. Cogranne, F. Retraint, I. Nikiforov, L. Fillatre, and P. Cornu, "Statistical decision methods in hidden information detection," in *Information Hiding*, ser. LNCS vol. 6958, 2011, pp. 163–177.

[7] A. D. Ker, P. Bas, R. Böhme, R. Cogranne, S. Craver, T. Filler, J. Fridrich, and T. Pevný, "Moving steganography and steganalysis from the laboratory into the real world," in *ACM Information hiding and multimedia security, IH&MMSec'13*, 2013, pp. 45–58.

[8] O. Dabeer, K. Sullivan, U. Madhow, S. Chandrasekaran, and B. Manjunath, "Detection of hiding in the least significant bit," *Signal Processing, IEEE Transactions on*, vol. 52, no. 10, pp. 3046 – 3058, oct. 2004.

[9] R. Cogranne and F. Retraint, "An asymptotically uniformly most powerful test for lsb matching detection," *Information Forensics and Security, IEEE Transactions on*, vol. 8, no. 3, pp. 464–476, March 2013.

[10] C. Zitzmann, R. Cogranne, L. Fillatre, I. Nikiforov, F. Retraint, and P. Cornu, "Hidden information detection based on quantized Laplacian distribution," in *Proc. of IEEE International Conference on Acoustics, Speech, and Signal Processing (ICASSP)*, March 2012.

[11] T. H. Thai, R. Cogranne, and F. Retraint, "Statistical model of quantized DCT coefficients : Application in the steganalysis of JSteg algorithm," *Image Processing, IEEE Transactions on*, vol. 23, no. 5, pp. 1980–1993, 2014.

[12] J. Fridrich and J. Kodovskỳ, "Rich models for steganalysis of digital images," *Information Forensics and Security, IEEE Transactions on*, vol. 7, no. 3, pp. 868 –882, june 2012.

[13] J. Kodovskỳ, J. Fridrich, and V. Holub, "Ensemble classifiers for steganalysis of digital media," *Information Forensics and Security, IEEE Transactions on*, vol. 7, no. 2, pp. 432–444, April 2012.

[14] R. Cogranne, T. Denemark, and J. Fridrich, "Theoretical model of the FLD ensemble classifier based on hypothesis testing theory," in *IEEE International Workshop on Information Forensics and Security (WIFS)*, 2014, pp. 167–172.

[15] R. Cogranne and J. Fridrich, "Modeling and extending the ensemble classifier for steganalysis of digital images using hypothesis testing theory," *(accepted for publication in) Information Forensics and Security, IEEE Transactions on*, 2015.

[16] A. D. Ker and T. Pevný, "A mishmash of methods for mitigating the model mismatch mess," *IS&T/SPIE Electronic Imaging conf.*, vol. 9028, Security, Steganography, and Watermarking of Multimedia Contents, 2014, pp. 1601–1615.

[17] A. D. Ker and T. Pevny, "The steganographer is the outlier: realistic large-scale steganalysis," *Information Forensics and Security, IEEE Transactions on*, vol. 9, no. 9, pp. 1424–1435, 2014.

[18] E. Page, "Continuous inspection schemes," *Biometrika*, pp. 100–115, 1954.

[19] A. Tartakovsky, I. Nikiforov, and M. Basseville, *Sequential Analysis: Hypothesis Testing and Changepoint Detection*, Taylor & Francis, 2014.

[20] B. K. Guépié, "Sequential detection of transient changes: application to water distribution network monitoring *in french*," Ph.D. dissertation, Troyes University of Technology (UTT), March 2013.

[21] B. K. Guépié, L. Fillatre, and I. Nikiforov, "Sequential detection of transient changes," *Sequential Analysis*, vol. 31, no. 4, pp. 528–547, 2012.

[22] P. Bas, T. Filler, and T. Pevný, "Break our steganographic system — the ins and outs of organizing boss," in in *Information Hiding*, ser. LNCS vol.6958, 2011, pp. 59–70.

[23] T. Denemark, V. Sedighi, V. Holub, R. Cogranne, and J. Fridrich, "Selection-channel-aware rich model for steganalysis of digital images," in *IEEE International Workshop on Information Forensics and Security (WIFS)*, 2014, pp. 48–53.

[24] V. Sedighi, J. Fridrich, and R. Cogranne, "Content-adaptive pentary steganography using the multivariate generalized gaussian cover model," vol. 9409, February 2015.

[25] V. Sedighi, R. Cogranne, and J. Fridrich, "Content-Adaptive Steganography by Minimizing Statistical Detectability," *Information Forensics and Security, IEEE Transactions on* (submitted).

[26] V. Holub and J. Fridrich, "Phase-aware projection model for steganalysis of JPEG images," *IS&T/SPIE Electronic Imaging conf.*, vol. 9409, Security, Steganography, and Watermarking of Multimedia Contents, 2015, pp. 94 090T–94 090T–11.

[27] L. Guo, J. Ni, and Y. Q. Shi, "An efficient jpeg steganographic scheme using uniform embedding," in *IEEE International Workshop on Information Forensics and Security (WIFS)*, Dec 2012, pp. 169–174.