Digital Signal Processing 24 (2014) 144–161

Contents lists available at ScienceDirect

Digital Signal Processing

www.elsevier.com/locate/dsp

Hidden information detection using decision theory and quantized samples: Methodology, difficulties and results

Rémi Cogranne [a,*], Florent Retraint [a,1], Cathel Zitzmann [a], Igor Nikiforov [a], Lionel Fillatre [b], Philippe Cornu [a]

[a] ICD, LM2S, Université de Technologie de Troyes, UMR STMR CNRS, 12, rue Marie Curie, CS 42060, 10004 Troyes cedex, France
[b] Université de Nice Sophia-Antipolis, Laboratoire d'Informatique, Signaux et Systèmes de Sophia-Antipolis, 06900 Sophia Antipolis, France

ARTICLE INFO

Article history:
Available online 3 October 2013

Keywords:
Hypothesis testing theory
LSB steganalysis
Quantized data
Optimal steganalysis
Nuisance parameters

ABSTRACT

This paper studies the detection of Least Significant Bits (LSB) steganography in digital media by using hypothesis testing theory. The main goal is threefold: first, it is aimed to design a test whose statistical properties are known, this especially allows the guaranteeing of a false alarm probability. Second, the quantization of samples is studied throughout this paper. Lastly, the use of a linear parametric model of samples is used to estimate unknown parameters and design a test which can be used when no information on cover medium is available. To this end, the steganalysis problem is cast within the framework of hypothesis testing theory and digital media are considered as quantized signals. In a theoretical context where media parameters are assumed to be known, the Likelihood Ratio Test (LRT) is presented. Its statistical performances are analytically established; this highlights the impact of quantization on the most powerful steganalyzer. In a practical situation, when image parameters are unknown, a Generalized LRT (GLRT) is proposed based on a local linear parametric model of samples. The use of such model allows us to establish GLRT statistical properties in order to guarantee a prescribed false-alarm probability. Focusing on digital images, it is shown that the well-known WS (Weighted-Stego) is close to the proposed GLRT using a specific model of cover image. Finally, numerical results on natural images show the relevance of theoretical findings.

1. Introduction

Information hiding concerns the transmission of a secret message buried in a host digital medium. It has recently received increasing interest driven by the large number of ensuing application such as watermark-based authentication and fingerprint tracing. Unfortunately, potential malicious uses, among which steganography, have also emerged. The "prisoners problem" [43] illustrates a typical scenario of steganography and steganalysis. Alice and Bob, two prisoners, communicate by embedding a secret binary message $\boldsymbol{M}$ into a cover-object $\boldsymbol{C}$ to obtain a stego-object $\boldsymbol{S}$ which is then sent through a public channel. Wendy, the warden, examines all their communications and tries to detect whether an inspected object $\mathbf{Z}$, either a cover-object $\boldsymbol{C}$ or a stego-object $\boldsymbol{S}$, contains a secret message $\boldsymbol{M}$ without being able to extract $\boldsymbol{M}$.

With many tools available in the public domain, steganography is within the reach of anyone, for legitimate or malicious purpose; it is thus of a crucial interest to be able to efficiently detect steganographic content among a (possibly huge) set of media.

1.1. Reliable detection of hidden bits as a main motivation

Many methods have recently been proposed to detect steganographic content [4,12,37] among which some are very efficient (see results of BOSS contest [2] for example). These methods can be roughly divided into four categories:

1. Only a few detectors rely on the hypothesis testing theory [7,9,13,16,32,48]. The performance of these detectors can be expressed analytically which allows the guaranteeing of a prescribed detection-error probability. Unfortunately, these tests lack an accurate cover medium model and therefore exhibit rather poor performances.
2. Many detectors belong to the class of *structural detectors* [26] which aims at detecting specific modifications of bit replacement using local pixels' correlation. The Regular–Singular (RS) [21] the Sample Pair Analysis (SPA) [14,34] and the triple/quadruple detectors [27] are, for instance, good representatives

* Corresponding author.
E-mail address: remi.cogranne@utt.fr (R. Cogranne).
URL: http://lm2s.utt.fr/fr/membres/cogrannr.html (R. Cogranne).
[1] With the financial support from the Prevention of and Fight against Crime Programme of the European Union European Commission—Directorate-General Home Affairs (2centre.eu project). Research funded, in part, by Troyes University of Technology (UTT) strategic program COLUMBO and by National Agency for Research (ANR) Program (Project ANR-07-SECU-004).

1051-2004/$ – see front matter
http://dx.doi.org/10.1016/j.dsp.2013.09.012

of this category. These methods usually achieve overall good performances but lack a statistical model of cover media which prevents the analytic calculation of detection performance.

3. The Weighted-Stego-image (WS) analysis, initially proposed in [20] to estimate the payload size and deeply studied in [28,29], forms a class apart which is known to have good detection performance. Similarly to the structural detectors, the WS detectors rely on a local autoregressive model of media. However, the statistical properties of WS algorithms remain unknown; therefore, probabilities of detection errors can only be measured empirically.
4. Lastly, *universal* or *blind* detectors aims at detecting any steganographic scheme in any kind of image using a set of selected features and supervised learning methods to train a classifier. As in all applications of supervised learning, a difficult problem is to choose an appropriate feature set which, for detection of hidden information, is usually done empirically. Moreover, the problem of measuring classification error probabilities remains open in the framework of statistical learning [41].

In certain operational context, for example a steganalysis tool for the law enforcement agencies, a good detection performance might not suffice. To avoid false alarm, the detector should be provided with an analytically predictable error probability in order to guarantee the respect of a prescribed false-alarm probability. In fact, among the four previously described categories of steganalyzers, only detectors based on decision theory can be statistically studied.

However, the application of hypothesis testing theory is not straightforward because of some fundamental difficulties. Firstly, the cover medium is quantized which prevents a direct application of classical hypothesis tests. Secondly, a digital medium exhibits a complex and structured content which acts as nuisance parameters as it has no interest for hidden information detection. Thirdly, no assumption on the size of hidden data (or more precisely on the relative payload) is available to the steganalyst, the tested hypothesis is thus composite.

In the present work the embedding scheme is assumed to belong to the commonly used family of LSB replacement scheme. Actually, as of December 2011, WetStone Technologies Inc. has 836 data hiding software among which 582 (70%) uses LSB replacement [22]. In addition, detection of mostly encountered algorithms is more important than the detection on seldom found ones. Thus, requirement of efficient and robust detection methods for the LSB replacement steganographic scheme is still a live research topic.

1.2. Contributions of this paper

The first step in the application of hypothesis testing theory for steganalysis has been done in [13] in 2004 using a simplistic model of independent and identically distributed (i.i.d.) samples. This methodology was latter proposed in [8,9] in 2011 to design an efficient test using a complex model of cover medium. During the same year, a theoretical approach was also used in [10,50] to design locally most powerful test (around a given hidden message length) which is difficult to apply in practice.

The fundamental originalities of the proposed article are that the content of cover medium as well as the impact of quantization are considered. In fact, it is necessary to take into account these two phenomena to apply hypothesis testing theory and to precisely establish the performance of proposed statistical tests.

To highlight the contributions of the presented methodology, let us first denote that the detection performance of steganalyzers is usually measured using a fixed hidden data length with a receiver operating characteristic (ROC) curve [15]. However, it is known that in practice, detection results do not only depend on relative payload or embedding rate. For instance, it has been shown that results depend on inspected medium content [5] and hidden data length [38]; the phenomena referred to as "camera mismatch" were also observed during BOSS contest [2]. Likewise, the detection performance of proposed detectors is not compared with an optimal bound or with the most powerful test.

In fact, the relation between cover medium properties and detection performance remains unclear for most of steganalyzers and optimal detection performance has only been established under the dubious condition of neglected quantization.

On the opposite, this paper proposes a novel methodology to detect information hidden in the LSB of a digital medium using hypothesis testing theory. The main contribution is threefold:

1. Throughout this paper, the quantization of pixels is thoroughly studied in order to measure its impact on hidden information detection. It is especially shown that data quantization and hidden information in the LSB plane heavily impact on the test performance.
2. In the ideal context when medium content is known, the statistical performance of the Likelihood Ratio Test (LRT) is then analytically established. This result provides an optimal upper-bound on the power function one can expect from any detector. Moreover, it is shown that detection performance depends on an "Insertion-to-Noise Ratio" (INR) which allows us to understand how some of cover medium properties impact the performance of the LRT.
3. Finally, when the structured content of inspected medium is unknown—that is the expectation and the variance of each sample—this nuisance parameter is explicitly taken into account using a linear parametric model. Based on this model of cover content, a sub-optimal Generalized Likelihood Ratio Test (GLRT) is proposed and its statistical properties are analytically given to guarantee, in practice, a prescribed false-alarm constraint.

The relevance of theoretical findings is emphasized through numerical experimentations on simulated data and on large databases of natural images.

1.3. Organization of the paper

The paper is organized as follows. Section 2 formally states the problem of steganalysis paying a careful attention to further difficulties. Then, Section 3 details the design of the LRT and the calculation of its performances in the ideal case when the cover medium parameters are assumed to be known. The more realistic case non-i.i.d. samples is studied in Section 4. Section 5 investigates the practical case when no information about cover medium is available. First, the linear parametric model used to estimated unknown parameters of inspected medium is presented. Then, the design of the GLRT is detailed together with the calculation of its statistical properties. Numerical results and comparisons with prior art steganography detectors are presented in Section 6. Finally, Section 7 concludes the paper.

2. Problem statement and difficulties

Let the vector $\boldsymbol{C} = \{c_n\}_{n=1}^{N}$ represents a digital cover medium of N samples. Each sample is usually quantized with b bits, hence $\boldsymbol{C} \in \mathcal{Z}^N$ with $\mathcal{Z} = \{0, 1, \ldots, 2^b - 1\}$. The value of a cover sample c_n is given by:

$$c_n = \mathcal{Q}_\Delta(y_n), \tag{1}$$