Detection of JSteg Algorithm Using Hypothesis Testing Theory and a Statistical Model with Nuisance Parameters.

Tong Qiao
ICD - LM2S - UMR CNRS 6281
Troyes University of Technology (UTT)
tong.qiao@utt.fr

Cathel Zitzmann
ICD - LM2S - UMR CNRS 6281
EPF Graduate School of Engineering
cathel.zitzmann@epf.fr

Rémi Cogranne
ICD - LM2S - UMR CNRS 6281
Troyes University of Technology (UTT)
remi.cogranne@utt.fr

Florent Retraint
ICD - LM2S - UMR CNRS 6281
Troyes University of Technology (UTT)
florent.retraint@utt.fr

ABSTRACT

This paper investigates the statistical detection of data hidden within DCT coefficients of JPEG images using a Laplacian distribution model. The main contributions is twofold. First, this paper proposes to model the DCT coefficients using a Laplacian distribution but challenges the usual assumption that among a sub-band all the coefficients follow are independent and identically distributed (i. i. d.). In this paper it is assumed that the distribution parameters change from DCT coefficient to DCT coefficient. Second this paper applies this model to design a statistical test, based on hypothesis testing theory, which aims at detecting data hidden within DCT coefficient with the JSteg algorithm. The proposed optimal detector carefully takes into account the distribution parameters as nuisance parameters. Numerical results on simulated data as well as on numerical images database show the relevance of the proposed model and the good performance of the ensuing test.

Categories and Subject Descriptors

H.1.1 [**Mathematics of Computing**]: Probability and Statistics—*Detection methodology*; I.4.10 [**Image Processing**]: Image Representation—*Statistical modelling*; H.1.1 [**Models and Principles**]: Systems and Information Theory—*Information theory*; D.2.11 [**Software Engineering**]: Software Architectures—*Information hiding*

Keywords

Steganography, Steganalysis, Hypothesis Testing Theory, Optimal Detection, Statistical Modelling, DCT coefficients.

IH&MMSec'14, June 11–13, 2014, Salzburg, Austria.

1. INTRODUCTION

Steganography nowadays refers to the technologies that exploit a covert channel to hide secret data within innocuous looking digital objects, without any perceptual modifications. Potentially a wide range of digital objects can be used to hide information such as images [16], videos [5], or network packets [1, 33]. On the opposite, passive steganalysis aims at discovering the use of such covert channel or at retrieving some information on hidden data, such as estimating the size of hidden data, the key or the algorithm used for embedding, for instance. During the last decade, steganographic algorithms and detection methods have been considerably improved, the reader is referred to [6] and [14] for detailed review of existing methods in steganography and steganalysis respectively.

However, as detailed in [16], a wide range of problems, theoretical as well as practical, remains uncovered and some prevent the moving of "steganography and steganalysis from the laboratory into the real world". This is especially the case in the field of Optimal Detection, see [16, Sec. 3.1], in which this paper lies. Roughly speaking, the goal of optimal detection in steganalysis, is to exploit an accurate statistical model of cover source, usually digital images, to design a statistical test which properties can be established; typically, in order to guarantee a false alarm probability and to calculate the optimal detection performance one can expect from the most powerful detector.

This optimal detection approach has been studied for the detection of data hidden within spatial domain of digital images since [13] and has been then considerably improved with more accurate statistical model of cover images [8, 31, 12]. For the detection of data hidden within the DCT coefficients of JPEG images, the application of hypothesis testing theory for designing optimal detectors, that are efficient in practice, is facing the problem of accurately modelling the statistical distribution of DCT coefficients. It can be noted that several models have been proposed in the literature to model statistically the DCT coefficients; among those models, the Laplacian distribution is probably the most widely used due to its simplicity and its fairly good accuracy [19]. More accurate models such as the Generalised Gaussian [23] and, more recently, the Generalised Gamma model [7] have been shown to provide much more accuracy at the cost of higher complexity. Some of those models have been exploited in the field of steganalysis, see [27, 4] for instance. In the framework of optimal detection, a first attempt has been made to design a statistical test modelling the DCT coefficient with the Laplacian distribution, see [34]. While the performance of this test has been analytically established and a threshold has been theoretically calculated to guarantee a prescribed false-alarm probability, a dramatic loss of performance has been empirically observed [34]. This can be explained by the inaccuracy of the Laplacian model that prevents the detection of such small changes as the embedding of hidden data cause.

It should be noted that other approaches have been proposed for the detection of data hidden within DCT coefficients of JPEG images, to cite few, the structural detection [18], the category attack [20], the WS detector [3], and universal or blind detectors [22, 26]. However, establishing the statistical properties of those detectors remains a difficult work which has not been studied yet. In addition, the most accurate detector based on statistical learning is sensitive to the so-called cover source-mismatch [2]: the training phase must be performed with caution.

In this paper, a novel approach is proposed to model the DCT coefficients of JPEG image and is exploited to design a powerful statistical test based on the concept of nuisance parameters. The key idea of this approach is to consider that all the DCT coefficients of a sub-band do not have the same distribution parameters which act as nuisance parameters for the detection of hidden data.

In fact most of the proposed model of DCT coefficients are based, or have at least their accuracy verified, on observations of histograms of all the DCT coefficients of a sub-band. This method implicitly assumes that, among a sub-band, all the DCT coefficients follow the same statistical distribution. In practice, it makes sense to consider that the content of the digital image will largely influence the distribution of those coefficients. This implies that the DCT coefficients may not all follow the same distribution because their distribution parameters depend on the local content of the image.

For simplicity and clarity, it is proposed in this paper to apply this methodology with the Laplacian model. A simple approach is proposed to estimate the expectation of each coefficient by denoising the image in spatial domain and transforming the denoised image back into the DCT domain. Then it is proposed to exploit the framework of hypothesis testing theory to design an optimal detector based on this model that takes into account the Laplacian distribution parameters as nuisance parameters .

The contributions of this paper are summarised below:

- First, a novel model that does not assume that all the DCT coefficients of a same sub-band are i. i. d. is proposed, contrary to almost every statistical model of DCT coefficients .

- Second, this statistical model of DCT coefficients is used to design an accurate test to detect data hidden within JPEG images with the JSteg algorithm. This statistical test takes into account distribution parameters of each DCT coefficient as nuisance parameters.

- Numerical results show the sharpness of the theoretically established results and the good performance that the proposed statistical test achieves. A comparison with the statistical test based on the Laplacian and on the assumption of i. i. d. coefficient, see [34], shows the relevance of the proposed methodology.

This paper is organised as follows. Section 2 formalises the statistical problem of detection of information hidden within the DCT coefficients of JPEG images. Then, Section 3 presents the optimal Likelihood Ratio Test (LRT) for detecting the JSteg algorithm based on the Laplacian distribution model. Section 4 presents the proposed approach for estimating the nuisance parameter in practice. Finally, Section 5 presents comparisons with other detector and Section 6 concludes the paper.

2. PROBLEM STATEMENT

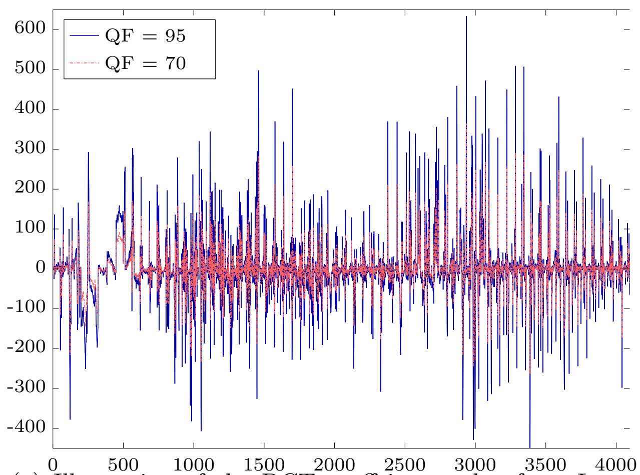
In this paper, a grayscale digital image is represented, in the spatial domain, by a single matrix $\mathbf{Z} = \{z_{i,j}\}, i \in \{1, \dots, I\}, j \in \{1, \dots, J\}$. The present work can be extended to colour image by analysing each colour channel separately. Most of the digital images are stored using the JPEG compression standard. This standard exploits the linear Discrete Cosine Transform (DCT), over blocks of 8×8 pixels to represent an image in the so-called DCT domain. In the present paper, we avoid the description of the imaging pipeline of a digital still camera; the reader is referred to [24] for a description of the whole imaging pipeline and to [25] for a detailed description of the JPEG compression standard.

Let us denote the DCT coefficients by the matrix $\mathbf{V} = \{v_{i,j}\}$. An alternative representation of those coefficients is usually adopted by gathering the DCT coefficients that corresponds to the same frequency sub-band. In this paper, this alternative representation is denoted by the matrix $\mathbf{U} = \{u_{k,l}\}, k \in \{1, \dots, K\}, l \in \{1, \dots, 64\}$ with $K \approx I \times J/64$[1].

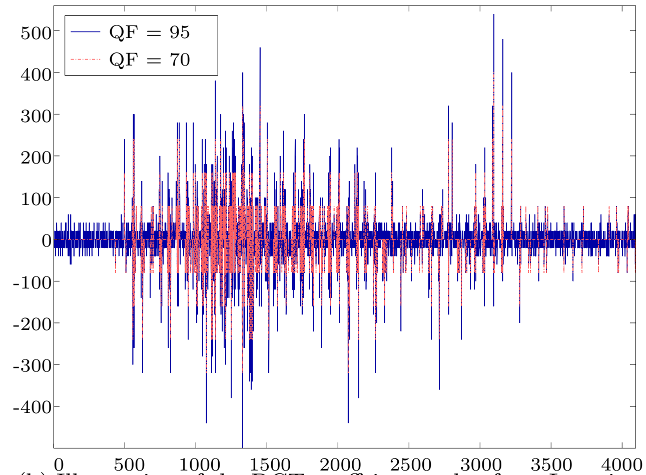
The coefficients from the first sub-band $u_{k,1}$, often referred to as DC coefficients, represent the mean of pixels value over k-th block of 8×8 pixels. The modification of those coefficients may be obvious and creates artifacts that can be detected easily, hence, they are usually not used for data hiding. Similarly, the JSteg algorithm does not use the coefficients from the others sub-bands, referred to as AC coefficients, if they equal 0 or 1. In fact, it is known that using the coefficients equal to 0 or 1 modifies significantly the statistical properties of AC coefficients; this creates a flaw that can be detected.

The JSteg algorithm embeds data within DCT coefficients of JPEG images using the well-known LSB (Least Significant Bit) replacement method, see details in [32]. In brief, this method consists in substituting the LSB of each DCT coefficient by a bit of the message it is aimed to hide. The number of bit hidden per coefficient, usually referred to as the payload, is denoted R. Since the JSteg algorithm does not use every DCT coefficient, the payload will in fact be measured in this paper as the number of bits hidden per *usable* coefficients (that is the number of bits divided by the number of AC coefficients that differ from 0 and 1).

[1]In this paper we assume, without loss of generality, that both the width and the height of the inspected image are multiple of 8.



(a) Illustration of the DCT coeffcients value from Lena images, sub-band (1,2).



(b) Illustration of the DCT coeffcients value from Lena images, sub-band (4,4).

Figure 1: Illustrative examples of the value of the DCT coefficients of two sub-bands from *lena* image. Those examples show that the assumption that DCT coefficients are i. i. d. within a sub-band hardly holds true in practice.

Let us assume that the DCT coefficients are independent and that they all follow the same probability distribution, denoted $\mathcal{P}_\theta$, parametrised by the parameter θ which may change among the coefficients. Since the DCT coefficients can only take value into a discrete set, the distribution $\mathcal{P}_\theta$ may be represented by its probability mass function (pmf) denoted $P_\theta = \{p_\theta[u]\}$; for simplicity[2], it is assumed in this paper that $u \in \mathbb{Z}$. Let us denote $\mathcal{Q}_\theta^R$ the probability distribution of *usable* DCT coefficients from the stego-image, after embedding a message with payload R. A short calculation shows that, see [13, 15, 35], the stego-image distribution may be represented with following the pmf $Q_\theta^R = \{q_\theta^R[u]\}_{u\in\mathbb{Z}}$ where

$$q_\theta^R[u] = (1 - R/2)p_\theta[u] + R/2 p_\theta[\bar{u}], \tag{1}$$

and $\bar{u} = u+(-1)^u$ represents the integer u with flipped LSB. For the sake of clarity, let us denote $\theta_{k,l}$ the distribution parameter of k-th DCT coefficient from l-th sub-band and let $\boldsymbol{\theta} = \{\theta_{k,l}\}, k \in \{1, \ldots, K\}, l \in \{2, \ldots, 64\}$ represents the distribution parameter of all the AC coefficients.

When inspecting a given JPEG image, more precisely its DCT coefficients matrix $\mathbf{U}$, in order to detect data hidden with the JSteg algorithm, the problem consists in choosing between the two following hypotheses $\mathcal{H}_0$: "*the coefficients* $u_{k,l}$ *follow the distribution* $\mathcal{P}_{\theta_{k,l}}$" and $\mathcal{H}_1$: "*the coefficients* $u_{k,l}$ *follow the distribution* $\mathcal{Q}_{\theta_{k,l}}^R$" which can be written formally:

$$\begin{cases} \mathcal{H}_0 : \left\{ u_{k,l} \sim \mathcal{P}_{\theta_{k,l}}, \forall k \in \{1, \ldots, K\}, \forall l \in \{2, \ldots, 64\} \right\}, \\ \mathcal{H}_1 : \left\{ u_{k,l} \sim \mathcal{Q}_{\theta_{k,l}}^R, \forall k \in \{1, \ldots, K\}, \forall l \in \{2, \ldots, 64\}, R > 0 \right\}. \end{cases} \tag{2}$$

A statistical test is a mapping $\delta : \mathbb{Z}^{I \cdot J} \mapsto \{\mathcal{H}_0, \mathcal{H}_1\}$ such that hypothesis $\mathcal{H}_i$ is accepted if $\delta(\mathbf{U}) = \mathcal{H}_i$ (see [21] for details on hypothesis testing). As previously explained, this paper focuses on the Neyman-Pearson bi-criteria approach: maximising the correct detection probability for a given false-alarm probability α_0. Let:

$$\mathcal{K}_{\alpha_0} = \left\{ \delta : \sup_{\boldsymbol{\theta}} \mathbb{P}_{\mathcal{H}_0}[\delta(\mathbf{U}) = \mathcal{H}_1] \leq \alpha_0 \right\}, \tag{3}$$

be the class of tests with a false alarm probability upper-bounded by α_0. Here $\mathbb{P}_{\mathcal{H}_i}(A)$ stands for the probability of event A under hypothesis $\mathcal{H}_i, i = \{0, 1\}$, and the supremum over $\boldsymbol{\theta}$ has to be understood as whatever the distribution parameters might be, in order to ensure that the false alarm probability α_0 can not be exceed.

Among all the tests in $\mathcal{K}_{\alpha_0}$, it is aimed at finding a test δ which maximises the power function, defined by the correct detection probability:

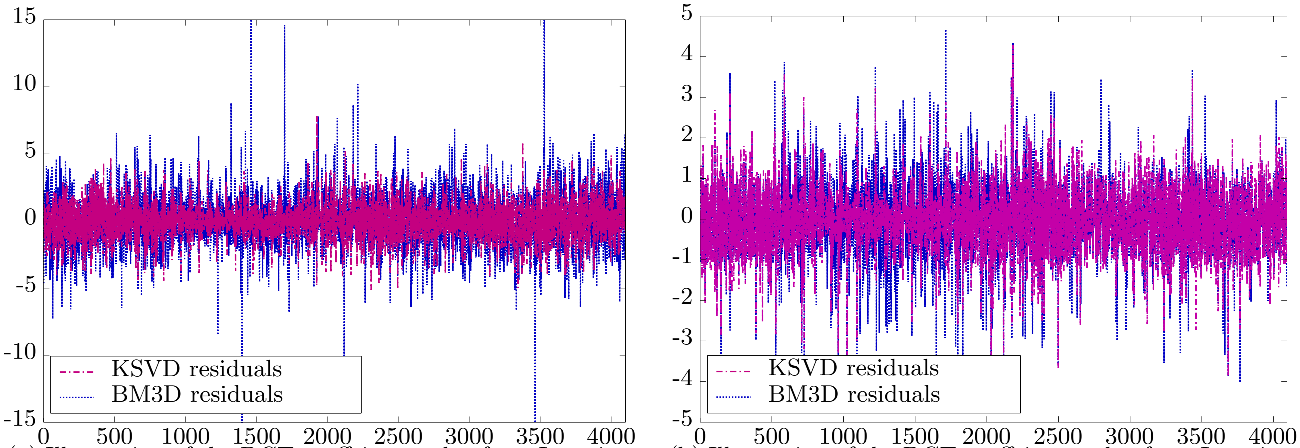
$$\beta_\delta = \mathbb{P}_{\mathcal{H}_1}[\delta(\mathbf{U}) = \mathcal{H}_1], \tag{4}$$

which is equivalent to minimising the missed detection probability $\alpha_1(\delta) = \mathbb{P}_{\mathcal{H}_1}[\delta(\mathbf{U}) = \mathcal{H}_0] = 1 - \beta_\delta$.

In order to design a practical *optimal detector*, as referred to in [16], for steganalysis in spatial domain, the main difficulty is to estimate the distribution parameters, that is the expectation and the variance of each pixel. On the opposite, in the case of DCT coefficients, the application of hypothesis testing theory to design an optimal detector has previously being attempted with the assumption that the distribution parameter remains the same for all the coefficients from a same sub-band. With this assumption, the estimation of the distribution parameters is not an issue because thousands of DCT coefficients are available. However which distribution model to choose remains an open problem.

The hypothesis testing theory has been applied for the steganalysis of JSteg algorithm in [34] using a Laplacian distribution model and using the assumption that DCT coefficients of each sub-band are i. i. d. However, this pioneer work does not allow the designing of an efficient test because a very important loss of performance has been observed when comparing results on real images and theoretically established ones. Such a result can be explained by the two following reasons: 1) the Laplacian model might be not accurate enough to detect steganagraphy and 2) the assumption that the DCT coefficients of each frequency sub-

[2] In practice, DCT coefficients belong to set $[-1024, \ldots, 1023]$, see [34].



(a) Illustration of the DCT coeffcients value from Lena images, sub-band (1,2).

(b) Illustration of the DCT coeffcients value from Lena images, sub-band (4,4).

Figure 2: Illustrative examples of the DCT coefficients of the residual noise, obtained by denoising image. The same two DCT sub-bands, as in Figure 1 are extracted the residual noise of *lena* image. On those examples the assumption of i. i. d. distribution seems to be more realistic.

band are i. i. d. may be wrong. Recently, it has been shown that the use of Generalised Gamma model or even more accurate model [28, 29] allows the designing of a test with very good detection performance. On the opposite, in this paper it is proposed to challenge the assumption that all the DCT coefficients of a sub-band are i. i. d.

A typical example is given by Figures 1 and 2. Figure 1a (resp. Figure 1b) represents the DCT coefficients of the sub-band (1,2) (resp. sub-band (4,4)) extracted from the image *lena*. Observing those two graphs, it is obvious that the assumption of all those coefficients being i. i. d. is doubtful. However, if it is assumed that each coefficient has a different expectation, one can estimate this expected value and compute the "residual noise", that is the difference between the observation and the computed expectation. Such results are shown in Figure 2, with two different models for estimating the expectation of DCT coefficients of the same two sub-bands from *lena*. Obviously, residual noises look much more i. i. d. than the original DCT coefficients.

In the following section, we detail the statistical test that takes into account both the expectation and the variance as nuisance parameters and we study the optimal detection when those parameters are known. A discussion on nuisance parameters is also provided in Section 4.

3. OPTIMAL DETECTION FRAMEWORK

When the payload R and the distribution parameters $\boldsymbol{\theta} = \{\theta_{k,l}\}, k \in \{1,\dots,K\}, l \in \{2,\dots,64\}$ are known, problem (2) is reduced to a statistical test between two simple hypotheses. In such a case, the Neyman-Pearson Lemma [21, theorem 3.2.1] states that the most powerful test in the class $\mathcal{K}_{\alpha_0}$ (3) is the LRT defined, on the assumption that DCT coefficients are independent, as:

$$\delta^{\mathrm{lr}}(\mathbf{U}) = \begin{cases} \mathcal{H}_0 \text{ if } \Lambda^{\mathrm{lr}}(\mathbf{U}) = \displaystyle\sum_{k=1}^{K}\sum_{l=2}^{64} \Lambda^{\mathrm{lr}}(u_{k,l}) < \tau^{\mathrm{lr}}, \\ \mathcal{H}_1 \text{ if } \Lambda^{\mathrm{lr}}(\mathbf{U}) = \displaystyle\sum_{k=1}^{K}\sum_{l=2}^{64} \Lambda^{\mathrm{lr}}(u_{k,l}) \geq \tau^{\mathrm{lr}}, \end{cases} \tag{5}$$

where the decision threshold τ^{lr} is the solution of the equation $\mathbb{P}_{\mathcal{H}_0}\left[\Lambda^{\mathrm{lr}}(\mathbf{U}) \geq \tau^{\mathrm{lr}}\right] = \alpha_0$, to ensure that the false alarm probability of the LRT equals α_0, and the log Likelihood Ratio (LR) for one observation is given, by definition, by:

$$\Lambda^{\mathrm{lr}}(u_{k,l}) = \log\left(\frac{q^R_{\theta_{k,l}}[u_{k,l}]}{p_{\theta_{k,l}}[u_{k,l}]}\right). \tag{6}$$

In practice, when the rate R is not known one can try to design a test which is locally optimal around a given payload rate, named Locally Asymptotically Uniformly Most Powerful (LAUMP) test, as proposed in [11, 35] but this lies outside the scope of this paper.

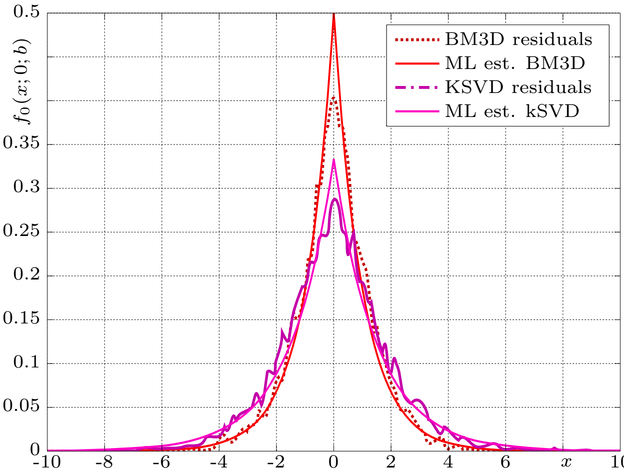
From the definition of $p_{\theta_{k,l}}[u_{k,l}]$ and $q^R_{\theta_{k,l}}[u_{k,l}]$ (1), it is easy to write the LR (6) as:

$$\Lambda^{\mathrm{lr}}(u_{k,l}) = \log\left(1 - \frac{R}{2} + \frac{R}{2}\frac{p_{\theta_{k,l}}[\bar{u}_{k,l}]}{p_{\theta_{k,l}}[u_{k,l}]}\right), \tag{7}$$

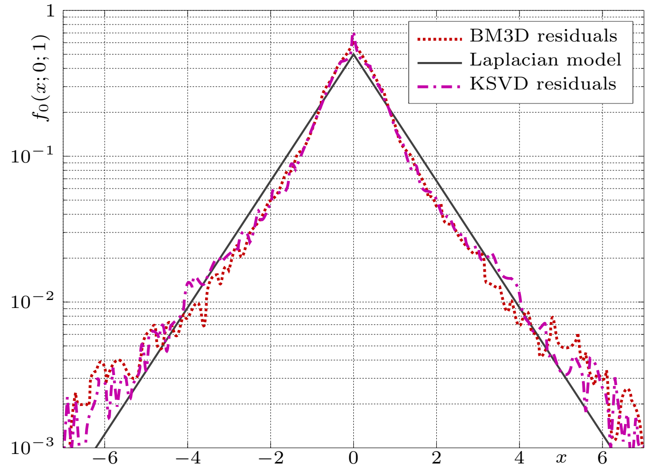
where, as previously defined, $\bar{u}_{k,l} = u_{k,l} + (-1)^{u_{k,l}}$ represents the DCT coefficient $u_{k,l}$ with flipped LSB.

Accepting, for a moment, that one is in this most favourable scenario, in which all the parameters are perfectly known, we can deduce some interesting results. Due to the fact that observations are considered to be independent, the LR $\Lambda^{\mathrm{lr}}(\mathbf{U})$ is the sum of random variables and some asymptotic theorems allow the establishing of its distribution when the number of coefficients become "sufficiently large". This asymptotic approach is usually verified in the case of digital images due to the very large number of pixels or DCT coefficients.

Let us denote $E_{\mathcal{H}_i}(\theta_{k,l})$ and $V_{\mathcal{H}_i}(\theta_{k,l})$ the expectation and the variance of the LR $\Lambda^{\mathrm{lr}}(u_{k,l})$ under hypothesis $\mathcal{H}_i, i = \{0,1\}$. Those quantity obviously depend on the parametrised distribution $\mathcal{P}_{\theta_{k,l}}$. The Lindeberg's central limit theorem (CLT) [21, theorem 11.2.5] states that as L tends to



(a) Illustration of the DCT coeffcients value from Lena images, sub-band (1,2).



(b) Illustration of the DCT coeffcients value from Lena images, sub-band (4,4).

Figure 3: Statistical distribution of the DCT coefficients of the residual noise plotted in Figure 2. For comparison, the Laplacian pdf, with parameters estimated by the Maximum Likelihood Estimation are also shown. Note that for a meaning comparison, Figure 3b show the results after normalisation by the estimated scale parameter $\widehat{b}$.

infinity it holds true that[3]:

$$\frac{\displaystyle\sum_{k=1}^{K}\sum_{l=2}^{64}\Lambda^{\mathrm{lr}}(u_{k,l}) - E_{\mathcal{H}_i}(\theta_{k,l})}{\left(\displaystyle\sum_{k=1}^{K}\sum_{l=2}^{64}V_{\mathcal{H}_i}(\theta_{k,l})\right)^{1/2}} \xrightarrow{d} \mathcal{N}(0,1)\ ,\ i=\{0,1\}\ , \quad (8)$$

where $\xrightarrow{d}$ represents the convergence in distribution and $\mathcal{N}(0,1)$ is the standard normal distribution, *i.e.* with zero mean and unit variance.

This theorem is of crucial interest to establish the statistical properties of the proposed test [8, 10, 31, 34]. In fact, once the moments have been calculated under both $\mathcal{H}_i, i=\{0,1\}$, one can normalise under hypothesis $\mathcal{H}_0$ the LR $\Lambda^{\mathrm{lr}}(\mathbf{U})$ as follows:

$$\begin{aligned}\overline{\Lambda}^{\mathrm{lr}}(\mathbf{U}) &= \frac{\Lambda^{\mathrm{lr}}(\mathbf{U}) - \sum_{k=1}^{K}\sum_{l=2}^{64}E_{\mathcal{H}_i}(\theta_{k,l})}{\left(\sum_{k=1}^{K}\sum_{l=2}^{64}V_{\mathcal{H}_i}(\theta_{k,l})\right)^{1/2}},\\ &= \frac{\sum_{k=1}^{K}\sum_{l=2}^{64}\Lambda^{\mathrm{lr}}(u_{k,l}) - E_{\mathcal{H}_i}(\theta_{k,l})}{\left(\sum_{k=1}^{K}\sum_{l=2}^{64}V_{\mathcal{H}_i}(\theta_{k,l})\right)^{1/2}}. \quad (9)\end{aligned}$$

Since this essentially consists in adding a deterministic value and scaling the LR, this operation of normalisation preserves the optimality of the LRT.

It immediately follows from Lindeberg's CLT (8) that $\overline{\Lambda}^{\mathrm{lr}}(\mathbf{U})$ asymptotically follows, as L tends to infinity, the normal distribution $\mathcal{N}(0,1)$. Hence, it is immediate to set the decision threshold that guarantee the prescribed false alarm probability:

$$\overline{\tau}^{\mathrm{lr}} = \Phi^{-1}\left(1-\alpha_0\right), \quad (10)$$

where Φ and Φ^{-1} respectively represent the cumulative distribution function (cdf) of the standard normal distribution and its inverse. Similarly, denoting

$$m_i = \sum_{k=1}^{K}\sum_{l=2}^{64}E_{\mathcal{H}_i}(\theta_{k,l}) \ \text{ and } \ \sigma_i^2 = \sum_{k=1}^{K}\sum_{l=2}^{64}V_{\mathcal{H}_i}(\theta_{k,l})\ , i=\{0,1\},$$

it is also straightforward to establish the detection function of the LRT given by:

$$\beta_{\overline{\delta}^{\mathrm{lr}}} = 1 - \Phi\left(\frac{\sigma_0}{\sigma_1}\Phi^{-1}\left(1-\alpha_0\right) + \frac{m_0-m_1}{\sigma_1}\right). \quad (11)$$

Equations (10) and (11) emphasise the main advantage of normalising the LR as described in relation (9): it allows to set a threshold that guarantee a false alarm probability independently from any distribution parameters and, this is particularly crucial because digital images are heterogeneous, their properties vary for each image. Second, the normalisation allows to easily establish the detection power which again, is achieved, for any distribution parameters and hence, for any inspected image.

3.1 Application with the Laplacian Distribution

In the case of the Laplacian distribution, the framework of hypothesis testing theory has been applied for the steganalysis of JSteg in [34] in which the moments of LR are calculated under the two following assumptions: 1) all the DCT coefficients from the same sub-band are i. i. d. and 2) the expectation of each DCT coefficient is zero.

The continuous Laplacian distribution has the following probability density function:

$$f_{\mu,b}(x) = \frac{1}{2b}\exp\left(-\frac{|x-\mu|}{b}\right) \quad (12)$$

where $\mu\in\mathbb{R}$, sometimes referred to as the location parameter, corresponds to the expectation, and $b>0$ is the so-called scale parameter. During the compression of JPEG images, the DCT coefficients are quantised. Hence, let us defined the discrete Laplacian distribution by the following pmf, see

[3]Note that we refer to the Lindeberg's CLT, whose conditions are easily verified in our case, because the random variable are independent but are not i. i. d.

details in Appendix A:

$$f_{\mu,b}[k] \stackrel{def.}{=} \mathbb{P}\Big[x \in [\Delta(k-1/2), \Delta(k+1/2)[\Big]$$
$$= \begin{cases} \exp\left(-\frac{|\Delta k-\mu|}{b}\right)\sinh\left(\frac{\Delta}{2b}\right) & \text{if } \frac{\mu}{\Delta} \notin [k-1/2; k+1/2[\\ 1-\frac{1}{2}\exp\Big(-\frac{\Delta(k+1/2)-\mu}{2b}\Big)-\frac{1}{2}\exp\Big(-\frac{\Delta(k-1/2)-\mu}{2b}\Big) & \text{otherwise} \end{cases} \tag{13}$$

where Δ is the quantisation step.
From the expression of the discrete Laplacian distribution (13) and from the expression of the LR (7), one can express the LR for the detection of JSteg under the assumption that DCT coefficients follow a Laplacian distribution, as follows, see Appendix B:

$$\log\left(1-\frac{R}{2}+\frac{R}{2}\exp\left[\frac{\Delta}{b}\text{sign}(\Delta k-\mu)(k-\bar{k})\right]\right). \tag{14}$$

It can be noted that this expression (14) of the LR is almost the same as the one obtained in [34] assuming that all DCT coefficients have a zero-mean, only the sign term $\text{sign}(\Delta k - \mu)$ becomes $\text{sign}(k)$ when assuming a zero-mean. It should also be noted that the log-LR equals 0 for every DCT coefficient whose value is 0 or 1 because the JSteg algorithm does not embed hidden data in those coefficients. In the present paper, the moments of the LR (14) are not analytically established, the reader interested is referred to [34].

4. DEALING WITH NUISANCE PARAMETERS

As already explained, most of the statistical model of DCT coefficients assume that within a sub-band the coefficients are i. i. d. However, as illustrated in Figure 1 and 2 this assumption is doubtful in practice. Another way to explain why the DCT coefficients may not be i. i. d. is to consider a block of 8×8 pixels in spatial domain, say the first, $\mathbf{z} = z_{i,j}\,, i \in \{1,\dots,8\}\,, j \in \{1,\dots,8\}$. The value of those pixels can be decomposed as:

$$z_{i,j} = x_{i,j} + n_{i,j},$$

where $x_{i,j}$ is a deterministic value that represents the expectation of pixel at location (i, j) and $n_{i,j}$ is the realisation of a random variable representing all noises corrupting the inspected image. Clearly, this decomposition can be done for the whole block $\mathbf{z} = \mathbf{x}+\mathbf{n}$, where $\mathbf{x} = \{x_{i,j}\}$ and $\mathbf{n} = \{n_{i,j}\}$. Since the DCT transformation is linear the DCT coefficient of any block may be expressed as :

$$\begin{aligned} \text{DCT}(\mathbf{z}) &= \mathbf{D}^T\mathbf{z}\mathbf{D} = \mathbf{D}^T(\mathbf{x}+\mathbf{n})\mathbf{D} \\ &= \mathbf{D}^T\mathbf{x}\mathbf{D} + \mathbf{D}^T\mathbf{n}\mathbf{D} = \text{DCT}(\mathbf{x}) + \text{DCT}(\mathbf{n}), \end{aligned} \tag{15}$$

where DCT represents the DCT transform and $\mathbf{D}$ is the change of basis matrix from spatial to DCT basis, often referred to as the DCT matrix.
It makes sense to assume that the expectation of the noise component $\mathbf{n}$ has a zero-mean in the spatial and in the DCT domain. On the opposite, it is difficult to justify that the DCT of pixels' expectation $\mathbf{x}$ should necessary be around zero. Actually, this assumption holds true if and only if the expectation is the same for of all the pixels from a block: $\forall i \in \{1,\dots,8\}\,, \forall j \in \{1,\dots,8\}\,, x_{i,j} = x$, see [30, 28, 29] for details.

On the opposite, in the paper, it is mainly aimed at estimating the expectation of each DCT coefficient. To this end, it is proposed to decompress a JPEG image $\mathbf{V}$ into the spatial domain to obtain $\mathbf{Z}$, then to estimate the expectation of each pixel $\widehat{\mathbf{Z}}$. Then this denoised image, which corresponds to estimated expectation of pixels in spatial domain, is transformed back into the DCT domain to finally obtain the estimated value of all DCT coefficients, denoted $\mathbf{V} = \{v_{i,j}\}\,, i \in \{1,\dots,I\}\,, j \in \{1,\dots,J\}$. Several methods have been tested to estimate the expectation of pixels in the spatial domain $\widehat{\mathbf{Z}}$, namely, the BM3D collaborative filtering, kSVD sparse dictionary learning and non-local weighted averaging method from NL-means.

In addition, the proposed model also assumes that the scale parameter $b_{k,l}$ is different for each DCT coefficient. The estimation of this parameter, for each DCT coefficient, is based on the WS Jpeg method to locally estimate the variance; that is, for coefficients $v_{i,j}$, it simply consists of the sample variance of the DCT coefficients of the same sub-band from neighbouring blocks:

$$\widehat{\sigma}^2_{i,j} = \frac{1}{7}\sum_{\substack{s=-1 \\ (s,t)\neq(0,0)}}^{1}\sum_{t=-1}^{1}\left(v_{i+8s,j+8t} - \bar{v}_{i,j}\right)^2, \tag{16}$$

where $\bar{v}_{i,j}$ is the sample mean: $\frac{1}{8}\sum_{\substack{s=-1 \\ (s,t)\neq(0,0)}}^{1}\sum_{t=-1}^{1} v_{i+8s,j+8t}$.
Let us recall that the Maximum Likelihood Estimation (MLE) of the scale parameter of Laplacian distribution from realisations $x1,\dots,x_N$ is given by $\hat{b} = N^{-1}\sum_{n=1}^{N}|x_n - \mu|$. The local estimation of the scale parameter it is proposed to use in this paper is given by:

$$\hat{b}_{i,j} = \frac{1}{8}\sum_{\substack{s=-1 \\ (s,t)\neq(0,0)}}^{1}\sum_{t=-1}^{1}\left|v_{i+8s,j+8t} - \bar{v}_{i,j}\right|, \tag{17}$$

where $\bar{v}_{i,j}$ is the sample mean previously defined. As in the WS Jpeg algorithm, this approach raises the problem of scale parameter estimation for blocks located on the sides of the image. In the present paper, as in the WS Jpeg method, it is proposed not to use those blocks in the test.

4.1 Design of the Proposed Detector

In Section 3 the framework of hypothesis testing theory has been presented assuming that distribution parameters are known for each DCT coefficient. To design a practical test, a usual solution consists in replacing the unknown parameter by its ML estimation. This leads to the construction of a Generalised LRT. A similar construction is adopted in this paper, using the *ad hoc* estimators presented at the beginning of section 4, instead of using the ML method to estimate the distribution parameters of each DCT coefficient. The proposed test is thus defined as:

$$\widehat{\delta}(\mathbf{U}) = \begin{cases} \mathcal{H}_0 \text{ if } \widehat{\Lambda}(\mathbf{U}) = \sum_{k=1}^{K}\sum_{l=2}^{64}\widehat{\Lambda}(u_{k,l}) < \widehat{\tau}, \\ \mathcal{H}_1 \text{ if } \widehat{\Lambda}(\mathbf{U}) = \sum_{k=1}^{K}\sum_{l=2}^{64}\widehat{\Lambda}(u_{k,l}) \geq \widehat{\tau}, \end{cases} \tag{18}$$

where the decision statistic $\widehat{\Lambda}(u_{k,l})$ for a single DCT coefficient is given by, see Equation (14):

$$\widehat{\Lambda}(u_{k,l}) = \log\left(1 + \frac{R}{2} + \frac{R}{2}\exp\left[\frac{\Delta}{\widehat{b}_{k,l}}\mathrm{sign}(\Delta k - \widehat{\mu}_{k,l})(k - \bar{k})\right]\right), \tag{19}$$

and, in order to have a normalised decision statistic for the whole image, $\widehat{\Lambda}(\mathbf{U})$ is defined as:

$$\widehat{\Lambda}(\mathbf{U}) = \tfrac{1}{S_L}\textstyle\sum_{k=1}^{K}\sum_{l=2}^{64}\widehat{\Lambda}(u_{k,l}) - E_{\mathcal{H}_0}(\widehat{\mu}_{k,l}, \widehat{b}_{k,l}) \tag{20}$$

$$\text{with} \qquad S_L^2 = \textstyle\sum_{k=1}^{K}\sum_{l=2}^{64} V_{\mathcal{H}_0}(\widehat{\mu}_{k,l}, \widehat{b}_{k,l}).$$

Finally, note that in practice the direct use of the estimated scale parameter $\widehat{b}_{k,l}$ may cause numerical instabilities because this term may be close to zero. A simple, yet efficient, technique to avoid this problem is to replace $\widehat{b}_{k,l}$ by the following term $\min\big(\widehat{b}_{k,l}, b_{\min}\big)$. That is when the estimate $\widehat{b}_{k,l}$ is smaller that than the fixed constant $b_{\min}$, the estimated is replaced by the constant. This technique is similar to the adding of a constant to the estimated variance used in the WS [15, 17].

4.2 Comparison with Prior-Art

The WS Jpeg, as well as the WS for spatial domain, is based on the underlying assumption that the observations follow a Gaussian distribution. As recently shown [11, 35], the WS implicitly assumes that the quantisation step is negligible. Let us rewrite the LR test for JSteg detection based on a Gaussian distribution model of DCT coefficients. Let X be a random variable following a quantised Gaussian distribution. Exploiting the assumption that the quantisation step is negligible compared to noise standard deviation allows the writing of:

$$\begin{aligned}\mathbb{P}[X = k] &= \int_{\Delta(k-1/2)}^{\Delta(k+1/2)} \frac{1}{\sigma\sqrt{2\pi}}\exp\left(-\frac{(x-\mu)^2}{2\sigma^2}\right)dx \\ &\approx \frac{\Delta}{\sigma\sqrt{2\pi}}\exp\left(-\frac{(\Delta k-\mu)^2}{2\sigma^2}\right)\end{aligned} \tag{21}$$

Putting this expression of the pmf under hypothesis $\mathcal{H}_0$ into the LR (2), and assuming that the quantisation step is negligible compared to the noise standard deviation, $\Delta << \sigma$, it is immediate to obtain the following expression of the LR under the assumption of Gaussian distribution of DCT coefficient

$$\begin{aligned}&\log\left(1 + \frac{R}{2} + \frac{R}{2}\frac{\exp\left(-\frac{(\Delta\bar{k}-\mu)^2}{2\sigma^2}\right)}{\exp\left(-\frac{(\Delta k-\mu)^2}{2\sigma^2}\right)}\right) \\ \approx &\frac{R\Delta}{\sigma^2} \quad (k-\bar{k}) \quad (\Delta k - \mu) \\ = &\overbrace{\;w_\sigma\;} \quad \overbrace{\;\pm 1\;} \quad \overbrace{(\Delta k - \mu)}\end{aligned} \tag{22}$$

see details in Appendix C.

This expression highlights the well known fact the WS consists in fact of three terms: 1) the term w_σ which is a weight so that pixels or DCT coefficients with highest variance have a smallest importance, 2) the term $(k - \bar{k}) = \pm 1$ according the LSB of k and 3) the term $(\Delta k - \mu)$.

In comparison, the expression of the LR for a Laplacian distribution model (14), as well as the expression of the proposed test with estimates (19) only depends on the term

$$\begin{aligned}&\frac{\Delta}{b} \quad (k-\bar{k}) \quad \mathrm{sign}(\Delta k - \mu) \\ = &\overbrace{\;w_b\;} \quad \overbrace{\;\pm 1\;} \quad \overbrace{\mathrm{sign}(\Delta k - \mu)}\end{aligned} \tag{23}$$

which is also made of three terms; the two first are roughly similar to the two first terms of the WS : 1) the term w_b is a weight so that DCT coefficients with highest "scale" b have a smallest importance, note that the variance is proportional to b^2, 2) the term $(k - \bar{k}) = \pm 1$ according to the LSB of k. However, in the expression of the LR based on the Laplacian model the term $(\Delta k - \mu)$ of the WS is replaced with its sign. This shows that the statistical tests based on Laplacian model and based on Gaussian model are essentially similar.

5. NUMERICAL RESULTS

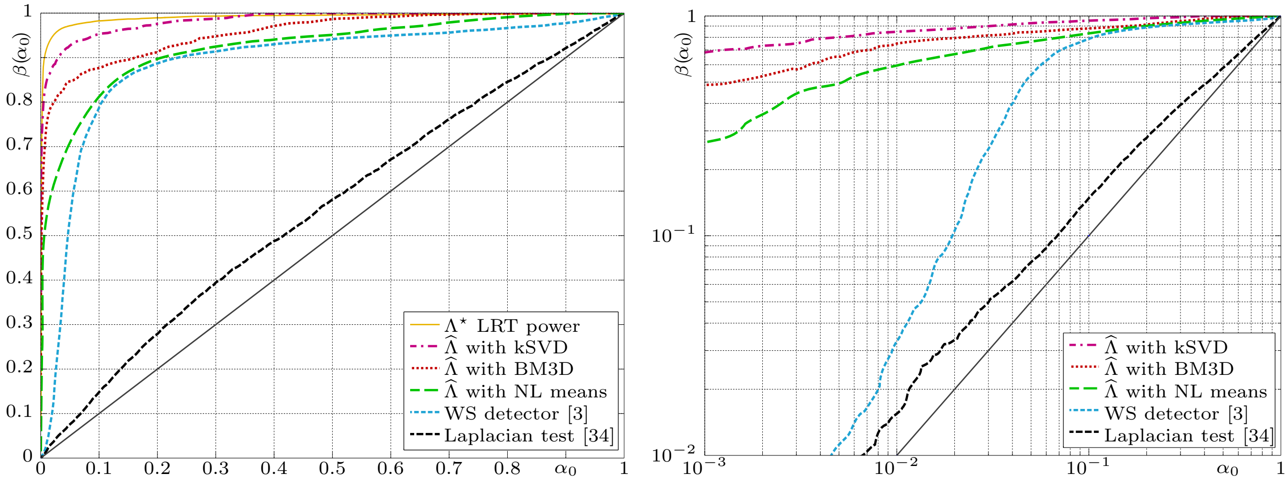
To verify the relevance of the proposed methodology, it is proposed to compare the proposed statistical test with two other detectors. The first chosen competitor is the statistical test proposed in [34] as it is also based on the Laplacian model but does not take into account the distribution parameters as nuisance parameters; it considers that DCT coefficients are i.i.d., following a Laplacian distribution with zero-mean. The comparison with this test is meaningful as it allows us to measure how much the detection performance is improved by removing the assumption that the DCT coefficients of each sub-band are i. i. d. The second chosen competitor is the WS [3] due to its similarity with the proposed statistical test, see details in Section 4.2.

For a large scale verification, it is proposed to use the BOSS database, made of 10 000 grayscale images of size 512×512 pixels, used with payload $R = 0.05$. Prior to our experiments, the images have been compressed in JPEG using the linux command convert which uses the standard quantisation table. Note also that all the JSteg steganography was performed using a Matlab source code we developed based on Phil Sallee's Jpeg Toolbox[4]. Three denoising methods have been tested to estimate the expectation of each DCT coefficient, namely the k-SVD denoising, the BM3D and the NL-means algorithms. The codes used for those three denoising methods have been downloaded from the Image Processing On-Line website[5].

Figure 4 shows the detection performance obtained over the BOSS database compressed with quality factor (QF) 70. The detection performances are shown as ROC curves, that is the detection power is plotted as a function of false-alarm probability. The Figure 4a particularly emphasises that the statistical test based on the Laplacian model does not perform well while the proposed methodology which takes into account the Laplacian distribution parameters as nuisance parameters allows us to largely improve the performance. Similarly the WS detector achieves overall good detection performance. However, it can be shown on Figure 4b, which presents the same results using a logarithmic scale, that for low false-alarm probabilities, the performance of the WS significantly decreases. On the opposite, the proposed statistical test still performs well.

[4]Phil Sallee's Jpeg Toolbox is available at : `http://dde.binghamton.edu/download/jpeg_toolbox.zip`

[5]Image Processing On-Line journal is available at: `http://www.ipol.im`



(a) Comparison of detection performance for BOSS database with quality factor 70 (linear scale).

(b) Comparison of detection performance for BOSS database with quality factor 70 (logarithmic scale).

Figure 4: Comparison of detection performance for BOSS database with quality factor 70

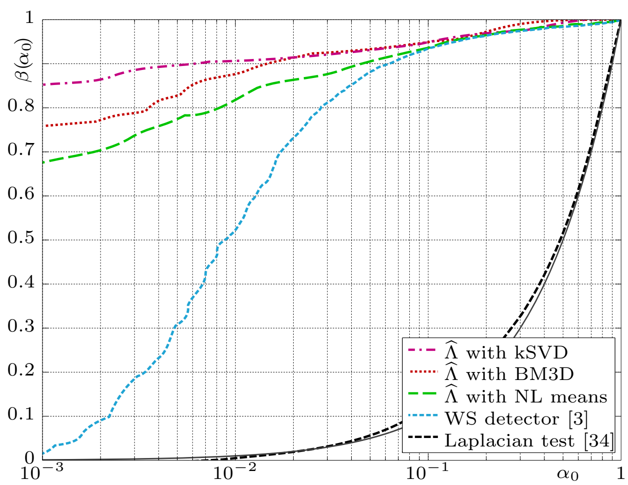


Figure 5: Comparison of detection performance for BOSS database with quality factor 85 (logarithmic scale).

Among the three denoising algorithms that have been tested, the k-SVD achieves the best performance but it can be observed on Figure 4 that the performance obtained using the BM3D and using the NL-means denoising methods are also very good.

To extend the results previously presented, a similar test has been performed over the BOSS database using the quality factor 85. The detection performance obtained by the proposed test and by the competitors are presented in Figure 5. Again, this figure shows that the statistical test based on the Laplacian model and assuming that DCT coefficients of a sub-band are i. i. d. has an unsatisfactory performance. It can also be noted that even though the WS performs slightly better for low false-alarm probability, compared to the results obtained with quality factor 70, it performs much worse than the proposed statistical test.

6. CONCLUSIONS

This paper aims at improving the optimal detection of data hidden within the DCT coefficients of JPEG images. Its main originality is that the usual Laplacian model is used as a statistical model of DCT coefficients but, opposed to what is usually proposed, it is not assumed that all DCT coefficients from a sub-band are i. i. d. This leads us to consider the Laplacian distribution parameters, namely the expectation μ and the scale parameter b, as nuisance parameters as they have no interest for the detection of hidden data, but must be carefully taken into account to design an efficient statistical test. Numerical results show that by estimating those nuisance parameters, the Laplacian model allows the designing of an accurate statistical test which outperforms the WS. The comparison with the optimal detector based on the Laplacian model and on the assumption that all DCT coefficients of a sub-band are i. i. d. shows the relevance of the proposed approach.

A possible future work would be to apply this approach with state-of-the-art statistical model of DCT coefficients, such as the Generalised Gaussian or the Generalised Gamma model. This could provide improvements in the detection performance at the cost of a higher complexity.

Acknowledgements

The work of R. Cogranne, F.Retraint and C. Zitzmann is funded by Troyes University of Technology (UTT) strategic program COLUMBO. The PhD thesis of Tong Qiao is funded by the China Scholarship Council (CSC) program.

7. REFERENCES

[1] V. Anantharam and S. Verdu. Bits through queues. *Information Theory, IEEE Transactions on*, 42(1):4–18, 1996.

[2] P. Bas, T. Filler, and T. Pevný. Break our steganographic system — the ins and outs of organizing boss. In *Information Hiding International Workshop*, LNCS vol.6958, pages 59–70, 2011.

[3] R. Böhme. Weighted stego-image steganalysis for jpeg covers. In *Information Hiding*, pages 178–194. Springer, 2008.

[4] R. Böhme and A. Westfeld. Breaking cauchy model-based jpeg steganography with first order statistics. In *Computer Security–ESORICS 2004*, pages 125–140. Springer, 2004.

[5] U. Budhia, D. Kundur, and T. Zourntos. Digital video steganalysis exploiting statistical visibility in the temporal domain. *Information Forensics and Security, IEEE Transactions on*, 2006.

[6] R. Böhme. *Advanced Statistical Steganalysis*. Springer Publishing Company, Incorporated, 1st edition, 2010.

[7] J.-H. Chang, J.-W. Shin, N. S. Kim, and S. Mitra. Image probability distribution based on generalized gamma function. *Signal Processing Letters, IEEE*, 12(4):325–328, 2005.

[8] R. Cogranne and F. Retraint. An asymptotically uniformly most powerful test for LSB matching detection. *Information Forensics and Security, IEEE Transactions on*, 8(3):464–476, 2013.

[9] R. Cogranne, F. Retraint, C. Zitzmann, I. Nikiforov, L. Fillatre, and P. Cornu. Detecting hidden information using decision theory: Methodology, results and difficulties. *Digital Signal Processing*, 24:144 – 161, 2014.

[10] R. Cogranne, C. Zitzmann, L. Fillatre, I. Nikiforov, F. Retraint, and P. Cornu. A cover image model for reliable steganalysis. In *Information Hiding International Workshop*, LNCS vol.6958, pages 178 – 192, , 2011, Springer-Verlag, New York.

[11] R. Cogranne, C. Zitzmann, L. Fillatre, F. Retraint, I. Nikiforov, and P. Cornu. Statistical decision by using quantized observations. In *IEEE International Symposium on Information Theory*, pages 1135 – 1139, August 2011.

[12] R. Cogranne, C. Zitzmann, F. Retraint, I. V. Nikiforov, P. Cornu, and L. Fillatre. A local adaptive model of natural images for almost optimal detection of hidden data. *Signal Processing*, 100:169 – 185, 2014.

[13] O. Dabeer, K. Sullivan, U. Madhow, S. Chandrasekaran, and B. Manjunath. Detection of hiding in the least significant bit. *Signal Processing, IEEE Transactions on*, 52(10):3046 – 3058, oct. 2004.

[14] J. Fridrich. *Steganography in Digital Media: Principles, Algorithms, and Applications*. Cambridge University Press, 1st edition edition, 2009.

[15] J. Fridrich and M. Goljan. On estimation of secret message length in LSB steganography in spatial domain. In Proc. SPIE, volume 5306, pages 23–34, 2004.

[16] 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 *Proceedings of the first ACM workshop on Information hiding and multimedia security*, IH&MMSec '13, pages 45–58, 2013.

[17] A. D. Ker and R. Böhme. Revisiting weighted stego-image steganalysis. In *Proc. SPIE 6819*, pages 501–517, 2008.

[18] J. Kodovsky and J. Fridrich. Quantitative structural steganalysis of JSteg. *Information Forensics and Security, IEEE Transactions on*, 5(4):681–693, 2010.

[19] E. Lam and J. Goodman. A mathematical analysis of the DCT coefficient distributions for images. *Image Processing, IEEE Transactions on*, 9(10):1661 –1666, oct 2000.

[20] K. Lee, A. Westfeld, and S. Lee. Category attack for lsb steganalysis of JPEG images. In *Digital Watermarking*, pages 35–48. Springer, 2006.

[21] E. Lehmann and J. Romano. *Testing Statistical Hypotheses, Second Edition*. Springer, 3rd edition, 2005.

[22] S. Lyu and H. Farid. Steganalysis using higher-order image statistics. *Information Forensics and Security, IEEE Transactions on*, 1(1):111 – 119, march 2006.

[23] F. Muller. Distribution shape of two-dimensional dct coefficients of natural images. *Electronics Letters*, 29(22):1935–1936, 1993.

[24] J. Nakamura. *Image sensors and signal processing for digital still cameras*. CRC Press, 2005.

[25] W. B. Pennebaker. *JPEG: Still image data compression standard*. Springer, 1992.

[26] T. Pevny and J. Fridrich. Multiclass detector of current steganographic methods for jpeg format. *Information Forensics and Security, IEEE Transactions on*, 3(4):635–650, 2008.

[27] P. Sallee. Model-based methods for steganography and steganalysis. *International Journal of Image and Graphics*, 5(1):167–189, jan. 2005.

[28] T. H. Thai, R. Cogranne, and F. Retraint. Steganalysis of Jsteg algorithm based on a novel statistical model of quantized DCT coefficients. In Proc. of IEEE*International Conference on Image Processing (ICIP)*, pages 4427 – 4431, 2013.

[29] 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*, 23(5):1980–1993, 2014.

[30] T. H. Thai, F. Retraint, and R. Cogranne. Statistical Model of Natural Images. In Proc. of IEEE *International Conference on Image Processing (ICIP)*, pages 2525 – 2528, 2012.

[31] T. H. Thai, F. Retraint, and R. Cogranne. Statistical detection of data hidden in least significant bits of clipped images. *Signal Processing*, 98:263 – 274, 2014.

[32] D. Upham. Jsteg. *Software available at* `http://zooid.org/~paul/crypto/jsteg/`, 2002.

[33] L. Yao, X. Zi, L. Pan, and J. Li. A study of on/off timing channel based on packet delay distribution. *Computers & Security*, 28(8):785–794, 2009.

[34] C. Zitzmann, R. Cogranne, L. Fillatre, I. Nikiforov, F. Retraint, and P. Cornu. Hidden information detection based on quantized Laplacian distribution. In IEEE *International Conference on Acoustics, Speech, and Signal Processing*, pages 1793–1796, 2012.

[35] C. Zitzmann, R. Cogranne, F. Retraint, I. Nikiforov, L. Fillatre, and P. Cornu. Statistical decision methods in hidden information detection. In *Information Hiding International Workshop*, LNCS vol.6958, pages 163 – 177, 2011.

APPENDIX

A. QUANTIZED LAPLACIAN PMF

Let X be a Laplacian random variable with expectation μ and variance b. Its pdf is thus, see (12):

$$f_{\mu,b}(x) = \frac{1}{2b}\exp\left(-\frac{|x-\mu|}{b}\right),$$

and a straightforward calculation shows that its cdf is given by:

$$F_{\mu,b}(x) = \frac{1}{2} + \frac{1}{2}\mathrm{sign}(x-\mu)\left(1-\exp\left(-\frac{|x-\mu|}{b}\right)\right), \quad (24)$$

$$= \begin{cases} \frac{1}{2}\exp\left(\frac{x-\mu}{b}\right) & \text{if } x<\mu, \\ 1-\frac{1}{2}\exp\left(-\frac{x-\mu}{b}\right) & \text{if } x\geq\mu. \end{cases} \quad (25)$$

Now consider the result from quantisation of this random variable $Y = \lfloor X/\Delta \rfloor$, it is immediate to establish the pmf of this random variable. Let us first consider the case $\Delta(k+1/2) < \mu$ (due to the symmetry of Laplacian pdf, the case $\Delta(k-1/2) > \mu$ is treated similarly).
The pmf of Y is given by:

$$\begin{aligned}
\mathbb{P}[Y=k] &= \mathbb{P}[\Delta(k-{}^1\!/_2) \leq X < \Delta(k+{}^1\!/_2)], \\
&= \frac{1}{2}\exp\left(\frac{\Delta(k+{}^1\!/_2)-\mu}{b}\right) - \frac{1}{2}\exp\left(\frac{\Delta(k-{}^1\!/_2)-\mu}{b}\right), \\
&= \frac{1}{2}\exp\left(\frac{\Delta k-\mu}{b}\right)\exp\left(\frac{\Delta}{2b}\right) \\
&\quad - \frac{1}{2}\exp\left(\frac{\Delta k-\mu}{b}\right)\exp\left(\frac{-\Delta}{2b}\right), \\
&= \exp\left(\frac{\Delta k-\mu}{b}\right)\sinh\left(\frac{\Delta}{2b}\right),
\end{aligned}$$

Applying similar calculations for case $\Delta(k-1/2) > \mu$, one gets:

$$\mathbb{P}[Y=k] = \exp\left(-\frac{|\Delta k-\mu|}{b}\right)\sinh\left(\frac{\Delta}{2b}\right), \quad (26)$$

which corresponds to the pmf given in Eq. (13). The case $\Delta(k-1/2) < \mu < \Delta(k+1/2)$ is treated similarly.

B. LOG-LIKELIHOOD RATIO CALCULATION

By putting the expression of quantised Laplacian pmf (26) into the expression of the LR (7), it is immediate to write:

$$\Lambda^{\mathrm{lr}}(u_{k,l}) = \log\left(1-\frac{R}{2}+\frac{R}{2}\frac{\exp\left(-\frac{|\Delta\bar{k}-\mu|}{b}\right)\sinh\left(\frac{\Delta}{2b}\right)}{\exp\left(-\frac{|\Delta k-\mu|}{b}\right)\sinh\left(\frac{\Delta}{2b}\right)}\right).$$

Let us study the term:

$$\begin{aligned}
\frac{\exp\left(-\frac{|\Delta\bar{k}-\mu|}{b}\right)\sinh\left(\frac{\Delta}{2b}\right)}{\exp\left(-\frac{|\Delta k-\mu|}{b}\right)\sinh\left(\frac{\Delta}{2b}\right)} &= \frac{\exp\left(-\frac{|\Delta\bar{k}-\mu|}{b}\right)}{\exp\left(-\frac{|\Delta k-\mu|}{b}\right)}, \\
&= \frac{\exp\left(-\frac{|\Delta k+\Delta(\bar{k}-k)-\mu|}{b}\right)}{\exp\left(-\frac{|\Delta k-\mu|}{b}\right)}, \\
&= \frac{\exp\left(-\frac{|\Delta k-\mu|}{b}\right)\exp\left(\frac{\mathrm{sign}(\Delta k-\mu)(k-\bar{k})}{b}\right)}{\exp\left(-\frac{|\Delta k-\mu|}{b}\right)}, \quad (27) \\
&= \exp\left(\frac{\mathrm{sign}(\Delta k-\mu)(k-\bar{k})}{b}\right).
\end{aligned}$$

From this Eq. (27), it is immediate to establish the expression (14):

$$\log\left(1-\frac{R}{2}+\frac{R}{2}\exp\left(\frac{\mathrm{sign}(\Delta k-\mu)(k-\bar{k})}{b}\right)\right).$$

C. LR BASED ON THE GAUSSIAN MODEL (WS)

Let X be a Gaussian random variable with expectation μ and variance σ^2. Define the quantized Gaussian random variable as follows $Y = \lfloor X/\Delta \rfloor$, its pmf is given by $P_{\mu,\sigma} = \{p_{\mu,\sigma}[k]\}_{k=-\infty}^{\infty}$ with:

$$p_{\mu,\sigma}[k] = \mathbb{P}[Y=k] = \int_{\Delta(k-1/2)}^{\Delta(k+1/2)} \frac{1}{\sigma\sqrt{2\pi}}\exp\left(-\frac{(x-\mu)^2}{2\sigma^2}\right)dx.$$

Assuming that the quantisation step Δ is "small enough" compared to the variance $\Delta << \sigma$, it holds true that [11, 9]:

$$p_{\mu,\sigma}[k] \approx \frac{\Delta}{\sigma\sqrt{2\pi}}\exp\left(-\frac{(\Delta k-\mu)^2}{2\sigma^2}\right), \quad (28)$$

and

$$p_{\mu,\sigma}[k] + p_{\mu,\sigma}[\bar{k}] \approx \frac{2\Delta}{\sigma\sqrt{2\pi}}\exp\left(-\frac{(\Delta({}^{k+\bar{k}}\!/_2)-\mu)^2}{2\sigma^2}\right). \quad (29)$$

Let us rewrite the LR for the detection of JSteg (7) as follows

$$\begin{aligned}
\Lambda^{\mathrm{lr}}(u_{k,l}) &= \log\left(1-\frac{R}{2}+\frac{R}{2}\frac{p_{\mu,\sigma}[\bar{k}]}{p_{\mu,\sigma}[k]}\right), \\
&= \log\left(1-R+\frac{R}{2}\frac{p_{\mu,\sigma}[\bar{k}]+p_{\mu,\sigma}[k]}{p_{\mu,\sigma}[k]}\right). \quad (30)
\end{aligned}$$

Using the expressions (28) and (29) let us study the following ratio:

$$\begin{aligned}
\frac{p_{\mu,\sigma}[\bar{k}]+p_{\mu,\sigma}[k]}{p_{\mu,\sigma}[k]} &= 2\frac{\exp\left(-\frac{(\Delta(k+\bar{k}/2)-\mu)^2}{2\sigma^2}\right)}{\exp\left(-\frac{(\Delta k-\mu)^2}{2\sigma^2}\right)}, \\
&= 2\frac{\exp\left(-\frac{(\Delta k-\mu+\Delta/2(\bar{k}-k))^2}{2\sigma^2}\right)}{\exp\left(-\frac{(\Delta k-\mu)^2}{2\sigma^2}\right)}, \\
&= 2\frac{\exp\left(-\frac{(\Delta k-\mu)^2}{2\sigma^2}\right)\exp\left(\frac{\Delta(\Delta k-\mu)(k-\bar{k})}{2\sigma^2}\right)\exp\left(-\frac{\Delta^2}{8\sigma^2}\right)}{\exp\left(-\frac{(\Delta k-\mu)^2}{2\sigma^2}\right)}, \\
&= 2\exp\left(\frac{\Delta(\Delta k-\mu)(k-\bar{k})}{2\sigma^2}\right)\exp\left(-\frac{\Delta^2}{8\sigma^2}\right). \quad (31)
\end{aligned}$$

Putting the expression (31) into the expression of the log-LR (30) immediately gives:

$$\Lambda^{\mathrm{lr}}(u_{k,l}) = \log\left(1 + R\left(\exp\left(\frac{\Delta(\Delta k - \mu)(k - \bar{k})}{2\sigma^2}\right)\exp\left(-\frac{\Delta^2}{8\sigma^2}\right) - 1\right)\right)$$

from which a Taylor expansion around $\Delta/\sigma = 0$, this results from the assumption that $\Delta << \sigma$, and finally gives the well-known expression of the WS:

$$\Lambda^{\mathrm{lr}}(u_{k,l}) = \frac{R\Delta}{\sigma^2}(k - \bar{k})(\Delta k - \mu) \tag{32}$$