IEEE COPYRIGHT AND CONSENT FORM

To ensure uniformity of treatment among all contributors, other forms may not be substituted for this form, nor may any wording of the form be changed. This form is intended for original material submitted to the IEEE and must accompany any such material in order to be published by the IEEE.Please read the form carefully and keep a copy for your files.

Author Online Use

6. Personal Servers. Authors and/or their employers shall have the right to post the accepted version of IEEE-copyrighted articles on their own personal servers or the servers of their institutions or employers without permission from IEEE, provided that the posted version includes a prominently displayed IEEE copyright notice and, when published, a full citation to the original IEEE publication, including a link to the article abstract in IEEE Xplore.Authors shall not post the final, published versions of their papers.

7. Classroom or Internal Training Use. An author is expressly permitted to post any portion of the accepted version of his/her own IEEE-copyrighted articles on the authors personal web site or the servers of the authors institution or company in connection with the authors teaching, training, or work responsibilities, provided that the appropriate copyright, credit, and reuse notices appear prominently with the posted material. Examples of permitted uses are lecture materials, course packs, e-reserves, conference presentations, or in-house training courses.

Statistical decision by using quantized observations

Rémi Cogranne, Cathel Zitzmann, Lionel Fillatre, Florent Retraint, Igor Nikiforov, Philippe Cornu
ICD - LM2S - Université de Technologie de Troyes - UMR STMR CNRS
12, rue Marie Curie - B.P. 2060 - 10010 Troyes cedex - France
Email: firstname.lastname@utt.fr

***Abstract*—In the last two decades substantial progress has been made in the detection of hidden information or hidden communication channels in media files or streams. Typically, it is necessary to reliably detect in a huge set of files (image, audio, and video) which of these files contain the hidden information. The goal of this paper is to study the problem of hypothesis testing based on quantized observations by using a parametric statistical model with nuisance parameters and to apply the obtained tests to the hidden information detection.**

I. Introduction and contribution

In a certain operational context of hidden information detection, the most important challenge is to get the detection algorithms with analytically predictable and bounded probabilities of false alarm and missed detection. These algorithms should be immediately applicable without any supervised learning methods using sets of training examples (i.e. without SVM-based algorithms).

A detailed analysis of this problem shows that the following theoretical challenges remain unsolved :

- How to deal with the quantized observations? How does the quantization impact the probabilities of false alarm and missed detection ?
- What is the benefits from using a parametric statistical model of cover media (or cover channel) for hidden information detection ?

The media files or streams are usually obtained by using digital recording device which obligatory includes a quantization. Physically, the parametric statistical model defines the media or stream in the continuous observation space but the decision should be done by using the quantized output. It is worth noting that the existence of a quantizer between the sensor and the estimation/decision algorithm leads to the increasing complexity of estimation/decision methods. Many results from the classical estimation theory are not applicable to quantized data (for example, the Gauss-Markov theorem). Some results on the statistical inference by using quantized observations are available in the literature (see for instance [1], [2], [3] for estimation and [4], [5], [1], [6], [7] for decision theory). Nevertheless, the problem of binary decision with quantized observations and nuisance parameters in the case of composite hypotheses remain unsolved.

The contribution of this paper with respect to previously published results is the following: 1) dealing with quantized observations in the presence of nuisance parameters; 2) a new model of useful signal (the information hidden in the least significant bit (LSB)); 3) the analysis of the ("non-fine") quantization impact on the probability of false alarm and missed detection. Let us also stress that the problem discussed in the paper is quite different from the previously published works in signal detection [4], [5], [1], [6], [7] by the fact that the quantizer cannot be optimized for hidden information detection because it is chosen by the designers of digital cameras, voice recorders, etc.

This work is supported by French National Agency (ANR) through ANR-CSOSG Program (Project ANR-07-SECU-004)

II. Statistical decision based on quantized observations

A. Model of quantized cover media

Let us assume that the observation vector $C_n = (c_1, \dots, c_n)^T$ which characterizes a cover media is defined in the following manner :

$$C_n = Q_1[Y_n], \quad Y_n \sim P_\theta, \tag{1}$$

where $Q_1[y_i] = \lfloor y_i \rfloor$ is the operation of uniform quantization (integer part of y_i) and the vector $Y_n = (y_1, \dots, y_n)^T$ follows the distribution P_θ parameterized by the parametric vector θ which describes the properties of media files or streams. In the framework of hidden information detection, θ is a nuisance parameter. The binary representation of c (the index is omitted to seek simplicity) is $c = Q_1[y] = \sum_{i=0}^{q-1} b_i 2^i$, where $b_i = \{0, 1\}$. A simplified model of quantization is used in this paper. It is assumed that the saturation is absent, i.e. the probability of the excess over the boundary 0 and $2^q - 1$ for the observation y is negligible.

B. Problem statement : test between two hypotheses

First, let us define two alternative hypotheses for one quantized observation z (seeking simplicity) :

$$\mathcal{H}_0 \ : \ z = c = Q_1[y] \sim Q_{Q_1} = [q_0, \dots, q_{2^q-1}]$$

and

$$\mathcal{H}_1 \ : z = \begin{cases} Q_2[y] + z_s & \text{with probability} \quad R, \quad z_s \in \{0, 1\}, \\ c = Q_1[y] & \text{with probability} \quad 1 - R, \end{cases}$$

where $Q_2[y] = \sum_{i=1}^{q-1} b_i 2^i$, is an uniform quantization by using 2^{q-1} thresholds, $Q_2[y] \sim Q_{Q_2}$, $z_s \sim Q_s = \mathrm{B}(1, p)$ is the Bernoulli distribution which defines the hidden information (usually $p = 0.5$). In the other words, to get the double quantization $Q_2[y]$ from $Q_1[y]$ it is assumed that the LSB is deleted, i.e. $b_0 \equiv 0$. Hence, under hypothesis $\mathcal{H}_1$, the LSB is used as a container of hidden information. In the rest of the paper it is assumed that $Q_2[z] = Q_2[y]$.

C. A known embedding rate. Two simple hypotheses : likelihood ratio test

Let us suppose that the distributions $Q_s(z_s) = 1/2$, Q_{Q_1}, Q_{Q_2} and the embedding rate R are exactly known. In this case the likelihood ratio (LR) for one observation is written as follows :

$$\Lambda_R(z) = R\frac{Q_{Q_2}\left(Q_2[z]\right)}{2Q_{Q_1}\left(z\right)} + (1-R). \tag{2}$$

The most powerful (MP) Neyman-Pearson test over the class

$$\mathcal{K}_{\alpha_0} = \{\delta : \mathbb{P}_0(\delta(Z_n) = \mathcal{H}_1) \leq \alpha_0\}$$

is given by the following decision rule :

$$\delta_R(Z_n)=\begin{cases} \mathcal{H}_0 \;\text{ if }\; \Lambda_R(Z_n)=\prod\limits_{i=1}^{n}\Lambda_R(z_i) \;<\; h \\ \mathcal{H}_1 \;\text{ if }\; \Lambda_R(Z_n)=\prod\limits_{i=1}^{n}\Lambda_R(z_i) \;\geq\; h \end{cases}. \tag{3}$$

The threshold h is defined as a solution of $\mathbb{P}_0(\Lambda_R(Z_n) \geq h) = \alpha_0$, where $\mathbb{P}_i(\ldots)$ denotes the probability under hypothesis $\mathcal{H}_i$, $i = 0, 1$. The MP test $\delta_R(Z_n)$ maximizes the power

$$\beta_{\delta_R} = 1 - \mathbb{P}_1(\delta_R(Z_n) = \mathcal{H}_0) = 1 - \alpha_1$$

over the class $\mathcal{K}_{\alpha_0}$.

D. The moments of approximate log likelihood ratio

Let us start with the simplest case of equation (2), where $Y_n \sim \mathcal{N}(\theta, \sigma^2)$. It is easy to see that for any R the LR given by (2) depends on the observations through the LR

$$\Lambda_1(z) = \frac{Q_{Q_2}\left(Q_2[z]\right)}{2Q_{Q_1}\left(z\right)}$$

computed under assumption that $R = 1$. The exact equation of this log LR is given by :

$$\begin{aligned}\log\Lambda_1(Z_n) \;=\; \sum_{i=1}^{n}\frac{1}{2\sigma^2}\Big[&-\left(Q_2[z_i]+1+\eta_{2,i}-\theta\right)^2 \\ &+\left(z_i+0.5+\eta_{1,i}-\theta\right)^2\Big].\end{aligned}$$

The exact expression of the log LR $\log\Lambda_1(Z_n)$ is complicated due to the corrective terms $\eta_{1,i}$ and $\eta_{2,i}$. The calculation shows that the impact of these terms on the log LR is usually negligible. The approximate (without the corrective terms) equation of the log LR is simpler

$$\log\widetilde{\Lambda}_1(Z_n)=\sum_{i=1}^{n}\frac{1}{2\sigma^2}\Big[-\left(Q_2[z_i]+1-\theta\right)^2+\left(z_i+0.5-\theta\right)^2\Big]. \tag{4}$$

Under hypothesis $\mathcal{H}_0$, the approximate log LR can be rewritten as follows

$$\log\widetilde{\Lambda}_1(Z_n) = \sum_{i=1}^{n}\left[\frac{\zeta_i(b_{0,i}-0.5)}{\sigma^2} - \frac{1}{8\sigma^2}\right],$$

where $\zeta_i = z_i + 0.5 - \theta$, $b_{0,i} = \mathrm{LSB}(z_i)$ and under hypothesis $\mathcal{H}_1$ is

$$\log\widetilde{\Lambda}_1(Z_n) = \sum_{i=1}^{n}\left[\frac{\xi_i(b_{0,i}-0.5)}{\sigma^2} + \frac{1}{8\sigma^2}\right],$$

where $\xi_i = Q_2[z_i] + 1 - \theta$ and $b_{0,i} = z_{s,i}$.

It follows from the central limit theorem that the ratio

$$\frac{\log\widetilde{\Lambda}_1(Z_n) - n\mathbb{E}_i\left(\log\widetilde{\Lambda}_1(z)\right)}{\sigma_i\sqrt{n}} \underset{n\to\infty}{\rightsquigarrow} \mathcal{N}(0,1), \;\; i = 0, 1,$$

where $\sigma_i^2 = \mathrm{Var}_i\left(\log\widetilde{\Lambda}_1(z)\right)$, will converge in distribution to the standard normal distribution as n goes to infinity. The expectation and variance are denoted by $\mathbb{E}_i(\ldots)$ and $\mathrm{Var}_i(\ldots)$ under $\mathcal{H}_i$, respectively. Hence, to compute the error probabilities it is necessary to get the expectations and variances of the approximate log LR. Under hypothesis $\mathcal{H}_0$, the expectation of the approximate log LR is given by the following expression

$$m_0 = \mathbb{E}_0\left[\log\widetilde{\Lambda}_1(z)\right] = -\frac{1}{8\sigma^2} + \frac{\varepsilon}{\sigma^2}, \tag{5}$$

where the coefficient ε defines the impact of the quantization. This coefficient is given by

$$\begin{aligned}\varepsilon=&\;\mathbb{E}_0\left[\zeta(b_0-0.5)\right] \\ =&\sum_{m=-\infty}^{\infty}\left[\Phi\left(\frac{2m+2-\theta}{\sigma}\right)-\Phi\left(\frac{2m+1-\theta}{\sigma}\right)\right]\frac{(2m+1.5-\theta)}{2} \\ -&\sum_{m=-\infty}^{\infty}\left[\Phi\left(\frac{2m+1-\theta}{\sigma}\right)-\Phi\left(\frac{2m-\theta}{\sigma}\right)\right]\frac{(2m+0.5-\theta)}{2}.\end{aligned} \tag{6}$$

Finally, the variance is given by

$$\sigma_0^2 = \mathrm{Var}_0\left[\log\widetilde{\Lambda}_1(z)\right] = \frac{\mathbb{E}_0\left[\zeta^2\right]-4\varepsilon^2}{4\sigma^4}, \tag{7}$$

where

$$\mathbb{E}_0\left[\zeta^2\right]=\sum_{m=-\infty}^{\infty}\left[\Phi\left(\frac{m+1-\theta}{\sigma}\right)-\Phi\left(\frac{m-\theta}{\sigma}\right)\right](m+0.5-\theta)^2.$$

Under hypothesis $\mathcal{H}_1$, the expectation and variance of the approximate log LR are given by the following expressions

$$m_1 = \mathbb{E}_1\left[\log\widetilde{\Lambda}_1(z)\right] = \frac{1}{8\sigma^2}, \tag{8}$$

$$\sigma_1^2 = \mathrm{Var}_1\left[\log\widetilde{\Lambda}_1(z)\right] = \frac{1}{4\sigma^4}\mathbb{E}_1\left[\xi^2\right], \tag{9}$$

where

$$\mathbb{E}_1\left[\xi^2\right]=\sum_{m=-\infty}^{\infty}\left[\Phi\left(\frac{2m+2-\theta}{\sigma}\right)-\Phi\left(\frac{2m-\theta}{\sigma}\right)\right](m+1-\theta)^2, \tag{10}$$

respectively. To illustrate the impact of the quantization, let us assume the following parameters of the Gaussian cover media model : $\widetilde{R} = 1$, $\theta \in [128; 132]$, $\sigma = 1$ and $n = 200$. The comparison of theoretical equations for α_1 with the Monte Carlo simulation (10^6 repetitions) are presented in Figure 1. This figure shows the probability of missed detection α_1 calculated with (solid line) and without (dashed line) taking

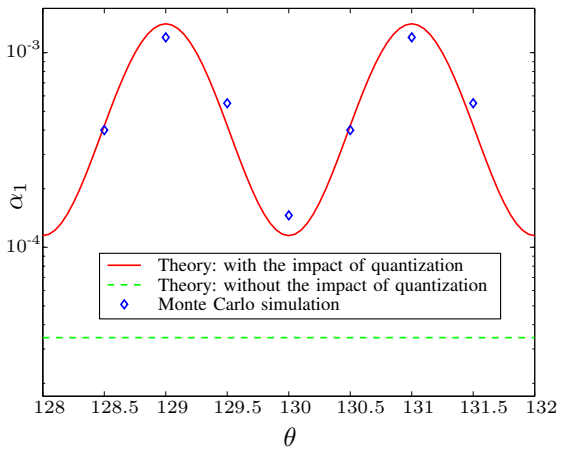

Figure 1. The impact of the quantization on the probability of missed detection α_1.

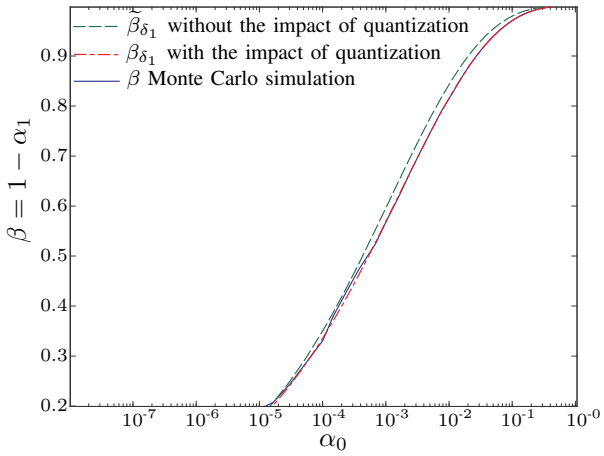

Figure 2. The power of the MP test given by (3) as a function of α_0 : with the impact of quantization β_{δ_1} (dash dotted line); without the impact of quantization $\widetilde{\beta}_{\delta_1}$ (dashed line); Monte Carlo simulation (solid line).

into account the impact of quantization for the prescribed significance level $\alpha_0 = 10^{-3}$. As it follows from Figure 1, the impact of quantization on the probability of missed detection α_1 is significant.

The following simplified equations can be proposed for the expectation and variance of the approximate log LR given by (4) without taking into account the impact of quantization

$$m_i = (-1)^{i+1}\frac{1}{8\sigma^2}, \quad \sigma_i^2 = \frac{1}{4\sigma^2}, \quad i = 0, 1. \tag{11}$$

Theorem 1: Let us assume that the true embedding rate takes an arbitrary value $\widetilde{R} \;: 0 < \widetilde{R} \leq 1$. The power β_{δ_1} of the MP test (3) with the log LR $\log \widetilde{\Lambda}_1(Z_n)$ given by (4) can be approximated by

$$\beta_{\delta_1} \simeq 1 - \Phi\left(\Phi^{-1}(1-\alpha_0)\frac{\sigma_0}{\sigma_{\widetilde{R}}} - \frac{(m_1 - m_0)\widetilde{R}\sqrt{n}}{\sigma_{\widetilde{R}}}\right) \tag{12}$$

for large n. The expectations m_i and variance σ_0^2 are computed by using equations (5) - (10) (resp. (11)) with (resp. without) taking into account the impact of quantization. The variance $\sigma^2_{\widetilde{R}}$ is also computed with taking into account the impact of quantization

$$\begin{aligned}\sigma^2_{\widetilde{R}} \;=\; & \frac{1}{4\sigma^2}\left[\left(\mathbb{E}_1\left[\xi^2\right] + \frac{1}{16}\right)\widetilde{R} + \left(\mathbb{E}_0\left[\zeta^2\right] + \frac{1}{16} - \varepsilon\right)(1-\widetilde{R})\right] \\ & - \left[m_1\widetilde{R} + m_0(1-\widetilde{R})\right]^2\end{aligned} \tag{13}$$

or without taking into account the impact of quantization

$$\sigma^2_{\widetilde{R}} = \frac{1 + \widetilde{R} - \widetilde{R}^2}{4\sigma^2}. \tag{14}$$

The comparison of Monte Carlo simulation (with 10^6 repetitions) of the test given by (3) with these two approaches is depicted in Figure 2 for $\theta = 129$, $\sigma = 1.5$ and $n = 100$. It is easy to see that the results from the Monte Carlo simulation perfectly coincide with equation (12) for β_{δ_1} taking into account of the impact of quantization.

III. An unknown embedding rate

A. Two composite hypotheses

Let us assume that the distributions Q_s, Q_{Q_1}, Q_{Q_2} are known, but the embedding rate R is unknown. The following alternative composite hypotheses have to be tested by using n observations Z_n representing the cover media :

$$\mathcal{H}_0 = \{R \leq r^*\} \;\text{ against }\; \mathcal{H}_1 = \{R > r^*\}. \tag{15}$$

Hence, the LR (2) becomes

$$\Lambda_{R_0,R_1}(Z_n) = \prod_{i=1}^{n} \frac{R_1 \frac{1}{2} Q_{Q_2}\left(Q_2[z_i]\right) + (1-R_1)Q_{Q_1}\left(z_i\right)}{R_0 \frac{1}{2} Q_{Q_2}\left(Q_2[z_i]\right) + (1-R_0)Q_{Q_1}\left(z_i\right)}, \tag{16}$$

where $R_0 \leq r^* < R_1$. The main difficulty is that the values of acceptable R_0 and unacceptable R_1 embedding rates are unknown. The ultimate challenge for anyone in the case of two composite hypotheses is to get a uniformly MP (UMP) test δ which maximizes the power function

$$\beta(R) = 1 - \mathbb{P}_R(\delta(Z_n) = \mathcal{H}_0),$$

where $\mathbb{P}_R(\ldots)$ denotes the probability under the assumption that the embedding rate is equal to R, for any $R > r^*$ over the class $\mathcal{K}_{\alpha_0} = \left\{\delta : \sup_{R \leq r^*} \mathbb{P}_R(\delta(Z_n) = \mathcal{H}_1) \leq \alpha_0\right\}$. An efficient solution is based on the asymptotic local approach proposed by L. Le Cam [8]. The idea of this approach is that the "distance" between alternative hypotheses depends on the sample size n in such a way that the two hypotheses get closer to each other when n tends to infinity. By using an asymptotic expansion of the log LR, a particular hypothesis testing problem can be locally reduced to a relatively simple UMP hypothesis testing problem between two Gaussian scalar means [8]. This approach is applied to the following model

$$Z_n \sim Q_R = \prod_{i=1}^{n} R\frac{1}{2}Q_{Q_2}\left(Q_2[z_i]\right) + (1-R)Q_{Q_1}\left(z_i\right).$$

Let us consider two converging sequences of hypotheses $\mathcal{H}_j(n) = \{R \in \mathbb{R}_j(n)\}$ $(j = 0, 1)$. The sets $\mathbb{R}_j(n)$ are of the form $\mathbb{R}_j(n) = r^* + \frac{1}{\sqrt{n}}\mu_r$. The rate of convergence is $\frac{1}{\sqrt{n}}$. Seeking simplicity, let us denote the hypotheses $\mathcal{H}_0(n) = \{R = r^*\}$ and $\mathcal{H}_1(n) = \{R = r^* + \frac{1}{\sqrt{n}}\mu_r\}$. The log LR $\log\Lambda\left(Z_n; \frac{1}{\sqrt{n}}\mu_r\right) = \log Q_{r^*+\frac{1}{\sqrt{n}}\mu_r}(Z_n) - \log Q_{r^*}(Z_n)$ possess the following asymptotic expansion :

$$\log\Lambda\left(Z_n; \frac{1}{\sqrt{n}}\mu_r\right) \simeq \frac{1}{\sqrt{n}}\mu_r\zeta_n(Z_n; r^*) - \frac{1}{2}\mu_r^2\mathcal{F}(r^*)$$

where $\mathcal{F}(R)$ is the Fisher information and

$$\zeta_n(Z_n; r^*) = \sum_{i=1}^{n} \left.\frac{\partial \log Q_R(z_i)}{\partial R}\right|_{R=r^*} \tag{17}$$

is the function of *efficient score* which is asymptotically Gaussian

$$\zeta_n(Z_n; r^*) \rightsquigarrow \begin{cases} \mathcal{N}(0, \mathcal{F}(r^*)) & \text{under } z_i \sim Q_{r^*} \\ \mathcal{N}(\mathcal{F}(r^*)\mu_r, \mathcal{F}(r^*)) & \text{under } z_i \sim Q_{r^*+\frac{\mu_r}{\sqrt{n}}} \end{cases}. \tag{18}$$

It can shown that the efficient score is given by

$$\zeta_n(Z_n; r^*) = \sum_{i=1}^{n} \zeta(z_i; r^*) = \sum_{i=1}^{n} \frac{\Lambda_1(z_i) - 1}{r^*\Lambda_1(z_i) + (1 - r^*)} \tag{19}$$

and the Fisher information $\mathcal{F}(R)$ is

$$\mathcal{F}(R) = \mathbb{E}_R\left[\frac{\Lambda_1(z) - 1}{R\Lambda_1(z) + (1 - R)}\right]^2.$$

Therefore, the local UMP test to chose between two alternative hypotheses (15) is given by the following rule :

$$\delta_{r^*}(Z_n) = \begin{cases} \mathcal{H}_0 & \text{if} \quad \zeta_n(Z_n; r^*) < h \\ \mathcal{H}_1 & \text{if} \quad \zeta_n(Z_n; r^*) \geq h \end{cases}. \tag{20}$$

B. Tractable likelihood ratio

As it follows from previous sections, in the case of arbitrary embedding rate R, an optimal solution is based on the log LR given by (16) if R_0 and R_1 are known or on the efficient score given by (19) if they are unknown but the value r^* is known. It is easy to see that in both cases the useful information obtained from observations Z_n of cover media (with or without a secret message) is concentrated in $\log\Lambda_1(z)$. Let us denote $y \stackrel{\text{def.}}{=} \zeta(z; r^*)$, hence

$$y = f(x; r^*) \stackrel{\text{def.}}{=} \frac{e^x - 1}{r^*e^x + 1 - r^*} \quad \text{with} \quad x \stackrel{\text{def.}}{=} \log\Lambda_1(z). \tag{21}$$

The asymptotic normality of $\zeta_n(Z_n; r^*) = \sum_{i=1}^{n}\zeta(z_i; r^*)$ is warranted due to Le Cam expansion (see equation (18)). Hence, to compute the loss of optimality of the MP test based on $\log\Lambda_1(Z_n)$ given by (3) and designed for $R = 1$ against the local MP test given by (20) with a certain value r^* and against the MP test based on $\log\Lambda_{\widetilde{R}}(Z_n)$ when the true embedding rate is $\widetilde{R}$ it is sufficient to compute first two moments of corresponding statistics under alternative hypotheses $\mathcal{H}_0$ and $\mathcal{H}_1$, see details in [9].

C. A more realistic model of cover media

As it follows from equation (12), the power β of an optimal test depends on the standard deviation σ of cover media for a given rate of false alarm α_0. Hence, to increase the power β, someone has to reduce the standard deviation σ by using a parametric model of cover media. As it is motivated in [10], the observation vector (pixels) extracted from the cover media file (digital image, for instance) by using a specially chosen segment or mask is characterized "block by block" by a regression model. Let us split the observation vector C in M statistically independent n dimensional sub-vectors C_j, i.e. $C^T = (C_1^T, \ldots, C_M^T)$. It is assumed that each segment C_j is approximated by the following regression model :

$$C_j = Q_1[Y_j],\ Y_j = Hx_j + \xi \sim \mathcal{N}(Hx_j, \sigma_j^2 I_n),\ j = 1, \ldots, M,$$

where H is a known $[n \times l]$ full rank matrix, $n > l$, I_n is an $(n \times n)$ identity matrix, $x_j \in \mathbb{R}^l$ is a nuisance parameter and σ^2 is the residual variance. The vector C_j (pixels) is extracted from the cover media file (digital image, for instance) by using a specially chosen segment or mask. The l columns of H span a column subspace $R(H)$ of the observation space $Y_j \in \mathbb{R}^n$. Such a parametric model is an efficient method to reduce the standard deviation σ_j. The new hypothesis testing problem with a parametric model of cover media consists in deciding between the following hypotheses

$$\mathcal{H}_0 \ : \ Z = C = Q_1[Y],\ \ Y = (Y_1^T, \ldots, Y_M^T)^T \in \mathbb{R}^{Mn} \tag{22}$$

$$\mathcal{H}_1 : z_i = \begin{cases} Q_2[y_i] + z_{s,i} & \text{with probability} \quad R \\ c_i = Q_1[y_i] & \text{with probability} \quad 1\text{–}R \end{cases}, \tag{23}$$

where $Y_j \sim \mathcal{N}(H_j x, \sigma_j^2 I_n)$. In practice, x_j and σ_j^2 are unknown. The theoretical aspects of dealing with nuisance parameters in the framework of statistical decision theory is discussed in [11]. An efficient approach to this problem is based on the theory of invariance in statistics. The optimal invariant tests and their properties in the context of image processing have been designed and studied in [12], [13]. The parameter vector x_j can be estimated by using $Q_2[Y_j]$ which is free from the embedded information. The "approximate" log GLR is given by

$$\log\widehat{\Lambda}_1(Z_j) \simeq \frac{1}{\sigma_j^2}\left[P_H^{\perp}Q_2[Z_j]\right]^T[B_0 - 0.5 \cdot \mathbf{1}_n] + \frac{n}{8\sigma_j^2}, \tag{24}$$

where $B_0 = (b_{0,1}, \ldots, b_{0,n})^T$, $\mathbf{1}_n = (1, \ldots, 1)^T$ and $P_H^{\perp} = I_n - H(H^TH)^{-1}H^T$ is a projection matrix.

Under hypothesis $\mathcal{H}_0$, the expectation and variance of the "approximate" log GLR for the total observation vector Y are given by the following expressions :

$$m_0 = \mathbb{E}_0\left[\sum_{j=1}^{M}\log\widehat{\Lambda}_1(Z_j)\right] \simeq \frac{M(2l - n)}{8\overline{\sigma}^2} \tag{25}$$

with $\frac{1}{\overline{\sigma}^2} = \frac{1}{M}\sum_{j=1}^{M}\frac{1}{\sigma_j^2}$ and

$$\sigma_0^2 = \mathrm{Var}_0\left[\sum_{j=1}^{M}\log\widehat{\Lambda}_1(Z_j)\right] \simeq M(n-l)\left[\frac{1}{4\overline{\sigma}^2} + \frac{1}{16\overline{\sigma}^4}\right] \tag{26}$$

with $\frac{1}{\overline{\sigma}^4}=\frac{1}{M}\sum_{j=1}^{M}\frac{1}{\sigma_j^4}$. Let us assume that the true embedding rate takes an arbitrary value $\widetilde{R} \;:\; 0<\widetilde{R}\leq 1$. Under hypothesis $\mathcal{H}_1$ with the true embedding rate $\widetilde{R}$, the expectation and variance of the "approximate" log GLR for the total observation vector Y are given by the following expressions :

$$m_{\widetilde{R}}=\mathbb{E}_{\widetilde{R}}\left[\sum_{j=1}^{M}\log\widehat{\Lambda}_1(Z_j)\right]\simeq\frac{M(2l-n+2\widetilde{R}(n-l))}{8\overline{\sigma}^2} \tag{27}$$

and

$$\sigma^2_{\widetilde{R}}=\mathrm{Var}_{\widetilde{R}}\left[\sum_{j=1}^{M}\log\widehat{\Lambda}_1(Z_j)\right]\simeq\frac{M(n-l)}{4\overline{\sigma}^2}+\frac{M(n-l)(1-R)^2}{16\overline{\sigma}^4} \tag{28}$$

Theorem 2: Let us assume that the Lindeberg's condition imposed on the log LR $\log\widehat{\Lambda}_1(Z_j)$ is satisfied. It follows from the central limit theorem that the following fraction

$$\frac{\sum_{j=1}^{M}\log\widehat{\Lambda}_1(Z_j)-\mathbb{E}_{\widetilde{R}}\left[\sum_{j=1}^{M}\log\widehat{\Lambda}_1(Z_j)\right]}{\sqrt{\mathrm{Var}_{\widetilde{R}}\left[\sum_{j=1}^{M}\log\widehat{\Lambda}_1(Z_j)\right]}}\underset{M\to\infty}{\rightsquigarrow}\mathcal{N}(0,1) \tag{29}$$

weakly converges to the standard normal distribution. For large M, the power β_{δ_1} of the test (3) with the log LR $\sum_{j=1}^{M}\log\widehat{\Lambda}_1(Z_j)$ given by (24) can be approximated

$$\beta_{\delta_1}\simeq 1-\Phi\left(\Phi^{-1}(1-\alpha_0)\frac{\sigma_0}{\sigma_{\widetilde{R}}}-\frac{(m_{\widetilde{R}}-m_0)}{\sigma_{\widetilde{R}}}\right) \tag{30}$$

where m_0, $m_{\widetilde{R}}$, σ_0 and $\sigma_{\widetilde{R}}$ are calculated by using equations (25) - (28).
If the residual variance σ_j^2 is unknown, then the estimation $\widehat{\sigma}_j^2=\frac{1}{n-l}\left\|P_H^{\perp}Q_2[Z_j]\right\|_2^2$ is used in equation (24).

D. Relation between the proposed and some known heuristic tests

The first right hand side term in equation (24) defines the sensitivity of the test because the second right hand side term $\frac{n}{8\widehat{\sigma}^2}$ does not depend on the embedded secret message. The first right hand side term in equation (24) represents an inner product of the vector of "residuals" $\varepsilon=P_HQ_2[Z_n]$, i.e. the vector of projection of Y_n on the orthogonal complement $R(H)^{\perp}$ of the column space $R(H)$, and the vector $[B_0-0.5\cdot\mathbf{1}_n]$ composed of $\mathrm{LSB}(z_i)-0.5$:

$$\sum_{i=1}^{n}\overbrace{\widehat{\sigma}^{-2}}^{=\text{"weight"}}\cdot\overbrace{(Q_2[y_i]-(H\widehat{x})_i+1)}^{=\text{"residual" }\varepsilon_i}\cdot\overbrace{(b_{0,i}-0.5)}^{=\mathrm{LSB}(z_i)-0.5}. \tag{31}$$

Let us now compare the last equation with the recently developed WS steganalysers reputed very efficient [14], [15]. These steganalysers are based on the following statistics :

$$\sum_{i=1}^{n}\overbrace{w_i}^{=\text{"weight"}}\cdot\overbrace{(z_i-\mathcal{F}(z)_i)}^{=\text{"residual" }\varepsilon_i}\cdot\overbrace{(z_i-\overline{z}_i)}^{=2\cdot(\mathrm{LSB}(z_i)-0.5)}, \tag{32}$$

where $\mathcal{F}(s)$ denotes a "filter" dedicated to estimate the cover-image by filtering the stego-image, the weight w_i is chosen as $\frac{1}{1+\sigma_i^2}$, σ_i^2 is the "local" variance and $\overline{z}_i$ denotes the nonnegative integer z_i with the LSB flipped. As it follows from equations (31) - (32), the steganalysers developed in [14], [15] coincide with the first term of the tractable log GLR (24). Nevertheless, the second right hand side term $\frac{n}{8\widehat{\sigma}^2}$ of (24) is also necessary to correctly calculate the threshold h as a solution of the following equation

$$\mathbb{P}_0(\log\widehat{\Lambda}_1(Z_n)\geq h)=\alpha_0.$$

IV. Conclusions

The problem of hypothesis testing using a parametric statistical model with nuisance parameters based on quantized observations has been discussed. In practice this problem is related to the detection of hidden information. Two new phenomena have been studied : *i)* the impact of observation quantization on the probabilities of false alarm and missed detection; *ii)* the benefits from using a parametric statistical model of cover media for hidden information detection.

References

[1] H. Poor, "Fine quantization in signal detection and estimation," *Information Theory, IEEE Transactions on*, vol. 34, no. 5, pp. 960 –972, sep 1988.

[2] L. Y. Wang and G. G. Yin, "Asymptotically efficient parameter estimation using quantized output observations," *Automatica*, vol. 43, pp. 1178–1191, July 2007.

[3] F. Gustafsson and R. Karlsson, "Statistical results for system identification based on quantized observations," *Automatica*, vol. 45, pp. 2794–2801, December 2009.

[4] S. Kassam, "Optimum quantization for signal detection," *Communications, IEEE Transactions on*, vol. 25, no. 5, pp. 479 – 484, may 1977.

[5] H. Poor and J. Thomas, "Applications of Ali-Silvey distance measures in the design generalized quantizers for binary decision systems," *Communications, IEEE Transactions on*, vol. 25, no. 9, pp. 893 – 900, sep 1977.

[6] B. Picinbono and P. Duvaut, "Optimum quantization for detection," *Communications, IEEE Transactions on*, vol. 36, no. 11, pp. 1254 – 1258, nov 1988.

[7] R. S. Blum, S. A. Kassam, and H. V. Poor, "Distributed detection with multiple sensors II. advanced topics," *Proceedings of the IEEE*, vol. 85, no. 1, pp. 64–79, 1997.

[8] L. Le Cam, *Asymptotic Methods in Statistical Decision Theory*. Series in Statistics, Springer, New York, 1986.

[9] C. Zitzmann, R. Cogranne, F. Retraint, I. Nikiforov, L. Fillatre, and P. Cornu, "Statistical decision methods in hidden information detection," in *Proceedings of the 13th Information Hiding Conference, Prague, May 18-20*, 2011, pp. 1–15.

[10] R. Cogranne, C. Zitzmann, L. Fillatre, F. Retraint, I. Nikiforov, and P. Cornu, "A cover image model for reliable steganalysis," in *Proceedings of the 13th Information Hiding Conference, Prague, May 18-20*, 2011, pp. 1–15.

[11] E. L. Lehmann, *Testing Statistical Hypotheses*. 2nd edn. Springer, New York, 1986.

[12] L. Scharf and B. Friedlander, "Matched subspace detectors," *IEEE Trans. Signal Process.*, vol. 42, no. 8, pp. 2146–2157, 1994.

[13] L. Fillatre, I. Nikiforov, and F. Retraint, "ε-optimal non-bayesian anomaly detection for parametric tomography," *IEEE Transactions on Image Processing*, vol. 17, no. 11, pp. 1985–1999, 2008.

[14] J. Fridrich and M. Goljan, "M.: On estimation of secret message length in LSB steganography in spatial domain," in *of Proc. SPIE*. Addison-Wesley, 2004, pp. 23–34.

[15] A. D. Ker and R. Böhme, "Revisiting weighted stego-image steganalysis," in *Proceedings SPIE, Electronic Imaging, Security, Forensics, Steganography, and Watermarking of Multimedia Contents X, volume 6819*. San Jose, CA, 27-31 January, 2008, pp. 5 1 – 5 17.