A Novel Interpretation of Content Authentication*

Pedro Comesaña[1,2], Félix Balado[1] and Fernando Pérez-González[2]

[1] Information Hiding Laboratory. University College Dublin, Dublin 4, Ireland
[2] Dept. Teoría de la Señal y Comunicaciones. ETSI Telecom., Universidad de Vigo, 36310 Vigo, Spain

ABSTRACT

This work deals with practical and theoretical issues raised by the information-theoretical framework for authentication with distortion constraints proposed by Martinian et al.[1] The optimal schemes proposed by these authors rely on random codes which bear close resemblance to the dirty-paper random codes which show up in data hiding problems. On the one hand, this would suggest to implement practical authentication methods employing lattice codes, but these are too easy to tamper with within authentication scenarios. Lattice codes must be randomized in order to hide their structure. One particular multimedia authentication method based on randomizing the scalar lattice was recently proposed by Fei et al.[2] We reexamine here this method under the light of the aforementioned information-theoretical study, and we extend it to general lattices thus providing a more general performance analysis for lattice-based authentication. We also propose improvements to Fei et al.'s method based on the analysis by Martinian et al., and we discuss some weaknesses of these methods and their solutions.

1. INTRODUCTION

In the last years multimedia editing tools have undergone an impressive evolution, putting powerful capabilities within reach of average unskilled users. This seeming advantage constitutes at the same time a serious threat. Indeed, using those advanced tools, the authenticity of multimedia contents can be effectively compromised by a much larger number of people than ever before. This new trend stresses the importance of developing multimedia authentication techniques aimed at solving this critical issue, especially if digital multimedia contents are to possess in the future the same forensic value —at least— than traditional analog data typically stored on paper or tape.

Multimedia authentication using data hiding significantly differs from traditional authentication based on cryptographic digital signatures. The two basic differences are: a) the *authenticating signal*, also called *authenticator*, is embedded within the original signal to be authenticated (host signal), rather than appended to it; and b) the *authenticated signal* thus obtained can be modified afterwards by an editor (or attacker) without altering its authenticity as long as the modifications respect the semantic meaning of the original; note that this in sharp contrast with traditional cryptography-based authentication, in which the modification of a single bit of the authenticated signal causes a negative authentication. Notice as well that the requirement of resilience to moderate distortion is already implied by the use of data hiding to embed the authenticator within the original host signal. This requirement motivates the more general name "authentication with distortion constraints", which embraces multimedia authentication when the host signal is a multimedia signal.

*This work was partially funded by *Xunta de Galicia* under projects PGIDT04 TIC322013PR and PGIDT04 PXIC32202PM; MEC project DIPSTICK, reference TEC2004-02551/TCM; FIS project IM3, reference G03/185 and European Comission 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.

Further author information: (Corresponding author: P.C.)
P.C.: E-mail: pcomesan@gts.tsc.uvigo.es, Telephone: +34 986 812683
F.B.: E-mail: fiz@ihl.ucd.ie, Telephone: +353 1 716 2454
F.P.-G.: E-mail: fperez@gts.tsc.uvigo.es, Telephone: +34 986 812124

Examples of semantics-preserving modifications are transcoding, low-power noise addition, mild filtering (enhancements, speckle and noise removal), and other typical signal processing operations. The decoder has to decide on the sole basis of the received signal whether it is authentic or not. That is, no reference to the original host signal is assumed, as in blind data hiding. As aforementioned, not every modification applied on the authenticated signal should be detected as a forgery, but just those which actually change the meaning of the authenticated content. This existence of certain allowed types of distortions raises a key problem for the designer, equivalent to define a distortion measure with perceptual meaning. Unfortunately, no such distortion measure exists.

To date, the application of data-hiding technologies to the multimedia authentication problem generically described above has not received much attention. This is clear when we consider the much larger number of works focussed on the application of data hiding to areas such as copyright protection, metadata embedding, and others. A sustained research thrust on data-hiding based multimedia authentication has only emerged lately, as it can be concluded from the increasing number of recent papers on the subject.[1–5] Following Martinian et al.,[1] two basic approaches encompass most prior data-hiding based authentication methods:

- Fragile (semifragile) embedding. In this type of method the embedded watermark (authenticator) is secretly agreed between the encoder and the decoder. The decoder then compares the extracted watermark against the known one by means of a (semantic) measure of authenticity. As we have discussed, defining the right measure is a difficult problem.
- Robust embedding (quantize-and-embed). This type of method is based on embedding distortion resilient relevant features of the content within the content itself. The decoder rebuilds those features from the possibly edited authenticated signal, and compares them to the decoded ones from the watermark. This procedure avoids the problem of defining a good perceptual measure for making the authentication decision.

Although these approaches give practical ways to tackle multimedia authentication, they are not based on solid information-theoretical foundations. Martinian et al.[1] have proposed such a framework which, as it happens in the quantize-and-embed case, does not require a semantic measure. Martinian's analysis also enables to show that quantize-and-embed approaches are suboptimal. Intuitively, this is due to the fact that some of the information required for authentication in this type of method is sent twice (inside the watermark and as part of the authenticated signal itself). The work of Martinian et al.[1] tackles the problem of authentication by constraining the allowed modifications by means of a *reference channel*, which establishes the only allowed processing that which will yield an authentic decision on the decoder. In the remainder of the paper we will review this framework, and study different issues raised by the conclusions of this work for the implementation of practical authentication schemes.

Notation and framework. The notation we will use in the remainder is introduced next. We will denote L-length random column vectors by capital boldface letters (e.g., $\mathbf{X}$), and their realizations with lower case boldface letters ($\mathbf{x}$); their unidimensional counterparts will be denoted as X and x, respectively. Furthermore, $\mathbf{X}$ will denote the zero-mean original host signal, with power $\sigma_X^2 = \frac{1}{L}\,\mathrm{E}[\mathbf{X}^T\mathbf{X}]$; $\mathbf{W}$ is the watermark which acts as authenticator, whose power is σ_W^2, and $\mathbf{Y} = \mathbf{X} + \mathbf{W}$ is the authenticated (watermarked) signal generated by the encoder. The distortion introduced by the editor or attacker will be modeled by the random variable $\mathbf{N}$, with power σ_N^2, yielding the received signal $\mathbf{Z} = \mathbf{Y} + \mathbf{N}$, which is the input to the decoder, i.e. the signal whose authenticity is checked. Encoder and decoder always share a secret codebook, which depends in general on a secret key. We will only denote explicitly this secret key where necessary. The authentication function $g(\cdot)$ can be considered as a binary decision device, as the task of the decoder is determining if the received signal $\mathbf{Z}$ is authentic (we will denote this hypothesis as $\mathcal{H}_0$), or forged (hypothesis $\mathcal{H}_1$).

Furthermore, we will focus on the following key measures of any authentication system:

- Probability of succesful attack (P_{sa}): it is defined as $P_{sa} \triangleq \Pr[g(\mathbf{Z}) = \mathcal{H}_0 \mid \mathcal{H}_1]$; clearly, the smaller P_{sa}, the better the authentication system will be.

- Embedding distortion ($D_e \triangleq \sigma_W^2$): in order to guarantee the imperceptibility of the embedding, it is common to put constraints on the embedding distortion, which is usually quantified by the power of the watermark.

2. INFORMATION-THEORETICAL FRAMEWORK

We review firstly in this section the main results due to Martinian et al.[1] As alredy mentioned, the criterion they establish for authenticity relies on narrowing the range of allowed edits (attacks) down to a single reference channel $\mathcal{R}$. The authors consider next the problem of obtaining an estimate of the original host signal $\mathbf{x}$ which is not influenced by $\mathbf{n}$ whenever the received signal $\mathbf{z}$ is authentic (i.e., whenever $\mathbf{y}$ has undergone the reference channel). The power of the associated estimation error is the so-called reconstruction distortion (D_r). The authors argue that the fundamental function to be considered in this problem is the achievable distortion region, defined by all distortion pairs (D_e,D_r) for which correct authentication (i.e., $P_{sa} = 0$) is asymptotically achievable. The distortion measures require the use of a non-negative valued distortion function $d(\cdot, \cdot)$.

Following this approach, and for the asymptotic case where the dimensionality of the problem goes to infinity, the authors state a coding theorem establishing that a given pair (D_e,D_r) is achievable if and only if encoding and reconstruction functions $f(\cdot, \cdot)$ and $r(\cdot)$ exist such that

$$I(Z;U) - I(X;U) \geq 0, \tag{1}$$

with U an auxiliar random variable satifying $\mathrm{E}[d(X, f(U, X))] \leq D_e$ and $\mathrm{E}[d(X, r(U))] \leq D_r$. The condition (1) clearly resembles the result by Gel'fand and Pinsker[6] on the capacity of a communication system with side information at the embedder. The essential difference lies on the fact that in the authentication scenario $I(Z;U) - I(X;U)$ is just constrained to be non-negative, rather than not smaller than the rate of the communication system.

Martinian et al. provide random coding solutions for the particular case of white Gaussian host signal and white Gaussian reference channel, using the quadratic distortion measure $d(a, b) = (a - b)^2$. For the low D_e regime, which is the most interesting for practical purposes, the optimal random coding scheme takes the form

$$U = X + T/\alpha \tag{2}$$

$$Y = U + (1 - \alpha)(X - U) = X + T, \tag{3}$$

where T is a zero-mean Gaussian random variable with variance D_e (choosing the MMSE estimate as the reconstruction function from U), and $\alpha \in (0, 1]$ an optimization constant. Now, the question is how to implement this scheme in a practical way. As (2) and (3) are completely equivalent to the random coding scheme for communications with side information at the encoder given by Costa,[7] the use of lattices appears as the obvious way towards a straightforward implementation. Nevertheless, we will see next and in Sections 3 and 4 that more considerations apply.

2.1. Considerations on Martinian's Analysis

We discuss next some aspects of the work of Martinian et al. which require further examination. Among them, we study the implications of using non-asymptotic schemes, the consequences of employing lattice implementations, and ways to tighten the security of the scheme.

Non-asymptotic analysis. The asymptotic assumption inherent in making the number of dimensions L tend to infinity plays a key role in the laxity of constraint (1) to be non-negative instead of strictly positive. As we will see next, it is possible to pursue expressions of $I(Z;U) - I(X;U)$ for finite L and a given probability of succesful attack.

Assume that the authentication system is based on the use of a codebook with rate R_1 bits per channel use, where R_1 is any achievable rate for the selected reference channel. The encoder agrees privately with the decoder a secret subset of $2^{L \cdot R_2}$ messages from the $2^{L \cdot R_1}$ possible messages, with $R_2 \leq R_1$. From the subset of codewords associated to the $2^{L \cdot R_2}$ messages, the encoder chooses as the transmitted codeword the one minimizing the embedding distortion. At the decoder side, the message embedded in $\mathbf{Z}$ is decoded, yielding one out of the

$2^{L \cdot R_1}$ possible messages. If the decoded message belongs to the secret subset, then the received signal is said to be authentic, and false otherwise.

If the received signal has been modified by a channel with power higher than that of the reference channel, then we will assume that the decoded message is uniformly distributed over the secret subset, as this is the situation of highest uncertainty for the attacker. In this case

$$P_{sa} = \frac{2^{L \cdot R_2}}{2^{L \cdot R_1}} = 2^{L(R_2 - R_1)},$$

and so it is straightforward to see that if a probability of succesful attack lower or equal to P_{sa} is to be ensured, then

$$R_2 \leq R_1 + \frac{1}{L} \log_2(P_{sa}), \tag{4}$$

establishing an upper bound to the cardinality of the set of messages chosen by the embedder. Furthermore, given that $R_2 \geq 0$, the expression (4) also establishes a lower bound on the rate of the chosen codebook conditioned to that P_{sa}, i.e.,

$$R_1 \geq -\frac{1}{L} \log_2(P_{sa}). \tag{5}$$

Be aware that when equality is achieved in (5), then $R_2 = 0$, i.e. only one message can be chosen by the embedder to identify the authentic signals.

The capacity of a system with side information at the embedder is given by the well-known result by Gel'fand and Pinsker[6] $C = \max_{p(\mathbf{u},\mathbf{w}|\mathbf{x})} I(\mathbf{Z};\mathbf{U}) - I(\mathbf{X};\mathbf{U})$, where the transition probabilities $p(\mathbf{z}|\mathbf{y})$ due to the reference channel have to be considered. As a particular case of (5), it is straightforward to see that

$$\max_{p(\mathbf{u},\mathbf{w}|\mathbf{x})} I(\mathbf{Z};\mathbf{U}) - I(\mathbf{X};\mathbf{U}) \geq -\frac{1}{L} \log_2(P_{sa}). \tag{6}$$

If we make L go to infinity, then we can see that the rightmost term in (6) goes to zero, and we just obtain the condition derived by Martinian et al.[1] for the achievability of a given distortion pair.

Assuming next the reference channel to be i.i.d. Gaussian noise with variance σ_N^2, we know that the capacity of a communication system with side information at the embedder is given by $\frac{1}{2}\log_2(1 + \mathrm{WNR})$, with $\mathrm{WNR} = D_e/\sigma_N^2$. This was established by Costa[7] for the Gaussian host and random codebook case, and recently shown to be also valid for the generic host and lattice-based schemes by Erez et al.[8] Therefore, considering the aforementioned reference channel, we can write (6) as

$$\frac{1}{2} \log_2(1 + \mathrm{WNR}) \geq -\frac{1}{L} \log_2(P_{sa}), \tag{7}$$

obtaining the following bound on the WNR needed to achieve a given P_{sa}

$$\mathrm{WNR} \geq (P_{sa})^{-2/L} - 1. \tag{8}$$

Given that the bound (8) can be achieved using lattices with Voronoi regions going to hyperspheres as the number of dimensions increases, it seems that those lattices are an asymptotically optimal option also for authentication applications. We will see later that further considerations apply when lattices are used for authentication.

Alternative criteria. It is important to realize that the analysis undertaken in[1] may lead to different optimal solutions if altenative criteria than the ones assumed in that work are established. Among them, we may take into account the following:

- Quality of the estimate of the original host signal. The importance of the estimate $\hat{\mathbf{X}}$ of $\mathbf{X}$, which is obtained from the observed signal $\mathbf{Z}$, relies on the fact that it allows us to guess what kind of tampering was performed by the attacker. As we saw at the start of this section, Martinian et al.[1] chose the variance of the estimation error (with quadratic $d(\cdot,\cdot)$) to quantify the quality of the estimate. Nevertheless, given that the estimate is constrained not to be influenced by the attacking distortion when the signal undergoes the test channel, closed-form results are not possible. In fact, the authors of[1] only provide bounds to the actual reconstruction distortion values.

  An alternative quality measure of the estimate is the uncertainty of $\mathbf{X}$ given the transmitted codeword $\mathbf{U}$, i.e. $h(\mathbf{X}|\mathbf{U})$. Be aware that, as long as the observed signal is authentic, the correct codeword $\mathbf{U}$ will be decoded, and so the estimate will be also independent of the distortion introduced by the attacker. Had the observed signal undergone a channel worse than the reference one, then it will be said with probability $1-P_{sa}$ that it was tampered, and $\hat{\mathbf{X}}$ will be not computed.

  If, as before, Costa's construction is followed, the quality of the estimate of the original host signal can be shown to be given by

$$h(\mathbf{X}|\mathbf{U}) = \frac{L}{2}\log_2\left(2\pi e\frac{\sigma_W^2\sigma_X^2}{\sigma_W^2+\alpha^2\sigma_X^2}\right), \tag{9}$$

  which requires the largest possible value of α to be minimized. Nevertheless, one must be aware that values of α in Costa's scheme are constrained by (6), if one wants the condition of the probability of succesful attack to be met. Therefore, the value of α chosen in this case will be the largest verifying (6); or, from another point of view, a trade-off exists between P_{sa} and $h(\mathbf{X}|\mathbf{U})$, achievable by modifying α.

- Security issues. Considering some recent works on data hiding security,[9–11] we propose to formalize security in the authentication scenario as

$$I(\mathbf{U}_i;\mathbf{U}_j) = 0, \text{ for all } 1 \le i,j \le |\mathcal{U}|, \quad i \ne j, \tag{10}$$

  where $\mathcal{U}$ is the set of authentication codewords, and $|\mathcal{U}|$ is its cardinality. The condition (10) implies that the knowledge of a given codeword must not disclose any knowledge about the remaining ones. If this condition does not hold, then an observer of a set of watermarked (authenticated) signals may infer information about the codewords not used to watermark those signals. This leakage may be used by the attacker to discover, even if approximately or partially, the position of the remaining codewords. With this information at hand he may find a codeword close to his tampered version of the authenticated signal, which will be considered as authentic by the decoder.[†]

  In this sense, the best an authentication system can do is to ensure the independence between its codewords, in such a way that the observation of a watermarked signal $\mathbf{Y}$ only provides information about the codeword used to produce this signal. This condition is verified by the random codebooks proposed by Martinian et al.[1] and then their analysis needs not be concerned with security issues. Nevertheless, the condition does not hold if, as suggested at the start of this section, the random codebook is implemented by means of lattices without any additional security mechanism.

Forgeries within the distortion constraint. The concept of authentication established by Martinian et al. involves the reconstruction of the original signal from the authenticated one within a distortion constraint. Although this is a plausible definition in the scenario considered, a pitfall of this approach is caused by the use of the MSE as the distortion constraint. The problem is that the flexibility of this distortion criterion may allow forgeries to be made within a given distortion constraint, which, hence, will be deemed authentic by the decoder. In order to see why, consider that the editor or attacker concentrates all admissible distortion in some coefficients of the authenticated signal. This may be enough to modify the semantics or the perceptual properties of the signal —when compared to an innocuous test channel— while keeping it within the authentication region

[†]The reader interested in the position estimation of codewords is referred to the work of Pérez-Freire et al.[12]

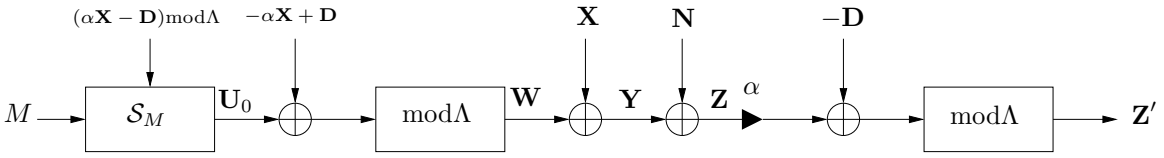

Figure 1. Scheme of the proposed method.

(reconstruction distortion region). Consequently, the probability of successful attack is in general higher than the one foreseen using the MSE.

There are two work-arounds to this problem. The first one would be considering a distortion measure that takes into account semantic or perceptual issues, which is obviously difficult to tackle in practice. A second, more practical, one entails refining the decision of the authentication scheme after an initial positive authentication. This amounts to adding a post-processing stage to the procedure proposed by Martinian et al. In order to accomplish this, it is possible to proceed as follows. If the decoder has decided positive authentication then the decoded codeword $\hat{\mathbf{U}}$ will be the one used by the encoder $\mathbf{U}$ with a high likelihood. Then, it is possible to use the encoding and reconstruction functions to build the residue signal $\hat{\mathbf{N}} \triangleq \mathbf{Z} - \hat{\mathbf{Y}} = \mathbf{Z} - f(\hat{\mathbf{U}}, r(\hat{\mathbf{U}}))$, which is an estimate of the channel undergone by the authenticated signal. Therefore, it is also possible to determine if that channel was compatible with the statistics of the test channel assumed. If this subsequent hypothesis test is negative then the signal at the input of the decoder will not be authenticated despite being inside the authentication region.

In the case of a Gaussian test channel, this amounts to ascertaining the Gaussianity of a realization of the residue $\hat{\mathbf{N}}$. This can be accomplished for small dimensionality L through the Shapiro-Wilk test,[13] or for larger L using the D'Agostino-Pearson K^2 test.[14] A cruder approach for general test channels would entail generating a histogram from $\hat{\mathbf{n}}$, and then observing the Kullback-Leibler distance with respect to a discretized version of the test channel distribution. In general, we will have to assume for this last operation to be legitimate that the test channel is memoryless and identically distributed, but tests such as Shapiro-Wilk allow to consider channels with memory.

Furthermore, note that if lattices are used to build the authentication codebook then the residue can never be truly Gaussian. In this case the test must be carried out for a modulo-lattice version of the test channel, with the consequence that any channel which yields a lattice-reduced Gaussian distribution when modularized will be validated. Although the effectiveness of attacks using such distributions will be of course limited, this observation shows another security weakness of authentication using lattices, related to the already commented vulnerability due to the impossibility of achieving (10) in this case.

3. A UNIFIED APPROACH FOR RANDOM AND LATTICE-BASED CONSTRUCTIONS

We have discussed that, although enticing, the sole use of lattices is not enough to implement an authentication scheme along the lines of Martinian et al.'s framework, due to the security pitfalls associated to regular codebooks. It is then necessary to put forward a strategy which, in a sense, is halfway between the impregnable but infeasible random codebooks and the vulnerable but implementable lattice-based methods. In order to shed some light in that direction, we investigate next a scheme that encompasses as particular cases both non-structured codebooks, such as random codebooks, and structured codebooks, such as those based on lattices.

Our strategy will be to show in which cases the scheme depicted in Fig. 1 approaches the capacity of an Additive White Gaussian Noise (AWGN) channel even in presence of a host signal $\mathbf{X}$ known by the embedder, but not known by the decoder. This result is in the same line that previous ones in the literature,[7,8] which in fact can be seen as the two particular cases previously mentioned. The scheme is based on the minimum distance quantization of the modulo-lattice-reduced signal $(\alpha\mathbf{X}-\mathbf{D})\mathrm{mod}\Lambda$, using a set of independently generated random codewords $\mathcal{S}_M$ (also termed bin) indexed by the transmitted symbol M, with $\mathbf{D}$ a dither vector. We will consider two different distributions of $\mathbf{D}$ corresponding to the two extreme cases considered:

1. $\mathbf{D} = \mathbf{0}$.
2. $\mathbf{D}$ uniformly distributed over the Voronoi region $\mathcal{V}(\Lambda)$ of the lattice used.

Each codeword $\mathbf{U}$ of each bin is uniformly distributed over the Voronoi region of the lattice used $\mathcal{V}(\Lambda)$, i.e. $\mathbf{U} \sim U(\mathcal{V}(\Lambda))$, so the variance per dimension of $\mathbf{U}$ coincides with the second moment per dimension associated with $\mathcal{V}(\Lambda)$; we will denote it as $\sigma^2_\Lambda \triangleq \frac{1}{L}\mathrm{E}\{||\mathbf{U}||^2\}$. Furthermore, the variance of the watermark $\mathbf{W}$, $\mathrm{Var}\{\mathbf{W}\}$, is constrained to be equal to σ^2_W, with $\sigma^2_W \leq \sigma^2_\Lambda$. On the other hand, the decoder tries to find out which codeword $\mathbf{U}_0$ was transmitted by the embedder, and taking into account the bin that codeword belongs to, i.e. $\mathcal{S}_{\hat{M}}$, he can estimate the embedded message $\hat{M}$. This estimation of the transmitted codeword is based on a first step of modulo lattice reduction, which significantly reduces the complexity of the decoder, and can be described as

$$\begin{aligned}\mathbf{Z}' &= [\alpha([\mathbf{U}_0 - \alpha\mathbf{X} + \mathbf{D}]\mathrm{mod}\Lambda + \mathbf{X} + \mathbf{N}) - \mathbf{D}]\mathrm{mod}\Lambda \\ &= [-(1-\alpha)([\mathbf{U}_0 - \alpha\mathbf{X} + \mathbf{D}]\mathrm{mod}\Lambda) + \alpha\mathbf{N} + \mathbf{U}_0]\mathrm{mod}\Lambda,\end{aligned}$$

so the maximum achievable rate of such system is

$$I(\mathbf{Z}'; M) = I(\mathbf{Z}'; \mathcal{S}_M) = h(\mathbf{Z}') - h(\mathbf{Z}'|\mathcal{S}_M);$$

given that $\mathcal{S}_M$ composed of $|\mathcal{S}_M|$ codewords, all of them with the same probability, since they are i.i.d., we can write

$$h(\mathbf{Z}'|\mathcal{S}_M) \leq h(\mathbf{Z}'|\mathbf{U}_0) + \log(|\mathcal{S}_M|).$$

For the cardinality of that set of codewords we will choose $|\mathcal{S}_M| = e^{I([\alpha\mathbf{X} - \mathbf{D}]\mathrm{mod}\Lambda; \mathbf{U}_0) + \delta}$, for an arbitrarily small $\delta > 0$, since with that value the probability of not finding a $\mathbf{U}_0$ jointly typical with $[\alpha\mathbf{X} - \mathbf{D}]\mathrm{mod}\Lambda$ goes to 0 when the dimensionality of the problem goes to infinity. Therefore, we can write

$$I(\mathbf{Z}'; M) \geq h(\mathbf{Z}') - h(\mathbf{Z}'|\mathbf{U}_0) - h(\mathbf{U}_0) + h(\mathbf{U}_0|[\alpha\mathbf{X} - \mathbf{D}]\mathrm{mod}\Lambda) - \delta, \tag{11}$$

where $h(\mathbf{Z}') \leq h(\mathbf{U}_0)$.

Given that (11) depends on distribution of the dither vector, in Sections 3.1 and 3.2 we will analyze the implications of considering the two proposed choices of the dither.

3.1. Dither vector uniformly distributed over the Voronoi region of the used lattice

Taking into account that

$$h(\mathbf{U}_0|[\alpha\mathbf{X} - \mathbf{D}]\mathrm{mod}\Lambda) = h([\mathbf{U}_0 - \alpha\mathbf{X} + \mathbf{D}]\mathrm{mod}\Lambda|\mathbf{U}_0) - h([-\alpha\mathbf{X} + \mathbf{D}]\mathrm{mod}\Lambda) + h(\mathbf{U}_0), \tag{12}$$

and given that $h([-\alpha\mathbf{X} + \mathbf{D}]\mathrm{mod}\Lambda)$ is equal to $h(\mathbf{U}_0)$ (since both random variables are uniformly distributed over the Voronoi region of the used lattice), we can see that $h(\mathbf{U}_0|[\alpha\mathbf{X} - \mathbf{D}]\mathrm{mod}\Lambda) = h(\mathbf{W}|\mathbf{U}_0)$. For the case where the dimensionality of the problem L goes to infinity, it is well-known that a sequence of lattices Λ^L exists such that $\lim_{L\to\infty} G(\Lambda^L) = \frac{1}{2\pi e}$, with $G(\Lambda^L)$ the normalized second moment of Λ^L. This yields an assymptotic Gaussian distribution for $\mathbf{U}_0$ and $(\alpha\mathbf{X} - \mathbf{D})\mathrm{mod}\Lambda$, and so both the distribution of $\mathbf{U}_0$ given $(\alpha\mathbf{X} - \mathbf{D})\mathrm{mod}\Lambda$ and $(\alpha\mathbf{X} - \mathbf{D} - \mathbf{U}_0)\mathrm{mod}\Lambda$ given $\mathbf{U}_0$ will be also Gaussian. Therefore, denoting $\mathrm{Var}\{\mathbf{W}|\mathbf{U}_0\} \triangleq \sigma^2_{W|U_0}$, it is straightforward to show that

$$\begin{aligned}I(\mathbf{Z}'; M) &\geq h(\mathbf{U}_0|[\alpha\mathbf{X} - \mathbf{D}]\mathrm{mod}\Lambda) - h(\mathbf{Z}'|\mathbf{U}_0) + h(\mathbf{Z}') - h(\mathbf{U}_0) - \delta \\ &\geq \frac{L}{2}\log(2\pi e\sigma^2_{W|U_0}) - \frac{L}{2}\log(2\pi e[(1-\alpha)^2\sigma^2_{W|U_0} + \alpha^2\sigma^2_N]) + h(\mathbf{Z}') - h(\mathbf{U}_0) - \delta.\end{aligned} \tag{13}$$

Given that $\frac{L}{2}\log(\sigma^2_{W|U_0}) - \frac{L}{2}\log((1-\alpha)^2\sigma^2_{W|U_0} + \alpha^2\sigma^2_N)$ is maximized choosing $\alpha^* = \frac{\sigma^2_{W|U_0}}{\sigma^2_{W|U_0} + \sigma^2_N}$, if one replaces that value in (13) will obtain that

$$I(\mathbf{Z}'; M) \geq \frac{L}{2}\log\left(1 + \frac{\sigma^2_{W|U_0}}{\sigma^2_N}\right) + h(\mathbf{Z}') - h(\mathbf{U}_0) - \delta.$$

Therefore, we can ensure that the capacity of the proposed scheme will be that of the AWGN channel without host signal interference, whenever $\mathbf{W}$ were independent of $\mathbf{U}_0$, since in that case $\sigma^2_{W|U_0} = \sigma^2_W$ and $h(\mathbf{Z}') = h(\mathbf{U}_0)$. This condition is verified when Erez and Zamir's construction is implemented.

3.2. No dither

When $\mathbf{D} = 0$, considering both (11) and (12), the capacity of the system can be lowerbounded as

$$I(\mathbf{Z}'; M) \geq h(\mathbf{Z}') - h(\mathbf{Z}'|\mathbf{U}_0) + h([\mathbf{U}_0 - \alpha\mathbf{X}] \text{mod} \Lambda | \mathbf{U}_0) - h([\alpha\mathbf{X}] \text{mod} \Lambda) - \delta.$$

A particular case of the proposed scheme is that due to Costa,[7] which is nothing but the previous scheme with $\sigma^2_\Lambda = \sigma^2_W + \alpha^2\sigma^2_X$, Gaussian host, $\mathbf{D} = \mathbf{0}$, the dimensionality going to infinity, and being $\mathbf{W}$ and $\mathbf{X}$ independent. The main apparent difference between the proposed scheme and the original one by Costa is the modulo reduction operation. Nevertheless, although one could think that the modulo reduction would be modifying the addition of the Gaussian signals in the presented scheme, the case is that the variance of all the signals in our scheme is smaller or equal than $\sigma^2_W + \alpha^2\sigma^2_X = \sigma^2_\Lambda$; therefore, when the number of dimensions goes to infinity, all the computed signals are inside the Voronoi region of the used lattice, so no modulo-reduction is really performed.

3.3. Application to Authentication Scenarios

The result introduced in the previous section is interesting from a theoretical point of view, and may be useful for increasing the security of data hiding applications (see[9–12] for further references on this topic), as it increases the number of parameters of the system to be estimated by an attacker. Nonetheless, its application to an authentication scenario poses additional issues. Indeed, one can see that adding a vector of the lattice $\lambda \in \Lambda$ to any legal content will lead to another content which will be deemed to be legally created, even when this modification could have a completely different meaning than the original. This is what Fei et al.[2] call security threat for authentication applications, and also the reason why the *encoding set* proposed in[2] depends on the particular Voronoi region of the coarse lattice. In this way, the knowledge of the location of a valid codeword on a given Voronoi region of the coarse lattice is not useful for determining the location of the codewords in a different Voronoi region.

The solution we propose here, which is based on the same principles as those suggested by Fei et al., is to generate a different codebook for quantizing $[\alpha\mathbf{X} - \mathbf{D}] \mod \Lambda$ depending on the Voronoi region of the lattice in which $\alpha\mathbf{X} - \mathbf{D}$ lies. Furthermore, the proposed decoding algorithm is based on a two-step procedure, in order to reduce the complexity of the decoder:[‡] 1) coarse quantization is performed in order to identify the Voronoi region of the lattice where the received signal $\mathbf{Z}'$ lies; and 2) the embedded symbol is estimated from $\mathbf{Z}'$ using the codebook corresponding to that Voronoi region.

An obvious source of errors of this scheme compared with the one analyzed at the beginning of this section, is that the received signal could lie in a different Voronoi region than $\alpha\mathbf{X} - \mathbf{D}$. In this case, the codebook used by the decoder to quantify $\mathbf{Z}'$ will be different than the one used in embedding, implying a decoding error. On the other hand, if we could ensure that the received signal were in the same Voronoi as $\alpha\mathbf{X} - \mathbf{D}$, then, both the encoding and decoding codebooks would coincide, and the fact that the codewords in other Voronoi regions were not the result of shifting the codewords in the current one by a lattice vector would not complicate the communication process; therefore, we could achieve the same reliable communication rate as in the previous section, i.e. the capacity of an AWGN channel.

In order to see the conditions that have to be verified to ensure that the received signal belongs to the same Voronoi region as $\alpha\mathbf{X} - \mathbf{D}$ let us recall the definition of *typical set*[15]

$$A^{(L)}_\epsilon = \left\{ (x_1, x_2, \cdots, x_L) : \left| -\frac{1}{L} \log f(x_1, x_2, \cdots, x_L) - h(X) \right| \leq \epsilon \right\},$$

[‡]This procedure is clearly suboptimal compared with the case where the complete set of codewords is considered in order to perform the quantization that estimates the embedded symbol.

which, for a Gaussian distribution with variance σ_X^2 becomes

$$A_\epsilon^{(L)} = \left\{(x_1, x_2, \cdots, x_L) : \left| \frac{||\mathbf{x}||^2}{L\sigma_X^2} - 1 \right| \leq 2\epsilon \right\};$$

Furthermore, taking into account the properties of the typical set it is well known that $P\{A_\epsilon^{(L)}\} > 1-\epsilon$ for L large enough; this implies that if we had a decreasing sequence ϵ_k, with $\epsilon_k > 0$, such that $\lim_{k\to\infty} \epsilon_k = 0$, then the vectors $\mathbf{x}$ belonging to $A_{\epsilon_k}^{(L)}$ would verify that $||x||^2 \to L\sigma_X^2$. Moreover, when L is large enough, the probability that a sample vector $\mathbf{x}$ does not belong to $A_{\epsilon_k}^{(L)}$, i.e. the probability that $||\mathbf{x}||^2 \notin [L\sigma_X^2(1-2\epsilon_k), L\sigma_X^2(1+2\epsilon_k)]$ is smaller than ϵ_k.

Therefore, when $L \to \infty$ and the Voronoi region of the lattice goes to a hypersphere, i.e. in the conditions where the capacity of the AWGN channel can be approached, we can see that for any $\epsilon > 0$

$$P\left\{ \frac{1}{L} ||\alpha\mathbf{W} + (\alpha\mathbf{X} - \mathbf{D}) \mathrm{mod}\Lambda + \alpha\mathbf{N}||^2 \in [\sigma_{eq}^2(1-2\epsilon), \sigma_{eq}^2(1+2\epsilon)] \right\} > 1-\epsilon, \tag{14}$$

where $\sigma_{eq}^2 = \mathrm{Var}\{\alpha\mathbf{W} + (\alpha\mathbf{X} - \mathbf{D})\mathrm{mod}\Lambda + \alpha\mathbf{N}\}$; from (14) we can write

$$P\left\{ \frac{1}{L} ||\alpha\mathbf{W} + (\alpha\mathbf{X} - \mathbf{D}) \mathrm{mod}\Lambda + \alpha\mathbf{N}||^2 \leq \sigma_{eq}^2(1+2\epsilon) \right\} > 1-\epsilon,$$

impliying that $\frac{1}{L}||\alpha\mathbf{W} + (\alpha\mathbf{X} - \mathbf{D})\mathrm{mod}\Lambda + \alpha\mathbf{N}||^2 \leq \sigma_\Lambda^2$ would be verified (and therefore the received signal would be in the same Voronoi region as $\alpha\mathbf{X} - \mathbf{D}$) with probability $1-\epsilon$, for arbitrarily small ϵ, whenever

$$\sigma_\Lambda^2 \geq (1+2\epsilon)\mathrm{Var}\{\alpha\mathbf{W} + (\alpha\mathbf{X} - \mathbf{D})\mathrm{mod}\Lambda + \alpha\mathbf{N}\}. \tag{15}$$

From the last inequality, and considering that

$$\begin{aligned} &\mathrm{Var}\{\alpha\mathbf{W} + (\alpha\mathbf{X} - \mathbf{D})\mathrm{mod}\Lambda + \alpha\mathbf{N}\} = \mathrm{Var}\{\alpha[\mathbf{U}_0 - \alpha\mathbf{X} + \mathbf{D}]\mathrm{mod}\Lambda + (\alpha\mathbf{X} - \mathbf{D})\mathrm{mod}\Lambda + \alpha\mathbf{N}\} \geq \\ &\mathrm{Var}\big\{\big[(1-\alpha)[\alpha\mathbf{X} - \mathbf{D} - \mathbf{U}_0]\mathrm{mod}\Lambda + \mathbf{U}_0\big]\mathrm{mod}\Lambda + \alpha\mathbf{N}\big\}, \end{aligned} \tag{16}$$

one realizes that for the case of uniform dither over the Voronoi region of the lattice, there is a trade-off between security and performance of the system in terms of probability of succesful attack (or equivalently, in terms of the achievable reliable rate). In fact, as we have previously shown, the capacity of the AWGN can be achieved by the proposed scheme when $\mathbf{U}_0$ and $\mathbf{W}$ are independent; but in that case (16) will take a value of $\sigma_\Lambda^2 + \alpha^2\sigma_N^2 \geq \sigma_\Lambda^2$, so (15) is not verified.

On the other hand, in Costa's construction,[7] where $\mathbf{X}$ is Gaussian, $\mathbf{D} = \mathbf{0}$, $\mathbf{W}$ and $\mathbf{X}$ are independent, $\sigma_\Lambda^2 = \sigma_W^2 + \alpha^2\sigma_X^2$ and $\alpha = \frac{\sigma_W^2}{\sigma_W^2 + \sigma_N^2}$, we can bound

$$\mathrm{E}\{||\alpha\mathbf{W} + (\alpha\mathbf{X} - \mathbf{D})\mathrm{mod}\Lambda + \alpha\mathbf{N}||^2\} \leq \alpha^2(\sigma_W^2 + \sigma_X^2 + \sigma_N^2) \leq \frac{\sigma_\Lambda^2}{1+2\epsilon},$$

and making $\alpha = \frac{\sigma_W^2}{\sigma_W^2 + \sigma_N^2}$, the last inequality can be written as

$$\alpha^2\sigma_X^2 + \frac{\sigma_W^4}{\sigma_W^2 + \sigma_N^2} \leq \frac{\alpha^2\sigma_X^2 + \sigma_W^2}{1+2\epsilon},$$

which will be verified for small enough values of ϵ. Therefore, in this case there is not a trade-off between security and performance; unfortunately, the cost is the high (in fact impractical) computational expense of this method.

Summarizing, if we want to design an authentication system, an equilibrium must be achieved among security, probability of succesful attack and computational cost.

4. PRACTICAL METHODS

As multimedia authentication is ultimately a practical application, it is fundamental to link the results introduced by Martinian et al.[1] with existing or new practical implementations. In the preceding section we have discussed the applicability of lattices to the implementation of authentication methods, and this section we will complete that analysis comparing some of the results introduced before, with those obtained for Fei et al.[2] scheme.

4.1. Fei's Lattice-based Authentication Method

Fei et al.,[2] independently of the insights gained from Martinian's analysis, have proposed an authentication system based on the use of randomized lattices. The basic idea is that the embedder, upon agreement with the detector, will choose a subset of centroids of an original lattice (in fact this choice is based on the decomposition of the original lattice in two nested lattices); if the received signal were in a given region (named admissible set, and that characterizes the distortions allowed by the system) around any of those chosen centroids, then the signal is said to be authentic, and non-authentic otherwise. If one wants this scheme to be hard to be attacked, it should be infeasible for the attacker to modify the watermarked signal with a non allowed distortion, and that such signal provides an *authentic* decision, even when the attacker has a complete knowledge of the parameters of the system, excepting the secret key. Taking these considerations into account, the authors motivate the need of a secure system. This security is provided using cryptographic functions, and different dither vectors depending on the outputs of those functions. The resulting technique is named *Nested Lattice Based MSB-LSB Scheme*, where MSB-LSB stands for *Most Significants Bits* and *Least Significant Bits*, respectively. Finally, the authors just marginally contemplate the possibility of using a distortion compensation technique.

In this section we will introduce some considerations about the penalties that Fei et al. scheme[2] pays to achieve its security requirement, focusing on the embedding distortion, and the maximum achievable rate (related to the probability of succesful attack).

Interpretation. The embedding function is based on quantizing the original host signal $\mathbf{x}$ to a valid codeword. The set of these valid codewords is chosen from the complete codebook depending on the secret key θ. On the other hand, the detection function is a binary hypothesis test deciding if the received signal has undergone an allowed *reference channel*, or if the editing operations are unlikely to follow such model. In[2] this reference channel is taken into account by defining a deterministic set, named the *admissible set* Ω_1, around each possible valid codeword; the union of the admissible sets centered at each valid codeword is the so-called *verification region* Ω_2.

In the general version of the scheme proposed in,[2] the complete codebook is a fine lattice Λ_1, whose fundamental Voronoi region $\mathcal{V}(\Lambda_1)$ contains the admissible set Ω_1. In order to choose those centroids of Λ_1 which are valid for the authenticacion application, a sublattice Λ_2 of Λ_1 is considered, yielding the nested lattice code (Λ_1, Λ_2). The lattices Λ_1 and Λ_2 are usually denoted as fine lattice and coarse lattice, respectively. The set of valid codewords for the authentication application is given by

$$\mathcal{C} = \{\ \mathbf{t} \in \mathbb{R}^L : \mathbf{t} = \mathbf{v} + \mathbf{d}(\theta, \mathbf{v}),\ \text{where } \mathbf{v} \in \Lambda_2\},$$

and $\mathbf{d}(\theta, \mathbf{v}) \in \Lambda_1/\Lambda_2$ is a key-dependent dither vector necessary for ensuring the security of the system, meaning that an attacker with access to some watermarked signals will not be able to recover the other authentication valid codewords. In this way, the embedding function can be written as

$$\mathbf{y} = g(\mathbf{x}, \theta) = \arg\min_{\mathbf{t}\in\mathcal{C}} ||\mathbf{x} - \mathbf{t}||^2 = \arg\min_{\mathbf{v}\in\Lambda_2} ||\mathbf{x} - \mathbf{v} - \mathbf{d}(\theta, \mathbf{v})||^2. \tag{17}$$

Due to its irregular structure, the encoding with this set of codewords would require exhaustive search over a set of neighbors, being unfeasible when L is increased. Therefore, the implementable embedding function proposed in[2] is defined as

$$\mathbf{y} = g(\mathbf{x}, \theta) = Q_{\Lambda_2}(\mathbf{x}) + \mathbf{d}(\theta, Q_{\Lambda_2}(\mathbf{x})). \tag{18}$$

Be aware that although the set of valid codewords remains unaltered, the embedding function is changed, meaning that a codeword different than that at minimum distance to the original host signal could be chosen for encoding. Obviously, this implies an increase in the embedding power, and subsequently a loss in the performance of the system, that we analyze in the next section.

4.2. Embedding distortion analysis

In this section we analyze the distortion introduced by the embedding function given by (18) in order to quantify the increase in that distortion due to reducing the complexity of the embedding function. Throughout this section, and based on the imperceptibility of the embedding process, we will assume that the power of the original host signal is much larger than the embedding distortion; this fact will enable us to consider the flat-host assumption.

First of all, when (18) is considered one can define the auxiliar random variable $\mathbf{V} \triangleq Q_{\Lambda_2}(\mathbf{X})$, yielding

$$\begin{aligned} \mathrm{E}\{||\mathbf{X}-\mathbf{Y}||^2\} &= \mathrm{E}\{||\mathbf{X}-\mathbf{V}+\mathbf{V}-\mathbf{Y}||^2\} = \mathrm{E}\{||\mathbf{X}-\mathbf{V}||^2\} + \mathrm{E}\{||\mathbf{V}-\mathbf{Y}||^2\} \\ &= L\cdot\sigma^2(\Lambda_2) + L\cdot[\sigma^2(\Lambda_2)-\sigma^2(\Lambda_1)] \\ &= 2\cdot L\cdot\sigma^2(\Lambda_2) - L\cdot\sigma^2(\Lambda_1), \end{aligned}$$

where $\sigma^2(\Lambda_2)$ denotes the second moment per dimension of Λ_2, and we have considered the fact that the dither vector, i.e. $\mathbf{Y}-\mathbf{V}$, is uniformly distributed on Λ_1/Λ_2 when one averages over the set of possible secrets keys θ, independently of the quantization error $\mathbf{X}-\mathbf{V}$, which in turn is uniformly distributed on $\mathcal{V}(\Lambda_2)$. Moreover, taking into account that a random variable $\mathbf{U}_2 \triangleq \mathbf{D}+\mathbf{U}_1$, where $\mathbf{U}_1 \sim U(\mathcal{V}(\Lambda_1))$ and $\mathbf{D} \sim U(\Lambda_1/\Lambda_2)$, verifies that $\mathbf{U}_2 \sim U(\mathcal{V}(\Lambda_2))$, one can write $\mathrm{Var}\{\mathbf{D}\} = L\cdot[\sigma^2(\Lambda_2)-\sigma^2(\Lambda_1)]$, for any pair of nested lattices (Λ_1,Λ_2).

Compared with the scheme by Erez and Zamir, Fei et al. spend an extra power of $L\cdot(\sigma^2(\Lambda_2)-\sigma^2(\Lambda_1))$, and the maximum rate of their scheme is $|\Lambda_1/\Lambda_2|$, i.e. the nesting ratio, which can be seen to be equal to $\left(\frac{\sigma^2(\Lambda_2)}{\sigma^2(\Lambda_1)}\right)^{L/2}$, establishing a clear relation between the excess of power required by the system and the achieved probability of succesful attack. Furthermore, in order to be able to properly decode the transmitted codeword the edition variance must be smaller than $\sigma^2(\Lambda_1)$. In this way, a large ratio $\frac{\sigma^2(\Lambda_2)}{\sigma^2(\Lambda_1)}$ implies a large penalty in the embedding power, but also leads a small probability of succesful attack, and the allowed editing distortions will have a small power.

5. CONCLUSIONS

In this paper we tried to provide an unified approach to different works on media authentication, trying to link different approaches in the literature.[1,2] In order to do so, we introduced a novel authentication scheme, inspired on the randomization of a lattice-based quantization scheme, where the number of codewords per Voronoi region of the used lattice can be larger than one, and distortion compensation is used (in contrast with Fei et al. approach,[2] where only one codeword per Voronoi region and no distortion compensation is considered in the performance analysis).

The obtained results show the trade-off between figures-of-merit as *probability of succesful attack* (closely related to the reliable achievable rate of the data hiding system), *security*, *embedding distortion* and *computational cost*. Whereas Martinian's approach[1] shows very good performance from the probability of succesful attack, security and embedding distortion point of view, it is computionally unfeasible; on the other hand, Fei's approach is computationally cheap, but at the cost of needing a significantly larger embedding distortion for a given probability of succesful attack. The proposed scheme tries to fill the gap between these two extreme approaches, providing a range of intermediate strategies. The performed analysis outlines the equilibrium that a system designer should take into account when devising an authentication scheme.

REFERENCES

1. E. Martinian, G. W. Wornell, and B. Chen, "Authentication with distortion criteria," *IEEE Transactions on Information Theory* **51**, pp. 2523–2542, July 2005.
2. C. Fei, D. Kundur, and R. H. Kwong, "Analysis and design of secure watemark-based authentication systems," *IEEE Transactions on Information Forensics and Security* **1**, pp. 43–55, March 2006.
3. Charles G. Boncelet, Jr., "The NTMAC for authentication of noisy messages," *IEEE Transactions on Information Forensics and Security* **1**, pp. 35–42, March 2006.
4. R. Ge, G. R. Arce, and G. D. Crescenzo, "Approximate message authentication codes for N-ary alphabets," *IEEE Transactions on Information Forensics and Security* **1**, pp. 56–67, March 2006.
5. V. Monga, A. Banerjee, and B. L. Evans, "A clustering based approach to perceptual image hashing," *IEEE Transactions on Information Forensics and Security* **1**, pp. 68–79, March 2006.
6. S. I. Gel'fand and M. S. Pinsker, "Coding for channel with random parameters," *Problems of Control and Information Theory* **9**(1), pp. 19–31, 1980.
7. M. H. M. Costa, "Writing on dirty paper," *IEEE Transactions on Information Theory* **29**, pp. 439–441, May 1983.
8. U. Erez and R. Zamir, "Achieving $\frac{1}{2}$ log(1+SNR) on the AWGN channel with lattice encoding and decoding," *IEEE Transactions on Information Theory* **50**, pp. 2293–2314, October 2004.
9. F. Cayre, C. Fontaine, and T. Furon, "Watermarking security: theory and practice," *IEEE Transactions on Signal Processing* **53**, pp. 3976–3987, October 2005.
10. P. Comesaña, L. Pérez-Freire, and F. Pérez-González, "Fundamentals of data hiding security and their application to Spread-Spectrum analysis," in *Information Hiding International Workshop*, M. Barni, J. H. Joancomartí, S. Katzenbeisser, and F. Pérez-González, eds., *Lecture Notes in Computer Science* **3727**, pp. 146–160, Springer, (Barcelona, Spain), June 2005.
11. L. Pérez-Freire, P. Comesaña, and F. Pérez-González, "Information-Theoretic Analysis of Security in Side-Informed Data Hiding," in *Information Hiding International Workshop*, M. Barni, J. H. Joancomartí, S. Katzenbeisser, and F. Pérez-González, eds., *Lecture Notes in Computer Science* **3727**, pp. 131–145, Springer, (Barcelona, Spain), June 2005.
12. 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* **1**, pp. 421–439, December 2006.
13. S. S. Shapiro, M. B. Wilk, and H. J. Chen, "Comparative study of various tests of normality," *Journal of the American Statistical Association* **63**, pp. 1343–1372, December 1968.
14. E. S. Pearson, R. B. D'Agostino, and K. . Bowman, "Tests for departure from normality: Comparison of powers," *Biometrika* **64**(2), pp. 231–246, 1977.
15. T. M. Cover and J. A. Thomas, *Elements of Information Theory*, Wiley series in Telecommunications, 1991.