Hindawi Publishing Corporation
EURASIP Journal on Information Security
Volume 2007, Article ID 45731, 14 pages
doi:10.1155/2007/45731

Research Article

Efficient Zero-Knowledge Watermark Detection with Improved Robustness to Sensitivity Attacks

Juan Ramón Troncoso-Pastoriza and Fernando Pérez-González

Signal Theory and Communications Department, University of Vigo, 36310 Vigo, Spain

Correspondence should be addressed to Juan Ramón Troncoso-Pastoriza, troncoso@gts.tsc.uvigo.es

Received 28 February 2007; Revised 20 August 2007; Accepted 18 October 2007

Recommended by Stefan Katzenbeisser

Zero-knowledge watermark detectors presented to date are based on a linear correlation between the asset features and a given secret sequence. This detection function is susceptible of being attacked by sensitivity attacks, for which zero-knowledge does not provide protection. In this paper, an efficient zero-knowledge version of the generalized Gaussian maximum likelihood (ML) detector is introduced. This detector has shown an improved resilience against sensitivity attacks, that is empirically corroborated in the present work. Two versions of the zero-knowledge detector are presented; the first one makes use of two new zero-knowledge proofs for absolute value and square root calculation; the second is an improved version applicable when the spreading sequence is binary, and it has minimum communication complexity. Completeness, soundness, and zero-knowledge properties of the developed protocols are proved, and they are compared with previous zero-knowledge watermark detection protocols in terms of receiver operating characteristic, resistance to sensitivity attacks, and communication complexity.

1. INTRODUCTION

Watermarking technology has emerged as a solution for authorship proofs or dispute resolving. In these applications, there are several requirements that watermarking schemes must fulfill, like imperceptibility, robustness to attacks that try to erase a legally inserted watermark or to embed an illegal watermark in some asset, and they must also be secure to the disclosure of information that could allow the breakage of the whole system by unauthorized parties.

The schemes that have been used up to now are symmetric, as they employ the same key for watermark embedding and watermark detection; thus, such key must be given to the party that runs the detector, which in most cases is not trusted. In order to satisfy the security requirements, two approaches have been proposed: the first one, called *asymmetric watermarking*, follows the paradigm of asymmetric cryptosystems, and employs different keys for embedding and detection; the second approach, *zero-knowledge watermarking*, makes use of zero-knowledge (ZK) protocols [1] in order to get a secure communication layer over a pre-existent symmetric protocol. In zero-knowledge watermark detection [2], a prover $\mathcal{P}$ tries to demonstrate to a verifier $\mathcal{V}$ the presence of a watermark in a given asset. Commitment schemes [3] are used to conceal the secret information, so that detection is performed without providing to $\mathcal{V}$ any information additional to the presence of the watermark.

Nevertheless, such minimum disclosure of information still allows for blind sensitivity attacks [4], that have arisen as very harmful attacks for methods that present simple detection boundaries. The ZK detection protocols presented to date—Adelsbach and Sadeghi [2] and Piva et al. [5]—are based on correlation detectors, for which blind sensitivity attacks are especially efficient.

In this paper, a new zero-knowledge blind watermark detection protocol is presented; it is based on the spread spectrum detector by Hernández et al. [6], which is optimal for additive watermarking in generalized Gaussian distributed host features (e.g., AC DCT coefficients of images). The robustness to sensitivity attacks comes from the complexity of the detection boundary for certain shape factors. Thus, when combined with zero-knowledge, it becomes secure and robust. This protocol will be compared in terms of performance and efficiency with the previous ZK protocols based

on additive spread-spectrum and Spread-Transform Dither Modulation (ST-DM), and rewritten in a form that greatly improves its communication and computation complexity.

The rest of the paper is organized as follows. In Section 2, some basics about zero-knowledge and watermark detection are reviewed, and the three studied detectors are compared, pointing out the improved robustness of the GG detector against sensitivity attacks. In Section 3, the needed ZK subprotocols are enumerated, along with their communication complexity and a detailed description of the developed proofs. Sections 4 and 5 detail the complete detection protocol and the improved version for a binary antipodal spreading sequence. Section 6 presents the security analysis for these protocols; complexity and implementation concerns are discussed in Section 7. Finally, some conclusions are drawn in Section 8.

2. NOTATION AND PREVIOUS CONCEPTS

In this section, some of the concepts needed for the development of the studied protocols are briefly introduced. Boldface lower-case letters will denote column vectors of length L, whereas boldface capital letters are used for matrices, and scalar variables will be denoted by italicized letters. Uppercase calligraphic letters represent sets or parties participating in a protocol.

2.1. Cryptographic primitives

2.1.1. Commitment schemes

Commitment schemes [3] are cryptographic tools that, given a common public parameter par_{com}, allow that one party of a protocol choose a determined value m from a finite set M and commit to his choice $C_m = \text{Com}(m, r, \text{par}_{\text{com}})$, such that he cannot modify it during the rest of the protocol; the committed value is not disclosed to the other party, thanks to the randomization produced by r, which constitutes the secret information needed to open the commitment.

The required security properties that the commit function must fulfill are *binding* and *hiding*; the first one guarantees that once produced a commitment C_m to a message m, the committer cannot open it to a different message m'; the second one guarantees that the distributions of the commitments to different messages are indistinguishable, so one commitment does not reveal any information about the concealed message. Each of these properties can be achieved either computationally or in an information-theoretic sense, but the information-theoretic version cannot be obtained for both properties at the same time.

The commitment scheme used in the present work is Damgård-Fujisaki's scheme [7], that provides statistically-hiding and computationally-binding commitments, based on Abelian groups of hidden order. Given the security parameters F, B, T, and k, the common parameters are a modulus n (that can be obtained as an RSA modulus), such that the order of $\mathbb{Z}_n^*$ can be upper bounded by 2^B, a generator h of a multiplicative subgroup of high order (the order must be F-rough) in $\mathbb{Z}_n^*$, and a value $g = h^\alpha$, such that the committer knows neither α nor the order of the subgroups. The commit function of a message $x \in [-T, T]$ with a random value $r \in [0, 2^{B+k}]$ takes the form $C_x = g^x h^r \bmod n$.

Additionally, this commitment scheme presents an additive homomorphism that allows computing the addition of two committed numbers ($C_{x+y} = C_x \cdot C_y \bmod n$) and the product of a committed number and a public integer ($C_{ax} = C_x^a \bmod n$).

2.1.2. Interactive proof systems

Interactive proof systems were introduced by Goldwasser et al. [1]; they are two party protocols in which a prover $\mathcal{P}$ tries to prove a statement x to a verifier $\mathcal{V}$, and both can make random choices. The two main properties that an interactive protocol must satisfy are *completeness* and *soundness*; the first one guarantees that a correct prover $\mathcal{P}$ can prove all correct statements to a correct verifier $\mathcal{V}$, and the second guarantees that a cheating prover $\mathcal{P}^*$ will only succeed in proving a wrong statement with negligible probability.

A special class of interactive protocols are proofs of knowledge [8], in which the proved statement is the knowledge of a witness that makes a given binary relation output a true value, such that a probabilistic algorithm called *knowledge extractor* exists, and it is able to output a witness for the common input x using any probabilistic polynomial time prover $\mathcal{P}^*$ as an oracle, in polynomial expected time (*weak soundness*).

2.1.3. Zero-knowledge protocols

In order for an interactive proof to be zero-knowledge [1], it must be such that the only knowledge disclosed to the verifier is the statement that is being proved. More formally, an interactive proof system $(\mathcal{P}, \mathcal{V})$ is statistically zero-knowledge if it exists a probabilistic polynomial algorithm (simulator) $S^{\mathcal{V}}$ such that the conversations produced by the real interaction between $\mathcal{P}$ and $\mathcal{V}$ are statistically indistinguishable from the outputs of $S^{\mathcal{V}}$.

2.2. Blind watermark detection

Given a host signal $\mathbf{x}$, a watermark $\mathbf{w}$, and a pair of keys $\{K_{\text{emb}}, K_{\text{det}}\}$ for embedding and detection (they are the same key in symmetric schemes), a digital blind watermark detection scheme consists of an *embedder* that outputs the watermarked signal $\mathbf{y} = \text{Embed}(\mathbf{x}, \mathbf{w}, K_{\text{emb}})$ and a *detector* that takes as parameters a possibly attacked signal $\mathbf{z} = \mathbf{y} + \mathbf{n}$, where $\mathbf{n}$ represents added noise, the watermark $\mathbf{w}$, and the detection key K_{det}, and it outputs a Boolean value indicating whether the signal $\mathbf{z}$ contains the watermark $\mathbf{w}$, without using the original host data $\mathbf{x}$.

Three detection algorithms will be compared in terms of their Receiver Operating Characteristic (ROC), namely, additive spread spectrum with a correlation-based detector (SS), spread-transform dither modulation without distortion compensation (ST-DM), and additive spread spectrum with a generalized Gaussian maximum likelihood (ML) detector (GG). In all of them, the host features $\mathbf{x}$ are considered

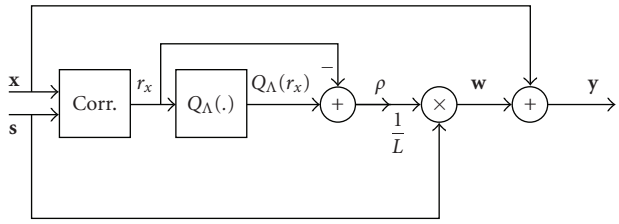

Figure 1: Block diagram of the watermark embedding process for ST-DM.

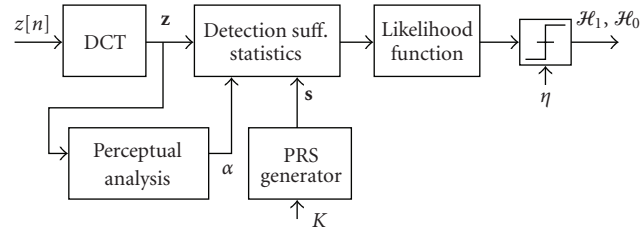

Figure 2: Block diagram of the watermark detection process for the GG detector.

i.i.d. with variance σ_X^2, the watermarked features are denoted by $\mathbf{y} = \mathbf{x}+\mathbf{w}$, and $\mathbf{z}$ represents the input to the receiver, which may be corrupted with AWGN noise $\mathbf{n}$, that is considered also i.i.d with variance σ_N^2. The binary hypothesis test that must be solved at the detector is

$$\begin{aligned} \mathcal{H}_0 &: \mathbf{z} = \mathbf{x} + \mathbf{n}, \\ \mathcal{H}_1 &: \mathbf{z} = \mathbf{x} + \mathbf{w} + \mathbf{n}. \end{aligned} \tag{1}$$

Table 1 summarizes the probabilities of false alarm (P_f) and missed detection (P_m) for the three detectors [9–11].

2.2.1. Additive spread spectrum with correlation-based detector

In SS, the watermark is generated as the product of a pseudorandom vector $\mathbf{s}$, that we will consider a binary sequence with values $\{\pm 1\}$ (with norm $\|\mathbf{s}\|^2 = L$) and a perceptual mask α (that is assumed to be constant to simplify the analysis), that controls the tradeoff between imperceptibility and distortion ($D_w = (1/L)\sum_{k=1}^{L} E\{w_k^2\} = E\{\alpha_k^2\} = \alpha^2$).

The maximum-likelihood detector for Gaussian distributed host features is a correlation-based detector:

$$r_z = \frac{1}{L}\sum_{k=1}^{L} z_k s_k \underset{\mathcal{H}_0}{\overset{\mathcal{H}_1}{\gtrless}} \eta, \tag{2}$$

where η is a threshold that depends on the probabilities of false alarm (P_f) and missed detection (P_m), as indicated in Table 1.

2.2.2. Spread transform dither modulation

Given the host features $\mathbf{x}$ and the secret spreading sequence $\mathbf{s}$, which will be considered here binary with values $\{\pm 1\}$, the embedding of the watermark in ST-DM [12] (similar to quantized projection QP [9, 10]) is done as indicated in Figure 1.

The host features $\mathbf{x}$ are correlated with the projection signal $\mathbf{s}$, and the result (r_x) is quantized with an Euclidean scalar quantizer $Q_\Lambda(\cdot)$ of step Δ, that controls the distortion, and with centroids defined by the shifted lattice $\Lambda \triangleq \Delta\mathbb{Z} + \Delta/2$. Let $\rho = (Q_\Lambda(r_x) - r_x)$; then the watermarked vector is given by

$$\mathbf{y} = \mathbf{x} + \mathbf{w} = \mathbf{x} + \frac{1}{L}\rho\mathbf{s}. \tag{3}$$

In order to detect the watermark, the host features, possibly degraded by AWGN noise $\mathbf{n}$, are correlated with the spreading sequence $\mathbf{s}$, and the resulting value $r_z = \sum_{k=1}^{L} z_k s_k$ is quantized and compared to a threshold η to determine whether the watermark is present:

$$\mid Q_\Lambda(r_z) - r_z \mid \underset{\mathcal{H}_0}{\overset{\mathcal{H}_1}{\lessgtr}} \eta. \tag{4}$$

Due to the Central Limit Theorem (CLT), the computed correlations can be accurately modeled by a Gaussian pdf.

2.2.3. Additive spread spectrum with generalized-Gaussian features

Figure 2 shows the detection scheme for this case. The host features are assumed to be the DCT coefficients of an image, what justifies the generalized Gaussian model with the following pdf:

$$\begin{aligned} f_X(x) &= Ae^{-|\beta x|^c}, \\ \beta &= \frac{1}{\sigma}\left(\frac{\Gamma(3/c)}{\Gamma(1/c)}\right)^{1/2}, \\ A &= \frac{\beta c}{2\Gamma(1/c)}. \end{aligned} \tag{5}$$

The embedding procedure is the same as the one described for SS. For detection, a preliminary perceptual analysis provides the estimation of the perceptual mask α that modulates the inserted secret sequence $\mathbf{s}$. The parameters c and β are also estimated from the received features. The likelihood function for detection is

$$l(\mathbf{y}) = \sum_k \beta^c \left(\mid Y_k \mid^c - \mid Y_k - \alpha_k s_k \mid^c \right) \underset{\mathcal{H}_0}{\overset{\mathcal{H}_1}{\gtrless}} \eta, \tag{6}$$

where η represents the threshold value used to make the decision.

Table 1: Probabilities of false alarm (P_f) and missed detection (P_m) for the three studied detectors.

	AddSS	ST-DM	GG
P_f	$Q(\sqrt{L}\eta/\sqrt{\sigma_X^2+\sigma_N^2})$	$\sum_{i=-\infty}^{\infty}[Q((\Delta(i+1/2)-\eta)/\sqrt{L(\sigma_X^2+\sigma_N^2)})-Q((\Delta(i+1/2)+\eta)/\sqrt{L(\sigma_X^2+\sigma_N^2)})]$	$Q((\eta+m_1)/\sigma_1)$
P_m	$Q(\sqrt{L}(\alpha-\eta)/\sqrt{\sigma_X^2+\sigma_N^2})$	$1-\sum_{i=-\infty}^{\infty}[Q((i\Delta-\eta)/\sqrt{L}\sigma_N)-Q((i\Delta+\eta)/\sqrt{L}\sigma_N)]$	$1-Q((\eta-m_1)/\sigma_1)$

As shown in [6], the pdfs of $l(Y)$ conditioned to hypotheses $\mathcal{H}_0$ and $\mathcal{H}_1$ are approximately Gaussian with the same variance σ_l^2, and respective means $-m_1$ and m_1, that can be estimated from the watermarked image [6].

2.2.4. Comparison

The three detectors can be compared in terms of robustness through their *Receiver Operating Characteristic* (ROC), taken from the formulas in Table 1. The correlation-based detector is only optimum when $c = 2$, and when $c \neq 2$, the generalized Gaussian detector outperforms it; ST-DM can outperform both for a sufficiently high DWR (Data to Watermark Ratio, $\mathrm{DWR} = 10\log_{10}(\sigma_X^2/\sigma_W^2)$), due to its host rejection capabilities. However, the performance of the generalized Gaussian detector and the ST-DM one are not much far apart when c is near 1 and the DWR in the projected domain ($\mathrm{DWR}_p = \mathrm{DWR} - 10\log_{10}L$) is low. Figure 3 shows a plot of the ROC for fixed DWR and WNR (Watermark to Noise Ratio, $\mathrm{WNR} = 10\log_{10}(\sigma_W^2/\sigma_N^2)$), with a features shape parameter of $c = 0.8$, that has been chosen as an example of a relatively common value for the distribution of AC DCT coefficients of most images. It is remarkable that even when the exact c is not used, and it is below 1, the performance of the GG detector with $c = 0.5$ is much better than that of the correlation-based one, and its ROC remains near the ST-DM ROC.

Regarding the resilience against sensitivity attacks, it can be shown that the correlation-based detector and the ST-DM one make the watermarking scheme very easy to break when the attacker has access to the output of the detector, as the detection boundaries for both methods are just hyperplanes; Figure 4 shows the two-dimensional detection regions for each of the three methods. On the other hand, the detection function in the GG detector when $c < 1$ (Figure 4(c)) presents the property that component-wise modifications produce bounded increments; that is, when modifying one component of the host signal Y, the increment produced in the likelihood function (6) is bounded by $|\alpha_k s_k|^c$ independently of the component $|Y_k|$ if $c < 1$:

$$\left|\,|Y_k|^c - |Y_k - \alpha_k s_k|^c\,\right| \leq |\alpha_k s_k|^c. \tag{7}$$

This means that it is not possible to get a signal in the boundary by modifying a single component (or a number N of components such that $\sum_N |\alpha_k s_k|^c$ is less than the gap to η), opposed to a correlation detector, in which just making one component big (or small) enough can get the signal out of the detection region. This property can make very difficult the task of finding a vector in the boundary given only one marked signal.

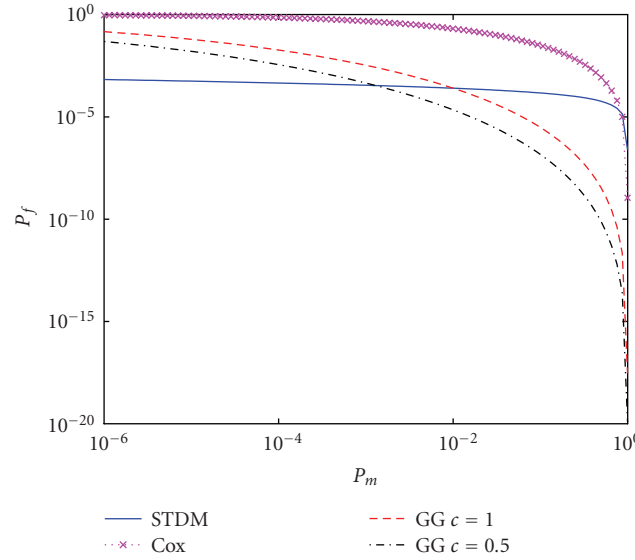

Figure 3: Theoretical ROC curves for the studied detectors under AWGN attacks, with DWR = 20 dB, WNR = 0 dB, $L = 1000$, and generalized Gaussian distributed host features with $c = 0.8$.

In order to quantitatively compare the resilience of the three detectors against sensitivity attacks, we will take as robustness criterion the number of calls to the detector needed for reaching an attack distortion equal to that of the watermark (NWR = 0 dB). This choice is supported by the fact that for an initially nonmarked host $\mathbf{x}$ in which a watermark $\mathbf{w}$ has been inserted, yielding $\mathbf{y}$, it is always possible to find a vector $\mathbf{z}$ in the boundary whose distortion with respect to $\mathbf{y}$ is less than the power of the watermark (e.g., taking the intersection between the detection boundary and the line that connects $\mathbf{x}$ and $\mathbf{y}$). Thus, a sensitivity attack can always reach a point with NWR = 0 dB. In general, it is not guaranteed that an attack can reach a lower NWR. Furthermore, given that for a blind detection the original nonmarked host is not known, imposing a more restrictive fidelity criterion for the attacker than for the embedder makes no sense. In light of the previous discussion, we can consider that a watermark has been effectively erased when a point $\mathbf{z}$ is found, whose distortion with respect to $\mathbf{y}$ is equal to the power of the embedded watermark $\mathbf{w}$; the number of iterations that a sensitivity attack needs to reach this point can thus be used for determining the robustness of the detector against the attack.

We have taken blind newton sensitivity attack (BNSA [4]; an *RRP*-compliant description of BNSA can be found in [13]) as a powerful representative of sensitivity attacks, and simulated its execution against the three studied detectors. Each iteration of this algorithm calls the detector a number

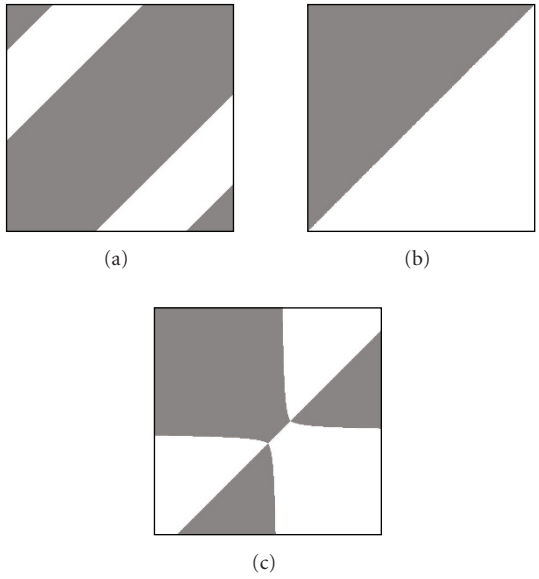

Figure 4: Two-dimensional detection boundaries for ST-DM (a), correlation-based detector (b), and GG detector (c).

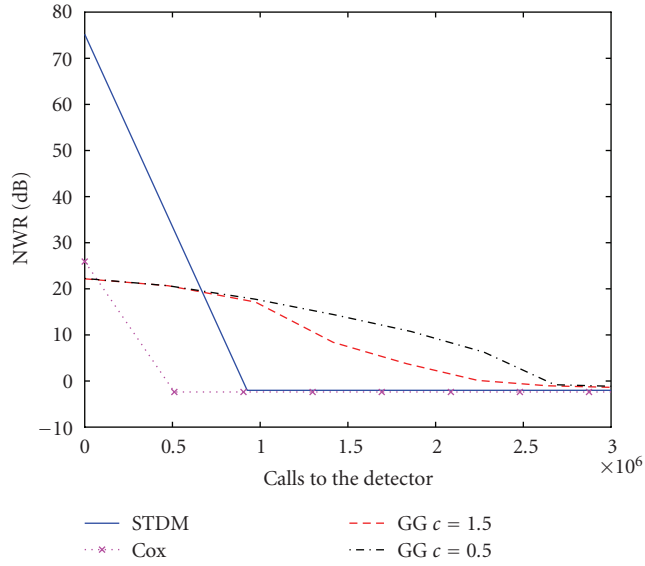

Figure 5: NWR for a sensitivity attack (BNSA) as a function of number of calls to the detector for correlation detector (Cox), ST-DM, and generalized Gaussian (GG) with $c = 0.5$, and $c = 1.5$ for DWR = 16 dB, $P_f = 10^{-4}$, and $L = 8192$.

of times proportional to the number of dimensions of the involved signals. The results show that both ST-DM and the correlation detector are completely broken in just one iteration of the algorithm, independently of the dimensionality of the signals, so the attack needs $\mathcal{O}(L)$ calls to the detector in order to succeed (achieving not only a point with NWR < 0 dB, but also convergence to the nearest point in the boundary). This is due to their simple detection boundaries, that have a constant gradient. Figure 5 shows the NWR of the attack as a function of the number of calls to the detector, for the three detectors, using DWR = 16 dB and $P_f = 10^{-4}$, as a result of averaging 100 random executions. The GG detector is used with two different shape factors, $c = 0.5$ and $c = 1.5$; the number of iterations needed to break the detector in both cases is bigger than for the correlation detectors, due to the more involved detection boundary, but this effect is more evident when $c < 1$, case in which the detector has the aforementioned property of bounded increments for component-wise modifications at the input.

The involved detection boundary of the generalized Gaussian ML detector makes the number of iterations needed for achieving convergence grow also with the dimensionality of the host. This means that the number of calls to the detector needed to get a certain target distortion is not only higher for the GG detector, but it also grows faster than for the other detectors with the dimensionality of the host (Figure 6) for fixed WNR and P_f. We have found empirically that the number of calls needed for reaching NWR = 0 dB is approximately $\mathcal{O}(L^{1.5})$. Furthermore, if we took as robustness criterion the absolute convergence of the algorithm (not only achieving NWR = 0 dB), the advantage of the GG detector is even better both in number of iterations and in number of calls to the detector; that is, while for the GG detector convergence is slowly achieved several iterations after reaching

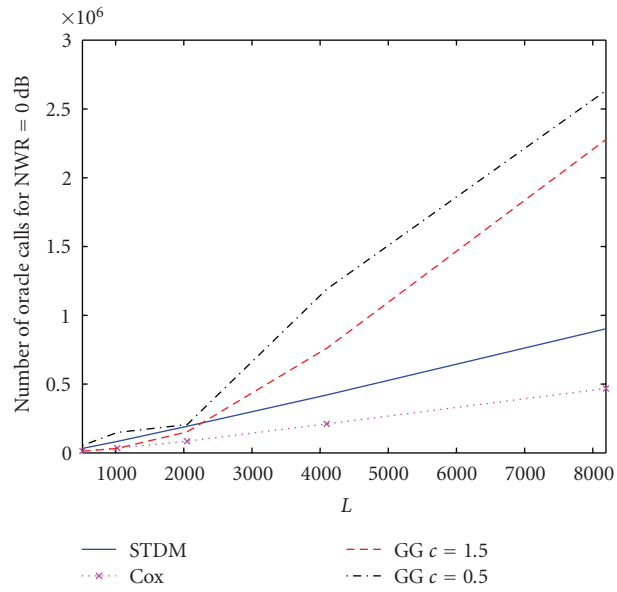

Figure 6: Number of calls to the detector for a sensitivity attack (BNSA) for reaching NWR = 0 dB as a function of the dimensionality of the watermark for correlation detector (Cox), ST-DM, and generalized Gaussian (GG) with $c = 0.5$ and $c = 1.5$ for DWR = 16 dB and $P_f = 10^{-4}$.

NWR = 0 dB, for correlation detectors BNSA achieves both NWR < 0 dB and convergence in just one iteration.

2.3. Zero-knowledge watermark detection

The use of zero-knowledge protocols in watermark detection was first issued by Craver [14], and later formalized

by Adelsbach et al. [2, 15]. The formal definition of a zero-knowledge watermark detection scheme concreted for a blind detection mechanism can be stated as follows.

Definition 1 (Zero-knowledge Watermark Detection). Given a secure commitment scheme with the operations Com() and Open(), and a blind watermarking scheme with the operations Embed() and Detect(), the watermarked host data $\mathbf{z}$ and the commitments on the watermark $C_{\mathbf{w}}$ and key C_{K_w} (for a keyed scheme), with their respective public parameters $\text{par}_{\text{com}} = (\text{par}_{\text{com}}^{\mathbf{w}}, \text{par}_{\text{com}}^{K_w})$, a zero-knowledge blind watermark detection protocol for this watermarking scheme is a zero-knowledge proof of knowledge between a prover $\mathcal{P}$ and a verifier $\mathcal{V}$ where on common input $x := (\mathbf{z}, C_{\mathbf{w}}, C_{K_w}, \text{par}_{\text{com}})$, $\mathcal{P}$ proves knowledge of a tuple $aux = (\mathbf{w}, K_w, r_{\text{com}}^{\mathbf{w}}, r_{\text{com}}^{K_w})$ such that

$$\begin{aligned}[\,&(\text{Open}(C_{\mathbf{w}}, \mathbf{w}, r_{\text{com}}^{\mathbf{w}}, \text{par}_{\text{com}}^{\mathbf{w}}) = \text{true}) \wedge \\ &(\text{Open}(C_{K_w}, K_w, r_{\text{com}}^{K_w}, \text{par}_{\text{com}}^{K_w}) = \text{true}) \wedge \\ &(\text{Detect}(\mathbf{z}, \mathbf{w}, K_w) = \text{true})\,].\end{aligned} \tag{8}$$

Adelsbach and Sadeghi introduced in [2] a zero-knowledge watermark detection protocol for the Cox et al. [16] detection scheme, that consists in a normalized correlation-detector for spread spectrum. In [17], they have studied the communication complexity of the non-blind protocol, that is much less efficient than the blind one, due to the higher number of committed operations that must be undertaken. Later, Piva et al. also developed a ZK watermark detection protocol for ST-DM in [5].

3. ZERO-KNOWLEDGE SUBPROOFS

The proofs that are employed in the previous zero-knowledge detectors and in the generalized Gaussian one are shown in Table 2 with their respective communication complexity, which has been calculated when applied to the Damgård-Fujisaki commitment scheme [7] as a function of the security parameters F, B, T and k, defined in Section 2.1.1.

The first five proofs are already existing zero-knowledge proofs for the opening of a commitment [7] (PK_{op}), the equality of two commitments [18] (PK_{eq}), the square of a commitment [18] (PK_{sq}), a commitment is inside an interval [18] (PK_{int}) and nonnegativity of a commitment [19] ($PK_{\geq 0}$).

All these proofs are just simple operations, but the lack of some operations like the computation of the absolute value or the square root, both necessary for the first implementation of the GG ML detector, led us to the development of the last two zero-knowledge proofs; PK_{sqrt} represents a proof that a committed integer is the rounded square root of another committed integer, and it is based on a mapping of quantized square roots into integers. PK_{abs} allows the application of the absolute value operator to a committed number, without disclosing the magnitude nor the sign of that number. Both proofs are described in the following.

3.1. Zero-knowledge proof that a committed integer is the rounded square root of another committed integer

Adelsbach et al. presented in [20] a proof for a generic function approximation whose inverse can be efficiently proven, covering, for example, divisions and square roots. Here, we present a specific protocol for proving a rounded square root that follows a similar philosophy, we study its communication complexity and propose a mapping (presented in Appendix A) that makes possible this zero-knowledge protocol to prove the correct calculation of square roots on committed integers (not necessarily perfect square residues):

$$PK_{\text{sqrt}}[\,y, r_1, r_2 : C_y = g^y h^{r_1} \bmod n \wedge C_{n\sqrt{y}} = g^{n\sqrt{y}} h^{r_2} \bmod n\,]. \tag{9}$$

Let C_y be the commitment to the integer whose square root must be calculated. The protocol that prover and verifier would follow is the next.

(1) First, the prover calculates the value $x = \text{round}(\sqrt{y})$, its commitment C_x, and the commitment to its squared value C_{x^2}, and sends both commitments and C_y to the verifier.

(2) The prover proves in zero-knowledge that C_{x^2} contains the squared value of the integer hidden in C_x, through $PK\{x, r_1, r_2 : C_x = g^x h^{r_1} \bmod n,\ C_{x^2} = g^{x^2} h^{r_2} \bmod n\}$.

(3) Then, the prover must prove that $x^2 \in [y - x, y + x]$, using a modified version of Boudot's proof [18] with hidden interval, that consists in considering also randomness in the commitments of the interval limits calculated by both parties at the first step of the proof. Using this interval instead of the one indicated in Appendix A, the zero values are also accepted with no ambiguity when the maximum allowable value for y is below the order of the group generated by g. The counterpart is that there are two possibilities for the square root of integers of the form $k^2 + k$, with k an integer, namely k and $k + 1$. The effect of this relaxation on the conditions imposed before is a small rise in the rounding error, smaller as k grows; if we take into account that the numbers that are considered integers are actually the quantization of real numbers using a step that is fixed by the precision of the system, the error is of the same order as this precision. Nevertheless, the need of working with null values without disclosing any information forces us to make this adaptation.

(4) At last, it is necessary to prove that $x \in [0, \sqrt{m}]$, if m is the order of the subgroup generated by g. If it is known—by the initialization of the commitment scheme—that $\log_2(m) = l$, then proving that $x \in [0, 2^{l/2-1}]$ is enough; if the working range for the committed integers is $[-T, T]$, with $T < \sqrt{m}$ (as it will be if the bit length of T is at most $l/2 - 1$), then it suffices with the proof that x is in the working range: $x \in [0, T]$.

Table 2: Zero-knowledge subproofs and their communication complexity.

Proof	Comp_{PK}(bits)
$PK_{\text{op}}[m, r : C_m = g^m h^r \bmod n]$	$3\lvert F\rvert + \lvert T\rvert + 2B + 3k + 2$
$PK_{\text{eq}}[m, r_1, r_2 : C_m^{(1)} = g_1^m h_1^{r_1} \bmod n \wedge C_m^{(2)} = g_2^m h_2^{r_2} \bmod n]$	$4\lvert F\rvert + \lvert T\rvert + 2B + 5k + 3$
$PK_{\text{sq}}[m, r_1, r_2 : C_m = g_1^m h_1^{r_1} \bmod n \wedge g_2^{m^2} h_2^{r_2} \bmod n]$	$4\lvert F\rvert + \lvert T\rvert + 3B + 5k + 3$
$PK_{\text{int}}[m, r : C_m = g^m h^r \bmod n \wedge m \in [a, b]]$	$25\lvert F\rvert + 5\lvert T\rvert + 10B + 27k + 2\lvert n\rvert + 20$
$PK_{\geq 0}[m, r : C_m = g^m h^r \bmod n \wedge m \geq 0]$	$11\lvert F\rvert + 4\lvert T\rvert + 12B + 14k + 9$
$PK_{\text{sqrt}}[m, r_1, r_2 : C_m = g^m h^{r_1} \bmod n \wedge C_{n\sqrt{m}} = g^{n\sqrt{m}} h^{r_2} \bmod n]$	$48\lvert F\rvert + 9\lvert T\rvert + 18B + 53k + 6\lvert n\rvert + 39$
$PK_{\text{abs}}[m, r_1, r_2 : C_m = g^m h^{r_1} \bmod n \wedge C_{\lvert m\rvert} = g^{\lvert m\rvert} h^{r_2} \bmod n]$	$19\lvert F\rvert + 6\lvert T\rvert + 16B + 24k + 15$

Claim 1. The presented interactive proof is computationally sound and statistically zero-knowledge in the random oracle model.

A sketch of the proof for this claim is given in Appendix C.

The communication complexity of this protocol is shown in Table 2.

3.2. Zero-knowledge proof that a committed integer is the absolute value of another committed integer

This proof is a zero-knowledge protocol that allows the application of the absolute value operator to a committed number, without disclosing the magnitude nor the sign of that number

$$PK_{\text{abs}}[x, r_1, r_2 : C_x = g_1^x h_1^{r_1} \bmod n \wedge C_{\lvert x\rvert} = g_2^{\lvert x\rvert} h_2^{r_2} \bmod n]. \tag{10}$$

As in a residue group $\mathbb{Z}_q$ there is no notion of "sign," we are using the commonly known mapping:

$$\text{sign}(x) = \begin{cases} 1, & x \in \left\{0, \left\lfloor \frac{q}{2} \right\rfloor\right\}, \\ -1, & x \in \left\{\left\lfloor \frac{q}{2} \right\rfloor + 1, n - 1\right\}; \end{cases}$$

taking into account that $-x \equiv q - x \bmod q$, the mapping is consistent.

Let $C_x = g_1^x h_1^{r_1} \bmod n$ be the commitment to a number x, whose sign is not known by the verifier, and $C_{\lvert x\rvert} = g_2^{\lvert x\rvert} h_2^{r_2} \bmod n$ the commitment to a number which is claimed to be the absolute value of x. The scheme of the protocol is as follows:

(1) both prover and verifier calculate the commitment to the opposite of x, with the help of the homomorphic properties of the commitment scheme:

$$C_{-x} = C_x^{-1}; \tag{11}$$

(2) next, the prover must demonstrate that the value hidden in $C_{\lvert x\rvert}$ corresponds to the value hidden in one of the previous commitments C_x, C_{-x}, using the ZK proof of knowledge described in Appendix B;

(3) at last, the prover demonstrates that the value hidden in $C_{\lvert x\rvert}$ is $\lvert x\rvert \geq 0$, using the protocol proposed by Lipmaa [19].

Claim 2. The presented interactive proof is computationally sound and statistically zero-knowledge in the random oracle model.

A sketch of the proof for this claim can be found in Appendix C.

The communication complexity of this protocol is given in Table 2.

4. ZERO-KNOWLEDGE GG WATERMARK DETECTOR

The zero-knowledge version of the generalized Gaussian detector conceals the secret pseudorandom signal s_k using the Damgård-Fujisaki scheme [7] C_{s_k}. The supposedly watermarked image Y_k is publicly available, so the perceptual analysis (α_k) and the extraction of the parameters β_k and c_k can be done in the public domain, as well as the estimation of the threshold η for a given point in the ROC. In this first implementation, only shape factors $c = 1$ or $c = 0.5$ are allowed, so the employed c_k will be the nearest to the estimated shape factor. The target is to perform the calculation of the likelihood function:

$$D = \sum_k \beta_k^{c_k} \left(\lvert Y_k \rvert^{c_k} - \underbrace{\lvert \overbrace{Y_k - \alpha_k s_k}^{A_k} \rvert^{c_k}}_{B_k} \right), \tag{12}$$

and the comparison with the threshold η, without disclosing s_k.

The protocol executed by prover and verifier so as to prove that the given image Y_k is watermarked with the sequence hidden in C_{s_k} is the following:

(1) prover and verifier calculate the commitment to $A_k = Y_k - \alpha_k s_k$ applying the homomorphic property of the Damgård-Fujisaki scheme:

$$C_{A_k} = \frac{g^{Y_k}}{C_{s_k}^{\alpha_k}}; \tag{13}$$

(2) next, the prover generates a commitment $C_{\lvert A_k\rvert}$ to the absolute value of A_k, sends it to the verifier, and proves in zero-knowledge that it hides the absolute value of the commitment C_{A_k}, through the developed proof PK_{abs} (Section 3.2);

(3) if $c = 1$ (Laplacian features) then the operation $\lvert A_k\rvert^c$ is not needed, so, just for the sake of notation $C_{B_k} = C_{\lvert A_k\rvert}$. If $c = 0.5$, the rounded square root of

$|A_k|$ must be calculated by the prover; then he generates the commitment $C_{B_k} = C_{\sqrt{|A_k|}}$, sends it to the verifier and proves in zero-knowledge the validity of the square root calculation, through the proof PK_{sqrt} (Section 3.1);

(4) both prover and verifier can independently calculate the value $\beta_k^{c_k}$ and $|Y_k|^{c_k}$, and complete the committed calculation of the sum $D = \sum_k \beta_k^{c_k}(|Y_k|^{c_k} - B_k)$, thanks to the homomorphic property of the used commitment scheme

$$C_D = \prod_k \left(\frac{g^{|Y_k|^{c_k}}}{C_{B_k}}\right)^{\beta_k^{c_k}}; \tag{14}$$

(5) finally, the prover must demonstrate in zero-knowledge that $D > \eta$, or equivalently, that $D - \eta > 0$, which can be done by running the proof of knowledge by Lipmaa [19] on $C_{\text{th}} = C_D g^{-\eta}$.

5. IMPROVED GG DETECTOR WITH BINARY ANTIPODAL SPREADING SEQUENCE (GGBA)

When the spreading sequence s_k is a binary antipodal sequence, so it takes only values $\{\pm s\}$, we can apply a trivial transformation to the detection function of the GG detector (6):

$$\begin{aligned} D &= \sum_k \beta_k^{c_k}\left(|Y_k|^{c_k} - |Y_k - \alpha_k s_k|^{c_k}\right) \\ &= \sum_k \beta_k^{c_k}\left(|Y_k|^{c_k} - \left(|Y_k - \alpha_k s|^{c_k}\cdot \mathbf{1}_{\{s\}}(s_k)\right.\right. \\ &\qquad \left.\left. + |Y_k + \alpha_k s|^{c_k}\cdot \mathbf{1}_{\{-s\}}(s_k)\right)\right) \\ &= \sum_k \beta_k^{c_k}\left(|Y_k|^{c_k} - \left(|Y_k - \alpha_k s|^{c_k}\cdot \frac{1}{2s}(s + s_k)\right.\right. \\ &\qquad \left.\left. + |Y_k + \alpha_k s|^{c_k}\cdot \frac{1}{2s}(s - s_k)\right)\right) \end{aligned} \tag{15}$$

$$\begin{aligned} &= \underbrace{\sum_k \beta_k^{c_k}\left(|Y_k|^{c_k} - \frac{1}{2}\left(|Y_k - s\alpha_k|^{c_k} + |Y_k + s\alpha_k|^{c_k}\right)\right)}_{G} \\ &\quad - \sum_k \underbrace{\frac{\beta_k^{c_k}}{2s}\left(|Y_k - s\alpha_k|^{c_k} - |Y_k + s\alpha_k|^{c_k}\right)}_{H_k} s_k. \end{aligned} \tag{16}$$

In (15), we use the fact that s_k can only be given a value s or $-s$ in order to substitute the indicator function $\mathbf{1}_{\{s\}}(s_k) = (1/2s)(s + s_k)$ and $\mathbf{1}_{\{-s\}}(s_k) = (1/2s)(s - s_k)$.

The factors termed as G and H_k in (16) can be computed in the clear-text domain, working with floating-point precision arithmetic, and then have their commitments generated. This implies that all the nonlinear operations are transferred to the clear-text domain, greatly reducing the communication overhead, as will be shown in Section 7; only additions and multiplications must be performed in the encrypted domain, and they can be undertaken through the homomorphic properties of the commitment scheme. This transference also diminishes the computational load, as clear-text operations are much more efficient than modular operations in a large ring.

The zero-knowledge protocol can be reduced to the following two steps.

(1) prover and verifier homomorphically compute $\text{th} = D - \eta$

$$C_{\text{th}} = \frac{g^{G-\eta}}{\prod_k C_{s_k}^{H_k}}. \tag{17}$$

(2) The prover demonstrates the presence of the watermark by running the zero-knowledge proof that $D - \eta > 0$.

The number of needed proofs during the protocol is reduced to only one, what propitiates the aforementioned reduction in computation and communication complexity, with the additional advantage that this scheme can be applied to any value of the shape parameter c_k, so it will be preferred to the previous one unless s_k is not binary antipodal.

6. SECURITY ANALYSIS FOR THE GG DETECTION PROTOCOLS

After presenting the protocols for the zero-knowledge implementation of the generalized Gaussian ML detector, we can state the following theorem.

Theorem 1. *The developed detection protocols for the generalized Gaussian detector are computationally sound and statistically zero-knowledge.*

A sketch of the proof for this theorem can be found in Appendix C.

The reformulation of the generalized Gaussian protocol deserves two comments concerning security. The first one involves the nonlinear operations that were performed under encryption in Section 4, which are now transferred to the public clear-text domain. Although this could seem at first sight a knowledge leakage, currently it is not; all those operations can be performed with the same public parameters as in Section 4 in a feasible time, so the parameters G and H_k that are publicly calculated in this protocol could also be obtained in the previous version, and their disclosure gives no *extra* knowledge.

The second comment deals with the correlation form of the reformulation, and its resilience to blind sensitivity attacks. Even when the operation performed in the encrypted domain is a correlation, the additive term (G) is what preserves the bounded-increment property, by virtue of which component-wise modifications of the input signal only produce bounded increments on the likelihood function:

$$-\alpha^c \le |Y_k|^c - |Y_k - \alpha s_k|^c \le \alpha^c, \quad c < 1. \tag{18}$$

The result of the addition is not disclosed during the protocol; thus, the correlation cannot be known even when the term G is public, and both terms cannot be decoupled, so

no extra knowledge is learned from G, and the difficulty for finding points in the detection boundary, that is a necessary step for sensitivity attacks, remains, as well as the shape of the detection regions, unaltered.

7. EFFICIENCY AND PRACTICAL IMPLEMENTATION

We will measure the efficiency of the developed protocols in terms of their communication complexity, as this parameter is what entails the bottleneck of the system, and it is easily quantifiable given the complexity measures calculated in the previous sections for each of the subprotocols.

Taking into account the plot of the raw protocol (Section 4), a total of $2L$ commitments (with a length $|n|$) are interchanged, namely the L commitments that correspond to the secret pseudorandom sequence $\mathbf{s}$ and the L commitments to $|A_k|$, while in the GGBA detector (Section 5) only the L commitments to $\mathbf{s}$ are sent; the rest of the commitments are either calculated using homomorphic computation or are already included in the complexity of the subprotocols.

Thus, the total communication complexity for the detector applied to Laplacian distributed features and $c = 0.5$ in the first scheme, as well as the complexity for the improved *GGBA* detector can be expressed as

$$\begin{aligned}
&\mathrm{Comp}_{\mathrm{ZKWD}_{\mathrm{GG}(c=1)}} \\
&= 2L|n| + L\cdot\left(\mathrm{Comp}_{PK_{\mathrm{abs}}} + \mathrm{Comp}_{PK_{\mathrm{op}}}\right) + \mathrm{Comp}_{PK_{\geq 0}}, \\
&\mathrm{Comp}_{\mathrm{ZKWD}_{\mathrm{GG}(c=0.5)}} \\
&= 2L|n| + L\cdot\left(\mathrm{Comp}_{PK_{\mathrm{abs}}} + \mathrm{Comp}_{PK_{\mathrm{op}}} + \mathrm{Comp}_{PK_{\mathrm{sqrt}}}\right) + \mathrm{Comp}_{PK_{\geq 0}}, \\
&\mathrm{Comp}_{\mathrm{ZKWD}_{\mathrm{GGBA}}} \\
&= (L+1)|n| + L\cdot\mathrm{Comp}_{PK_{\mathrm{op}}} + \mathrm{Comp}_{PK_{\geq 0}}.
\end{aligned} \tag{19}$$

In every calculation, L proofs of knowledge of the opening of the initial commitments have been added, as even when they are not explicitly mentioned in the sketch of the protocols, they are needed to protect the verifier.

In order to reduce the total time spent during the interaction, it is possible to convert the whole protocol in a non-interactive one, following the procedure described in [21], keeping the condition that the parameters for the commitment scheme must not be chosen by the prover, or he would be able to fake all the proofs. In addition to the reduction in interaction time, the use of this technique also overcomes the necessity of a honest verifier that some subprotocols impose.

The calculated complexity for Piva et al.'s ST-DM detector and Adelsbach and Sadeghi's blind correlation-based detector is the following:

$$\begin{aligned}
&\mathrm{Comp}_{\mathrm{ZKWD}_{\mathrm{STDM}}} \\
&\quad = (L+1)|n| + L\cdot\mathrm{Comp}_{PK_{\mathrm{op}}} + \mathrm{Comp}_{PK_{\mathrm{int}}}, \\
&\mathrm{Comp}_{\mathrm{ZKWD}_{\mathrm{SS}}} \\
&\quad = (L+1)|n| + L\cdot\mathrm{Comp}_{PK_{\mathrm{op}}} + 2\mathrm{Comp}_{PK_{\geq 0}} + \mathrm{Comp}_{PK_{\mathrm{sq}}}.
\end{aligned} \tag{20}$$

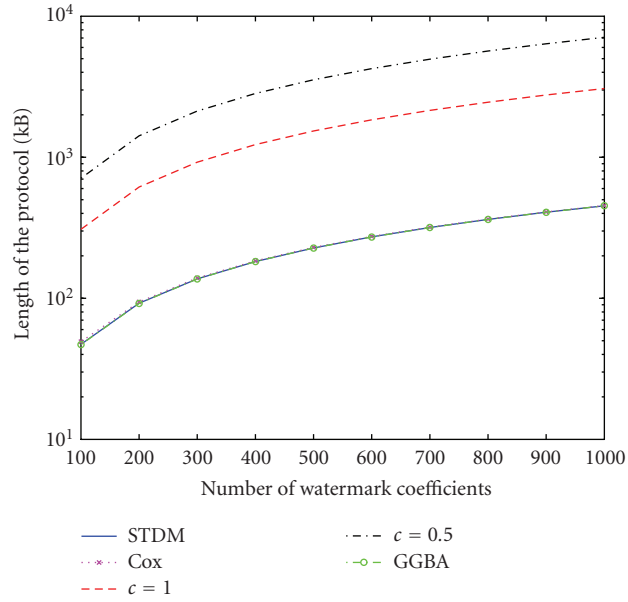

Figure 7: Communication complexity in kB for the studied protocols.

As a numeric example, in Figure 7 the evolution of the communication complexity for every protocol is compared using $|F| = 80$, $|n| = 1024$, $B = 1024$, $T = 2^{256}$ and $k = 40$, for growing L. All the protocols have complexity $\mathcal{O}(L)$. The two protocols for generalized Gaussian host features with $c = 1$ and $c = 0.5$ have a higher complexity, due to the operations that cannot be computed by making use of the homomorphic property of the commitment scheme (absolute value and square root). Nevertheless, their complexity is comparable to that of the zero-knowledge non-blind detection protocol developed by Adelsbach et al. [17].

On the other hand, the zero-knowledge *GGBA* detector achieves the lowest communication complexity of all the studied protocols, even lower than the previous correlation-based protocols, with the increased protection against blind sensitivity attacks when $c < 1$ is used, being this the first benefit of the reformulated algorithm.

Furthermore, the communication complexity of the protocol is constant if we discard the initial transmission of the commitments for the spreading sequence and their corresponding proofs of opening; once this step is performed, the protocol can be applied to several watermarked works for proving the presence of the same watermark with a (small) constant communication complexity.

Regarding computation complexity, the original detection algorithm (without the addition of the zero-knowledge protocol) for the generalized Gaussian is more expensive than ST-DM or Cox's (normalized) linear correlator, due to its nonlinear operations. The use of zero-knowledge produces an increase in computation complexity, as, additionally to the calculation and verification of the proofs, homomorphic computation involves modular products and exponentiations in a large ring, so clear-text operations have almost negligible complexity in comparison with encrypted operations.

The second benefit of the presented GGBA zero-knowledge protocol is that all the nonlinear operations are transferred from the encrypted domain (where they must be performed using proofs of knowledge) to the clear-text public domain; thus, all the operations that made the symmetric protocol more expensive than the correlation-based detectors can be neglected in comparison with the encrypted operations, so the computation complexity of the zero-knowledge GGBA protocol will be roughly the same as the one for the correlation-based zero-knowledge detectors.

8. CONCLUSIONS

The presented zero-knowledge watermark detection protocol based on generalized Gaussian ML detector outperforms the previous correlation-based zero-knowledge detectors implemented to date in terms of robustness against blind sensitivity attacks, while improving on the ROC of the correlation-based spread-spectrum detector with a performance that is near that of ST-DM.

If the employed spreading sequence is a binary antipodal sequence, the protocol can be restated in a much more efficient way, reaching a communication complexity that is even lower than that of the previous correlation-based protocols, while keeping its robustness against sensitivity attacks.

Two zero-knowledge proofs for square root calculation and absolute value have been presented. They serve as building blocks for the zero-knowledge implementation of the generalized Gaussian ML detector, and also allow for the encrypted execution of these two nonlinear operations in other high level protocols.

Finally, the use of the technique shown in [21] makes the whole protocol noninteractive, so that it does not need a honest verifier to achieve the zero-knowledge property. In order to get protection against cheating provers, the proofs shown in [22] can be employed to prove some statistical properties of the inserted watermark, resulting in an increase in communication complexity.

APPENDICES

A. MAPPING FOR ROUNDED SQUARE ROOT

Current cryptosystems are based in modular operations in a group of high order. Although simple operations like addition or multiplication have a direct mapping from quantized real numbers to modular arithmetic (provided that the number of elements inside the used group is big enough to avoid the effect of the modulus), when trying to cope with non-integer operations, like divisions or square roots, problems arise.

In the following, a mapping that represents quantized square roots inside integers in the range $\{1,\ldots,n-1\}$ is presented, and existence and uniqueness of the solutions for this mapping are derived. The target is to find which conditions must be satisfied by the input and the output to keep this operation secure when the arguments are concealed.

The mapping must be such that if $y \in \mathbb{Z}^+$ and $x = \sqrt{y} \in \mathbb{R}$, then ${}_n\sqrt{y} := \text{round}(x)$. For this mapping to behave like the conventional square root for positive reals, it is necessary to bound the domain where it can be applied. The formalization of the mapping would be as follows:

$$\begin{aligned} {}_n\sqrt{\cdot} : A &= \{y \in \mathbb{Z}^+ \mid y < n\} \longrightarrow B = \{x \in \mathbb{Z}^+ | x < \text{round}(\sqrt{n})\} \\ &\qquad y \longrightarrow x = {}_n\sqrt{y} = \text{round}(\sqrt{y}). \end{aligned} \tag{A.1}$$

In order for this definition to be valid, and given that the elements with which this mapping works are just the representatives of the residue classes of $\mathbb{Z}_n$ in the interval $\{1,\ldots,n-1\}$, we can state the following lemma.

Lemma 1 (Existence and uniqueness of a solution). *A unique $x \in [1, x_m] \cap \mathbb{Z}^+$ exists, such that for all $y \in \{1,\ldots,\min(x_m^2 + x_m, n-1)\}$, $x_m \leq \lceil \sqrt{n} \rceil - 1$,*

$$x^2 \bmod n \in [y - x, y + x)_n, \quad x \leq y, \tag{A.2}$$

where $[,)_n$ represents the modular reduction of the given interval.

Proof.

Existence. Given $y \in \mathbb{Z}^+$, its real square root admits a unique decomposition as an integer and a decimal in this way:

$$\sqrt{y} = x + d, \quad x = \text{round}(\sqrt{y}) \in \mathbb{Z}^+,\ d \in [-0.5, 0.5). \tag{A.3}$$

Squaring the previous expression, both sides of the equality must be integers, so,

$$\begin{aligned} (\sqrt{y})^2 &= x^2 + d^2 + 2dx \\ x^2 &= y - 2dx - d^2, \end{aligned} \tag{A.4}$$

and taking into account that y is integer, $2dx + d^2$ must be also an integer, and it is bounded by

$$2dx + d^2 \in [-x + 0.25, x + 0.25) \Longrightarrow 2dx + d^2 \in [-x + 1, x]. \tag{A.5}$$

Substituting this last equation in the previous one gives the desired result:

$$x^2 \in [y - x, y + x - 1]. \tag{A.6}$$

Thus, the modular reduction of x^2 is inside the modular reduction of the interval, and x exists.

Uniqueness. Here uniqueness is concerned with modular operations, and the possibility that the interval $[y-x, y+x)$ include integers out of the initial representing range $\{0,\ldots,n-1\}$, which would result in ambiguities after applying the mod operator. In the following, all the operations are modular, and thus, the mod operator is omitted. The intervals also represent their modular reduction.

The proof is based on reductio ad absurdum. Let $y \in \{1,\ldots,x_m^2 + x_m\}$, and let $x, x' \in [1, x_m] \cap \mathbb{Z}^+$ two different

integers such that both fulfill $x = \sqrt[n]{y}$, $x' = \sqrt[n]{y}$. This means that

$$\begin{aligned} x^2 &\in [y - x, y + x) \cap \mathbb{Z}, \\ x'^2 &\in [y - x', y + x') \cap \mathbb{Z}. \end{aligned} \tag{A.7}$$

Combining the previous relations, x and x' must be such that

$$x^2 - x'^2 \in (-x - x', x + x') \cap \mathbb{Z}. \tag{A.8}$$

Let us suppose, without loss of generality, that $x > x'$. If both x, x' are less than $x_m \leq \lceil \sqrt{n} \rceil - 1$, then their squares are below n, and follow the same behavior as if no modular operation were applied. Squares in $\mathbb{Z}$ can be represented by the following recursive formula:

$$\begin{aligned} y_k = k^2 = y_{k-1} + k + k - 1 \Longrightarrow \\ y_k - y_i = k^2 - i^2 = \begin{cases} \sum_{l=1}^{k-i-1} 2(k-l) + k + i, & k > i \\ 0, & k = i, \end{cases} \end{aligned} \tag{A.9}$$

what means that in order for x^2 and x'^2 to be spaced less that $x + x'$ the next inequality must be satisfied:

$$\sum_{l=1}^{x'-x-1} 2(x-l) + x + x' < x + x' \Longrightarrow \sum_{l=1}^{x-x'-1} 2(x-l) < 0. \tag{A.10}$$

Thus, the only solution is $x = x'$.

If, on the other hand, $x = x_m$, and taking into account that

$$x^2 \in [y - x, y + x - 1] \Longleftrightarrow y \in [x^2 - x + 1, x^2 + x], \tag{A.11}$$

there are two possibilities.

(1) $y \in \{x^2 - x + 1, \ldots, n - 1\}$: if $x \neq x'$, then $x' < \text{round}(\sqrt{n})$, so the range $(x'^2 - x', x'^2 + x']$ cannot include y, and x is the only admissible solution.

(2) $y \in \{1, \ldots, x^2 + x - n\}$: this is only possible if $x_m^2 + x_m > n$; in such case, given the condition imposed on x_m, then

$$y \leq x_m^2 + x_m - n \leq \sqrt{n}^2 - 1 + x_m - n = x_m - 1. \tag{A.12}$$

As $x = x_m$, this means that $y < x$, which violates one of the conditions established at the beginning. □

One issue in the previous exposition is that it is possible that the mapping is not defined over the entire set $\{1, \ldots, n - 1\}$. Instead, if the modulus is not public, the full working range is not known, and it becomes necessary to upper bound the integers with which the system will work. In this case, the upper bound can be set to $y_m = x_m^2 + x_m$, and the mapping can be applied to the full working range; furthermore, the condition that $x \leq y$ can be eliminated, as $x \in \{1, \ldots, x_m\}$ already guarantees that there is no ambiguity.

A similar reasoning can be applied when the working range includes negative numbers:

$$\left\{ -\left\lfloor \frac{n}{2} \right\rfloor, \ldots, 0, \ldots, \left\lceil \frac{n}{2} \right\rceil - 1 \right\}. \tag{A.13}$$

In this case, it is enough if $x \in \{1, \ldots, \text{round}(\sqrt{n/2})\}$, and $y \in \{1, \ldots, \lceil n/2 \rceil - 1\}$, as x^2 covers all the range of positive numbers in which y is included, and there are no ambiguities with the mod operation, as the overlap in intervals can only be produced with negative numbers, already discarded by the previous conditions.

Limiting the working range is the biggest issue of this method; with sequential modular additions and multiplications in $\mathbb{Z}_n$, it is only needed that the result of applying the same sequence of operations (without applying the modulus) in $\mathbb{Z}$ belongs to the interval $\{1, \ldots, n - 1\}$ to reach the same value with modular operations. In the case of the defined square root, it is necessary that the operations made before applying a root also return a number inside the interval $\{1, \ldots, n - 1\}$, and it is not enough that the final result of all the computation is in this interval.

B. ZERO-KNOWLEDGE PROOF THAT A COMMITMENT HIDES THE SAME VALUE AS ONE OF TWO GIVEN COMMITMENTS

This proof constitutes a mixture of a variation of the proof of equality of two commitments [18] and the technique shown in [23] to produce an OR proof through the application of secret sharing schemes.

Given three commitments $C_{x_1} = g_1^{x_1} h_1^{r_1}$, $C_{x_2} = g_2^{x_2} h_2^{r_2}$ and $C_x = g^x h^r$, the prover states that $x = x_1$ or that $x = x_2$. The notation used for the security parameters $(B, T, k, F = C(k))$ is the same as in Section 2.1.1; the structure of the proof is the following.

(1) Let us suppose that $x_i = x$, and $x_j \neq x$, with $i, j \in \{1, 2\}$, $i \neq j$. Then, for x_j, the prover must generate the values

$$\begin{aligned} W_{j1} &= g_j^{u_j} h_j^{u_{j1}} C_{x_j}^{-e_j}, \\ W_{j2} &= g^{u_j} h^{u_{j2}} C_x^{-e_j}, \end{aligned} \tag{B.1}$$

such that e_j is a randomly chosen t-bit integer ($e_j \in [0, C(k))$), u_j is randomly chosen in $[0, C(k)T2^k)$ and u_{j1} and u_{j2} are randomly chosen in $[0, C(k)2^{B+2k})$.

For x_i, the prover chooses at random $y_i \in [1, C(k)T2^k)$ and $r_{i3}, r_{i4} \in [0, C(k)2^{B+2k})$, and constructs

$$\begin{aligned} W_{i1} &= g_i^{y_i} h_i^{r_{i3}}, \\ W_{i2} &= g^{y_i} h^{r_{i4}}. \end{aligned} \tag{B.2}$$

Then, the prover sends to the verifier the values W_{11}, W_{12}, W_{21}, W_{22}.

(2) The verifier generates a random t-bit number $s \in [0, C(k))$, and sends it to the prover.

(3) The prover calculates the remaining challenge applying an XOR $e_i = e_j \oplus s$, and then generates the following values:

$$\begin{aligned} u_i &= y_i + e_i x, \\ u_{i1} &= r_{i3} + e_i r_i, \\ u_{i2} &= r_{i4} + e_i r, \end{aligned} \tag{B.3}$$

and sends to the verifier $e_1, u_1, u_{11}, u_{12}, e_2, u_2, u_{21}, u_{22}$.

(4) The verifier checks that the challenges e_1, e_2 are consistent with his random key s ($s = e_1 \oplus e_2$), and then checks, for $k = \{1, 2\}$, the proofs

$$\begin{aligned} g_1^{u_k} h_1^{u_{k1}} C_{x_k}^{-e_k} &= W_{k1}, \\ g^{u_k} h^{u_{k2}} C_x^{-e_k} &= W_{k2}. \end{aligned} \tag{B.4}$$

The completeness of the proof follows from its definition, as if one of the x_k is equal to x, then all the subproofs will succeed.

The soundness of the protocol resides in the key s, that is generated by the verifier. This protocol can be decomposed in two parts, each one consisting in the proof that $x = x_i$ for each x_i. Both are based in a protocol that is demonstrated to be sound [18]. So, without access to e_i at the first stage, the only way for the prover to generate the correct values with nonnegligible probability is that $x_i = x$; if $x_i \neq x$, he must generate e_i in advance for making that the proof succeeds. With this premise, one of the e_i must be fixed by the prover, and he indirectly commits to it in the first stage of the protocol; but the other value e_j is determined by e_i and by the random choice of the verifier s, so for the prover it is as random as s, guaranteeing that the second proof will only succeed with negligible probability when $x_j = x$.

The protocol is witness hiding, due to the followed procedure for developing it [23]; thanks to the statistically hiding property of the commitments, all the values generated for the false proof will be indistinguishable from those of the true proof. Furthermore, the protocol is also zero-knowledge, as a simulator can be built that given the random choices (s) of the verifier can construct both proofs applying the same trick as for the false proof, and the distribution of the resulting commitments will be statistically indistinguishable from that of the real interactions; in fact, the original protocol was honest-verifier zero-knowledge, but adding the additional XOR on the verifier's random choice for the true proof makes that the resulting value is completely random, at least if one of the parties is honest (it is like a fair coin flip), so the zero-knowledge property is gained in this process.

Applying the technique shown in [21], the previous protocol can be transformed in a noninteractive zero-knowledge proof of knowledge, by using a hash function H, so that $s = H(W_{11} \| W_{12} \| W_{21} \| W_{22})$, and eliminating the transmission of $W_{11}, W_{12}, W_{21}, W_{22}$. This way, the verifier checks that

$$\begin{aligned} &e_1 \oplus e_2 \\ &= s = H\left(g_1^{u_1} h_1^{u_{11}} C_{x_1}^{-e_1} \| g^{u_1} h^{u_{12}} C_x^{-e_1} \| g_2^{u_2} h_2^{u_{21}} C_{x_2}^{-e_2} \| g^{u_2} h^{u_{22}} C_x^{-e_2}\right). \end{aligned} \tag{B.5}$$

C. SECURITY PROOFS

In this appendix, we have included the sketches of the security proofs for the developed protocols.

C.1. Sketch of the proof for Claim 1

Completeness and soundness of the protocol in Section 3.1 are held upon the validity of the mapping of Appendix A.

Proof.

Completeness. If both prover and verifier behave according to the protocol in Section 3.1, then the verifier will accept all the subproofs and all its tests will succeed. If x is generated as the rounded square root of y, the square proof and both range proofs will be accepted because of the validity of the mapping of Appendix A and the completeness of these subproofs.

Soundness. Taking into account the consideration about integers of the form $k^2 + k$, the binding property of the commitment guarantees that the prover cannot open the generated C_x and C_{x^2} to incorrect values; thus, appealing to the uniqueness property of the mapping of Appendix A, the computational soundness of the range and squaring subproofs guarantees that a proof for a value that does not fulfill that mapping will only succeed with negligible probability.

Zero-knowledge. We can construct a simulator S^{V^*} for the verifier's view of the interaction. S^{V^*} must generate values C_x and C_{x^2} as commitments to random values, that will be statistically indistinguishable from the true commitments, due to the statistically hiding property of the commitment scheme. Furthermore, the statistical zero-knowledge property of the squaring and range subproofs guarantees that simulators for these proofs exist and generate the correct views, and the generation of C_x and C_{x^2} does not affect these views, due to their indistinguishability with respect to the true commitments, and that the simulators do not need knowledge of the committed values in order to succeed. □

C.2. Sketch of the proof for Claim 2

Proof.

Completeness. If both parties adhere to the protocol, then when $C_{|x|}$ hides the absolute value of the number concealed in C_x, the protocol always succeeds due to the completeness of the OR proof and the nonnegativity proof.

Soundness. Due to the binding property of the commitments, the prover cannot open C_x and $C_{|x|}$ to incorrect values. Furthermore, due to the soundness of the subproofs, if $C_{|x|}$ hides a negative number, the proof in step (3) will fail, so the complete protocol will fail (except with negligible probability); on the other hand, if $C_{|x|}$ does not hide a number with the same absolute value as the one hidden by C_x, the proof in step (2) will also fail (except with negligible probability). Thus, the whole protocol will only succeed for a non-valid input with a negligible probability given by the soundness error of the proofs in steps (2) and (3).

Zero-knowledge. We can construct a simulator S^{V^*} such that the real interactions have a probability distribution indistinguishable from that of the outputs of the simulator. The

statistical zero-knowledge property of the OR and nonnegativity subproofs guarantees that simulators exist that can produce sequences that are statistically indistinguishable from these protocols' outputs, so the only quantity that the simulator S^{V^*} has to produce is C_{-x}, whose true value can be generated directly from C_x due to the homomorphic property of the used commitment scheme. Thus, the whole protocol is statistically zero-knowledge. □

C.3. Sketch of the proof for Theorem 1

Proof.

Completeness. Let us assume that both parties behave according to the protocol. The values C_{A_k} calculated by the correct prover and the correct verifier coincide. For correctly produced $C_{|A_k|}$, the completeness of the absolute value subproof guarantees the acceptance of the verifier; equally, the completeness of the rounded square root subproof guarantees the acceptance for a correctly calculated C_{B_k}. Next, the values of C_D computed by both parties coincide, and, finally, due to the completeness of the nonnegativity proof, the verifier will accept the whole proof in case the signal $\{Y_k\}$ is inside the detection region. For the case of a binary antipodal spreading sequence (Section 5), if the values G, H_k and C_{th} are correctly calculated, the completeness of the nonnegativity proof guarantees the acceptance when $\{Y_k\}$ is inside the detection region. This concludes the completeness proof.

Soundness. The binding property of the commitments assures that the prover will not be able to open the commitments that he calculates (C_{A_k}, $C_{|A_k|}$, C_{B_k}, C_D, C_{th}) to wrong values. Furthermore, the statistical soundness of the used subproofs (absolute value, rounded square root, and nonnegativity) guarantees that an incorrect input in any of them will only succeed with negligible probability. This fact, together with the homomorphic properties of the commitments, that makes impossible for the prover to fake the arithmetic operations performed in parallel by the verifier, propitiates that the probability that a signal $\{Y_k^*\}$ that is not inside the detection region succeeds the proof be negligible.

Zero-knowledge. We can construct a simulator S^{V^*} such that the real interactions have a probability distribution indistinguishable from that of the outputs of the simulator. The statistical zero-knowledge property of the absolute value, rounded square root and nonnegativity subproofs guarantee the existence of simulators for their outputs; thus, S^{V^*} can generate C_{A_k}, C_D, and C_{th} as in a real execution of the protocol, thanks to the homomorphic properties of the commitment scheme. On the other hand, it must generate $C_{|A_k|}$ and C_{B_k} as commitments to random numbers; the statistical hiding property of the commitments guarantees that the distribution of these random commitments be indistinguishable from the true commitments. Furthermore, these generated values will not affect the indistinguishability of the simulators for the subproofs, as these simulators do not need knowledge of the committed values in order to succeed. Thus, the output of S^{V^*} is indistinguishable from true interactions of an accepting protocol, and the whole protocol is statistically zero-knowledge. □

ACKNOWLEDGMENTS

This work was partially funded by Xunta de Galicia under projects PGIDT04 TIC322013PR and PGIDT04 PXIC32202PM, Competitive Research Units Program Ref. 150/2006, MEC project DIPSTICK, Ref. TEC2004-02551/TCM, MEC FPU grant, Ref. AP2006-02580, and European Commission through the IST Program under Contract IST-2002-507932 ECRYPT. ECRYPT disclaimer: the information in this paper is provided as is, and no guarantee or warranty is given or implied that the information is fit for any particular purpose. The user thereof uses the information at its sole risk and liability. This work was partially presented at ACM Multimedia and Security Workshop 2006 [24] and Electronic Imaging 2007 [25].

REFERENCES

[1] S. Goldwasser, S. Micali, and C. Rackoff, "The knowledge complexity of interactive proof systems," *SIAM Journal on Computing*, vol. 18, no. 1, pp. 186–208, 1989.

[2] A. Adelsbach and A.-R. Sadeghi, "Zero-knowledge watermark detection and proof of ownership," in *Proceedings of the 4th International Workshop on Information Hiding (IH '01)*, vol. 2137 of *Lecture Notes in Computer Science*, pp. 273–288, Springer, Pittsburgh, Pa, USA, April 2001.

[3] I. Damgård, "Commitment schemes and zero-knowledge protocols," in *Lectures on Data Security: Modern Cryptology in Theory and Practice*, vol. 1561 of *Lecture Notes in Computer Science*, pp. 63–86, Springer, Aarhus, Denmark, July 1998.

[4] P. Comesaña, L. Pérez-Freire, and F. Pérez-González, "Blind newton sensitivity attack," *IEE Proceedings on Information Security*, vol. 153, no. 3, pp. 115–125, 2006.

[5] A. Piva, V. Cappellini, D. Corazzi, A. De Rosa, C. Orlandi, and M. Barni, "Zero-knowledge ST-DM watermarking," in *Security, Steganography, and Watermarking of Multimedia Contents VIII*, E. J. Delp III and P. W. Wong, Eds., vol. 6072 of *Proceedings of SPIE*, pp. 1–11, San Jose, Calif, USA, January 2006.

[6] J. R. Hernández, M. Amado, and F. Pérez-González, "DCT-domain watermarking techniques for still images: detector performance analysis and a new structure," *IEEE Transactions on Image Processing*, vol. 9, no. 1, pp. 55–68, 2000.

[7] I. Damgård and E. Fujisaki, "A statistically-hiding integer commitment scheme based on groups with hidden order," in *Proceedings of the 8th International Conference on the Theory and Application of Cryptology and Information Security: Advances in Cryptology (ASIACRYPT '02)*, vol. 2501 of *Lecture Notes In Computer Science*, pp. 125–142, Springer, Queenstown, New Zealand, December 2002.

[8] M. Bellare and O. Goldreich, "On defining proofs of knowledge," in *Proceedings of the 12th Annual International Cryptology Conference on Advances in Cryptology (CRYPTO '92)*, vol. 740 of *Lecture Notes in Computer Science*, pp. 390–420, Springer, Santa Barbara, Calif, USA, August 1992.

[9] L. Pérez-Freire, P. Comesaña, and F. Pérez-González, "Detection in quantization-based watermarking: performance and security issues," in *Security, Steganography, and Watermarking of Multimedia Contents VII*, E. J. Delp III and P. W. Wong, Eds., vol. 5681 of *Proceedings of SPIE*, pp. 721–733, San Jose, Calif, USA, January 2005.

[10] F. Pérez-González, F. Balado, and J. R. Hernández Martin, "Performance analysis of existing and new methods for data

hiding with known-host information in additive channels," *IEEE Transactions on Signal Processing*, vol. 51, no. 4, pp. 960–980, 2003.

[11] M. Barni and F. Bartolini, *Watermarking Systems Engineering. Signal Processing and Communications*, Marcel Dekker, New York, NY, USA, 2004.

[12] B. Chen and G. W. Wornell, "Quantization index modulation: a class of provably good methods for digital watermarking and information embedding," *IEEE Transactions on Information Theory*, vol. 47, no. 4, pp. 1423–1443, 2001.

[13] P. Comesaña and F. Pérez-González, "Breaking the BOWS watermarking system: key guessing and sensitivity attacks," to appear in *EURASIP Journal on Information Security*.

[14] S. Craver, "Zero knowledge watermark detection," in *Proceedings of the 3rd International Workshop on Information Hiding (IH '99)*, vol. 1768 of *Lecture Notes in Computer Science*, pp. 101–116, Springer, Dresden, Germany, September 2000.

[15] A. Adelsbach, S. Katzenbeisser, and A.-R. Sadeghi, "Watermark detection with zero-knowledge disclosure," in *Multimedia Systems*, vol. 9, pp. 266–278, Springer, Berlin, Germany, 2003.

[16] I. J. Cox, J. Kilian, T. Leighton, and T. Shamoon, "A secure, robust watermark for multimedia," in *Proceedings of the 1st International Workshop on Information Hiding (IH '96)*, vol. 1174 of *Lecture Notes in Computer Science*, pp. 185–206, Springer, Cambridge, UK, May-June 1996.

[17] A. Adelsbach, M. Rohe, and A.-R. Sadeghi, "Non-interactive watermark detection for a correlation-based watermarking scheme," in *Proceedings of the 9th IFIP TC-6 TC-11 International Conference on Communications and Multimedia Security (CMS '05)*, vol. 3677 of *Lecture Notes in Computer Science*, pp. 129–139, Springer, Salzburg, Austria, September 2005.

[18] F. Boudot, "Efficient proofs that a committed number lies in an interval," in *Proceedings of the International Conference on the Theory and Application of Cryptographic Techniques: Advances in Cryptology (EUROCRYPT '00)*, vol. 1807 of *Lecture Notes in Computer Science*, pp. 431–444, Springer, Bruges, Belgium, May 2000.

[19] H. Lipmaa, "On diophantine complexity and statistical zero-knowledge arguments," in *Proceedings of the 9th International Conference on the Theory and Application of Cryptology and Information Security: Advances in Cryptology (ASIACRYPT '03)*, vol. 2894 of *Lecture Notes in Computer Science*, pp. 398–415, Springer, Taipei, Taiwan, November-December 2003.

[20] A. Adelsbach, M. Rohe, and A.-R. Sadeghi, "Complementing zero-knowledge watermark detection: proving properties of embedded information without revealing it," *Multimedia Systems*, vol. 11, no. 2, pp. 143–158, 2005.

[21] M. Bellare and P. Rogaway, "Random oracles are practical: a paradigm for designing efficient protocols," in *Proceedings of the 1st ACM Conference on Computer and Communications Security (CCS '93)*, pp. 62–73, ACM Press, Fairfax, Va, USA, November 1993.

[22] A. Adelsbach, M. Rohe, and A.-R. Sadeghi, "Overcoming the obstacles of zero-knowledge watermark detection," in *Proceedings of the Workshop on Multimedia and Security (MM&Sec '04)*, pp. 46–54, Magdeburg, Germany, September 2004.

[23] R. Cramer, I. Damgård, and B. Schoenmakers, "Proofs of partial knowledge and simplified design of witness hiding protocols," in *Proceedings of the 14th Annual International Cryptology Conference on Advances in Cryptology (CRYPTO '94)*, vol. 839 of *Lecture Notes In Computer Science*, pp. 174–187, Santa Barbara, Calif, USA, August 1994.

[24] J. R. Troncoso-Pastoriza and F. Pérez-González, "Zero-knowledge watermark detector robust to sensitivity attacks," in *Proceedings of the 8th Workshop on Multimedia and Security (MM&Sec '06)*, pp. 97–107, Geneva, Switzerland, September 2006.

[25] J. R. Troncoso-Pastoriza and F. Pérez-González, "Efficient non-interactive zero-knowledge watermark detector robust to sensitivity attacks," in *Security, Steganography, and Watermarking of Multimedia Contents IX*, E. J. Delp III and P. W. Wong, Eds., vol. 6505 of *Proceedings of SPIE*, pp. 1–12, San Jose, Calif, USA, January 2007.