Optimal Detection of OutGuess using an Accurate Model of DCT Coefficients

Thanh Hai Thai, Rémi Cogranne, Florent Retraint
ICD - ROSAS - LM2S - Troyes University of Technology - UMR 6281, CNRS
Troyes - France

***Abstract*—This paper presents an optimal statistical test for the detection of OutGuess steganographic algorithm using an accurate statistical model of Discrete Cosine Transform (DCT) coefficients. First, this paper presents the proposed novel statistical model of quantized DCT coefficients. Then, this model is applied to design an optimal statistical test for the detection of OutGuess data hiding scheme. To this end, the detection of hidden data is then cast within the framework of hypothesis testing theory. The optimal Likelihood Ratio Test (LRT) is first presented. Then, for a practical application, a Generalized LRT is proposed using Maximum Likelihood Estimations of unknown parameters. Large scale numerical results show that the proposed approach allows the reliable and efficient detection of OutGuess scheme.**

***Index Terms*—DCT modelling, Discrete Cosine Transform, OutGuess, Steganalysis, Hypothesis Testing.**

I. Introduction

Digital communications nowadays play a crucial role in private and professional life. Because digital communication may be susceptible to eavesdropping, security and privacy have become more important than ever. In order to hide the very existence of a secret message, steganographic methods, that aim at hiding the very existence of the secret message by embedding into a innocuous-looking cover object, have been developed. On the opposite, steganalysis methods, that aim at detecting the presence of hidden data have also been largely studied, see [1] for a complete introduction.

Most of the steganographic algorithms and most of the steganalysis methods focuses on digital images. A very large majority of digital images are compressed using the JPEG compression standard, hence, the design of specific steganalysis methods for this compression standard is crucial.

The recent steganalysis methods proposed in the past years can be divided into the two following categories:

- The machine learning based methods are the most popular due to their efficiency. They are based on the extraction of features, that can potentially detect data hidden with a large set of steganographic algorithms, which are then provided to a machine learning algorithm which is trained to find the best decision rule. The Support Vector Machine (SVM) have been mainly used [2], [3] for supervised learning while the design of features set remains the most challenging problem in this field.
- An alternative approach relies on the use of statistical detection theory and exploits a statistical model of cover and stego-images. Such so-called optimal detectors have been applied, first, for the detection of LSB replacement [4], [5] and for LSB matching in [6]; their extension for the simple Jsteg embedding scheme in JPEG domain have been recently proposed in [7], [8].

The main advantages of machine learning based steganalyzer are their efficiency and relative simplicity. However, the statistical performance of machine learning based detector remains very difficult to established and hence, is usually only evaluated empirically. On the opposite, the optimal detectors usually does not achieve good performance in practice, but their properties are analytically established. This provides the possibility to guarantee a prescribed false-alarm probability, which is crucial for application in practice when a large number of images have to be inspected, and an upper bound on the detection performance one can expect.

It should be noted that the application of optimal detection for JPEG compressed images is a challenging difficulty as providing the required accurate statistical model of DCT coefficient remains an open problem. In the literature, several models have been empirically proposed: the Laplacian distribution [9], remains a popular choice for its high simplicity and relative accuracy, the Gaussian mixture [10] the Cauchy distribution [11], the Generalized Gaussian [12] and the Generalized Gamma [13] are example of more sophisticated yet more accurate models. The latter have been empirically shown to be among the most accurate.

One of the first application of optimal detection to steganalysis of DCT coefficients have been presented in [7] using the Laplacian model. This approach has poor performance in practice as the relative accuracy of this statistical model does not allow to detect as small as those cause by data hiding.

The goal of this paper is twofold. First, it aims at improving the current state-of-the-art of optimal steganalysis of DCT coefficient by using an accurate and well-founded statistical model. Second, it aims to apply optimal steganalysis to non-trivial embedding scheme, as currently only simple LSB replacement and LSB matching steganalysis have been studied. To this end, this paper focuses on the detection of data hidden with the OutGuess embedding scheme.

The present paper is organized as follows. First, Section II presents the proposed statistical model of DCT coefficients used in this paper. Then, Section III studies the statistical modeling of OutGuess data hiding scheme. Section IV presents the optimal statistical test, namely the Likelihood Ratio Test (LRT), for the detection of OutGuess and the proposed Generalized LRT, which addressed the case when the distribution parameters are unknown. Finally, Section V presents numerical results and comparisons with state-of-the-art DCT coefficients models and Section VI concludes the paper.

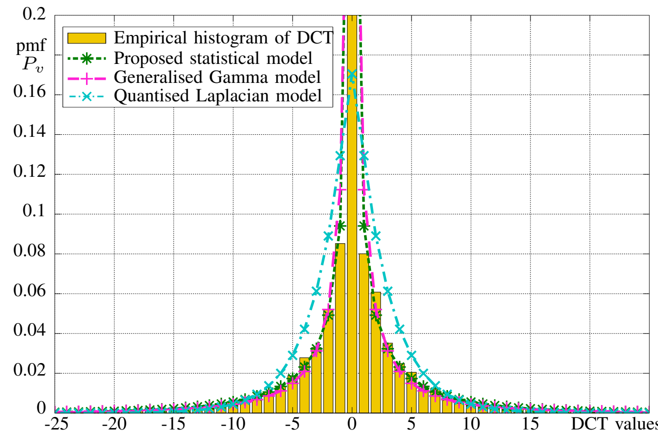

Fig. 1. Comparison between the quantized Laplacian, quantized GΓ and proposed model for quantized AC coefficient

II. Statistical Model of Quantized AC Coefficients

A. Statistical Model of AC coefficients

For brevity, we avoid recalling the whole imaging pipeline in the present paper. For clarity, only one color channel is considered in this paper, the three colors can be processed independently. It is assumed that after the post-acquisition processes the uncompressed "grayscale" digital image, denoted $\mathbf{Z} = \{z_{k,l}\}$ is well modeled as the realization of correlated Gaussian random variables, see [8], [14] for details.

The JPEG compression standard mainly relies on the Discrete Cosine Transform (DCT) and then on the quantization of the DCT coefficients. The DCT operation is applied blockwise over the all digital image by transforming each block of 8×8 pixels of $\mathbf{Z}$ into the DCT domain as follows[1]:

$$I_{p,q} = \frac{T_p T_q}{4} \sum_{m=0}^{7} \sum_{n=0}^{7} z_{m,n} \cos\left(\frac{(2m+1)p\pi}{16}\right) \cos\left(\frac{(2n+1)q\pi}{16}\right) \tag{1}$$

where $z_{m,n}$, $0 \leq m \leq 7$, $0 \leq n \leq 7$ denotes a pixel within a 8×8 block of $\mathbf{Z}$, and $I_{p,q}$, $0 \leq p \leq 7$, $0 \leq q \leq 7$ denotes the DCT coefficient of the corresponding block, for the sub-band at location (p, q). The normalizing terms T_p (resp. T_q) equal 1 except for $p = 0$ (resp. for $q = 0$) for which $T_0 = 1/\sqrt{2}$.

The coefficient at location $(0, 0)$, called the DC coefficient, represents the mean value of pixels in the 8×8 block. The remaining 63 coefficients are called AC coefficients.
The steganographic algorithms usually only embed hidden data in AC coefficients whose differ from 0 and 1 as using other coefficients is believed be potentially more easily detectable. In order to model AC coefficients, the variability of block variance due to the heterogeneity in the image has to be taken into account. Based on the doubly stochastic model [15], the probability density function (pdf) of AC coefficient I is:

$$f_I(x) = \int_0^{\infty} f_{I|\sigma^2}(x|t) f_{\sigma^2}(t) dt \quad x \in \mathbb{R}, \tag{2}$$

where σ^2 denotes the block variance which is also a random variable (RV). Assuming that the pixels $z_{m,n}$ are identically distributed within a 8×8 block [15] allows us to approximate the AC coefficient I as a zero-mean Gaussian RVs in virtue of Central Limit Theorem (CLT) for correlated RVs [16]:

$$f_{I|\sigma^2}(x|t) = \frac{1}{\sqrt{2\pi t}} \exp\Big(-\frac{x^2}{2t}\Big). \tag{3}$$

It is shown in [8], [14] that, from the modeling of the imaging pipeline, the block variance σ^2 can be accurately modeled by a Gamma distribution $\mathcal{G}(\eta, \nu)$ whose pdf is defined by:

$$f_{\sigma^2}(t) = \frac{t^{\eta-1}}{\nu^{\eta}\Gamma(\nu)} \exp\Big(-\frac{t}{\nu}\Big), \tag{4}$$

with η a positive shape parameter, ν a positive scale parameter, and $\Gamma(\cdot)$ the gamma function.

By putting the Gamma statistical model for the block variance (4) into (2) allows the writing of the AC DCT coefficients as follows:

$$f_I(x) = \frac{1}{\sqrt{2\pi}\nu^{\eta}\Gamma(\eta)} \int_0^{\infty} \exp\Big(-\frac{t}{\nu} - \frac{x^2}{2t}\Big) t^{\eta-\frac{3}{2}} dt. \tag{5}$$

From [17], the integral representation of the modified Bessel $K_\nu(\cdot)$ yields to

$$f_I(x) = \sqrt{\frac{2}{\pi}} \frac{\Big(|x|\sqrt{\frac{\nu}{2}}\Big)^{\eta-\frac{1}{2}}}{\nu^{\eta}\Gamma(\eta)} K_{\eta-\frac{1}{2}}\Big(|x|\sqrt{\frac{2}{\nu}}\Big). \tag{6}$$

The equation (6) is the pdf of the proposed statistical model of DCT coefficients; it is parametrized by only two parameters (η, ν) for which efficient and robust estimations are proposed below.

B. Statistical Model of Quantized AC coefficients

In the JPEG compression standard, the DCT coefficients are then quantized before a non-destructive (i.e reversible) compression scheme is applied. For the sake of clarity, from this section, the quantized DCT coefficients are arranged into 64 vectors of coefficients. Let $V_k = (v_{k,1}, \ldots, v_{k,n})^T$, $k \in \{1, \ldots, 64\}$ denote the vector of length n representing the quantized DCT coefficients of the k-th sub-band and $\mathbf{U}^T$ denotes the transpose of the matrix $\mathbf{U}$. Let also Δ_k denote the quantization step associated with the k-th sub-band. For clarity, the index of sub-band, or mode, is omitted in the next

[1]For simplicity, width and height of images are considered multiples of 8.

calculations as the same approach holds for all the modes.
It immediately follows from the definition of uniform quantization and from the pdf (6) that the pmf, denoted $P_V(l)$, $l \in \mathbb{Z}$, of the AC coefficients after quantization with step Δ is given by:

$$P_V(l) = \mathbb{P}\big[V = l\big] = \int_{\Delta(l-\frac{1}{2})}^{\Delta(l+\frac{1}{2})} f_I(x)dx. \tag{7}$$

The integral in (7) can be analytically calculated, see details in [8], [17], and finally leads to the following definition of the pmf P_V:

$$P_V(l) = \begin{cases} G\big(|l|\big) - G\big(|l|-1\big) & \forall l \in \mathbb{Z}_* \\ 2G(0) & l = 0. \end{cases} \tag{8}$$

where the function $G(\cdot)$ is defined by:

$$\begin{aligned} G(l) = \frac{1}{2} g(l) \Big[& K_{\eta-\frac{1}{2}}(g(l)) \mathbf{L}_{\eta-\frac{3}{2}}(g(l)) \\ & + K_{\eta-\frac{3}{2}}(g(l)) \mathbf{L}_{\eta-\frac{1}{2}}(g(l)) \Big], \end{aligned} \tag{9}$$

with $g(l) = \Delta(l+\frac{1}{2})\sqrt{\frac{2}{\nu}}$ and $\mathbf{L}_\nu(\cdot)$ the modified Struve function [17].

Again, it can be noted that the proposed statistical model of quantized DCT coefficients has only two unknown parameters because the quantization step Δ is available is the JPEG file header. The estimations of the unknown distribution parameters from DCT coefficients of a given JPEG image is addressed in the following Subsection.

C. Estimation of Distribution Parameters

To estimate the parameter of the proposed statistical model from DCT coefficients of a JPEG image, it is proposed to use the Maximum Likelihood Estimation (MLE) approach. Because the MLE can not be solved analytically, it is proposed to use a numerical optimization method and, for improving accuracy and efficiency, the estimations from the methods of moments are used as a starting point. Hence, let us first briefly describe the estimations of parameters from the method of moments.
Because the proposed distribution of the DCT coefficients has a pdf, or pmf after quantization, which is an even function, it is immediate that the odd moments equal to 0. Based on the definitions of the expectation, the second and fourth moments of the distribution of the non-quantized DCT coefficients (6) are given by, assuming a "fine" independent from signal and uniformly distributed (?):

$$\begin{aligned} \mathbb{E}_{V_k}\big[V_k^2\big] &= \frac{1}{\Delta_k^2}\mathbb{E}_{I_k}\big[I_k^2\big] + \mathbb{E}_{\epsilon_k}\big[\epsilon_k^2\big] \\ &= \frac{\eta_k \nu_k}{\Delta_k^2} + \frac{1}{12} \end{aligned} \tag{10}$$

$$\begin{aligned} \mathbb{E}_{V_k}\big[V_k^4\big] &= \frac{1}{\Delta_k^4}\mathbb{E}_{I_k}\big[I_k^4\big] + \frac{6}{\Delta_k^2}\mathbb{E}_{I_k}\big[I_k^2\big]\mathbb{E}_{\epsilon_k}\big[\epsilon_k^2\big] + \mathbb{E}_{\epsilon_k}\big[\epsilon_k^4\big] \\ &= \frac{3}{\Delta_k^4}\eta_k\nu_k^2(\eta_k+1) + \frac{1}{2\Delta_k^2}\eta_k\nu_k + \frac{1}{80}. \end{aligned} \tag{11}$$

It immediately follows that the estimation of the parameters (η_k, ν_k) are obtained from the method of moments by:

$$\widehat{\eta}_k^{mom} = \frac{\Big(\widehat{m}_{k,2} - \frac{1}{12}\Big)^2}{\frac{1}{3}\widehat{m}_{k,4} - \widehat{m}_{k,2}^2 + \frac{1}{360}} \tag{12}$$

$$\widehat{\nu}_k^{mom} = \frac{\Delta_k^2\Big(\widehat{m}_{k,2} - \frac{1}{12}\Big)}{\widehat{\eta}_k^{mom}}, \tag{13}$$

where $\widehat{m}_{k,2} = \frac{1}{n}\sum_{i=1}^n v_{k,i}^2$ and $\widehat{m}_{k,4} = \frac{1}{n}\sum_{i=1}^n v_{k,i}^4$ correspond to the empirical second and fourth moments of V_k.

By definition, the MLE of (η_k, ν_k) are defined as the solution of the following equation:

$$\Big(\widehat{\eta}_k^{MLE}, \widehat{\nu}_k^{MLE}\Big) = \operatorname*{arg\,max}_{(\eta_k,\nu_k)} \sum_{i=1}^n \log P_V\big(v_{k,i}\big). \tag{14}$$

Since the problem (14) does not have analytical solutions, it is proposed to use the the Nelder-Mead optimization method [18]. For accuracy and efficiency, the estimations obtained from the method of moments $\big(\widehat{\eta}_k^{mom}, \widehat{\nu}_k^{mom}\big)$ are used at starting point of the algorithm. Similarly, in the present paper, the estimations of the parameters of the quantized Laplacian and for the Quantized Generalized Γ (GΓ) model are obtained with the same approach, estimations of parameters with the method of moments used as starting points for solving numerically the MLE.

The Fig. 1 illustrates the empirical distribution of the DCT coefficients at location $(2,2)$ extracted from the first image of BOSS Base [19], compressed with quality factor 75. The comparison with the proposed statistical model as well as the Laplacian and GΓ model show the accuracy of the proposed statistical model, especially in comparison to the popular Laplacian model. Note that the Matlab source codes used to estimate the parameters of the proposed statistical model, and hence to obtain results presented in Figure 1, are available on the Internet at: http://remi.cogranne.pagesperso-orange.fr/.

III. OutGuess Data Hiding Scheme: Description and Statistical Modeling

The OutGuess steganographic scheme hides data within the AC coefficients of JPEG images using the Least Significant Bits (LSB) replacement method. That is, when the hidden bit differs from the LSB of a given AC coefficient, the LSB of this coefficient is flipped to match the hidden bit [20], [21]. As briefly described before, the OutGuess only uses AC coefficients which differ from 0 and 1; the use of others coefficients are considered more easily detectable. Let C_k, $k = \{1, \ldots, 64\}$ denote the values of the DCT coefficients of a given JPEG image re-arranged as 64 (column) vectors, each correspond to a specific DCT sub-band. The pmf of the quantized DCT coefficient C_k is denoted by $P_{\eta_k,\nu_k;\Delta_k}$ characterized by the parameters (η_k, ν_k) and Δ_k the known quantization step. Let n_k denote the number of non-zero AC coefficients in the k-th sub-band and let R be the payload, measured as the number of hidden bits per non-zero AC coefficient (bpnzAC):

$$R = \frac{L}{\sum_{k=2}^{64} n_k}. \tag{15}$$

with L the length, in bits, of the hidden message. The OutGuess embedding algorithm operated in two steps. First, the message is hidden using the LSB replacement mechanism. Then, a second step is applied to ensure that the global histogram, of all the AC coeffcients, is strictly the same after embedding. [2] More formally, let $h_c(x)$, $x \in \mathbb{Z}$ denotes the empirical histogram of the AC coefficients of the cover image. After the first step of embedding the histogram of the stego-image AC coefficients is denoted $h_s(x)$. To correct the impact of data hiding the OutGuess will then, during it second step, flipped LSB of extra DCT coefficients such that $\forall x \in \mathbb{Z}, h_s(x) = h_c(x)$. Hence the final pmf of the AC coefficient of sub-band k, denoted $S^{(R)}_{\eta_k,\nu_k;\Delta_k}$, is given by:

$$\begin{aligned} S^{(R)}_{\eta_k,\nu_k;\Delta_k}(x) = \Big(1-\frac{R}{2}\Big) P_{\eta_k,\nu_k;\Delta_k}(x) + \frac{R}{2} P_{\eta_k,\nu_k;\Delta_k}(\overline{x}) \\ + p_x\big(P_{\eta_k,\nu_k;\Delta_k}(x) - P_{\eta_k,\nu_k;\Delta_k}(\overline{x})\big), \end{aligned} \tag{16}$$

where $\overline{x} = x + (-1)^x$ denotes the integer x with flipped LSB and p_x is the probability of modification due to the correction, in the second step of OutGuess, see details in [21]. The two first terms in (16) account for the data hiding while the last term models the modification due to the correction. Note that because the correction proposed in OutGuess applies for the global histogram, p_x differs from actual probability of modification in each sub-band; it is this flaw that the proposed method aims at detecting.

When inspecting a given JPEG image, the problem of steganalysis can thus be cast with the framework of hypothesis testing theory with the goal of choosing between the the following hypotheses:

$$\begin{cases} \mathcal{H}_0 = \Big\{ v_{k,i} \sim P_{\eta_k,\nu_k;\Delta_k}, (\eta_k,\nu_k;) \in \mathbb{R}^2_+ \Big\} \\ \mathcal{H}_1 = \Big\{ v_{k,i} \sim S^{(R)}_{\eta_k,\nu_k;\Delta_k}, (\eta_k,\nu_k) \in \mathbb{R}^2_+, R \in (0,1] \Big\}. \end{cases} \tag{17}$$

The hypothesis testing problem (17) of OutGuess detection highlights the main difficulty. First, the alternative hypothesis $\mathcal{H}_1$ is composite when the payload R is unknown and, in such a situation, a test that is optimal for any payload may scarcely exists. Besides, the main problem in practice is that the distribution parameters (η_k,ν_k) are unknown and have to be estimated from the observed AC coefficients.

In this paper, it is first assumed that the payload R and the distribution parameters (η_k,ν_k) are known and the optimal Likelihood Ratio Test is presented. Then this paper focuses on the estimation of distribution parameters (η_k,ν_k) to design a Generalized LRT. It is worth noting that when the embedding rate R is unknown, the design of a test which is asymptotically optimal around a prescribed targeted payload has been studied in [22], [23], using local asymptotic normality. However, this paper focuses on the design the proposed GLRT, and assumes that payload R is known.

[2] For simplicity, it is assumed that AC coefficients belongs to $\mathbb{Z}$ while in practice they lie in the range $[-1024; 1023]$.

IV. Statistical Test for OutGuess Detection

When distribution parameters (η_k,ν_k) and the payload R is known, it follows from Neyman-Pearson lemma [24, theorem 3.2.1] that is the class

$$\mathcal{K}_{\alpha_0} = \Big\{ \delta : \mathbb{P}_0\big[\delta(\mathbf{V}) = \mathcal{H}_1\big] \le \alpha_0 \Big\}$$

of all the tests with a false-alarm probability upper bounded by α_0, the most powerful test δ is the LRT given by the following decision rule

$$\delta = \begin{cases} \mathcal{H}_0 & \text{if} \quad \Lambda(\mathbf{V}) = \sum_{k=2}^{64} \Lambda(V_k) < \tau \\ \mathcal{H}_1 & \text{if} \quad \Lambda(\mathbf{V}) = \sum_{k=2}^{64} \Lambda(V_k) \ge \tau \end{cases} \tag{18}$$

provided that the threshold τ is the solution of the equation

$$\mathbb{P}_0\Big[\Lambda(\mathbf{V}) \ge \tau\Big] = \alpha_0 \tag{19}$$

to ensure that the LR test δ is in the class $\in \mathcal{K}_{\alpha_0}$. Here, $\Lambda(V_k) = \sum_{i=1}^{n_k} \Lambda(v_{k,i})$ denotes the LR for the k-th sub-band, which from the assumption that AC coefficient of neighboring blocks are statistically independent, is given by the sum of LR each AC coefficients $\Lambda(v_{k,i})$ defined by

$$\begin{aligned} \Lambda(v_{k,i}) &= \log \frac{S^{(R)}_{\eta_k,\nu_k;\Delta_k}\big(v_{k,i}\big)}{P_{\eta_k,\nu_k;\Delta_k}\big(v_{k,i}\big)} \\ &= \log\left[1 - \frac{R}{2} + p_x + \left(\frac{R}{2} - p_x\right) \frac{P_{\eta_k,\nu_k;\Delta_k}\big(\overline{v}_{k,i}\big)}{P_{\eta_k,\nu_k;\Delta_k}\big(v_{k,i}\big)}\right], \end{aligned} \tag{20}$$

with, again, $\overline{v}_{k,i} = v_{k,i} + (-1)^{v_{k,i}}$ is the integer $v_{k,i}$ with flipped LSB.

Let define the function $d_k(m)$ as

$$d_k(m) = \log\left[1 - \frac{R}{2} + p_x + \left(\frac{R}{2} - p_x\right) \frac{P_{\eta_k,\nu_k;\Delta_k}\big(\overline{m}\big)}{P_{\eta_k,\nu_k;\Delta_k}\big(m\big)}\right]. \tag{21}$$

The expectation and the variance of the LR $\Lambda(v_{k,i})$ under hypothesis $\mathcal{H}_0$ are given by:

$$\mu_{k,0} = \mathbb{E}_0\Big[\Lambda(v_{k,i})\Big] = \sum_{m \in \mathbb{Z}} d_k(m) P_{\eta_k,\nu_k;\Delta_k}\big(m\big) \tag{22}$$

$$\sigma^2_{k,0} = \mathrm{Var}_0\Big[\Lambda(v_{k,i})\Big] = \sum_{m \in \mathbb{Z}} \big(d_k(m) - \mu_{k,0}\big)^2 P_{\eta_k,\nu_k;\Delta_k}(m) \tag{23}$$

with $\mathbb{E}_j[\cdot]$ and $\mathrm{Var}_j[\cdot]$ the expectation and variance under hypothesis $\mathcal{H}_j$, $j = \{0,1\}$ respectively. It is worth noting that the number of usable coefficients n_k is modeled as a random variable following binomial distribution $\mathcal{B}(n, p^*_k)$ with parameter

$$p^*_k = 1 - P_{\eta_k,\nu_k;\Delta_k}(0) - P_{\eta_k,\nu_k;\Delta_k}(1) \tag{24}$$

which remains the same under hypothesis $\mathcal{H}_0$ and $\mathcal{H}_1$. Taking into account that the number of usable coefficients is a random variable, it finally follows from Wald's identity [25] that the

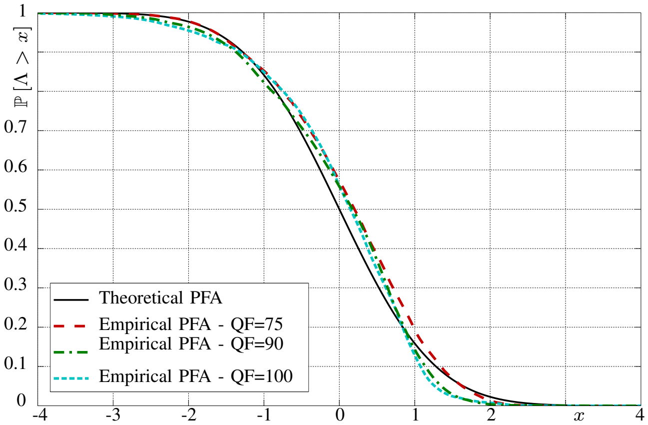

Fig. 2. Comparison between theoretical and empirical false-alarm probabilities for different quality factors.

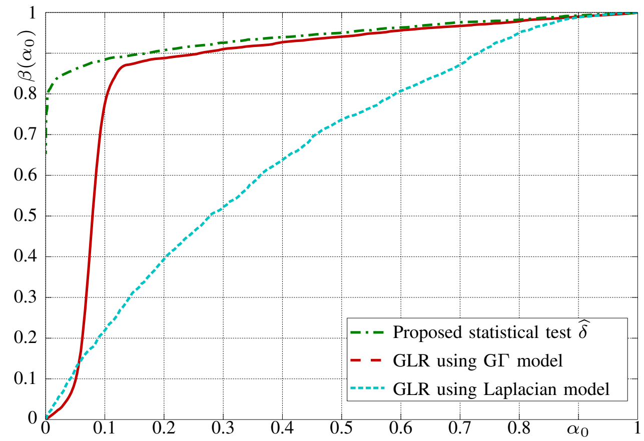

Fig. 3. Comparison between the proposed statistical test and the GLRT tests based on GΓ and Laplacian model for QF 75 and payload $R = 0.05$.

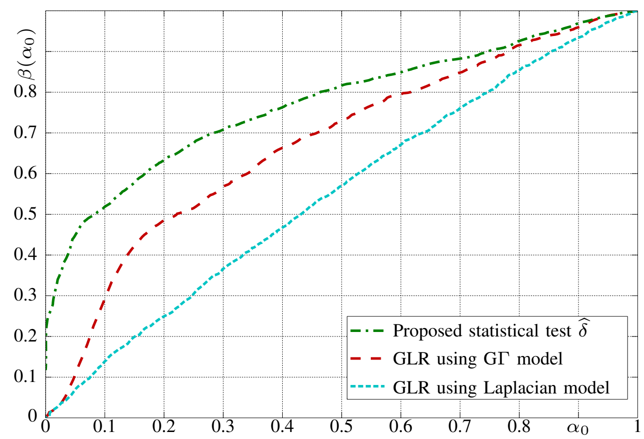

Fig. 4. Comparison between the proposed statistical test and the GLRT tests based on GΓ and Laplacian model for QF 100 and payload $R = 0.10$.

expectation and variance of the random sum $\Lambda(V_k)$ are defined by

$$\mathbb{E}_0\Big[\Lambda(V_k)\Big] = \mathbb{E}_0[n_k]\mathbb{E}_0\Big[\Lambda(v_{k,i})\Big] = np_k^*\mu_{k,0} \tag{25}$$

$$\begin{aligned}\mathrm{Var}_0\Big[\Lambda(V_k)\Big] &= \mathbb{E}_0[n_k]\mathrm{Var}_0\Big[\Lambda(v_{k,i})\Big] + \mathbb{E}_0^2\Big[\Lambda(v_{k,i})\Big]\mathrm{Var}_0[n_k] \\ &= np_k^*\sigma_{k,0}^2 + np_k^*(1-p_k^*)\mu_{k,0}^2. \end{aligned}\tag{26}$$

In virtue of the Lindeberg CLT [24, theorem 11.2.5], the LR for the k-th sub-band, $\Lambda(V_k)$, follows a Gaussian distribution with expectation $np_k^*\mu_{k,0}$ and variance $np_k^*\sigma_{k,0}^2 + np_k^*(1-p_k^*)\mu_{k,0}^2$. Hence it immediately follows that the LR for all the AC coefficients $\Lambda(\mathbf{V})$ asymptotically, as the number of AC coefficients tends to infinity, the following Gaussian distribution under hypothesis $\mathcal{H}_0$:

$$\Lambda(\mathbf{V}) \xrightarrow{D} \mathcal{N}(\mu_0, \sigma_0^2) \tag{27}$$

$$\text{with} \quad \mu_0 = \sum_{k=2}^{64} np_k^*\mu_{k,0} \tag{28}$$

$$\text{and} \quad \sigma_0^2 = \sum_{k=2}^{64} \Big[np_k^*\sigma_{k,0}^2 + np_k^*(1-p_k^*)\mu_{k,0}^2\Big]. \tag{29}$$

To apply easily the proposed statistical test in practice, it is proposed to normalize the LR $\Lambda(\mathbf{V})$ as follows

$$\Lambda^\star(\mathbf{V}) = \frac{\Lambda(\mathbf{V}) - \mu_0}{\sigma_0}. \tag{30}$$

such that under $\mathcal{H}_0$ the normalized LR convergences to the standard Gaussian distribution: $\Lambda^\star(\mathbf{V}) \xrightarrow{D} \mathcal{N}(0,1)$. This is particularly important as the normalization of the LR allows to set easily the decision threshold $\tau^\star$ to meet a prescribed false-alarm probability constraint as follows:

$$\tau^\star = \Phi^{-1}(1-\alpha_0). \tag{31}$$

The relation (31) holds true whatever the distribution parameters might be. Hence, thanks to the proposed normalization, the guarantee of a prescribed false-alarm probability is theoretically straightforward for any inspected image.

In practice, because the distribution parameters (ν_k, η_k) are unknown the proposed GLR is defined by the normalized LR test except that the distribution parameters used are estimated from the given image using the MLE described in section II-C. Hence the proposed statistical test is formally defined as:

$$\widehat{\delta} = \begin{cases} \mathcal{H}_0 & \text{if} \quad \widehat{\Lambda}^\star(\mathbf{V}) = \sum_{k=2}^{64} \widehat{\Lambda}(V_k) < \tau^\star \\ \mathcal{H}_1 & \text{if} \quad \widehat{\Lambda}^\star(\mathbf{V}) = \sum_{k=2}^{64} \widehat{\Lambda}(V_k) \geq \tau^\star \end{cases} \tag{32}$$

$$\text{with } \widehat{\Lambda}^\star = \frac{\widehat{\Lambda}(\mathbf{V}) - \widehat{\mu_0}}{\widehat{\sigma_0}}. \tag{33}$$

Here $\widehat{\Lambda}(\mathbf{V})$ corresponds to the Generalized LR (20) when the distribution parameters are replaced with their estimation $(\widehat{\nu}_k, \widehat{\eta}_k)$, see Eq. (14), and $\widehat{\mu_0}$ and $\widehat{\sigma_0}$ corresponds to its two first moments also obtained by using estimated distribution parameters in (29)

V. Numerical Results and Comparisons

One of the first goal of this paper is to provide a statistical test with analytically predictable results. Hence, Figure 2 presents a comparison between the theoretically established probability of false-alarm (PFA), as a function of the detection

TABLE I
DETECTION PERFORMANCE, MEASURED AS TOTAL PROBABILITY OF ERROR P_E FOR DIFFERENT QF AND DIFFERENT PAYLOADS.

	Proposed Model	GΓ model	Laplacian model
Quality Factor 75			
$R = 0.025$	0.1995	0.2251	0.4364
$R = 0.05$	0.1102	0.1311	0.3774
$R = 0.10$	0.0699	0.0968	0.2810
Quality Factor 90			
$R = 0.025$	0.2287	0.3091	0.4469
$R = 0.05$	0.1506	0.2314	0.3999
$R = 0.10$	0.1308	0.1785	0.3220
Quality Factor 100			
$R = 0.025$	0.3938	0.4360	0.4882
$R = 0.05$	0.3317	0.3725	0.4792
$R = 0.10$	0.2920	0.3565	0.4625

threshold, and the empirically obtained detection threshold. The experimental results have been obtained using the images from BOSS [19]. Even though the empirical false-alarm probability does not exactly match the theoretically established one, the Figure 2 shows that the obtained PFA are very closed to the established ones.

Regarding to the detection performance, because one of the main originality of the paper lies in the proposed statistical model of AC coefficients, it is proposed to compare the results of the proposed GLR test with the results obtained with GLR tests based on Generalized Gamma (GΓ) and the usual quantized Laplacian distribution. Figure 3 and 4 present a comparison between the empirically obtained performance of those three tests over all the images from BOSS base for which the embedding is possible. In fact, due to the histogram correction, the OutGuess first checks the possibility of embedding the required payload (including the correction), if not then the image is discarded from the results. Figure 3 presents the comparison between those three three tests for quality factor (QF) 75 and with payload $R = 0.05$ while the Figure 4 presents the comparison for QF 100 and payload $R = 0.1$. It is worth noting that the proposed statistical test outperforms its two competitors, and especially for law false-alarm probabilities, which it the most crucial case in practice. In fact, the Generalized Gamma model seems to be inaccurate for a non-negligible number of images, as it has also been shown in [8], this explains its poor performance very low false-alarm probabilities.
Finally, the table V present the overall detection performance, measure using the total probability of error under equal Bayesian priors, denoted P_E, for those three model for QF 75, 90, and 100 and for three different payloads $R = 0.025$, $R = 0.05$ and $R = 0.1$.

VI. CONCLUSION

This paper presents a novel methodology for detection of OutGuess based on hypothesis testing theory and on a novel model of quantized DCT coefficients. The exploitation of accurate statistical model for detection of non-trivial data hiding scheme open the way for the detection of more sophisticated steganographic system. The proposed statistical model is shown to allows the achieving of much higher performance than the popular Laplacian model as well as the state-of-the-art generalized gamma model. In addition, numerical results show that the proposed optimal detector allows the guaranteeing of a prescribed false-alarm probability.

VII. ACKNOWLEDGEMENTS

The authors would like to thank Jessica Fridrich for providing code of OutGuess simulator and for the fruitful discussions. This work has been funded, in part, by Troyes University of Technology (UTT) strategic program COLUMBO and by Champagne-Ardenne state project STEG-DETECT for scholars international mobility.

REFERENCES

[1] J. Fridrich, *Steganography in Digital Media : Principles, Algorithms, and Applications*. New york: Cambridge Univ. Press, 2009.
[2] Y. Wang and P. Moulin, “Optimized feature extraction for learning-based image steganalysis,” *Information Forensics and Security, IEEE Transactions on*, vol. 2, no. 1, pp. 31–45, March 2007.
[3] T. Pevny, J. Fridrich, and A. Ker, “From blind to quantitative steganalysis,” *Information Forensics and Security, IEEE Transactions on*, vol. 7, no. 2, pp. 445–454, April 2012.
[4] 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*, vol. 100, pp. 169 – 185, 2014.
[5] T. H. Thai, F. Retraint, and R. Cogranne, “Statistical detection of data hidden in least significant bits of clipped images,” vol. 98, pp. 263 – 274, May 2014.
[6] 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.
[7] 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.
[8] 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. 1–14, 2014.
[9] R. Reininger and J. Gibson, “Distributions of the two-dimensional DCT coefficients for images,” *Communications, IEEE Transactions on*, vol. 31, no. 6, pp. 835–839, Jun. 1983.
[10] T. Eude, R. Grisel, H. Cherifi, and R. Debrie, “On the distribution of the DCT coefficients,” in *Acoustics, Speech, and Signal Processing, IEEE International Conference on*, vol. v, Apr. 1994, pp. 365–368.
[11] J. E. Eggerton and M. D. Srinath, “Statistical distributions of image DCT coefficients,” *Computers and Electrical Engineering*, vol. 12, no. 3-4, pp. 137–145, Jan. 1986.
[12] F. Muller, “Distribution shape of two-dimensional DCT coefficients of natural images,” *Electronics Letters*, vol. 29, no. 22, pp. 1935–1936, Oct. 1993.
[13] J.-H. Chang, J.-W. Shin, N. S. Kim, and S. K. Mitra, “Image probability distribution based on generalized gamma function,” *Signal Processing Letters, IEEE*, vol. 12, no. 4, pp. 325–328, Apr. 2005.
[14] T. H. Thai, F. Retraint, and R. Cogranne, “Statistical model of natural images,” in *Image Processing (ICIP), 19th IEEE International Conference on*, 2012, pp. 2525–2528.
[15] E. Y. Lam and J. W. Goodman, “A mathematical analysis of the DCT coefficient distributions for images,” *Image Processing, IEEE Transactions on*, vol. 9, no. 10, pp. 1661–1666, Oct. 2000.
[16] M. Blum, “On the central limit theorem for correlated random variables,” *Proceedings of the IEEE*, vol. 52, no. 3, pp. 308–309, Mar. 1964.
[17] I. M. Ryzhik and I. S. Gradshteyn, *Tables of Integrals, Series, and Products*. United Kingdom: Elsevier, 2007.
[18] J. A. Nelder and R. Mead, “A simplex method for function minimization,” *The Computer Journal*, vol. 7, pp. 308–313, 1965.
[19] P. Bas, T. Filler, and T. Pevný, “Break our steganographic system — the ins and outs of organizing boss,” in *Information Hiding, International Workshop*, May 2011.

[20] N. Provos, "Defending against statistical steganalysis." in *Usenix Security Symposium*, vol. 10, 2001, pp. 323–336.

[21] J. Fridrich, M. Goljan, and D. Hogea, "Attacking the OutGuess," in *Proceedings of the ACM Workshop on Multimedia and Security*, vol. 2002. Juan-les-Pins, France, 2002.

[22] R. Cogranne, C. Zitzmann, L. Fillatre, F. Retraint, I. Nikiforov, and P. Cornu, "Statistical decision by using quantized observations," in *Information Theory Proceedings, IEEE International Symposium on*, Aug. 2011, pp. 1210–1214.

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

[24] E. L. Lehmann and J. P. Romano, *Testing Statistical Hypotheses*, 3rd ed. New york: Springer, 2005.

[25] A. Wald, "Some generalizations of the theory of cumulative sums of random variable," *The Annals of Mathematical Statistics*, pp. 287–293, Sep. 1945.