Security of Lattice-Based Data Hiding Against the Watermarked Only Attack

Luis Pérez-Freire, Fernando Pérez-González

Abstract

This paper presents a security analysis for data hiding methods based on nested lattice codes, extending the analysis provided by previous works to a more general scenario. The security is quantified as the difficulty of estimating the secret key used in the embedding process, assuming the attacker has available several signals watermarked with the same secret key. The theoretical analysis accomplished in the first part of this paper quantifies security in an information-theoretic sense by means of the mutual information between the watermarked signals and the secret key, addressing important issues such as the possibility of achieving perfect secrecy and the impact of the embedding rate in the security level. In the second part, a practical algorithm for estimating the secret key is proposed, and the information extracted is used for implementing a reversibility attack on real images.

Index Terms

Watermarking security, quantization index modulation, lattice data hiding, nested lattice codes, mutual information, equivocation, perfect secrecy, set-membership estimation, tree search.

EDICS Category: WAT-BINM, WAT-THEO, WAT-OTHA

I. Introduction

Watermarking security has emerged in the last years as a new research topic, whose basics can be found in [1],[2],[3] and the references therein. The framework for security analysis adopted in these works follows a cryptanalytic approach: all the parameters of the watermarking scheme are assumed to be public, and the security relies only on a secret key. Similarly to a cryptographic key, the latter will be used, in general, for watermarking several pieces of content.

Luis Pérez-Freire and Fernando Pérez-González are with the Signal Theory and Communications Department, ETSI Telecommunications, University of Vigo, 36310 Vigo, Spain (e-mail: {lpfreire,fperez}@gts.tsc.uvigo.es)

This work was supported in part by *Xunta de Galicia* under Projects PGIDT04 TIC322013PR, PGIDT04 PXIC32202PM, 07TIC012322PR (FACTICA), 2007/149 (REGACOM); MEC Project DIPSTICK, reference TEC2004-02551/TCM, MEC Project SPROACTIVE, reference TEC2007-68094-C02- 01/TCM, and by the European Commission through the IST Programme 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.

Depending on the considered application, the number of contents watermarked with the same key may be very large, constituting a security risk. Thus, the main target of the security analysis is to determine whether the watermarking scheme conceals properly the secret key; if it is not the case, then we are interested in assessing the security level of the scheme, defined as the number of observations (i.e. signals watermarked with the same key) needed to achieve an estimate of the secret key up to a certain accuracy [2]. As in cryptanalysis, the development of practical attacks for estimating secret keys must be considered as an important part of the security analysis. The importance of the proposed framework lies in that a successful attack to security offers a complete break of the watermarking system. If the attacker manages to accurately estimate the secret key, then he has total access to the watermarking channel [1],[4] for embedding and decoding hidden information at will, and even for removing watermarks with small distortion. As an illustrative example of a scenario that fits in the framework described in this paper we can mention fingerprinting applications, where a seller embeds identification codes of the buyers in the digital contents to be sold, with the purpose of enabling their traceability. In this case, each content contains different embedded information, but all the contents will have been watermarked with the secret key of the seller. Thus, a possible approach for the buyers in order to implement a collusion attack (although many other approaches are possible) is to estimate first the secret key and then to remove the watermark of each content. In this case, a security analysis could provide an approximate figure of the observations needed for successfully implementing such collusion attack, giving a certain advantage to the seller.

In this paper we focus on the security analysis of data hiding schemes based on nested lattice codes [5], usually known as lattice DC-DM schemes, which have been widely studied in the last years, mainly due to the connection between Costa's result [6] and recent works on lattice codes [7]. However, their study from the security viewpoint has not been addressed until very recently in [8]. The analysis in [8] was mainly restricted to the so-called "Known Message Attack" (KMA) scenario, where the messages embedded in each watermarked signal were assumed to be known by the attacker. Though the KMA scenario provides important insights into the security of lattice DC-DM, for a complete security analysis it is necessary to consider a more general setup: the "Watermarked Only Attack" (WOA) scenario [1], where the attacker no longer knows anything about the embedded messages. This paper extends the theory and algorithms developed in [8] to WOA, presenting in a more rigorous manner the results reported in [9]. As in [8],[9], the tool for measuring the security is the mutual information between the observations and the secret key (a.k.a. information leakage), which provides a lower bound on the key estimation error (see [8, Sect. III-D]). Intuitively, a smaller information leakage implies a higher security level, because in that case the attacker needs a larger number of observations in order to achieve the same key estimation error. The first part of this paper measures the information leakage about the key for lattice DC-DM schemes, putting special emphasis in the comparison between KMA and WOA scenarios. The possibility of achieving perfect secrecy is also addressed. The second part of this paper shows how the information about the key provided by the observations can be extracted and used in practical scenarios, proposing a reversibility attack based on the estimated key. The proposed estimation algorithm works with

any arbitrary nested lattice code, and is applicable to high embedding rate scenarios.

The main notational conventions used in this paper are the following: random variables are denoted by capital letters, and vectors are represented by means of boldface letters. Calligraphic letters are reserved for sets. The volume of a bounded set $\mathcal{X} \in \mathbb{R}^n$ is denoted by $\text{vol}(\mathcal{X})$, whereas the cardinality of a discrete set $\mathcal{C}$ with a countable number of elements is denoted by $|\mathcal{C}|$. $H(\cdot)$ and $h(\cdot)$ denote entropy and differential entropy [10], respectively, both expressed in natural units. The indicator function, denoted by $\phi_{\mathcal{B}}(\cdot)$ and defined as

$$\phi_{\mathcal{B}}(\mathbf{z}) = \begin{cases} 1, & \mathbf{z} \in \mathcal{B} \\ 0, & \text{otherwise,} \end{cases} \tag{1}$$

will be widely used throughout the text. The probability density function of a random variable X is denoted by $f(x)$, and the expectation of a function $\psi(X)$ over X is denoted by $E_X[\psi(X)]$.

The p-ary alphabet that represents the messages to be transmitted is defined as $\mathcal{M} \triangleq \{0, 1, \ldots, p-1\}$. The "messsage space", defined as $\mathcal{M}^{N_o} \triangleq \mathcal{M} \times \ldots \times \mathcal{M}$, represents the p^{N_o} possible message sequences that can be formed with such an alphabet in N_o channel uses. The sequences in $\mathcal{M}^{N_o}$ can be arranged using any arbitrary ordering (their value in base p, for instance), not relevant for our analysis. The notation $\mathbf{m}^{(k)} = [m_1^k, \ldots, m_{N_o}^k]$ will be used for indexing the kth sequence in $\mathcal{M}^{N_o}$.

II. Theoretical model

This section introduces the mathematical model for lattice data hiding and the lattice constructions used in the paper. As this introduction is not intended to be exhaustive, the interested reader is referred to [11] for a comprehensive description of lattices, and to [5] for a more complete description of lattice codes for data hiding. The assumptions about the attacker are also stated at the end of this section.

A. Lattices and nested lattice codes

Algebraically, a lattice Λ is defined as a discrete subgroup of $\mathbb{R}^n$ endowed with the natural addition operation. A lattice in n-dimensional space can be generated by any integer combination of a set of n linearly independent basis vectors $\{\mathbf{v}_1, \ldots, \mathbf{v}_n\}$, which form the "generating matrix" of Λ. Every lattice has an associated nearest neighbor lattice quantizer which maps any vector $\mathbf{x} \in \mathbb{R}^n$ to the nearest lattice point of Λ, and is defined as

$$Q_\Lambda(\mathbf{x}) = \arg\min_{\mathbf{u} \in \Lambda} \{||\mathbf{x} - \mathbf{u}||\}, \tag{2}$$

where $||\cdot||$ denotes the Euclidean norm. Ties in (2) can be broken arbitrarily but systematically. The fundamental Voronoi region of Λ, a concept that will be frequently recalled in this paper, is defined as [11]

$$\mathcal{V}(\Lambda) \triangleq \{\mathbf{x} \in \mathbb{R}^n : Q_\Lambda(\mathbf{x}) = \mathbf{0}\},$$

and corresponds to the n-dimensional polytope wherein all points are quantized to $\mathbf{0}$. The volume of a lattice Λ is defined as the volume of its Voronoi region, and will be denoted by $\text{vol}(\mathcal{V}(\Lambda))$. We will also define the modulo-Λ

reduction of a vector $\mathbf{x} \in \mathbb{R}^n$ as

$$\mathbf{x} \mod \Lambda \triangleq \mathbf{x} - Q_\Lambda(\mathbf{x}),$$

and the modulo-Λ reduced vector will be denoted by $\tilde{\mathbf{x}}$. Notice that $\tilde{\mathbf{x}} \in \mathcal{V}(\Lambda)$, and it can be regarded as the quantization error resulting from the quantization operation.

The application of lattices to data hiding is based on the concept of "lattice partitioning." Given a certain lattice Λ, we can define a sublattice Λ' which is a subset of the points in Λ (i.e. $\Lambda' \subset \Lambda$) and is itself a lattice. The set $\{\mathbf{u}+\Lambda'\}$, with $\mathbf{u} \in \Lambda$, is known as a coset of Λ'. Due to the periodic structure of lattices, there exist infinite $\mathbf{u}$ that yield the same coset. The vector $\mathbf{u}$ with the smallest Euclidean norm is termed a "coset leader." From the definitions of lattice and Voronoi region, it follows that the coset leaders always belong to $\mathcal{V}(\Lambda')$. The set of all cosets of Λ' with respect to Λ is called the "partition" of Λ induced by Λ', and it carries a group structure with the natural addition operation. It can be proved that the number of different cosets is given by the so-called "nesting ratio" $\frac{\text{vol}(\mathcal{V}(\Lambda'))}{\text{vol}(\mathcal{V}(\Lambda))} = p$. The union of the p cosets yields the lattice Λ, i.e. $\bigcup_{k=0}^{p-1} \mathbf{d}_k + \Lambda' = \Lambda$, where $\mathbf{d}_k$ denote the coset leaders.

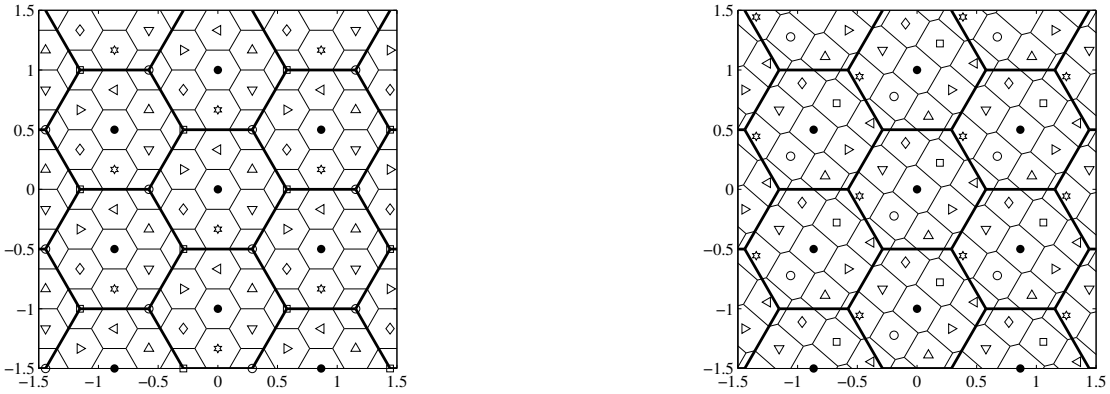
A nested lattice code is defined by two parameters: a shaping (coarse) lattice Λ and a fine lattice Λ_f such that $\Lambda \subset \Lambda_f$. The pair (Λ, Λ_f) defines a partition which in turn yields a set of p coset leaders or codewords $\mathcal{C}_p = \{\mathbf{d}_k,\ k \in \mathcal{M}\}$. Each letter $k \in \mathcal{M}$ is mapped to one coset leader $\mathbf{d}_k \in \mathcal{C}_p$, and thus to the kth coset of Λ. Although the mapping between elements of $\mathcal{M}$ and $\mathcal{C}$ can be arbitrary, we will assume that the letter 0 corresponds to $\mathbf{d}_0 = \mathbf{0}$. Nested lattice codes can be constructed in a number of ways. Although most results of this paper are quite general, we focus in some cases on two particular constructions, given their importance: the self-similar lattice construction [12] and the so-called "Construction A" [11],[13]. In the self-similar construction, the nested code is obtained as follows:[1]

1) Define a positive integer $p^{\frac{1}{n}} \in \mathbb{N}$, where n is the dimensionality of Λ.
2) Compute the fine lattice as $\Lambda_f \triangleq p^{-\frac{1}{n}}\Lambda$. It follows that the lattice Λ is a sublattice of Λ_f, resulting in a nesting ratio $\frac{\text{vol}(\mathcal{V}(\Lambda))}{\text{vol}(\mathcal{V}(\Lambda_f))} = p$, and an embedding rate $R = \log(p)/n$.
3) Obtain the set of coset leaders $\mathcal{C}_p$ as $\Lambda_f \cap \mathcal{V}(\Lambda)$.

In Construction A, the nested lattice code is completely specified by a "generating vector" and the lattice Λ. It is summarized as follows:

1) Define a positive integer p and a generating vector $\mathbf{g} \in \mathbb{Z}_p^n$, where $\mathbb{Z}_p = \{0, 1, \ldots, p-1\}$. Compute the codebook $\mathcal{Q} \triangleq \{\mathbf{c} \in \mathbb{Z}_p^n : \mathbf{c} = k \cdot \mathbf{g} \mod p,\ k \in \mathcal{M}\}$, which is contained in the hypercube $[0, p)^n$. Then, construct the lattice $\Lambda' = p^{-1}\mathcal{Q} + \mathbb{Z}^n$.
2) Define the generating matrix $\mathbf{G} \in \mathbb{R}^{n \times n}$ (where each column is a basis vector) of the coarse (shaping) lattice Λ. Apply the linear transformation $\Lambda_f = \mathbf{G}\Lambda'$. It immediately follows that Λ is a sublattice of Λ_f and the nesting ratio is $\frac{\text{vol}(\mathcal{V}(\Lambda))}{\text{vol}(\mathcal{V}(\Lambda_f))} = p$, resulting in a coding rate $R = \log(p)/n$.

[1]More general self-similar partitions consider also rotations of Λ [14], but we will restrict our attention to those obtained through scaling.



(a) (b)

Fig. 1. Nested lattice codes of rate $R = \log(9)/2$ with hexagonal shaping lattice, obtained by means of self-similar construction (a) and with Construction A (b). The Voronoi regions of Λ_f and Λ are represented by thin and thick lines, respectively.

3) The set of coset leaders $\mathcal{C}_p$ is given by $\Lambda_f \cap \mathcal{V}(\Lambda)$, or equivalently, $p^{-1}\mathbf{G}\mathcal{Q} \mod \Lambda$. Notice that the mapping between the letters of the alphabet and the coset leaders follows directly from the construction procedure.

The advantage of Construction A over the self-similar construction is that it allows to build codes of arbitrary rate (without the restriction $p^{\frac{1}{n}} \in \mathbb{N}$). Moreover, if p is chosen as a prime number and Λ is a good lattice for MSE quantization, then good (asymptotic) properties of the code are ensured [13].

Examples of 2-dimensional nested lattice codes are shown in Fig. 1, using a hexagonal shaping lattice and $p = 9$. For Construction A, the generating vector $\mathbf{g} = [1, 2]^T$ was chosen. In both cases, the lattice points belonging to the same coset are represented by the same symbol. Notice that (as in Fig. 1(a)) a coset leader may fall exactly in the frontier between several quantization cells.

B. Embedding and decoding

The mathematical model for lattice data hiding considered in this paper is shown in Fig. 2. First, the host signal is partitioned into non-overlapping blocks $\mathbf{X}_k$ of length n. The message to be embedded may undergo channel coding, yielding the symbols $M_k \in \mathcal{M}$. In our setup, the messages embedded in different blocks are assumed to be equiprobable in $\mathcal{M}$ and independent from each other, unless otherwise stated. The parameter $\mathbf{T}$ is a n-dimensional vector, termed "secret dither," used to randomize the embedding and decoding functions. This vector plays the role of secret key. In the lattice DC-DM scheme, each letter M_k is embedded in one block $\mathbf{X}_k$ by means of a randomized lattice quantizer as shown in Fig. 2. First, using $\mathbf{T}$, M_k and Λ (the shaping lattice), the coset $\mathcal{U}_{M_k,\mathbf{T}} = \Lambda + \mathbf{d}_{M_k} + \mathbf{T}$ is obtained, where $\mathbf{d}_{M_k}$ is the coset leader associated to M_k. Thereafter, the block $\mathbf{X}_k$ is quantized to the nearest point in $\mathcal{U}_{M_k,\mathbf{T}}$ and the resulting quantization error is computed. Finally, this quantization error is scaled by the "distortion compensation parameter" $\alpha \in [0, 1]$, and added back to $\mathbf{X}_k$ in order to obtain the watermarked block $\mathbf{Y}_k$. Mathematically,

$$\mathbf{Y}_k = \mathbf{X}_k + \alpha(Q_{\mathcal{U}_{M_k,\mathbf{T}}}(\mathbf{X}_k) - \mathbf{X}_k), \tag{3}$$

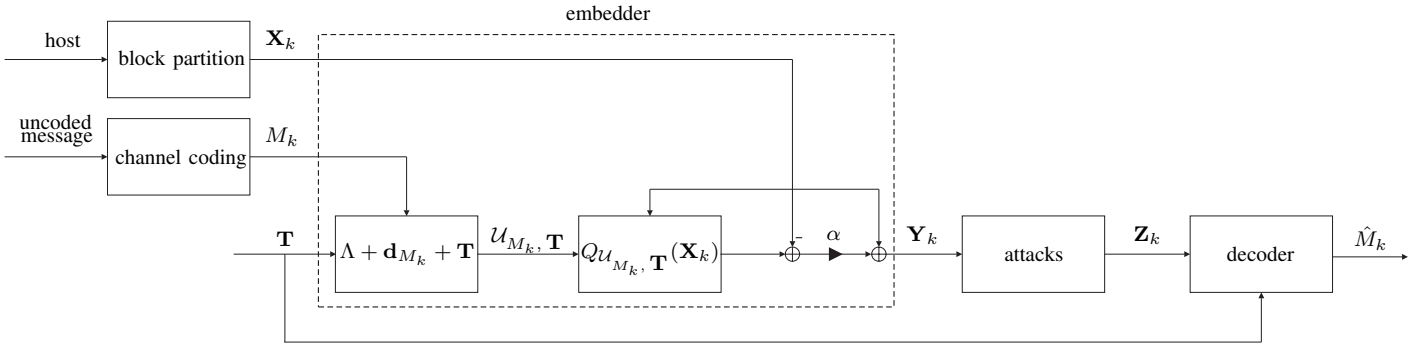


Fig. 2. Block diagram showing the lattice data hiding model. Parameter $\mathbf{T}$ is the secret dither.

where $Q_{\mathcal{U}_{M_k,\mathbf{T}}}(\mathbf{x})$ is a nearest-neighbor quantizer whose centroids are distributted according to $\mathcal{U}_{M_k,\mathbf{T}}$. Taking into account that $Q_{\Lambda+\mathbf{p}}(\mathbf{x}) = Q_\Lambda(\mathbf{x}-\mathbf{p})+\mathbf{p}$, the embedding function (3) is usually implemented in practice by means of a "dithered" lattice quantizer as follows:

$$\mathbf{Y}_k = \mathbf{X}_k + \alpha(Q_\Lambda(\mathbf{X}_k - \mathbf{d}_{M_k} - \mathbf{T}) - \mathbf{X}_k + \mathbf{d}_{M_k} + \mathbf{T}). \tag{4}$$

Notice that Eq. (4) is equivalent to (3) and to the embedder depicted in Fig. 2. The distortion caused by the embedding process can be computed by resorting to the usual assumption (a.k.a. "flat-host assumption") that the variance of the components of $\mathbf{X}_k$ is sufficiently large, in such a way that the host distribution is uniform inside each quantization cell. Thus, from (4), the embedding distortion per dimension in a mean-squared-error sense results in

$$D_w = \frac{1}{n} E_{\mathbf{X}_k,\mathbf{Y}_k}\left[||\mathbf{X}_k - \mathbf{Y}_k||^2\right] = \alpha^2 P(\Lambda),$$

where $P(\Lambda)$ denotes the second-order moment per dimension of the Voronoi region of Λ [11].

The most popular decoders are those termed "lattice decoders", where the embedded message is estimated by choosing the coset which is closest to the received (attacked) vector $\mathbf{Z}_k = \Upsilon(\mathbf{Y}_k)$, where $\Upsilon(\cdot)$ represents the transformation applied by the attacker to $\mathbf{Y}_k$. Hence, the lattice decoder can be mathematically formulated as

$$\hat{M}_k = \min_{m\in\mathcal{M}} \left\{||Q_\Lambda(\mathbf{Z}_k - \mathbf{d}_m - \mathbf{T}) - \mathbf{Z}_k + \mathbf{d}_m + \mathbf{T}||\right\}, \tag{5}$$

where $||\cdot||$ denotes the Euclidean norm. Bear in mind that the decoder needs the correct realization of $\mathbf{T}$ for successful performance. Traditionally, the attacker designs the function $\Upsilon(\cdot)$ without taking into account the secret dither $\mathbf{T}$. In this paper, however, the attacker focuses his strategy on $\mathbf{T}$, as explained below.

C. Assumptions on the attacker's strategy

With regard to the attacker's strategy, it is assumed that he manages to gather an ensemble of watermarked blocks $\{\mathbf{Y}_k,\ k = 1,\ldots,N_o\}$ (hereinafter, "observations"), which may belong to different host signals, but all of them were watermarked with the same secret key $\mathbf{T}$. Notice that this is a reasonable assumption, since a user is very unlikely to change his secret key every time he watermarks an object; in fact, depending on the considered scenario, the number of available observations N_o could be fairly large.

We further assume, sticking to Kerckhoffs' principle [15], that the attacker knows the embedding parameters being used, i.e. Λ, $\mathcal{C}_p$, and α, whereas he ignores the host blocks $\mathbf{X}_k$, the embedded symbols M_k, and $\mathbf{T}$. The first step performed by the attacker is the modulo reduction of the watermarked blocks as $\tilde{\mathbf{Y}}_k \triangleq \mathbf{Y}_k \mod \Lambda$. Under the flat-host assumption introduced above (Sect. II-B), such modulo reduction does not imply any loss of information for the attacker, as discussed in [8]. Note that (4) can be rewritten as

$$\mathbf{Y}_k = \mathbf{d}_{M_k} + \mathbf{T} + \mathbf{Q}_\Lambda(\mathbf{U}_k) + (1-\alpha)\mathbf{N}_k, \tag{6}$$

where $\mathbf{U}_k \triangleq \mathbf{X}_k - \mathbf{d}_{M_k} - \mathbf{T}$ and $\mathbf{N}_k \triangleq \mathbf{U}_k - Q_\Lambda(\mathbf{U}_k)$. Thus, the modulo-$\Lambda$ reduced observations seen by the attacker are given by

$$\tilde{\mathbf{Y}}_k = (\mathbf{d}_{M_k} + \mathbf{T} + (1-\alpha)\mathbf{N}_k) \mod \Lambda. \tag{7}$$

From (7), it becomes clear that the secret dither $\mathbf{T}$ is concealed by the transmitted message M_k and by the host-interference $\mathbf{N}_k$. The parameter α controls the amount of host-interference or "self-noise", thus affecting the robustness of the lattice data hiding scheme. However, from a security standpoint, one can take benefit of the randomness introduced by the self-noise for achieving good secrecy and complicating the attacker's task, as will be seen in Section III-A. Finally, the secret dither $\mathbf{T}$ is assumed to be uniformly distributed in the Voronoi region $\mathcal{V}(\Lambda)$, which turns out to be the worst case for the attacker [8].

The statistical distribution of the observations seen by the attacker can be easily identified by recalling the flat-host assumption, which makes $\mathbf{N}_k$ uniformly distributed in $\mathcal{V}(\Lambda)$. Hence, it follows from (7) that

$$\varphi(\mathbf{x}) \triangleq f(\tilde{\mathbf{y}}_k | m_k = 0, \mathbf{t} = \mathbf{0}) = (\text{vol}(\mathcal{Z}(\Lambda)))^{-1} \cdot \phi_{\mathcal{Z}(\Lambda)}(\mathbf{x}), \tag{8}$$

with $\mathcal{Z}(\Lambda) \triangleq (1-\alpha)\mathcal{V}(\Lambda)$, and $\phi_{\mathcal{Z}(\Lambda)}(\mathbf{x})$ defined in (1). Using again the flat-host assumption, we can simply write

$$f(\tilde{\mathbf{y}}_k | m_k, \mathbf{t}) = \varphi(\tilde{\mathbf{y}}_k - \mathbf{d}_{m_k} - \mathbf{t} \mod \Lambda).$$

Hence, $f(\tilde{\mathbf{y}}_k|\mathbf{t}) = \frac{1}{p}\sum_{m_k=0}^{p-1} f(\tilde{\mathbf{y}}_k|m_k, \mathbf{t})$, and $f(\tilde{\mathbf{y}}_k) = \int f(\tilde{\mathbf{y}}_k|\mathbf{t}) f(\mathbf{t}) d\mathbf{t}$. Under the assumption of uniform $\mathbf{T}$, it is elementary to prove that each watermarked signal $\tilde{\mathbf{Y}}_k$ is uniformly distributed over $\mathcal{V}(\Lambda)$.

III. Theoretical security analysis

The amount of information that leaks from the observations is quantified by means of the mutual information $I(\tilde{\mathbf{Y}}_1, \ldots, \tilde{\mathbf{Y}}_{N_o}; \mathbf{T})$ [2]. Making use of the chain rule for entropies [10], this mutual information can be rewritten in a more illustrative manner as

$$\begin{aligned} I(\tilde{\mathbf{Y}}_1, \ldots, \tilde{\mathbf{Y}}_{N_o}; \mathbf{T}) &= I(\tilde{\mathbf{Y}}_1, \ldots, \tilde{\mathbf{Y}}_{N_o}; \mathbf{T} | M_1, \ldots, M_{N_o}) + I(\tilde{\mathbf{Y}}_1, \ldots, \tilde{\mathbf{Y}}_{N_o}; M_1, \ldots, M_{N_o}) \\ &- I(\tilde{\mathbf{Y}}_1, \ldots, \tilde{\mathbf{Y}}_{N_o}; M_1, \ldots, M_{N_o} | \mathbf{T}). \end{aligned} \tag{9}$$

The first term in the right hand side of (9) is the information leakage for KMA, which was studied in [8]. We recall here two fundamental properties of the KMA scenario, that will be frequently used in the remaining of this paper:

1) Under the assumption of uniform $\mathbf{T}$, the conditional pdf of the dither signal is [8, Sect. II]

$$f(\mathbf{t}|\tilde{\mathbf{y}}_1,\ldots,\tilde{\mathbf{y}}_{N_o},m_1^k,\ldots,m_{N_o}^k) = (\mathrm{vol}(\mathcal{S}_{N_o}(\mathbf{m}^{(k)})))^{-1}\cdot\phi_{\mathcal{S}_{N_o}(\mathbf{m}^{(k)})}(\mathbf{t}) \tag{10}$$

where $\mathbf{m}^{(k)} \triangleq (m_1^k,\ldots,m_{N_o}^k)$, and

$$\mathcal{S}_{N_o}(\mathbf{m}^{(k)}) \triangleq \bigcap_{j=1}^{N_o}\mathcal{D}_j(m_j^k), \quad \text{with} \quad \mathcal{D}_j(m_j^k) \triangleq (\tilde{\mathbf{y}}_j - \mathbf{d}_{m_j^k} - \mathcal{Z}(\Lambda)) \mod \Lambda. \tag{11}$$

$\mathcal{S}_{N_o}(\mathbf{m}^{(k)})$ denotes the "feasible region" for the secret dither, conditioned on the observations and the embedded message sequence $\mathbf{m}^{(k)}$.

2) As shown in [8, Sect. III], the feasible region is always modulo-Λ convex if $\alpha \geq 0.5$ (see [8, Sect. III] for the definition of modulo-Λ convexity). In essence, this means that the feasible region is not composed of disjoint regions if $\alpha \geq 0.5$, thus simplifying the analysis and also the design of practical dither estimators.

The third term in the right hand side of (9) represents the achievable rate for a fair user, i.e. knowing the secret dither $\mathbf{T}$, whereas the second term is the rate achievable by unfair users (which is not null, in general) that do not know $\mathbf{T}$. In this section we consider, first, the possibility of achieving perfect secrecy about $\mathbf{T}$. Thereafter, we study the asymptotic behavior of the information leakage about $\mathbf{T}$ in other situations where perfect secrecy is not achieved, and finally we analyze a practical lattice data hiding scheme.

A. Theoretical and practical perfect secrecy

If the lattice data hiding scheme fulfills the condition

$$I(\tilde{\mathbf{Y}}_1,\ldots,\tilde{\mathbf{Y}}_{N_o};\mathbf{T}) = h(\tilde{\mathbf{Y}}_1,\ldots,\tilde{\mathbf{Y}}_{N_o}) - h(\tilde{\mathbf{Y}}_1,\ldots,\tilde{\mathbf{Y}}_{N_o}|\mathbf{T}) = h(\mathbf{T}) - h(\mathbf{T}|\tilde{\mathbf{Y}}_1,\ldots,\tilde{\mathbf{Y}}_{N_o}) = 0,\ \forall\, N_o, \tag{12}$$

then it is said to provide "perfect secrecy", meaning that no information about $\mathbf{T}$ can be obtained from the observations, no matter the computational effort employed by the attacker. As mentioned in Sect. II-C, each watermarked signal $\tilde{\mathbf{Y}}_k$ is uniformly distributed over $\mathcal{V}(\Lambda)$, since we are assuming uniform $\mathbf{T}$. Hence, the closer is the distribution of $\tilde{\mathbf{Y}}_k$ conditioned on $\mathbf{T}$ to the uniform over $\mathcal{V}(\Lambda)$, the more secure is the scheme. Once the shaping lattice is fixed, the two free parameters for tuning such distribution are the set of coset leaders $\mathcal{C}_p$ and the parameter α. The possibility of achieving perfect secrecy is considered below in Lemma 1 and Proposition 1. We recall that the messages embedded in different observations are assumed to be independent.

First, let us denote by $\mathcal{B}(\mathbf{c},r)$ the n-dimensional closed hypersphere of radius r centered in $\mathbf{c}$. Define the "covering radius" of the lattice Λ as

$$r_c(\Lambda) \triangleq \min\{r : \mathcal{V}(\Lambda) \subseteq \mathcal{B}(\mathbf{0},r)\}. \tag{13}$$

Lemma 1: Consider a sequence of nested lattice codes $(\Lambda,\mathcal{C}_p^*)$ such that $r_c(\Lambda_f) \to 0$ as $p \to \infty$. If $\alpha < 1$, this sequence of lattice codes asymptotically achieves perfect secrecy as $p \to \infty$, and the statistical distribution of $\tilde{\mathbf{Y}}_k$ conditioned on $\mathbf{T}$ converges to the uniform over $\mathcal{V}(\Lambda)$, $\forall\, k = 1,\ldots,N_o$.

Proof: See Appendix A.[2] ■

Lemma 1 shows that the information leakage can be reduced down to zero if the size of the alphabet is properly increased. However, bear in mind that if the condition $r_c(\Lambda_f) \to 0$ does not hold (i.e. if the coset leaders do not "cover" the whole Voronoi region of Λ), the resulting code does not necessarily offer good secrecy, even for high embedding rates (see sections III-C and III-D, which deal with a lattice repetition code). It must be observed that an infinite alphabet size is not affordable in practice. However, this choice of alphabet is used in [7] for showing that a nested lattice code asymptotically achieves the capacity of the modulo-lattice Gaussian channel. Thus, Lemma 1 shows, in conjunction with [7], that simultaneous maximization of robustness and security is theoretically (asymptotically) possible. Finally, we would like to remark that the result of asymptotic perfect secrecy holds for any distribution of the secret dither $\mathbf{T}$, not necessarily the uniform over $\mathcal{V}(\Lambda)$ (cf. Appendix A). In the next proposition, a nested code achieving perfect secrecy and realizable in practice is proposed.

Proposition 1: Any self-similar lattice code of rate $R = \log(p)/n$ and distortion compensation parameter $\alpha = 1 - p^{-\frac{1}{n}}$ achieves perfect secrecy. In that case $\tilde{\mathbf{Y}}_k$ conditioned on $\mathbf{T}$ is uniform over $\mathcal{V}(\Lambda)$, $\forall\, k = 1, \ldots, N_o$.

Proof: Similarly to the proof of Lemma 1, the proof of perfect secrecy can be reduced to showing that the resulting scheme fulfills the condition $I(\tilde{\mathbf{Y}}_1; \mathbf{T}) = h(\tilde{\mathbf{Y}}_1) - h(\tilde{\mathbf{Y}}_1|\mathbf{T}) = 0$.

For $\alpha = 1 - p^{-\frac{1}{n}}$, we have $\mathcal{Z}(\Lambda) = p^{-\frac{1}{n}}\mathcal{V}(\Lambda) = \mathcal{V}(\Lambda_f)$, so the pdf defined in (8) is given by $\varphi(\mathbf{x}) = p \cdot (\mathrm{vol}(\mathcal{V}(\Lambda)))^{-1} \cdot \phi_{\mathcal{V}(\Lambda_f)}(\mathbf{x})$. Hence, under the assumption of equiprobable symbols we can write

$$f(\tilde{\mathbf{y}}_1|\mathbf{T} = \mathbf{t}) = \frac{1}{p}\sum_{i=0}^{p-1} \varphi(\tilde{\mathbf{y}}_1 - \mathbf{t} - \mathbf{d}_i \mod \Lambda) = (\mathrm{vol}(\mathcal{V}(\Lambda)))^{-1} \sum_{i=0}^{p-1} \phi_{\mathcal{V}(\Lambda_f)}(\tilde{\mathbf{y}}_1 - \mathbf{t} - \mathbf{d}_i \mod \Lambda). \tag{14}$$

Taking into account that self-similar partitions fulfill the next "covering property":

$$\bigcup_{i=0}^{p-1} (\mathcal{V}(\Lambda_f) - \mathbf{t} - \mathbf{d}_i) \mod \Lambda = \mathcal{V}(\Lambda), \quad \bigcap_{i=0}^{p-1} (\mathcal{V}(\Lambda_f) - \mathbf{t} - \mathbf{d}_i) \mod \Lambda = \emptyset, \tag{15}$$

it follows that for every $\tilde{\mathbf{y}}_1 \in \mathcal{V}(\Lambda)$ there exists exactly one $\mathbf{d}_i$ such that $(\tilde{\mathbf{y}}_1 - \mathbf{t} - \mathbf{d}_i) \mod \Lambda \in \mathcal{V}(\Lambda_f)$. Hence, (14) becomes $f(\tilde{\mathbf{y}}_1|\mathbf{T} = \mathbf{t}) = (\mathrm{vol}(\mathcal{V}(\Lambda)))^{-1}\ \forall\ \tilde{\mathbf{y}}_1 \in \mathcal{V}(\Lambda)$, so $h(\tilde{\mathbf{Y}}_1|\mathbf{T} = \mathbf{t}) = \log(\mathrm{vol}(\mathcal{V}(\Lambda)))$. Since the entropy of a continuous random variable with bounded support is upper bounded by the log-volume of its support set, we have

$$h(\tilde{\mathbf{Y}}_1|\mathbf{T} = \mathbf{t}) \leq h(\tilde{\mathbf{Y}}_1) \leq \log(\mathrm{vol}(\mathcal{V}(\Lambda))).$$

Thus, we have $h(\tilde{\mathbf{Y}}_1) = h(\tilde{\mathbf{Y}}_1|\mathbf{T}) = \log(\mathrm{vol}(\mathcal{V}(\Lambda)))$, resulting in a null information leakage, and $\tilde{\mathbf{Y}}_1$ is necessarily uniform over $\mathcal{V}(\Lambda)$ regardless the distribution of $\mathbf{T}$. ■

Some important remarks to this result are given below.

Remark 1: The proof of perfect secrecy in Proposition 1 relies on the lattice code itself rather than on the statistical distribution of $\mathbf{T}$. Actually, the result holds for any distribution of the secret dither $\mathbf{T}$. However, it must be taken

[2]The condition $\alpha < 1$ imposed in Lemma 1 comes from the necessity of having a continuous, Riemann integrable pdf for our proof to be valid. The case $\alpha = 1$, for which $f(\tilde{\mathbf{y}}_1|\mathbf{t})$ becomes a probability mass function, cannot be dealt with using the arguments of Appendix A.

into account that the value of α that provides perfect secrecy can be conflicting with other requirements (e.g. error probability), so it is important to properly choose the distribution of $\mathbf{T}$ for maximizing the security when perfect secrecy cannot be attained. This distribution has been shown in [8] to be the uniform over $\mathcal{V}(\Lambda)$, which yields $\tilde{\mathbf{Y}}_k$ also uniform over $\mathcal{V}(\Lambda)$.

Remark 2: It is possible to show that for $\alpha_k = 1 - kp^{-\frac{1}{n}},\ k = 1, \ldots, p^{\frac{1}{n}} - 1$, the condition of perfect secrecy still holds. However, for $k > 1$ there are overlaps between adjacent symbols that produce nonzero error probability even in the absence of noise. This makes necessary the use of channel coding (i.e. error correcting codes) to recover the embedded message reliably, thus breaking the hypothesis of independence between messages embedded in different blocks. As a result, perfect secrecy about $\mathbf{T}$ cannot be assured. To see this, consider a simple example with 2 observations $\{\tilde{\mathbf{Y}}_1, \tilde{\mathbf{Y}}_2\}$, where the embedding parameters fulfill the conditions for perfect secrecy stated in Proposition 1. The mutual information between observations and secret dither is written as

$$I(\tilde{\mathbf{Y}}_1, \tilde{\mathbf{Y}}_2|\mathbf{T}) = I(\tilde{\mathbf{Y}}_1; \mathbf{T}) + I(\tilde{\mathbf{Y}}_2; \mathbf{T}|\tilde{\mathbf{Y}}_1), \tag{16}$$

whereas the rightmost term of (16) can be written as $I(\tilde{\mathbf{Y}}_2; \mathbf{T}|\tilde{\mathbf{Y}}_1) = h(\tilde{\mathbf{Y}}_2|\tilde{\mathbf{Y}}_1) - h(\tilde{\mathbf{Y}}_2|\tilde{\mathbf{Y}}_1, \mathbf{T})$. Given only $\tilde{\mathbf{Y}}_1$, no information about $\mathbf{T}$ is leaked, so $h(\tilde{\mathbf{Y}}_2|\tilde{\mathbf{Y}}_1) = h(\tilde{\mathbf{Y}}_2)$. Given $\mathbf{T}$, we have that $\tilde{\mathbf{Y}}_1$ and $\tilde{\mathbf{Y}}_2$ are independent because M_1, M_2 are independent. Hence, $h(\tilde{\mathbf{Y}}_2|\tilde{\mathbf{Y}}_1, \mathbf{T}) = h(\tilde{\mathbf{Y}}_2|\mathbf{T})$, resulting $I(\tilde{\mathbf{Y}}_2; \mathbf{T}|\tilde{\mathbf{Y}}_1) = I(\tilde{\mathbf{Y}}_2; \mathbf{T}) = 0$. Now, consider the case where M_1, M_2 are not independent. Given only $\tilde{\mathbf{Y}}_1$, no information about $\mathbf{T}$ nor M_1 is obtained, so we have $h(\tilde{\mathbf{Y}}_2|\tilde{\mathbf{Y}}_1) = h(\tilde{\mathbf{Y}}_2)$ again. However, given $\tilde{\mathbf{Y}}_1$ and $\mathbf{T}$, information about M_1 is leaked. Since M_2 is dependent on M_1, we have $h(\tilde{\mathbf{Y}}_2|\tilde{\mathbf{Y}}_1, \mathbf{T}) \leq h(\tilde{\mathbf{Y}}_2|\mathbf{T})$, resulting $I(\tilde{\mathbf{Y}}_2; \mathbf{T}|\tilde{\mathbf{Y}}_1) \geq 0$.

Remark 3: The proof of the proposition resorts to the flat-host assumption to show null information leakage. This means that, in practice, small information leakages may exist due to the finite variance of the host signal, which causes the host distribution to not be strictly uniform in each quantization cell. However, this information leakage seems to be hardly exploitable in practical attacks.

Remark 4: Perfect secrecy about $\mathbf{T}$ does not necessarily mean perfect secrecy about the embedded messages. Using (9), the rate for the attacker under the condition of perfect secrecy is given by

$$I(\tilde{\mathbf{Y}}_1, \ldots, \tilde{\mathbf{Y}}_{N_o}; M_1, \ldots, M_{N_o}) = I(\tilde{\mathbf{Y}}_1, \ldots, \tilde{\mathbf{Y}}_{N_o}; M_1, \ldots, M_{N_o}|\mathbf{T}) - I(\tilde{\mathbf{Y}}_1, \ldots, \tilde{\mathbf{Y}}_{N_o}; \mathbf{T}|M_1, \ldots, M_{N_o}). \tag{17}$$

This "unfair" rate is studied in Section III-C for the repetition coding lattice scheme.

B. Asymptotic analysis and comparison with KMA

When perfect secrecy is not attained, a closed-form expression for the information leakage cannot be given, in general. We are interested here in studying the general behavior of the information leakage for large N_o and comparing it with the KMA scenario. By following a similar reasoning to that in [8, Sect. II], it can be shown that the mutual information in (9) is concave and increasing with N_o. The second and third terms of Eq. (9) define the gap between

the security level of KMA and WOA scenarios. More specifically, we will define the gap as

$$\begin{aligned}\delta(N_o) &\triangleq I(\tilde{\mathbf{Y}}_1,\ldots,\tilde{\mathbf{Y}}_{N_o};M_1,\ldots,M_{N_o}|\mathbf{T}) - I(\tilde{\mathbf{Y}}_1,\ldots,\tilde{\mathbf{Y}}_{N_o};M_1,\ldots,M_{N_o}) \\ &= H(M_1,\ldots,M_{N_o}|\tilde{\mathbf{Y}}_1,\ldots,\tilde{\mathbf{Y}}_{N_o}) - H(M_1,\ldots,M_{N_o}|\tilde{\mathbf{Y}}_1,\ldots,\tilde{\mathbf{Y}}_{N_o},\mathbf{T}). \end{aligned} \quad (18)$$

This gap, which is always positive, quantifies the information about $\mathbf{T}$ that is lost due to the a priori ignorance of the embedded messages. As can be seen from (18), the gap can be formulated as "the information that $\mathbf{T}$ provides about the embedded messages upon the observation of watermarked signals". The next lemma shows that without the knowledge of $\mathbf{T}$ there exists an irreducible ambiguity in the estimation of the embedded message sequence.

Lemma 2: Given N_o observations, consider the a priori message space $\mathcal{M}^{N_o}$ where all the message sequences are assumed to be a priori equiprobable. For $\mathbf{m}^{(1)},\mathbf{m}^{(2)} \in \mathcal{M}^{N_o}$, let us define the equivalence relation

$$\mathbf{m}^{(1)} \sim \mathbf{m}^{(2)} \text{ if } [\mathbf{d}_{m_1^2},\ldots,\mathbf{d}_{m_{N_o}^2}] = [(\mathbf{d}_{m_1^1}+\mathbf{d}_j) \mod \Lambda,\ldots,(\mathbf{d}_{m_{N_o}^1}+\mathbf{d}_j) \mod \Lambda], \text{ for some } j \in \mathcal{M}. \quad (19)$$

Each equivalence class is composed of p elements, and the sequences belonging to the same equivalence class have all the same a posteriori probability.

Proof: By the additive structure of Λ_f, the operation $\mathbf{d}_l = (\mathbf{d}_k+\mathbf{d}_j) \mod \Lambda$, with $l,k,j \in \mathcal{M}$, defines a bijective mapping $\mathbf{d}_k \to \mathbf{d}_l$. Then, for a message sequence $\mathbf{m}^{(1)} = [m_1^1,\ldots,m_{N_o}^1]$, the operation

$$[(\mathbf{d}_{m_1^1}+\mathbf{d}_j) \mod \Lambda,\ldots,(\mathbf{d}_{m_{N_o}^1}+\mathbf{d}_j) \mod \Lambda]$$

yields p different sequences of length N_o when j is varied from 0 to $p-1$. Thus, each equivalence class as defined in (19) is composed of p elements. The a posteriori probability of a message sequence is derived in Appendix B. For a sequence $\mathbf{m}^{(1)} \in \mathcal{M}^{N_o}$, combining equations (51) and (55) of Appendix B we arrive at

$$\Pr(\mathbf{m}^{(1)}|\tilde{\mathbf{y}}_1,\ldots,\tilde{\mathbf{y}}_{N_o}) = \Pr(\mathbf{d}_{m_1^1},\ldots,\mathbf{d}_{m_{N_o}^1}|\tilde{\mathbf{y}}_1,\ldots,\tilde{\mathbf{y}}_{N_o}) = \frac{\mathrm{vol}(\mathcal{S}_{N_o}(\mathbf{m}^{(1)}))}{(\mathrm{vol}(\mathcal{Z}(\Lambda)))^{N_o}\cdot\mathrm{vol}(\mathcal{V}(\Lambda))} \cdot \frac{\Pr(\mathbf{m}^{(1)})}{f(\tilde{\mathbf{y}}_1,\ldots,\tilde{\mathbf{y}}_{N_o})}. \quad (20)$$

Now, let us define $\hat{\mathcal{S}}_{N_o}(\mathbf{m}^{(1)}) = \bigcap_{k=1}^{N_o}(\tilde{\mathbf{y}}_k - \mathbf{d}_{m_k^1} - \mathbf{u} - \mathcal{Z}(\Lambda)) \mod \Lambda$, with $\mathbf{u}$ an arbitrary vector constant for all k. Clearly, $\hat{\mathcal{S}}_{N_o}(\mathbf{m}^{(1)}) = (\mathcal{S}_{N_o}(\mathbf{m}^{(1)}) - \mathbf{u}) \mod \Lambda$, so $\mathrm{vol}(\hat{\mathcal{S}}_{N_o}(\mathbf{m}^{(1)})) = \mathrm{vol}(\mathcal{S}_{N_o}(\mathbf{m}^{(1)}))$. Hence, if $\mathbf{u} = \mathbf{d}_j$, then

$$[(\mathbf{d}_{m_1^1}+\mathbf{u}) \mod \Lambda,\ldots,(\mathbf{d}_{m_{N_o}^1}+\mathbf{u}) \mod \Lambda] = [(\mathbf{d}_{m_1^1}+\mathbf{d}_j) \mod \Lambda,\ldots,(\mathbf{d}_{m_{N_o}^1}+\mathbf{d}_j) \mod \Lambda] = [\mathbf{d}_{m_1^2},\ldots,\mathbf{d}_{m_{N_o}^2}],$$

and it follows that $\mathrm{vol}(\mathcal{S}_{N_o}(\mathbf{m}^{(1)})) = \mathrm{vol}(\mathcal{S}_{N_o}(\mathbf{m}^{(2)}))$. Inserting this result in (20) and recalling that the message sequences are assumed to be a priori equiprobable, the lemma follows. ■

The structure of the equivalence classes depends on the mapping between the elements of $\mathcal{M}$ and $\mathcal{C}_p$ and on the lattice code itself. However, for lattice codes obtained through Construction A (following the procedure described in Sect. II-A), the equivalence relation (19) can be simply expressed as

$$\mathbf{m}^{(1)} \sim \mathbf{m}^{(2)} \text{ if } \mathbf{m}^{(2)} = (\mathbf{m}^{(1)} + j\cdot\mathbf{1}) \mod p, \text{ for some } j \in \mathcal{M}, \quad (21)$$

where $\mathbf{1}$ denotes the vector with its components equal to 1, and the modulo operation is applied componentwise. The proof directly follows from (19) simply by observing that $(\mathbf{d}_k+\mathbf{d}_j) \bmod \Lambda = \mathbf{d}_{(k+j) \bmod p}$ for Construction A.

By virtue of Lemma 2, if the messages embedded in different blocks are mutually independent, then the attacker can aspire (at most) at reducing the uncertainty about the embedded message sequence to a set of p equiprobable sequences. The following theorem states how this ambiguity affects the information leakage about $\mathbf{T}$ for large N_o.

Note that for any nested lattice code there exists a value α_0 such that for $\alpha > \alpha_0$ we can assure $\mathcal{Z}(\Lambda) \subset \mathcal{V}(\Lambda_f)$, so error-free decoding is guaranteed in the absence of noise. Obviously, $\alpha_0 > 0.5$ in any case, although it will depend on the lattice code, in general.

Theorem 1: For any nested lattice code, if α is chosen such that $\mathcal{Z}(\Lambda) \subset \mathcal{V}(\Lambda_f)$, then

$$\lim_{N_o \to \infty} \frac{1}{n} \delta(N_o) = \frac{1}{n} \log(p) = R.$$

Proof: See Appendix C. ■

The result of Theorem 1 has two main implications:

1) For low embedding rates (i.e. low R) the information per dimension that leaks about $\mathbf{T}$ is approximately the same as if the attacker knew the embedded messages. This case is highly relevant in practice, since in practical scenarios the watermarker usually resorts to low embedding rates that allow to recover the embedded message without the use of complex channel coding schemes.
2) As shown in Appendix C, when $N_o \to \infty$, the feasible regions associated to the only message sequences with nonnull probability converge to p different vectors of the form $(\mathbf{t} - \mathbf{d}_j) \mod \Lambda,\ j \in \mathcal{M}$, which are equiprobable. Thus, unambiguous estimation of the secret dither vector is not possible in the WOA scenario.

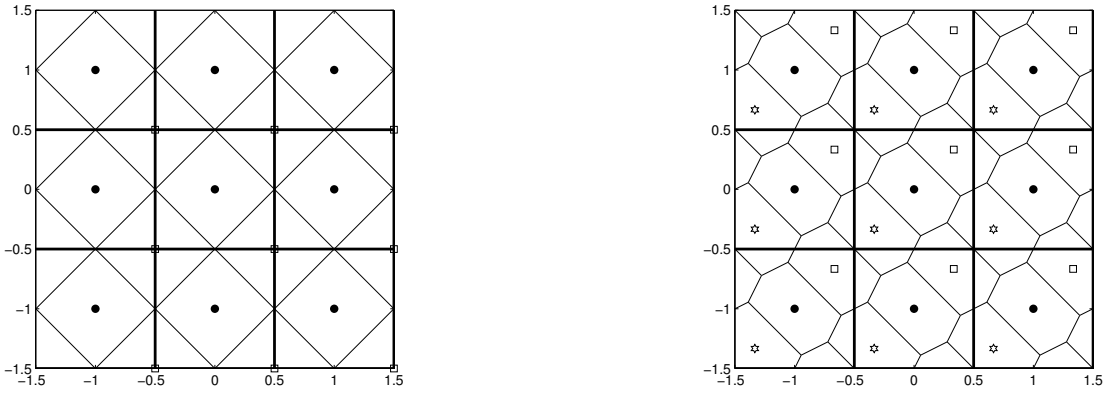
The values of α for which Theorem 1 holds guarantee that no decoding errors occur in the absence of noise. However, in some cases it is advantageous to choose smaller values of α, e.g. when a certain degree of attacking noise is expected [7],[16]. In such cases, the gap function does not necessarily achieve the value R, but this seems to be valid as a lower bound, as we will check in Section III-C for the repetition coding lattice scheme.

Final remark: If the messages conveyed by different observations were not independent, the residual uncertainty expressed in Theorem 1 would be further reduced. This is the case, for instance, when a channel code is applied, since it could provide the attacker with information about the a priori probabilities of each message sequence. This consideration is important when high robustness against noise is sought or when α is small, since the use of channel codes is mandatory in these cases for lowering the decoding error probability.

C. Theoretical results for cubic shaping lattices with repetition coding

DC-DM with repetition coding using scalar quantizers, also known as Scalar Costa Scheme (SCS) with repetition coding [16], is one of the most popular schemes for lattice data hiding. Redundant embedding of the information is performed by repeatedly embedding the same message in n different host samples, using a pair of scalar lattices $\Lambda = \Delta\mathbb{Z}$, $\Lambda_f = \Delta\mathbb{Z}/p$, which yield $\mathcal{V}(\Lambda) = [-\Delta/2, \Delta/2)$, and $d_k = (\Delta k/p) \mod \Delta,\ k = 0, \ldots, p-1$. For a n-dimensional host vector $\mathbf{X}_k$, the embedding function (4) is particularized to

$$Y_{k,i} = X_{k,i} + \alpha(Q_\Lambda(X_{k,i} - d_{M_k} - T_i) - X_{k,i} + d_{M_k} + T_i),\ i = 1, \ldots, n, \tag{22}$$



(a) (b)

Fig. 3. Illustration of repetition codes with $n = 2$ and $\Delta = 1$. Fig. (a) represents a code with $p = 2$, where dots and squares represent the cosets for $m = 0$ and $m = 1$, respectively, and the fine lattice Λ_f is the well-known "checkerboard lattice" [11]. Fig. (b) represents a code with $p = 3$. Notice that the coset leaders fall in the main diagonal of $\mathcal{V}(\Lambda)$.

where the subindex i indicates the ith component of the n-dimensional vector. This simple coding scheme, which results in a code of rate $R = \log(p)/n$, is equivalent to a code obtained through Construction A with $\mathbf{g} = [1, \ldots, 1]^T$ and $\Lambda = \Delta\mathbb{Z}^n$. Bear in mind that, due to the redundant embedding of the message, the repetition scheme provides the attacker with more information about the embedded message than a scheme where n independent scalar embeddings are performed in parallel. Thus, one can intuitively realize that the repetition scheme analyzed here is less secure than the simple SCS. Figure 3 shows two examples of lattice repetition codes for $n = 2$. Although the robustness of this data hiding code has been analyzed in depth in [17], here we are interested in analyzing its security properties.

In order to obtain the information leakage, we rewrite the third term of Eq. (9) as

$$\begin{aligned} I(\tilde{\mathbf{Y}}_1, \ldots, \tilde{\mathbf{Y}}_{N_o}; M_1, \ldots, M_{N_o}|\mathbf{T}) &= H(M_1, \ldots, M_{N_o}) - H(M_1, \ldots, M_{N_o}|\tilde{\mathbf{Y}}_1, \ldots, \tilde{\mathbf{Y}}_{N_o}, \mathbf{T}) \\ &= N_o \cdot (\log(p) - H(M_1|\tilde{\mathbf{Y}}_1, \mathbf{T})). \end{aligned} \tag{23}$$

In turn, the second term in the right hand side of (9) can be expressed as

$$I(\tilde{\mathbf{Y}}_1, \ldots, \tilde{\mathbf{Y}}_{N_o}; M_1, \ldots, M_{N_o}) = N_o \cdot \log(p) - H(M_1, \ldots, M_{N_o}|\tilde{\mathbf{Y}}_1, \ldots, \tilde{\mathbf{Y}}_{N_o}).$$

By combining the two above equations with the information leakage for cubic lattices in the KMA scenario [8, Sect. III-A], the information leakage per dimension reads as

$$\begin{aligned} \frac{1}{n} I(\tilde{\mathbf{Y}}_1, \ldots, \tilde{\mathbf{Y}}_{N_o}; \mathbf{T}) &= \frac{1}{n}\left(N_o \cdot H(M_1|\tilde{\mathbf{Y}}_1, \mathbf{T}) - H(M_1, \ldots, M_{N_o}|\tilde{\mathbf{Y}}_1, \ldots, \tilde{\mathbf{Y}}_{N_o})\right) \\ &- \log(1-\alpha) + \sum_{i=2}^{N_o} \frac{1}{i}, \quad \text{for } N_o \geq 2,\ \alpha \geq 0.5. \end{aligned} \tag{24}$$

Eq. (24) does not admit a closed-form expression, although it is possible to accurately obtain the entropies of interest in a numerical manner.

1) Computation of the achievable rate for fair users: The first term in the right hand side of (24) represents the uncertainty about the embedded message when the secret dither is known. Under the flat-host assumption,

$$H(M_1|\tilde{\mathbf{Y}}_1,\mathbf{T}) = H(M_1|\tilde{\mathbf{Y}}_1,\mathbf{T}=\mathbf{0}) = E_{\tilde{\mathbf{Y}}_1}\left[H(M_1|\tilde{\mathbf{Y}}_1=\tilde{\mathbf{y}},\mathbf{T}=\mathbf{0})\right]. \tag{25}$$

The a posteriori probability of a certain message m is given by

$$\begin{aligned}\Pr(m|\tilde{\mathbf{y}},\mathbf{t}=\mathbf{0}) &= \frac{f(\tilde{\mathbf{y}}|m,\mathbf{t}=\mathbf{0})\cdot\Pr(m)}{f(\tilde{\mathbf{y}}|\mathbf{t}=\mathbf{0})} = \frac{\Pr(m)}{f(\tilde{\mathbf{y}}|\mathbf{t}=\mathbf{0})}\prod_{i=1}^{n} f(\tilde{y}_i|m,t_i=0)\\ &= \frac{\Pr(m)}{f(\tilde{\mathbf{y}}|\mathbf{t}=\mathbf{0})}\prod_{i=1}^{n}\varphi((\tilde{y}_i-d_m) \mod \Lambda), \text{ for } \tilde{y}_i\in\bigcup_{k=0}^{p-1}(d_k+\mathcal{Z}(\Lambda) \mod \Lambda),\end{aligned} \tag{26}$$

where $\tilde{y}_i,\ i=1,\ldots,n$, are the components of $\tilde{\mathbf{y}}$. From (26), we can see that the feasible messages (i.e. with non-null probability) are those whose coset leader is contained in the interval given by $\bigcap_{i=1}^{n}(\tilde{y}_i-d_m-\mathcal{Z}(\Lambda) \mod \Lambda)$, and that the feasible messages are equiprobable. Given the symmetry of the pdf of $\tilde{y}_i|t_i=0$, we can write

$$H(M_1|\tilde{\mathbf{Y}}_1,\mathbf{T}=\mathbf{0}) = E_{Z_i}\left[\log\left(\sum_{k=0}^{p-1}\phi_{\mathcal{H}}((\Delta\cdot k/p) \mod \Lambda)\right)\right], \text{ for } \alpha\geq 0.5, \tag{27}$$

where $\phi_{\mathcal{H}}(\cdot)$ is the indicator function defined in Eq. (1), and

$$\mathcal{H} \triangleq \left[\max_{i=1,\ldots,n}\{Z_i\}-(1-\alpha)\Delta/2,\ \min_{i=1,\ldots,n}\{Z_i\}+(1-\alpha)\Delta/2\right),$$

with $Z_i \sim U((1-\alpha)[-\Delta/2,\Delta/2))$.[3] The expectation (27) is obtained numerically by Monte Carlo integration (averaging over the realizations of Z_i).

2) Computation of the achievable rate for unfair users: The second term in the right hand side of (24) is given by

$$\frac{1}{n}E_{\tilde{\mathbf{Y}}_1,\ldots,\tilde{\mathbf{Y}}_{N_o}}\left[H(M_1,\ldots,M_{N_o}|\tilde{\mathbf{Y}}_1=\tilde{\mathbf{y}}_1,\ldots,\tilde{\mathbf{Y}}_{N_o}=\tilde{\mathbf{y}}_{N_o})\right], \tag{28}$$

The a posteriori probability distribution of the message sequences can be obtained by combining the equations (51) and (55) of Appendix B. For a message sequence $\mathbf{m}^{(k)}=[m_1^k,\ldots,m_{N_o}^k]$,

$$\Pr(m_1^k,\ldots,m_{N_o}^k|\tilde{\mathbf{y}}_1,\ldots,\tilde{\mathbf{y}}_{N_o}) = \Pr(m_1^k,\ldots,m_{N_o}^k)\cdot\frac{f(\tilde{\mathbf{y}}_1,\ldots,\tilde{\mathbf{y}}_{N_o}|m_1^k,\ldots,m_{N_o}^k)}{\sum_{i=1}^{p^{N_o}} f(\tilde{\mathbf{y}}_1,\ldots,\tilde{\mathbf{y}}_{N_o},\mathbf{m}^{(i)})} = \frac{\mathrm{vol}(\mathcal{S}_{N_o}(m_1^k,\ldots,m_{N_o}^k))}{\sum_{i=1}^{p^{N_o}}\mathrm{vol}(\mathcal{S}_{N_o}(\mathbf{m}^{(i)}))}. \tag{29}$$

For $\alpha\geq 0.5$, the feasible regions involved in the calculation of (29) are always modulo-Λ convex hypercubes [8], and as such they can be easily computed componentwise. The entropy (28) is obtained by Monte Carlo, computing the probability of the all the message sequences in a large set of realizations of $\tilde{\mathbf{Y}}_1,\ldots,\tilde{\mathbf{Y}}_{N_o}$. Notice that, although the cardinality of the message space grows exponentially with N_o, only the message sequences with non-null probability need to be taken into account, making the problem computationally feasible.

[3]Hence, in the case of repetition coding the problem of computing $H(M_1|\tilde{\mathbf{Y}}_1,\mathbf{T})$ can be seen as the dual of the problem of computing $h(T|\tilde{Y}_1,\ldots,\tilde{Y}_{N_o},M_1,\ldots,M_{N_o})$ for a scalar lattice, which was addressed in [8, Sect. III-A].

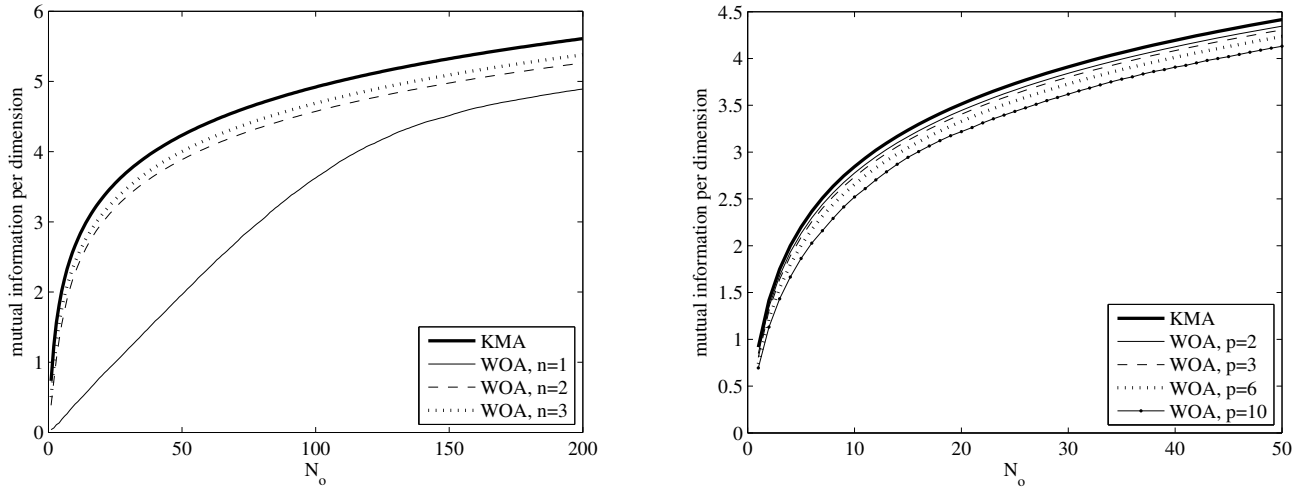
D. Numerical results

The results shown in this section support some of the conclusions drawn in sections III-A and III-B. Fig. 4 illustrates the information leakage about $\mathbf{T}$ for the repetition code and provides a comparison with the results obtained for the KMA scenario. Fig. 4(a) shows the negative impact on the security level of increasing the dimensionality whilst keeping constant the size of the alphabet. Fig. 4(b) shows the security improvement brought about by the increase of the alphabet size, although this improvement is not very significant. Finally, notice in both Fig. 4(a) and Fig. 4(b) that the gap between the information leakage for KMA and WOA tends asymptotically to a constant. For the case of Fig. 4(a), Theorem 1 holds, so the asymptotic value of the gap is actually $\log(p)/n$. In Fig. 4(b), Theorem 1 holds only for $p = 2$, although for the larger values of p the gap can still be seen to be approximately $\log(2)/n$. The gap is more deeply considered in the next paragraph.

Fig. 5 shows the gap function defined in (18) for two instances of the lattice repetition code with different parameters. The code considered in Fig. 5(a) is for $n = 1$ and $p = 2$, which is equivalent to binary SCS [16], the simplest lattice code. By Eq. (7), the watermarked signal observed by the attacker is $\tilde{Y}_k = (d_{M_k} + T + (1-\alpha)N_k) \mod \Lambda$, where N_k is uniform over the interval $[-\Delta/2, \Delta/2)$, with variance $(1-\alpha)^2\Delta^2/12$. For this code, the equivocation for a fair user is always 0 in a noiseless scenario whenever $\alpha \geq 0.5$ [16]. Thus, in this case the gap function is equivalent to the equivocation about the embedded messages for the attacker, i.e. $H(M_1, \ldots, M_{N_o}|\tilde{\mathbf{Y}}_1, \ldots, \tilde{\mathbf{Y}}_{N_o})$. It can be seen that the gap tends in all cases to $\log(2)$ as N_o is increased, as stated in Theorem 1, except in the case $\alpha = 0.5$, where the gap increases indefinitely. The explanation for this behavior is simple: according to the scheme proposed in Proposition 1, the combination $n = 1$, $p = 2$ and $\alpha = 0.5$ provides perfect secrecy about $\mathbf{T}$, so the secret dither cannot be disclosed, and consequently the embedded message sequence cannot be reliably estimated. However, as mentioned in Remark 4 of Sect. III-A, perfect secrecy about $\mathbf{T}$ does not imply perfect secrecy about the message. In fact, if perfect secrecy about the message were achieved, then the gap should increase linearly in N_o, but we can check in Fig. 5(a) that it is far from being the case. Finally, Fig. 5(b) shows the gap for a repetition code with $p = 8$, $n = 10$. Bear in mind that this code provides error-free decoding only if $\alpha \geq 1 - 1/8$. It can be seen that for values of α sufficiently large, the gap nearly approaches $R = \log(8)/10 \approx 0.21$ nats. However, for smaller values of α, the gap becomes slightly larger.

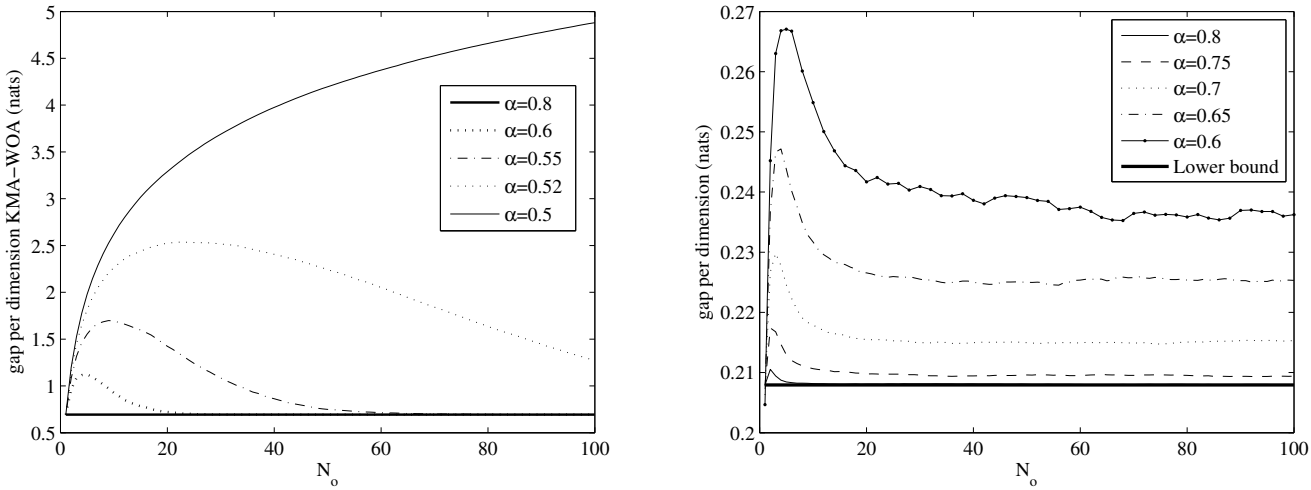
IV. A PRACTICAL DITHER ESTIMATOR

A practical dither estimator for the WOA scenario is proposed in this section. Since the core of the estimation procedure is the estimator devised in [8, Sect. IV], the latter is briefly recalled in Section IV-A. Throughout this section, the secret dither will be assumed to be uniformly distributed in $\mathcal{V}(\Lambda)$. Furthermore, no channel coding across different blocks takes place, so the message sequences will be considered to be a priori equiprobable. Notice that otherwise the attacker could exploit the a priori probability of each message sequence, simplifying the estimation.



(a) $\alpha = 0.52$, $p = 2$ (b) $\alpha = 0.6$, $n = 10$

Fig. 4. Information leakage per dimension for DC-DM with repetition coding. Impact of the repetition rate (a), and of the alphabet size (b).



(a) $p = 2$, $n = 1$ (b) $p = 8$, $n = 10$

Fig. 5. Gap per dimension between KMA and WOA for DC-DM with repetition coding.

A. Set-membership estimator for the KMA scenario

In case the embedded message $\mathbf{m}^{(k)} = (m_1^k, \ldots, m_{N_o}^k)$ is known by the attacker, the Maximum Likelihood (ML) estimate of the secret dither given N_o observations is given by any point in $\mathcal{S}_{N_o}(\mathbf{m}^{(k)})$, the feasible region defined in (11). The estimation algorithm proposed in [8] works by computing $\mathcal{S}_{N_o}(\mathbf{m}^{(k)})$ and picking the center of this region as the estimate of $\mathbf{t}$.[4] The only working assumption is that $\alpha \geq 0.5$, in order to guarantee the modulo-Λ convexity of the intersections [8, Sect. III]. After adding a certain offset to the observations, the feasible region defined by the ith observation (Eq. (11)) is redefined as

$$\mathcal{D}_i(m_i^k) \triangleq \tilde{\mathbf{v}}_i(m_i^k) + \mathcal{Z}(\Lambda),\ i = 1, \ldots, N_o, \quad \text{where} \quad \tilde{\mathbf{v}}_i(m_i^k) \triangleq (\tilde{\mathbf{y}}_i - \mathbf{d}_{m_i^k} - \tilde{\mathbf{y}}_1 + \mathbf{d}_{m_1^k}) \mod \Lambda. \tag{30}$$

[4]Under the flat-host assumption, this estimate is equivalent to the conditional mean estimator, and thus it minimizes the mean squared error between the estimate and the secret dither vector [8].

Notice that in (30) we make explicit the dependence of $\tilde{\mathbf{v}}_i$ with the embedded message m_i^k. The aim of introducing the offset $-\tilde{\mathbf{y}}_1+\mathbf{d}_{m_1^k}$ in every observation is to get a convex $\mathcal{S}_{N_o}(\mathbf{m}^{(k)})$ for all k, as discussed in [8]. Obviously, this offset must be removed from the final dither estimate. Instead of directly computing $\mathcal{S}_{N_o}(\mathbf{m}^{(k)})$, the estimation algorithm computes an outer bound using the "inner polytope" or the "optimal volume ellipsoid" (OVE) method. In this paper we will consider only the former method, where $\mathcal{S}_{N_o}(\mathbf{m}^{(k)})$ is described by means of a set of linear inequalities which are the inputs to an optimization algorithm that computes an ellipsoid $\mathcal{E}_{N_o}(\mathbf{m}^{(k)})$ that bounds $\mathcal{S}_{N_o}(\mathbf{m}^{(k)})$. This ellipsoid is completely defined by a vector $\mathbf{c} \in \mathbb{R}^n$ and a positive-definite matrix $\mathbf{P} \in \mathbb{R}^{n\times n}$. Thus, the ellipsoids allow us to describe the feasible regions with a reduced and constant number of parameters, independently of the complexity of the underlying region.

B. Joint Bayesian and set-membership estimation for the WOA scenario

The ML estimate of the secret dither in the WOA scenario can be expressed as

$$\hat{\mathbf{t}}_{ML} = \arg \max_{\mathbf{t}\in\mathcal{V}(\Lambda)} f(\tilde{\mathbf{y}}_1,\ldots,\tilde{\mathbf{y}}_{N_o}|\mathbf{T}=\mathbf{t}) = \arg \max_{\mathbf{t}\in\mathcal{V}(\Lambda)} \sum_{k=1}^{p^{N_o}} f(\tilde{\mathbf{y}}_1,\ldots,\tilde{\mathbf{y}}_{N_o},\mathbf{m}^{(k)}|\mathbf{T}=\mathbf{t}). \tag{31}$$

Estimation based on Eq. (31) is impractical due to the number of summation terms, which is exponentially increasing with the number of observations. In order to keep an affordable complexity for the estimation algorithm we resort to the usual "Viterbi approximation", where the summation in (31) is approximated by the value of the maximum term. This way, approximate ML estimation of the dither amounts to estimating the most probable message sequence, and then performing dither estimation as in Sect. IV-A using the estimate of the embedded message. Mathematically, it can be written as

$$\hat{\mathbf{t}} = \arg \max_{\mathbf{t}\in\mathcal{V}(\Lambda)} f(\tilde{\mathbf{y}}_1,\ldots,\tilde{\mathbf{y}}_{N_o},\mathbf{m}^{(\hat{k})}|\mathbf{t}), \text{ with } \hat{k} = \arg \max_{k=1,\ldots,p^{N_o}} f(\tilde{\mathbf{y}}_1,\ldots,\tilde{\mathbf{y}}_{N_o}|\mathbf{m}^{(k)}). \tag{32}$$

At this point we recall that the a priori path space $\mathcal{M}^{N_o}$ is divided in equivalence classes (with p elements each) defined by the relation (19). Since the paths belonging to the same equivalence class have the same a posteriori probability (according to Lemma 2), the path $\mathbf{m}^{(\hat{k})}$ in (32) is not unique. In order to get rid of this ambiguity, we will reduce the search space to one representative per equivalence class. At the same time, this strategy reduces the cardinality of the search space by a factor p. Notice that this complexity reduction does not imply any loss in performance, since the whole set of feasible paths can be recovered from the set of equivalence classes.

It is important to clarify that the two-stage estimator just defined is suboptimal, in general, since its performance is subject to the correct estimate of the embedded message. However, it is not the final purpose of the security analysis to present the best estimator, but rather to show that the information leakages identified in the theoretical part can be exploited in practice. As for the estimator presented here, if α fulfills the condition imposed in Theorem 1, then no loss of optimality is incurred for large N_o. The reason is that for sufficiently large N_o there exist only p equiprobable feasible message sequences (which belong to the same equivalence class), as explained in the proof of the theorem.

Since we are restricting the search to one representative per equivalence class, it is clear that the summation (31) will equal the value of the maximum term.

Hereinafter, we will use the term "path" for denoting each message sequence $\mathbf{m}^{(k)} = [m_1^k, \ldots, m_{N_o}^k],\ k = 1, \ldots, p^{N_o}$. From App. B, we know that the a posteriori probability of the observations given $\mathbf{m}^{(k)}$ is given by

$$f(\tilde{\mathbf{y}}_1, \ldots, \tilde{\mathbf{y}}_{N_o} | \mathbf{m}^{(k)}) = \frac{\text{vol}(\mathcal{S}_{N_o}(\mathbf{m}^{(k)}))}{(\text{vol}(\mathcal{Z}(\Lambda)))^{N_o} \cdot \text{vol}(\mathcal{V}(\Lambda))}. \tag{33}$$

The only term of (33) that depends on the hypothesized path is $\mathcal{S}_{N_o}(\mathbf{m}^{(k)})$. This implies that, in practical terms, the most probable paths are those with the largest feasible region. Hence, we can define the "score" of the path $\mathbf{m}^{(k)}$ as

$$\lambda(\mathbf{m}^{(k)}) \triangleq \text{vol}(\mathcal{S}_{N_o}(\mathbf{m}^{(k)})). \tag{34}$$

which can be used to compare the probabilities of different paths as long as they have the same length. It follows that, given N_o observations, the ML estimate of the most probable path is simply given by $\mathbf{m}^{(\hat{k})}$, where $\hat{k} = \arg\max_k \lambda(\mathbf{m}^{(k)})$. The search for the most probable path can be carried out by means of a tree search where each branch of the tree represents a hypothesized path with an associated score. For saving computational resources, the outer bound introduced in Section IV-A and a "beam search" strategy [18, Chapt. 12] will be applied during the tree search. The steps of the proposed dither estimation algorithm are summarized in Algorithm 1 below. The input data are the observations $\{\tilde{\mathbf{y}}_i,\ i = 1 \ldots, N_o\}$ and the parameters of the nested lattice code.

Algorithm 1: Secret dither estimation

1) Initialization: $\mathbf{m}^{(1)} = 0$, $\mathcal{D}_1(\mathbf{m}^{(1)}) = (1-\alpha)\mathcal{V}(\Lambda)$, $K_1 = 1$, with K_1 denoting the number of feasible paths for the first observation (1 in our case).
2) For $i = 2, \ldots, N_o$
   a) Let $\{\mathbf{m}^{(k)},\ k = 1, \ldots, K_{i-1}\}$ be the set of feasible paths for the $i-1$ first observations. Construct a set of candidate paths as $\{\mathbf{m}^{(k,l)} \triangleq [\mathbf{m}^{(k)},\ l],\ k = 1, \ldots, K_{i-1},\ l = 0, \ldots, p-1\}$.
   b) Compute the ellipsoids $\mathcal{E}_i(\mathbf{m}^{(k,l)}) \supseteq \mathcal{S}_i(\mathbf{m}^{(k,l)})$ using $\tilde{\mathbf{v}}_r(m_r^{(k,l)}) = (\tilde{\mathbf{y}}_r - \mathbf{d}_{m_r^{(k,l)}} - \tilde{\mathbf{y}}_1) \mod \Lambda,\ r = 1, \ldots, i$, where $m_r^{(k,l)}$ denotes the rth element of $\mathbf{m}^{(k,l)}$.
   c) Compute the score $\lambda(\mathbf{m}^{(k,l)})$ of each path as $\text{vol}(\mathcal{E}_i(\mathbf{m}^{(k,l)}))$. Arrange the paths in order of descending score as $\mathbf{m}^{(1)}, \ldots, \mathbf{m}^{(p \cdot K_{i-1})}$. Compute now $K_i \triangleq \max_q q + 1$, subject to the constraint $\lambda(\mathbf{m}^{(1)})/\lambda(\mathbf{m}^{(q)}) < \beta$, $q = 1, \ldots, p \cdot K_{i-1} \leq K_{max}$. The parameters $\beta > 0$, $K_{max} \in \mathbb{N}^+$ are termed "beam factors". The set of $K_i < K_{max}$ "surviving paths" for the next iteration is $\{\mathbf{m}^{(1)}, \ldots, \mathbf{m}^{(K_i - 1)}\}$.
3) Let $\mathbf{m}^{(1)}$ be the path with the highest score resulting from Step 2, and let $\hat{\mathbf{t}}_1$, defined as the center of $\mathcal{E}_{N_o}(\mathbf{m}^{(1)})$, the dither estimate associated to $\mathbf{m}^{(1)}$ (this choice minimizes the mean squared error of the estimate [8]). The p paths belonging to the equivalence class $[\mathbf{m}^{(1)}]$ can be computed according to (19), and the p corresponding dither estimates are given by

$$\hat{\mathbf{t}}_k = (\hat{\mathbf{t}}_1 - \mathbf{d}_k + \tilde{\mathbf{y}}_1) \mod \Lambda,\ k \in \mathcal{M}. \tag{35}$$

Note that $\tilde{\mathbf{y}}_1$ is added in order to cancel the offset introduced in Step 2b.

Remark 1: Through the variation of the "beam factor" β one can control the tradeoff between computational complexity and accuracy. If $\beta = 1$, only the most probable path is retained in each iteration of Algorithm 1. Hence, complexity reaches its minimum for $\beta = 1$, but the probability of missing the correct path may be very high. As β is increased, the number of surviving paths per iteration increases, in general. In the case $\beta \to \infty$, all the paths are retained in each iteration, making the complexity unaffordable, in general, but reducing to 0 the probability of missing the correct path. The impact of varying β is shown in Sect. IV-C1. Notice that β does not limit the absolute number of paths to be considered in each iteration. This is why the parameter K_{max} is also introduced in Step 2.c, for limiting the complexity in absolute terms.

Remark 2: It is possible that Step 2.c results in $K_i = 0$. If this is the case, then it means that the true path has been discarded at some previous iteration of the algorithm due to too restrictive beam factors β and/or K_{max}. If this happens, Algorithm 1 must be restarted after increasing the values of the beam factors.

Remark 3: The outer bounding of the feasible regions may impact negatively the estimator performance, due to the introduction of spurious paths and variation of the scores.

Remark 4: When the shaping lattice $\mathcal{V}(\Lambda)$ is cubic, the feasible regions are hyperrectangles. In such case, they can be easily computed componentwise, since they are simply defined by n real segments. Thus, there is no need to apply the inner polytope algorithm when the shaping lattice is cubic.

The computation of all the outer bounding ellipsoids in Step 2.b is the most time-consuming task of the estimation algorithm. Clearly, this step can be sped up if we can use a fast algorithm to discard the "unfeasible" paths. Formally, a certain path $\mathbf{m}^{(k)},\ k = 1, \ldots, p^{N_o}$, is said to be "unfeasible" or "inconsistent" with the observations if the associated feasible region $\mathcal{S}_{N_o}(\mathbf{m}^{(k)})$ is an empty set; otherwise, the path is said to be "feasible" or "consistent". That is, the unfeasible paths are those that yield a null score (i.e. null a posteriori probability), so it is not worth keeping them for the next iteration. Thus, Step 2.b in Algorithm 1 is broken down in two steps: Step 2.b.i, that checks the feasibility of the candidate paths, and Step 2.b.ii, which is the same as the original 2.b of Algorithm 1, but computing only the feasible region of the feasible paths.

For the $i-1$ first observations, we have the pairs $\{\mathbf{m}^{(k)}, \mathcal{E}_{i-1}(\mathbf{m}^{(k)}),\ k = 1, \ldots, K_{i-1}\}$. In order to check the feasibility of a certain candidate path $\mathbf{m}^{(k,l)},\ k = 1, \ldots, K_{i-1},\ l = 0, \ldots, p-1$, we need to check whether the intersection $\mathcal{E}_{i-1}(\mathbf{m}^{(k)}) \bigcap \mathcal{D}_i(l)$ is empty or not. To this end, we have used an algorithm based on the OVE algorithm proposed in [19], which is described below.

Algorithm 2: Check unfeasible paths

According to [8, Sect. IV], assume the feasible region for the ith observation can be specified by a matrix $\mathbf{\Phi} \in \mathbb{R}^{n \times n_f/2}$ and a vector $\boldsymbol{\gamma} \in \mathbb{R}^{n_f/2 \times 1}$ such that $\mathcal{D}_i(m_i^{(k,l)}) = \bigcap_{j=1}^{n_f/2} \mathcal{F}_{i,j}$, where

$$\mathcal{F}_{i,j} = \{\mathbf{z} \in \mathbb{R}^n : |\tilde{\mathbf{v}}_i(m_i^{(k,l)})^T \boldsymbol{\phi}_j - \mathbf{z}^T \boldsymbol{\phi}_j| \leq \gamma_j\}, \tag{36}$$

being $\boldsymbol{\phi}_j$ the jth column of $\boldsymbol{\Phi}$, $\gamma_j \triangleq \boldsymbol{\phi}_j^T \mathbf{z}_{0,j}$ is the jth element of $\boldsymbol{\gamma}$, and $\mathbf{z}_{0,j}$ is a vector in the jth facet of $\mathcal{Z}(\Lambda)$. For $k = 1, \ldots, K_i$, and $l = 0, \ldots, p-1$,

1) Compute
$$\eta_j = \frac{\tilde{\mathbf{v}}_i(m_i^{(k,l)}) + \gamma_j - \boldsymbol{\phi}_j^T \mathbf{c}_{i-1}}{\sqrt{\boldsymbol{\phi}_j^T \mathbf{P}_{i-1} \boldsymbol{\phi}_j}}, \qquad \zeta_j = \frac{\gamma_j}{\sqrt{\boldsymbol{\phi}_j^T \mathbf{P}_{i-1} \boldsymbol{\phi}_j}}, \quad j = 1, \ldots, n_f/2, \tag{37}$$
where $\mathbf{P}_{i-1}$ and $\mathbf{c}_{i-1}$ are the positive-definite matrix and center defining the ellipsoid $\mathcal{E}_{i-1}(\mathbf{m}^{(k)})$.
2) If $\eta_j \notin [-1, 1 + 2\zeta_j]$, for some $j = 1, \ldots, n_f/2$, then the hypothesized path $\mathbf{m}^{(k,l)}$ is unfeasible. Otherwise, the path is declared as feasible.[5]

C. Experimental results

This section presents the results of applying Algorithm 1 over some practical lattice DC-DM schemes. The experiments have been carried out under the following assumptions: the host signals follow a i.i.d. Gaussian distribution with zero mean and variance $\sigma_X^2 = 10$, and the DWR is 30 dB in all cases (DWR $\triangleq 10 \log_{10}(\sigma_X^2 / D_w)$); the embedded messages are equiprobable (i.e., no coding is applied along different blocks), and the attacker knows the parameters of the nested lattice code being used, as stated in Section II. In order to assess the performance of the dither estimator without ambiguities (due to the result of Lemma 2), it is assumed that the message conveyed by the first observation corresponds to the symbol 0. The parameter K_{max} of Algorithm 1 has been set to 250 in all cases.

1) Tradeoff complexity-accuracy: One interesting performance measure is the resulting probability of decoding error when the decoder uses the dither estimate, instead of the true dither vector. If this measure is represented in terms of the "beam factor" (β) of Algorithm 1, then the tradeoff between complexity and accuracy becomes patent. This tradeoff is illustrated in Fig. 6(a) for a cubic shaping lattice and repetition coding (see Section III-C) with $n = 10$ and $p = 6$. Using dither estimates obtained with $N_o = 100$ observations, the numerically computed symbol error rate (SER) is shown in Fig. 6(a) for different values of α. For reference, Fig. 6(a) also shows in dashed lines the SER obtained by a fair decoder, i.e. knowing the true dither signal. As can be seen, the SER is always decreased as β is increased, achieving the same decoding performance as the fair decoder in the cases of $\alpha = 0.6$ and $\alpha = 0.7$. However, for $\alpha = 0.5$, the SER of the unfair decoder is slightly larger. The reason is that in this case the probability of deciding a wrong dither is not negligible even for high β.

2) Estimation error: Now we measure the performance of the estimator in terms of the mean squared error (MSE) per dimension between the dither estimate and the actual dither vector. Three different shaping lattices have been considered. In all cases, a beam factor $\beta = 45$ dB has been used.

Fig. 6(b) shows the results obtained for a scheme using a cubic shaping lattice in 10 dimensions and repetition coding with $\alpha = 0.6$. It can be seen that for $p = 4$ it is still possible to attain the same accuracy as in the KMA

[5]Notice that, due to the outer bounding, we may label as feasible some paths that are actually unfeasible. However, it is important to realize that the converse never occurs.

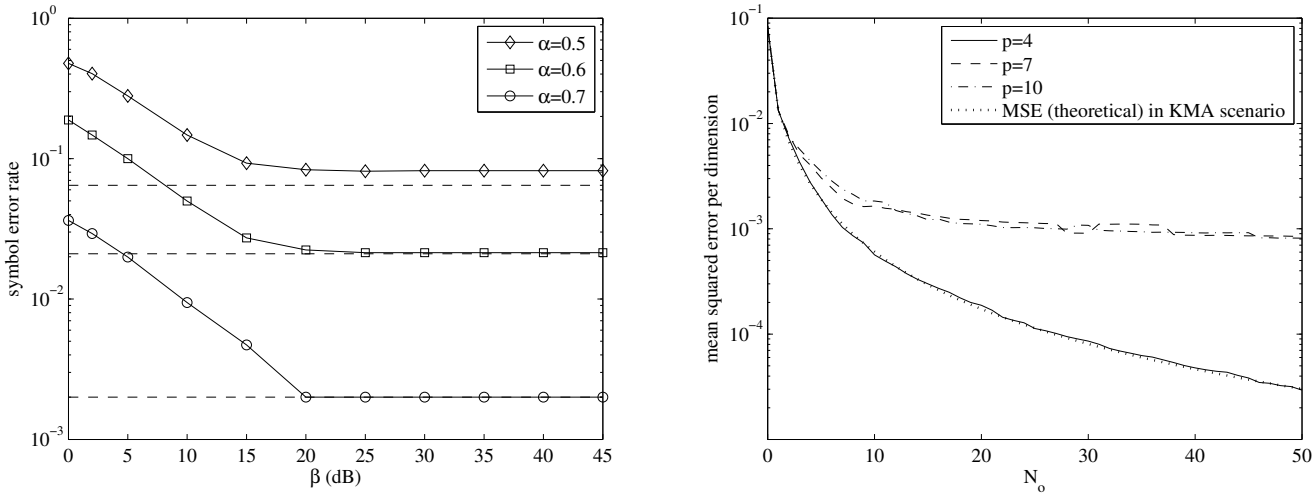


Fig. 6. Estimation results for a cubic lattice with repetition coding , with $n = 10$ and DWR = 30 dB. Figure (a) shows the symbol error rate for a dither estimated with $N_o = 100$ observations versus the beam factor β in dB, for $p = 6$. Figure (b) shows the MSE per dimension in the dither estimation obtained for different embedding rates and $\alpha = 0.6$.

scenario, whereas for $p = 7$ and $p = 10$ a significant degradation of the MSE is observed. This degradation is a consequence of the fact that, as p is increased, the probability of correctly retrieving the embedded path decreases when α is kept constant (even knowing the true value of the dither).[6] In the experiments, the probability of choosing an incorrect path has been found to be around 0.05 and 0.1 for $p = 7$ and $p = 10$, respectively.

Fig. 7(a) shows the results obtained for a hexagonal shaping lattice and $\alpha = 0.7$. Notice that, although α is higher than in the former case, the maximum embedding rate considered now is substantially larger ($\frac{1}{2}\log_2(9)$ bits vs. $\frac{1}{10}\log_2(10)$ bits). Similarly as above, we can see that increasing p degrades the MSE: for $p = 4$ it is still possible to achieve the same accuracy as in the KMA scenario, but for $p = 4$ the MSE is increased, and for $p = 9$ the MSE is not reduced with N_o. Finally, Fig. 7(b) shows the results obtained for the E_8 shaping lattice [11], the best lattice quantizer in 8 dimensions. It can be seen that in this case, even with large alphabets (e.g. $p = 12$), the estimator achieves its optimal performance. The results of Fig. 7 correspond to lattice codes obtained through Construction A. The inner polytope algorithm has been used for computing the approximate feasible regions.

D. Reversibility attack

An accurate dither estimate (subjected to an unknown modulo-Λ shift, as the one obtained here) allows to implement a number of harmful attacks. As an illustrative example, we present here a reversibility attack, consisting in producing an estimate of the original host signal. In the context of lattice-data hiding methods, our attack is based on the fact that the embedding function is reversible whenever $\alpha < 1$ and we know both $\mathbf{T}$ and the embedded message. Using

[6]Recall that, due to the Viterbi approximation, the performance of the estimator is degraded when the embedded message cannot be correctly decoded, as explained in IV-B.

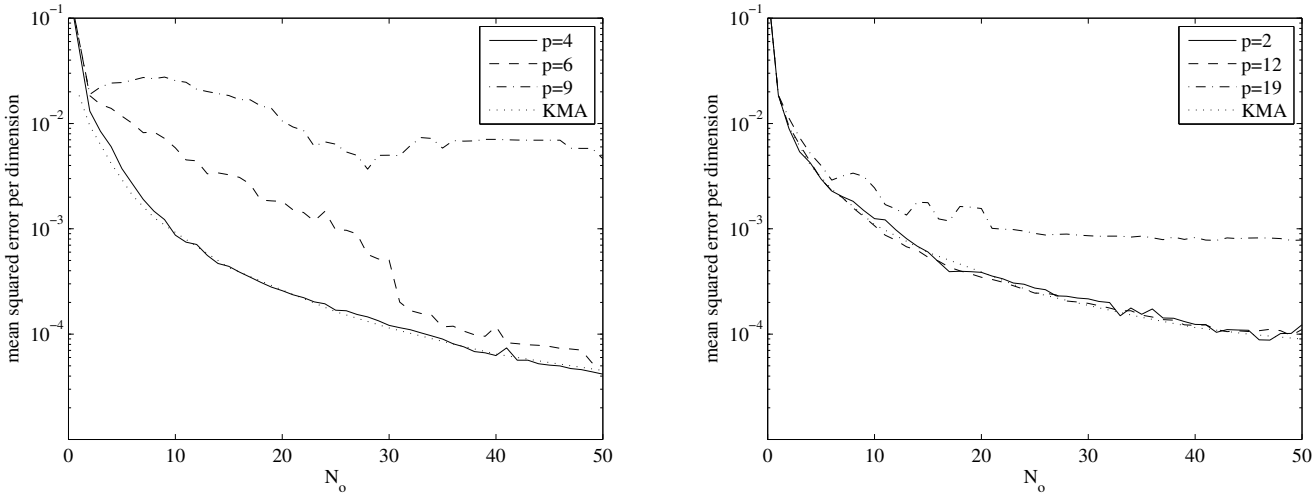


Fig. 7. MSE per dimension for $\alpha = 0.7$ and different embedding rates. Results for $n = 2$ and $n = 8$ using hexagonal (a) and E_8 (b) shaping lattices, respectively. DWR = 30 dB in both cases.

our dither and path estimates $\hat{\mathbf{t}}$, $\mathbf{m}^{(\hat{k})}$, the host estimate corresponding to the ith watermarked block is computed as[7]

$$\hat{\mathbf{x}}_i = \mathbf{y}_i - \frac{\alpha}{1-\alpha}(Q_\Lambda(\mathbf{y}_i - \mathbf{d}_{m_i^{\hat{k}}} - \hat{\mathbf{t}}) - \mathbf{x}_i + \mathbf{d}_{m_i^{\hat{k}}} + \hat{\mathbf{t}}). \tag{38}$$

It is interesting to notice that the ambiguity in the estimated message does not affect negatively the host estimation whenever the estimated path $\mathbf{m}^{(\hat{k})}$ belongs to the equivalence class (defined in (19)) of the actual embedded path. The reason is that the dither estimate associated to any path in $[\mathbf{m}]$ yields the same fine lattice Λ_f, and thus it is valid for performing a successful reversibility attack. This can be readily seen if we realize that the dither estimates associated to the equivalence class of $\mathbf{m}^{(\hat{k})}$, given by (35), differ only in the term $\mathbf{d}_k$, and that $\Lambda_f = \Lambda_f - \mathbf{d}_k$, for $k \in \mathcal{M}$.

Fig. 8 shows the result of implementing the proposed reversibility attack on a real watermarked image. The parameters of the watermarking algorithm are $\Lambda = E_8$, $\alpha = 0.7$, $p = 10$, and the coset leaders were obtained through Construction A. The watermark is embedded in the low frequency coefficients of 8×8 non-overlapping DCT blocks, resulting in a PSNR after embedding of 38.2 dB. The resulting host estimate, using only the message and dither estimates from the first 50 DCT blocks, is shown in Fig. 8(b) and presents a PSNR of approximately 56 dB. If the estimated host is quantized back to integers, then PSNR $\rightarrow \infty$, meaning that the host has been estimated perfectly.

V. Conclusions

We have presented in this paper an investigation of the security provided by data hiding schemes based on nested lattice codes randomized by means of secret dithering. Although it has been shown that it is theoretically possible to achieve perfect secrecy, the security level of many practical scenarios (i.e., simple shaping lattices, low embedding rates, etc.) can be fairly low. In fact, the security weaknesses of the data hiding schemes studied in this paper have been

[7]The same reversibility function had been proposed in [16] in the context of scalar lattices.

(a) (b)

Fig. 8. Illustration of a reversibility attack based on dither estimate according to Eq. (38). Image watermarked using $\Lambda = E_8$, $\alpha = 0.7$, $p = 10$ and PSNR = 38.2 dB (a), and estimate of the original image with PSNR = 55.9 dB (b).

shown to be exploitable in practice with affordable complexity, allowing for instance to obtain host estimates with high fidelity. One obvious strategy for minimizing security risks is to reuse the secret key as few times as possible, but this may introduce serious synchronization problems. Another strategy is to look for more secure forms of randomization or choosing the embedding parameters that maximize the security: in general, the information leakage about the secret dither can be reduced by increasing the embedding rate or decreasing α, but this solution demands for more powerful error correcting codes (ECC) if one wants to guarantee reliable transmission. A possible drawback, as noted in this paper, is that the use of ECCs may introduce statistical dependence between different observations that could be exploited by an attacker, particularly for simple ECCs. The complexity of exploiting the information leakage provided by ECCs deserves further study in the future.

ACKNOWLEDGEMENTS

The authors want to thank one anonymous reviewer and to Dr. Kalker, whose comments helped to improve significantly the present manuscript, and to Dr. Pedro Comesaña for his useful comments on the proof of Lemma 1.

APPENDIX A
PROOF OF LEMMA 1

First, one must realize that if the assumption of independence between embedded messages holds, then null information leakage for one observation implies perfect secrecy for all N_o. Thus, the proof of perfect secrecy can be reduced to show that the condition $I(\tilde{\mathbf{Y}}_1; \mathbf{T}) = h(\tilde{\mathbf{Y}}_1) - h(\tilde{\mathbf{Y}}_1|\mathbf{T}) = 0$ is fulfilled.

With a little abuse of notation, in this appendix we will denote the pdf of $\tilde{\mathbf{y}}_1$ conditioned on $\mathbf{t}$ by $f_p(\tilde{\mathbf{y}}_1|\mathbf{t})$, for making clear the dependence with p, the alphabet size. We will first consider the term $h(\tilde{\mathbf{Y}}_1|\mathbf{T})$. Due to the flat-host assumption, we have $h(\tilde{\mathbf{Y}}_1|\mathbf{T}) = h(\tilde{\mathbf{Y}}_1|\mathbf{T}=\mathbf{t})$ for any $\mathbf{t}$. Hence, we need to calculate

$$\lim_{p\to\infty} h(\tilde{\mathbf{Y}}_1|\mathbf{T}=\mathbf{t}) = -\lim_{p\to\infty}\int_{\mathcal{V}(\Lambda)} f_p(\tilde{\mathbf{y}}_1|\mathbf{t})\log(f_p(\tilde{\mathbf{y}}_1|\mathbf{t}))d\tilde{\mathbf{y}}_1. \tag{39}$$

Computation of the integral in (39) is unaffordable. However, by virtue of the bounded convergence theorem [20, Chapt. 4], integral sign and limit can be interchanged if the integrand converges pointwise and

$$|f_p(\tilde{\mathbf{y}}_1|\mathbf{t})\log(f_p(\tilde{\mathbf{y}}_1|\mathbf{t}))| \leq g(\tilde{\mathbf{y}}_1)\ \forall\ p, \tag{40}$$

where $g(\tilde{\mathbf{y}}_1)$ is any function such that $\int_{\mathcal{V}(\Lambda)}|g(\tilde{\mathbf{y}}_1)|d\tilde{\mathbf{y}}_1 < \infty$. In our case it suffices to choose a constant function $g(\tilde{\mathbf{y}}_1) = \text{vol}(\mathcal{Z}(\Lambda))^{-1}\log(\text{vol}(\mathcal{Z}(\Lambda))^{-1})\ \forall\ \tilde{\mathbf{y}}_1 \in \mathcal{V}(\Lambda)$. The integral of $|g(\tilde{\mathbf{y}}_1)|$ for this choice is finite, since $\alpha < 1$ by assumption. Thus, (39) can be computed as

$$\lim_{p\to\infty} h(\tilde{\mathbf{Y}}_1|\mathbf{T}=\mathbf{t}) \quad = \quad -\int_{\mathcal{V}(\Lambda)}\lim_{p\to\infty} f_p(\tilde{\mathbf{y}}_1|\mathbf{t})\log(f_p(\tilde{\mathbf{y}}_1|\mathbf{t}))d\tilde{\mathbf{y}}_1. \tag{41}$$

We turn, for the moment, our attention to the computation of the limit of the conditioned pdf. The pdf of $\tilde{\mathbf{y}}_1$ conditioned on $\mathbf{t}$ is given by

$$f_p(\tilde{\mathbf{y}}_1|\mathbf{t}) = \frac{1}{p}\sum_{k=0}^{p-1}\varphi(\tilde{\mathbf{y}}_1 - \mathbf{t} - \mathbf{d}_k \mod \Lambda) = \frac{1}{\text{vol}(\mathcal{V}(\Lambda))}\sum_{k=0}^{p-1}\varphi(\tilde{\mathbf{y}}_1 - \mathbf{t} - \mathbf{d}_k \mod \Lambda)\cdot\text{vol}(\mathcal{V}(\Lambda_f)), \tag{42}$$

where the second equality follows from the definition of nesting ratio (cf. Sect. II-A). By recalling the definition of the function $\varphi(\mathbf{x})$ in (8), Eq. (42) can be further rewritten as

$$f_p(\tilde{\mathbf{y}}_1|\mathbf{t}) = \frac{1}{\text{vol}(\mathcal{V}(\Lambda))}\cdot\frac{1}{\text{vol}(\mathcal{Z}(\Lambda))}\sum_{k=0}^{p-1}\phi_{\mathcal{Z}(\Lambda)}(\tilde{\mathbf{y}}_1 - \mathbf{t} - \mathbf{d}_k \mod \Lambda)\cdot\text{vol}(\mathcal{V}(\Lambda_f)), \tag{43}$$

where $\mathcal{Z}(\Lambda) = (1-\alpha)\mathcal{V}(\Lambda)$. Let us analyze the sum in (43). Each term is given by

$$\phi_{\mathcal{Z}(\Lambda)}(\tilde{\mathbf{y}}_1 - \mathbf{t} - \mathbf{d}_k \mod \Lambda) = \begin{cases} 1, & \text{for } \mathbf{d}_k \in (\tilde{\mathbf{y}}_1 - \mathbf{t} - \mathcal{Z}(\Lambda)) \mod \Lambda \\ 0, & \text{elsewhere} \end{cases} \tag{44}$$

Hence, the sum $\sum_{k=0}^{p-1}\phi_{\mathcal{Z}(\Lambda)}(\tilde{\mathbf{y}}_1 - \mathbf{t} - \mathbf{d}_k \mod \Lambda)$ is equivalent to counting the points $\mathbf{d}_k \in \mathcal{D}_p$ that fall inside the region $(\tilde{\mathbf{y}}_1 - \mathbf{t} - \mathcal{Z}(\Lambda)) \mod \Lambda$. Let us denote by $\mathcal{F}$ the subset of points of $\mathcal{D}_p$ for which the indicator function (44) is nonnull. The set $\mathcal{F}$ induces a partition (according to Λ_f) of the region $(\tilde{\mathbf{y}}_1 - \mathbf{t} - \mathcal{Z}(\Lambda) \mod \Lambda$. Since each term in the sum is multiplied by $\text{vol}(\mathcal{V}(\Lambda_f))$, the result is an approximation of the volume of $-\mathcal{Z}(\Lambda)$ (which is equal to $\mathcal{Z}(\Lambda)$ up to a set of measure zero), i.e.

$$\sum_{k=0}^{p-1}\phi_{\mathcal{Z}(\Lambda)}(\tilde{\mathbf{y}}_1 - \mathbf{t} - \mathbf{d}_k \mod \Lambda)\cdot\text{vol}(\mathcal{V}(\Lambda_f)) = \text{vol}(\mathcal{Z}(\Lambda)) + \varepsilon(p), \tag{45}$$

where $\varepsilon(p)$ represents the discrepancy between $\text{vol}(\mathcal{Z}(\Lambda))$ and the value of the sum, which comes from the points of $\mathcal{F}$ close to the boundary of $\mathcal{Z}(\Lambda)$. Recall that, by definition of covering radius (cf. Eq. (13)),

$$\mathcal{V}(\Lambda_f) \subset \mathcal{B}(\mathbf{0}, r_c(\Lambda_f)), \tag{46}$$

where $\mathcal{B}(\mathbf{0}, r_c(\Lambda_f))$ is the n-dimensional closed ball of radius $r_c(\Lambda_f)$. Hence, if $r_c(\Lambda_f) \to 0$ as $p \to \infty$, the term $\varepsilon(p)$ goes to 0 as well, and we have the definition of n-dimensional Riemann integral:[8]

$$\lim_{p\to\infty} \sum_{k=0}^{p-1} \phi_{\mathcal{Z}(\Lambda)}(\tilde{\mathbf{y}}_1 - \mathbf{t} - \mathbf{d}_k \mod \Lambda) \cdot \text{vol}(\mathcal{V}(\Lambda_f)) = \int_{\mathcal{Z}(\Lambda)} d\mathbf{d} = \text{vol}(\mathcal{Z}(\Lambda)),\ \forall\, \tilde{\mathbf{y}}_1 \in \mathcal{V}(\Lambda). \tag{47}$$

Finally, by substituting (47) into (43) we arrive at

$$\lim_{p\to\infty} f_p(\tilde{\mathbf{y}}_1|\mathbf{t}) = \frac{1}{\text{vol}(\mathcal{V}(\Lambda))},\ \forall\, \tilde{\mathbf{y}}_1 \in \mathcal{V}(\Lambda). \tag{48}$$

In order to compute the limit in (41), we observe that the function $x\log(x)$ is continuous in $[0, Q]$, with finite Q, so it is uniformly continuous on that interval. Therefore, (41) can be computed as

$$\lim_{p\to\infty} h(\tilde{\mathbf{Y}}_1|\mathbf{T}=\mathbf{t}) \;=\; -\int_{\mathcal{V}(\Lambda)} \text{vol}(\mathcal{V}(\Lambda))^{-1} \log(\text{vol}(\mathcal{V}(\Lambda))^{-1} d\tilde{\mathbf{y}}_1 = \log(\text{vol}(\mathcal{V}(\Lambda))). \tag{49}$$

As for the other term involved in the mutual information, $h(\tilde{\mathbf{Y}}_1)$, we know that the entropy of a continuous random variable with bounded support can be upper bounded by the log-volume of its support set. Thus, we can write

$$h(\tilde{\mathbf{Y}}_1|\mathbf{T}) \le h(\tilde{\mathbf{Y}}_1) \le \log(\text{vol}(\mathcal{V}(\Lambda))). \tag{50}$$

Since $\lim_{p\to\infty} h(\tilde{\mathbf{Y}}_1|\mathbf{T}) = \log(\text{vol}(\mathcal{V}(\Lambda)))$ it is immediate, by (49) and (50), that $\lim_{p\to\infty} h(\tilde{\mathbf{Y}}_1) = \log(\text{vol}(\mathcal{V}(\Lambda)))$, fulfilling the condition of perfect secrecy regardless the distribution of $\mathbf{T}$, and Lemma 1 follows.

APPENDIX B
A POSTERIORI PROBABILITY OF THE MESSAGE SEQUENCES

In order to compute the a posteriori probability of a message sequence $\mathbf{m}^{(i)} = [m_1^i, \ldots, m_{N_o}^i]$ (hereinafter, a "path"), this probability is first rewritten using Bayes' rule:

$$\Pr(m_1^i, \ldots, m_{N_o}^i | \tilde{\mathbf{y}}_1, \ldots, \tilde{\mathbf{y}}_{N_o}) = \frac{f(\tilde{\mathbf{y}}_1, \ldots, \tilde{\mathbf{y}}_{N_o}|m_1^i, \ldots, m_{N_o}^i) \cdot \Pr(m_1^i, \ldots, m_{N_o}^i)}{f(\tilde{\mathbf{y}}_1, \ldots, \tilde{\mathbf{y}}_{N_o})}. \tag{51}$$

We will focus on the probability a posteriori of the observations, which can be written as

$$\begin{aligned} f(\tilde{\mathbf{y}}_1, \ldots, \tilde{\mathbf{y}}_{N_o}|m_1^i, \ldots, m_{N_o}^i) &= \int_{\mathcal{V}(\Lambda)} f(\tilde{\mathbf{y}}_1, \ldots, \tilde{\mathbf{y}}_{N_o}|m_1^i, \ldots, m_{N_o}^i, \mathbf{t}) \cdot f(\mathbf{t}) d\mathbf{t} \\ &= \int_{\mathcal{V}(\Lambda)} f(\tilde{\mathbf{y}}_{N_o}|m_{N_o}^i, \mathbf{t}) \cdot f(\tilde{\mathbf{y}}_1, \ldots, \tilde{\mathbf{y}}_{N_o-1}|m_1^i, \ldots, m_{N_o-1}^i, \mathbf{t}) \cdot f(\mathbf{t}) d\mathbf{t}, \end{aligned} \tag{52}$$

where the second equality follows from the mutual independence between the observations when the secret dither $\mathbf{t}$ is known. Eq. (52) can be rewritten as

$$\begin{aligned} &f(\tilde{\mathbf{y}}_1, \ldots, \tilde{\mathbf{y}}_{N_o}|m_1^i, \ldots, m_{N_o}^i) \\ &\quad= f(\tilde{\mathbf{y}}_1, \ldots, \tilde{\mathbf{y}}_{N_o-1}|m_1^i, \ldots, m_{N_o-1}^i) \int_{\mathcal{V}(\Lambda)} f(\tilde{\mathbf{y}}_{N_o}|m_{N_o}^i, \mathbf{t}) \cdot f(\mathbf{t}|\tilde{\mathbf{y}}_1, \ldots, \tilde{\mathbf{y}}_{N_o-1}, m_1^i, \ldots, m_{N_o-1}^i) d\mathbf{t}. \end{aligned} \tag{53}$$

[8]Notice that the volume of $\mathcal{Z}(\Lambda)$ can be computed by means of a Riemann integral, because for $\alpha < 1$, $\mathcal{Z}(\Lambda)$ is compact, and it is the linear (and invertible) image of a n-dimensional hypercube.

If the same procedure is applied recursively to the leftmost term in the right hand side of (53), we arrive at the following factorization for the a posteriori probability:

$$\begin{aligned} f(\tilde{\mathbf{y}}_1,\ldots,\tilde{\mathbf{y}}_{N_o}|m_1^i,\ldots,m_{N_o}^i) &= \prod_{k=1}^{N_o} f(\tilde{\mathbf{y}}_k|m_1^i,\ldots,m_k^i,\tilde{\mathbf{y}}_1,\ldots,\tilde{\mathbf{y}}_{k-1}) \\ &= \prod_{k=1}^{N_o}\int_{\mathcal{V}(\Lambda)} f(\tilde{\mathbf{y}}_k|m_k^i,\mathbf{t})\cdot f(\mathbf{t}|\tilde{\mathbf{y}}_1,\ldots,\tilde{\mathbf{y}}_{k-1},m_1^i,\ldots,m_{k-1}^i)d\mathbf{t}. \end{aligned} \tag{54}$$

In order to compute each factor of (54), we recall the flat-host assumption, which implies that $f(\tilde{\mathbf{y}}_k|m_k^i,\mathbf{t}) = \varphi(\tilde{\mathbf{y}}_k - \mathbf{d}_{m_k^i} - \mathbf{t} \mod \Lambda)$. Thus, each factor of (54) can be seen as a circular convolution over $\mathcal{V}(\Lambda)$. Furthermore, under the assumption that $\mathbf{T}\sim U(\mathcal{V}(\Lambda))$, the conditional pdf of the dither is given by Eq. (10). By combining (8) and (10), it can be seen that the integrand of the kth factor in (54) is

$$\begin{cases} (\text{vol}(\mathcal{Z}(\Lambda))\cdot\text{vol}(\mathcal{S}_{k-1}(\mathbf{m}^{(i)})))^{-1}, & \text{for } \mathbf{t}\in\mathcal{S}_{k-1}(\mathbf{m}^{(i)}) \text{ such that } (\tilde{\mathbf{y}}_k - \mathbf{d}_{m_k^i} - \mathbf{t}) \mod \Lambda \in \mathcal{Z}(\Lambda) \\ 0, & \text{otherwise.} \end{cases}$$

The condition on $\mathbf{t}$ in the equation above is equivalent to $\mathbf{t}\in S_{k-1}(\mathbf{m}^{(i)})$ such that $\mathbf{t}\in(\tilde{\mathbf{y}}_k - \mathbf{d}_{m_k^i} - \mathcal{Z}(\Lambda)) \mod \Lambda$, so each factor in (54) is proportional to the volume of $\mathcal{S}_k(\mathbf{m}^{(i)}) = \mathcal{S}_{k-1}(\mathbf{m}^{(i)})\cap\mathcal{D}_k(m_k^i)$. Finally, Eq. (54) can be succinctly expressed as

$$\begin{aligned} f(\tilde{\mathbf{y}}_1,\ldots,\tilde{\mathbf{y}}_{N_o}|m_1^i,\ldots,m_{N_o}^i) &= \begin{cases} \displaystyle\prod_{k=1}^{N_o}\frac{\text{vol}(\mathcal{S}_k(m_1^i,\ldots,m_k^i))}{\text{vol}(\mathcal{Z}(\Lambda))\cdot\text{vol}(\mathcal{S}_{k-1}(m_1^i,\ldots,m_{k-1}^i))}, & \text{for } \tilde{\mathbf{y}}_1,\ldots,\tilde{\mathbf{y}}_{N_o}\in\mathcal{S}_{N_o}(\mathbf{m}^{(i)}) \\ 0, & \text{otherwise} \end{cases} \\ &= \begin{cases} \dfrac{\text{vol}(\mathcal{S}_{N_o}(m_1^i,\ldots,m_{N_o}^i))}{(\text{vol}(\mathcal{Z}(\Lambda)))^{N_o}\cdot\text{vol}(\mathcal{V}(\Lambda))}, & \text{for } \tilde{\mathbf{y}}_1,\ldots,\tilde{\mathbf{y}}_{N_o}\in\mathcal{S}_{N_o}(\mathbf{m}^{(i)}) \\ 0, & \text{otherwise} \end{cases} \end{aligned} \tag{55}$$

APPENDIX C
PROOF OF THEOREM 1

As shown in Eq. (51) and (55) from App. B, the a posteriori probability of a message sequence is nonnull only if its feasible region is not an empty set. Using the definition of feasible region (cf. (11), we can state that a certain message sequence $\mathbf{m}^{(i)}$ has nonnull a posteriori probability if

$$\bigcap_{k=1}^{N_o}(\tilde{\mathbf{y}}_k - \mathbf{d}_{m_k^i} - \mathcal{Z}(\Lambda)) \mod \Lambda \neq \emptyset. \tag{56}$$

We say that the sequences fulfilling the above condition are "feasible" given the set of observations. The proof of the theorem is based on the concept of feasibility.

First, let us denote by $\mathbf{m}^{(1)}$ the message sequence embedded in the observations. We will arrange, without loss of generality, the message space $\mathcal{M}^{N_o}$ in two disjoint subsets: $\{\mathbf{m}^{(i)},\ i=1,\ldots,p\}$, which will represent the sequences in the equivalence class of $\mathbf{m}^{(1)}$, and $\{\mathbf{m}^{(i)},\ i=p+1,\ldots,p^{N_o}\}$, which will represent the remaining sequences in $\mathcal{M}^{N_o}$. For the sake of clarity, the proof will be conducted in 4 steps.

Step 1) In this step we will show that all the sequences in the equivalent class of $\mathbf{m}^{(1)}$ are feasible.

Notice that from Eq. (7), taking into account that $\mathbf{N}_k$ is uniformly distributed over $\mathcal{V}(\Lambda)$, we can write

$$(\tilde{\mathbf{y}}_k - \mathbf{d}_{m_k^i}) \mod \Lambda \in (\mathcal{Z}(\Lambda) + \mathbf{t} + \mathbf{d}_{m_k^1} - \mathbf{d}_{m_k^i}) \mod \Lambda, \text{ for all } k = 1, \ldots, N_o,$$

or equivalently,

$$(\mathbf{t} + \mathbf{d}_{m_k^1} - \mathbf{d}_{m_k^i}) \mod \Lambda \in (\tilde{\mathbf{y}}_k - \mathbf{d}_{m_k^i} - \mathcal{Z}(\Lambda)) \mod \Lambda, \text{ for all } k = 1, \ldots, N_o. \tag{57}$$

By Lemma 2, for $i = 1, \ldots, p$, and for all k, we have $\mathbf{d}_{m_k^i} = (\mathbf{d}_{m_k^1} + \mathbf{d}_l) \mod \Lambda$, for some fixed $l \in \mathcal{M}$. Hence, $(\mathbf{t} + \mathbf{d}_{m_k^1} - \mathbf{d}_{m_k^i}) \mod \Lambda = (\mathbf{t} - \mathbf{d}_l) \mod \Lambda$, for all $k = 1, \ldots, N_o$. Substituting in (57), we obtain

$$(\mathbf{t} - \mathbf{d}_l) \mod \Lambda \in (\tilde{\mathbf{y}}_k - \mathbf{d}_{m_k^i} - \mathcal{Z}(\Lambda)) \mod \Lambda, \text{ for all } k = 1, \ldots, N_o. \tag{58}$$

This proves that, for the message sequences belonging to the equivalence class of $\mathbf{m}^{(1)}$, the intersection (56) is always non-empty, since at least the point $(\mathbf{t} - \mathbf{d}_l) \mod \Lambda$ is contained in the intersection.

Step 2) In this step we prove that the feasible regions of the message sequences that belong to the equivalence class of $\mathbf{m}^{(1)}$ asymptotically converge to $(\mathbf{t} - \mathbf{d}_l) \mod \Lambda$, $l \in \mathcal{M}$.

First, we prove the next lemma, concerned with the asymptotic convergence of $\mathcal{S}_{N_o}(\mathbf{m}^{(1)})$.

Lemma 3: If the secret dither takes the value $\mathbf{t}$, then $\mathcal{S}_{N_o}(\mathbf{m}^{(1)})$ converges almost surely to $\mathbf{t}$ as $N_o \to \infty$.

Proof: The sequence $\mathbf{m}^{(1)}$, which for simplicity will be denoted by $\mathbf{m}$, equals the true message sequence embedded in $\{\tilde{\mathbf{Y}}_k, k = 1, \ldots, N_o\}$. Assume that the secret dither, which is fixed for all k, takes the value $\mathbf{t}$. The random variable defined as $\tilde{\mathbf{V}}_k \triangleq (\tilde{\mathbf{Y}}_k - \mathbf{d}_{M_k}) \mod \Lambda$ is uniformly distributed over $(\mathbf{t} + \mathcal{Z}(\Lambda)) \mod \Lambda$. For $\alpha \geq 0.5$, the feasible region $\mathcal{S}_{N_o}(\mathbf{M})$ is a modulo-Λ convex set. For any lattice Λ, $\mathcal{Z}(\Lambda)$ can be upper bounded by $\mathcal{B}(\mathbf{0}, (1-\alpha)r_c(\Lambda))$, where $r_c(\Lambda)$ is the covering radius of Λ defined in (13), and $\mathcal{B}(\mathbf{c}, r)$ denotes the n-dimensional closed hypersphere of radius r centered in $\mathbf{c}$. Therefore,

$$\mathcal{S}_{N_o}(\mathbf{M}) \subseteq \bigcap_{k=1}^{N_o} \mathcal{B}(\tilde{\mathbf{V}}_k, (1-\alpha)r_c(\Lambda)) \mod \Lambda.$$

The intersection between two hyperspheres of radius r becomes arbitrarily small as the distance between the centers of both spheres approaches $2r$. In the limit, their intersection is the unique point equidistant to the two centers. This means that the feasible region $\mathcal{S}_{N_o}(\mathbf{M})$ converges to $\mathbf{t}$ if the maximum distance between the modulo-Λ reduced observations $\tilde{\mathbf{V}}_k$ approaches $2(1-\alpha)r_c(\Lambda)$. For this condition to hold, at least one observation $\tilde{\mathbf{V}}_i$ must be arbitrarily close to a certain vertex $\mathbf{a}$ of $(\mathbf{t} + \mathcal{Z}(\Lambda)) \mod \Lambda$, and at least another observation $\tilde{\mathbf{V}}_j$, with $j \neq i$, must be arbitrarily close to another vertex $\mathbf{b}$ at distance $2(1-\alpha)r_c(\Lambda)$. Intuitively, the probability of finding such a pair of observations goes to 1 almost surely as $N_o \to \infty$.

Let us define the random variable $D_{N_o} \triangleq \max\{\delta_{N_o}(\mathbf{a}), \delta_{N_o}(\mathbf{b})\}$, with $\delta_{N_o}(\mathbf{a}) = \min_k |\mathbf{a} - \tilde{\mathbf{V}}_k|$ and $\delta_{N_o}(\mathbf{b}) = \min_k |\mathbf{b} - \tilde{\mathbf{V}}_k|$. Formally, we want to show that $D_{N_o} \xrightarrow{\text{a.s.}} 0$, which is equivalent to show that for all $\epsilon > 0$

$$\Pr\left(\lim_{N_o \to \infty} |D_{N_o}| > \epsilon\right) = 0. \tag{59}$$

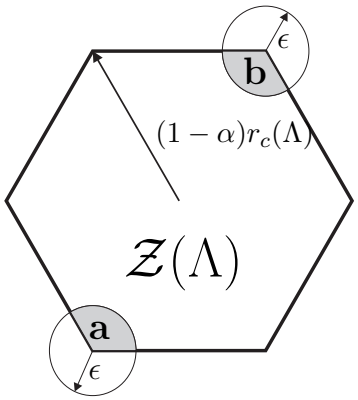


Fig. 9. Geometrical interpretation of the variables involved in the proof of Lemma 3. The value $(1-\alpha)r_c(\Lambda)$ is the covering radius of $\mathcal{Z}(\Lambda)$. The regions of $\mathcal{Z}(\Lambda)$ within distance ϵ of the vertices of interest, a and b, are shaded.

In order to prove almost sure convergence, it is sufficient to prove that the sum $\sum_{N_o=1}^{\infty} \Pr\left(|D_{N_o}| > \epsilon\right)$ is finite. In such case, almost sure convergence follows by virtue of the Borel-Cantelli lemma [21]. The probability of interest can be obtained as

$$\Pr\left(|D_{N_o}| > \epsilon\right) = \Pr\left(D_{N_o} > \epsilon\right) = \Pr\left(\delta_{N_o}(\mathbf{a}) > \epsilon \bigcup \delta_{N_o}(\mathbf{b}) > \epsilon\right) \leq \Pr\left(\delta_{N_o}(\mathbf{a}) > \epsilon\right) + \Pr\left(\delta_{N_o}(\mathbf{b}) > \epsilon\right)$$

$$= \Pr\left(\bigcap_{k=1}^{N_o} |\mathbf{a} - \tilde{\mathbf{V}}_k| > \epsilon\right) + \Pr\left(\bigcap_{k=1}^{N_o} |\mathbf{b} - \tilde{\mathbf{V}}_k| > \epsilon\right) = \Pr\left(|\mathbf{a} - \tilde{\mathbf{V}}_k| > \epsilon\right)^{N_o} + \Pr\left(|\mathbf{b} - \tilde{\mathbf{V}}_k| > \epsilon\right)^{N_o} \quad (60)$$

The probabilities in (60) can be easily computed by taking into account the geometrical interpretation provided by Figure 9, using an hexagonal lattice as example. For the first probability in the right hand side of (60) we have

$$\Pr\left(|\mathbf{a} - \tilde{\mathbf{V}}_k| > \epsilon\right) = \frac{\text{vol}(\mathcal{Z}(\Lambda)) - \text{vol}(\mathcal{B}(\mathbf{a}, \epsilon) \cap \mathcal{Z}(\Lambda))}{\text{vol}(\mathcal{Z}(\Lambda))} = 1 - \frac{\text{vol}(\mathcal{B}(\mathbf{a}, \epsilon) \cap \mathcal{Z}(\Lambda))}{\text{vol}(\mathcal{Z}(\Lambda))} = 1 - \rho_{\mathbf{a}}(\epsilon),$$

with $0 < \rho_{\mathbf{a}}(\epsilon) \leq 1$. Following a similar calculation for the other term involved in (60), we can write

$$\Pr\left(|D_{N_o}| > \epsilon\right) \leq (1 - \rho_{\mathbf{a}}(\epsilon))^{N_o} + (1 - \rho_{\mathbf{b}}(\epsilon))^{N_o},$$

and consequently $\sum_{N_o=1}^{\infty} \Pr(|D_{N_o}| > \epsilon) \leq 1/\rho_{\mathbf{a}}(\epsilon) + 1/\rho_{\mathbf{b}}(\epsilon) < \infty$. Since this sum is finite, the result (59) follows as a consequence of the Borel-Cantelli lemma. Hence, we have proved that the feasible region of $\mathbf{m}^{(1)}$ converges to a single point as $N_o \to \infty$. ■

As shown in Lemma 2, the feasible region for the sequences in the equivalence class of $\mathbf{m}^{(1)}$ only differ in their centers: for $i = 1, \ldots, p$, $\mathcal{S}_{N_o}(\mathbf{m}^{(i)}) = (\mathcal{S}_{N_o}(\mathbf{m}^{(1)}) - \mathbf{d}_l) \mod \Lambda$. Combining this result with Lemma 3 and with the result of Step 1), it follows that the feasible regions of the sequences in the equivalence class of $\mathbf{m}^{(1)}$ converge to $(\mathbf{t} - \mathbf{d}_l) \mod \Lambda$, $l \in \mathcal{M}$, as $N_o \to \infty$.

Step 3) Now we consider the sequences not belonging to the equivalence class of $\mathbf{m}^{(1)}$. We will prove that these sequences are all unfeasible for sufficiently large N_o.

Among the elements of this set of sequences, consider a sequence $\mathbf{m}^{(j)}$ which, for some $i = 1, \ldots, p$, equals $\mathbf{m}^{(i)}$ for all k but for a certain $k = k_0$. Consider the feasible region of this sequence, which is given by

$$\mathcal{S}_{N_o}(\mathbf{m}^{(j)}) = \left(\bigcap_{k \setminus k_0} (\tilde{\mathbf{y}}_k - \mathbf{d}_{m_k^j} - \mathcal{Z}(\Lambda)) \mod \Lambda \right) \bigcap \left((\tilde{\mathbf{y}}_{k_0} - \mathbf{d}_{m_{k_0}^j} - \mathcal{Z}(\Lambda)) \mod \Lambda \right) = \mathcal{H}_1 \bigcap \mathcal{H}_2. \quad (61)$$

For all $k \setminus k_0$, $(\mathbf{t} + \mathbf{d}_{m_k^1} - \mathbf{d}_{m_k^i}) \mod \Lambda = (\mathbf{t} - \mathbf{d}_l)$, for some fixed $l \in \mathcal{M}$. Hence, using the result of Step 2),

$$\lim_{N_o \to \infty} \mathcal{H}_1 = (\mathbf{t} - \mathbf{d}_l) \mod \Lambda. \quad (62)$$

For $k = k_0$, $(\mathbf{t} + \mathbf{d}_{m_k^1} - \mathbf{d}_{m_k^i}) = (\mathbf{t} - \mathbf{d}_c)$, with $c \neq l$. Therefore, $(\mathbf{t} - \mathbf{d}_c) \mod \Lambda \in \mathcal{H}_2$. Since $\mathcal{Z}(\Lambda) \subset \mathcal{V}(\Lambda_f)$ by assumption, then $(\mathbf{t} - \mathbf{d}_l) \mod \Lambda$, with $c \neq l$, cannot belong to $\mathcal{H}_2$. Thus,

$$\lim_{N_o \to \infty} \mathcal{S}_{N_o}(\mathbf{m}^{(j)}) = \lim_{N_o \to \infty} \mathcal{H}_1 \bigcap \mathcal{H}_2 = \emptyset. \quad (63)$$

This shows that the sequences that do not belong to the equivalence class of $\mathbf{m}^{(1)}$ cannot contain in their feasible region any of the points $(\mathbf{t} - \mathbf{d}_l) \mod \Lambda$, $l \in \mathcal{M}$, when $N_o \to \infty$. It only remains to be proved that these points are the only feasible regions when $N_o \to \infty$. This is a consequence of the next result.

Claim: For any point $\mathbf{z} \in \mathcal{V}(\Lambda)$ such that $\mathbf{z} \neq (\mathbf{t} - \mathbf{d}_l) \mod \Lambda$, $l \in \mathcal{M}$, the probability of observing one $\tilde{\mathbf{y}}_k$ such that $\mathbf{z}$ does not belong to any of the regions $(\tilde{\mathbf{y}}_k - \mathbf{d}_i - \mathcal{Z}(\Lambda)) \mod \Lambda$, $i \in \mathcal{M}$, goes to 1 as $N_o \to \infty$.

Sketch of the proof: For fixed $\mathbf{t}$, the support of the observations is $\mathcal{R} \triangleq \bigcup_{l \in \mathcal{M}} (\mathbf{t} + \mathbf{d}_l + \mathcal{Z}(\Lambda)) \mod \Lambda$. Note that $\mathcal{R}$ does not cover the whole Voronoi region $\mathcal{V}(\Lambda)$, since $\mathcal{Z}(\Lambda) \subset \mathcal{V}(\Lambda_f)$ by assumption. Let us denote by $\overline{\mathcal{R}}$ the complement of $\mathcal{R}$ in $\mathcal{V}(\Lambda)$. If $\mathbf{z} \in \overline{\mathcal{R}}$, then it suffices to observe $\tilde{\mathbf{y}}_k = (\mathbf{t} + \mathbf{d}_l) \mod \Lambda$, for some $l \in \mathcal{M}$. On the other hand, if $\mathbf{z} \in \mathcal{R}$, then it suffices to observe $\tilde{\mathbf{y}}_k = (\mathbf{t} + \mathbf{d}_l - \mathbf{e}) \mod \Lambda$, for some $l \in \mathcal{M}$, where $\mathbf{e}$ is the shortest norm vector such that $(\mathbf{z} + \mathbf{e}) \mod \Lambda \in \overline{\mathcal{R}}$. Since for fixed $\mathbf{t}$ the observations are uniformly distributed over $\mathcal{R}$, the probability of observing such $\tilde{\mathbf{y}}_k$ goes to 1 as $N_o \to \infty$. ■

Therefore, any sequence not belonging to the equivalence class of $\mathbf{m}^{(1)}$ is unfeasible for $N_o \to \infty$.

Step 4) The three previous steps have shown that the only message sequences in $\mathcal{M}^{N_o}$ that have nonnull probability when $N_o \to \infty$ are those belonging to the equivalence class of $\mathbf{m}^{(1)}$. We know, by Lemma 2, that all these sequences are equiprobable. Hence, if the attacker has enough computational power for checking an exponentially increasing number of intersections, then his uncertainty about the embedded message sequence becomes $\log(p)$ for $N_o \to \infty$, and the theorem follows.

REFERENCES

[1] F. Cayre, C. Fontaine, and T. Furon, "Watermarking security: theory and practice," *IEEE Trans. Signal Processing*, vol. 53, no. 10, oct 2005.

[2] P. Comesaña, L. Pérez-Freire, and F. Pérez-González, "Fundamentals of data hiding security and their application to Spread-Spectrum analysis," in *7th Information Hiding Workshop, IH05*, ser. Lecture Notes in Computer Science, vol. 3727. Barcelona, Spain: Springer Berlin / Heidelberg, June 2005, pp. 146–160.

[3] L. Pérez-Freire, P. Comesaña, J. R. Troncoso-Pastoriza, and F. Pérez-González, "Watermarking security: a survey," *Transactions on Data Hiding and Multimedia Security I*, vol. 4300, pp. 41–72, October 2006.

[4] T. Kalker, "Considerations on watermarking security," in *Fourth IEEE Workshop on Multimedia Signal Processing (MMSP)*, Cannes, France, October 2001, pp. 201–206.

[5] P. Moulin and R. Koetter, "Data hiding codes," *Proceedings of IEEE*, vol. 93, no. 12, pp. 2083–2126, December 2005.

[6] M. H. M. Costa, "Writing on dirty paper," *IEEE Transactions on Information Theory*, vol. 29, no. 3, pp. 439–441, May 1983.

[7] U. Erez and R. Zamir, "Achieving $\frac{1}{2}\log(1+\text{SNR})$ over the Additive White Gaussian Noise Channel with Lattice Encoding and Decoding," *IEEE Transactions on Information Theory*, vol. 50, no. 10, pp. 2293–2314, October 2004.

[8] L. Pérez-Freire, F. Pérez-González, T. Furon, and P. Comesaña, "Security of lattice-based data hiding against the Known Message Attack," *IEEE Transactions on Information Forensics and Security*, vol. 1, no. 4, pp. 421–439, December 2006.

[9] L. Pérez-Freire and F. Pérez-González, "Exploiting security holes in lattice data hiding," in *9th Information Hiding Workshop, IH07*, ser. Lecture Notes in Computer Science, vol. 4567. Saint Malo, France: Springer Berlin / Heidelberg, June 2008, pp. 159–173.

[10] T. M. Cover and J. A. Thomas, *Elements of Information Theory*. Wiley series in Telecommunications, 1991.

[11] J. H. Conway and N. J. A. Sloane, *Sphere packings, lattices and groups*, 3rd ed., ser. Comprehensive Studies in Mathematics. New York: Springer-Verlag, 1999, vol. 290.

[12] G. D. Forney, "Multidimensional constellations - Part II: Voronoi constellations," *IEEE Journal of Selected Areas in Communications*, vol. 7, no. 6, pp. 941–958, August 1989.

[13] U. Erez, S. Litsyn, and R. Zamir, "Lattices which are good for (almost) everything," *IEEE Transactions on Information Theory*, vol. 51, no. 10, pp. 3401–3416, October 2005.

[14] J. H. Conway, E. M. Rains, and N. J. A. Sloane, "On the existence of similar sublattices," *Canadian Journal of Mathematics*, vol. 51, pp. 1300–1306, 1999.

[15] A. Kerckhoffs, "La cryptographie militaire," *Journal des sciences militaires*, vol. 9, pp. 5–38, January 1883.

[16] J. J. Eggers, R. Bäuml, R. Tzschoppe, and B. Girod, "Scalar Costa Scheme for information embedding," *IEEE Transactions on Signal Processing*, vol. 51, no. 4, pp. 1003–1019, April 2003, special Issue on Signal Processing for Data Hiding in Digital Media and Secure Content Delivery.

[17] P. Comesaña, F. Pérez-González, and F. Balado, "On distortion-compensated dither modulation data-hiding with repetition coding," *IEEE Transactions on Signal Processing*, vol. 54, no. 2, pp. 585–600, February 2006.

[18] X. Huang, A. Acero, and H.-W. Hon, *Spoken Language Processing: A Guide to Theory, Algorithm and System Development*. Prentice Hall, 2001.

[19] M. F. Cheung, S. Yurkovich, and K. M. Passino, "An optimal volume ellipsoid algorithm for parameter set estimation," *IEEE Transactions on Automatic Control*, vol. 38, no. 8, pp. 1292–1296, August 1993.

[20] H. L. Royden, *Real analysis*, 3rd ed. Prentice Hall, 1988.

[21] W. Feller, *An Introduction to Probability Theory and Its Applications*, 3rd ed. New York: John Wiley and Sons, 1968.