Active Content Fingerpriting

Farzad Farhadzadeh, *Student Member, IEEE,* Sviatoslav Voloshynovskiy, *Senior Member, IEEE*

***Abstract*—Content fingerprinting and digital watermarking are techniques that are used for content protection and distribution monitoring and, more recently, for interaction with physical objects. Over the past few years, both techniques have been well studied and their shortcomings understood. In this paper, we introduce a new framework called *active content fingerprinting* which takes the best from two worlds of content fingerprinting and digital watermarking, in order to overcome some of the fundamental restrictions of these techniques in terms of performance and complexity. The proposed framework extends the encoding process of conventional content fingerprinting in a way similar to digital watermarking, thus allowing the extraction of fingerprints from the modified cover data. We consider several encoding strategies, examine the performance of the proposed schemes in terms of bit error rate, the probabilities of correct identification and false acceptance and compare it with those of conventional fingerprinting and digital watermarking. Finally, we extend the proposed framework to the multidimensional case based on lattices and demonstrate its performance on both synthetic data and real images.**

***Index Terms*—Content identification, digital fingerprint, digital watermark, bit error rate, probability of correct identification, probability of false acceptance.**

I. Introduction

In today's world, digital reproduction tools and user-generated content (UGC) websites, such as YouTube, which enable massive distribution, sharing and storage of multimedia contents, have undergone an impressive evolution, providing professional solutions to various groups of users. In addition, the portable devices (e.g. tablets, smartphones) are developed to such an extent as to allow a content consumer to interact with contents they are consuming, such as TV shows, movies, musics, sport broadcasts and video games. Some extra data are displayed on a portable device synchronized with the content being viewed or listed. Such applications are also known as *2nd screen* [2]. In recent years, portable devices are also used to interact with physical objects (packaging, magazines, identity documents, banknotes, etc.) via some encoded modalities (watermarks, graphic design, glyphs) [3]; with forensic properties of printed data [4] or with microstructures of physical surfaces and information carriers [5]. Such an interaction is used to provide more information about products to be consumed and to be used for security purposes, e.g., product identification, tracking and tracing.

In spite of such obvious advantages, these modern tools provide unprecedented possibilities for counterfeiters to virtually reproduce any physical or digital items, i.e., images, videos, audiofiles, documents in electronic or printed forms. Thus, the issue of content identification becomes a critical one, demanding an urgent solution for various applications.

Content identification systems are facing numerous requirements related to identification accuracy, complexity, privacy, security and memory storage [6]. The trade-off between these requirements is quite a complex problem that still remains largely unsolved. To address this trade-off *content fingerprints* are used [7], [8]. A content fingerprint represents a short, robust and distinctive content description. The main idea behind content fingerprinting consists in the extraction of a lower dimensional content representation that can be accomplished as follows [6], [7], [8]. First, a lower dimensional data representation from a content or its extracted feature is obtained using dimensionality reduction or feature extraction. Secondly, to address complexity, security, privacy and memory storage requirements, the transformed data are quantized and indexed enabling efficient search for similar fingerprints.

In the conventional content fingerprinting, the fingerprint is extracted directly from the original content and does not require any content modification to preserve the original content quality and integrity. In this sense, it can be considered as a *passive content fingerprinting* (pCFP).

Another approach to content protection and identification is based on *digital watermarking* (DWM). These days, DWM is a well-studied domain where a great deal of investigation was given to its performance [9], [10] and more recently its security [11] [1]. The essential difference between pCFP and DWM is that, in fingerprinting, a content owner only assigns some ID number to the content, while in digital watermarking one can mark every individual copy of the content by embedding a unique message or mark. DWM possesses two advantages over the pCFP: (a) each copy of a content can be marked independently and (b) there is no need for complex search procedures due to the usage of structured error correction codes (ECC), as there is with the random fingerprint in pCFP.

Meanwhile, the practical digital watermarking techniques face the problem of host interference. To cancel the host interference, special binning or quantization techniques such as quantization index modulation (QIM) are used that are demonstrated to be very insecure compared to the spread spectrum based methods which, in turn, suffer from host interference [12]. This recalls the trade-off that must exist between performance, expressed either in terms of probability of error or achievable rates and security, which is prone to leak due to an embedded message and a key.

Unfortunately, there is little known about the security of practical pCFP. With only a few exceptions, such as [13], the secret key estimation in the pCFP is not a well studied problem. However, it is obvious that since a content is not modified in the pCFP framework and if the database is handled properly,

Preliminary results from this work were presented in the IEEE International Workshop on Information Forensics and Security, 2012 [1].

F. Farhadzadeh and S. Voloshynovskiy are with the Department of Computer Science, University of Geneva, 7, Route de Drize, CH-1227 Carouge (GE), Switzerland (Email: {Farzad.Farhadzadeh, svolos}@unige.ch) and S. Voloshynovskiy is the corresponding author.

[1] We do not pretend here to provide an exhaustive overview of DWM achievements that is out of scope of this paper.

the attacker obtains much less information for the secret key estimation as opposed to digital watermarking. Furthermore, the pCFP does not require embedding of messages into host data. Thus there is no need to cancel host interference that leads to content quality degradation. This can be considered as an advantage of pCFP over DWM. Therefore, it is interesting to investigate new strategies in content identification that would benefit from the strengths of DWM and pCFP.

In this paper, we introduce a new hybrid technique that combines pCFP and DWM to achieve a better trade-off between performance, complexity and potential security. More particularly, we will address the performance of the proposed technique and investigate low-complexity identification strategies. We refer to this technique as *active content fingerprinting* (aCFP). The aCFP essentially obeys the structure of pCFP with the only difference occurring at the enrollment part, where both fingerprint and modified content are generated. We will extend the identification of standard pCFP to more elaborated strategies that benefit from the statistics of modulated contents.

It is important to point out that a content modulated by the aCFP does not carry out any embedded message. The sole purpose of this modulation at the enrollment stage is to increase the overall content identification system performance and reduce the complexity of identification.

In this paper, we consider several modulation strategies based on scalar and vector modulations. We will analyze a group of scalar modulation functions and highlight the connections to DWM. We also investigate a class of modulation functions that leads to enhanced identification performance for the fixed embedding distortion. Then we demonstrate possible identification strategies based on *bounded distance decoding* (BDD) that benefit from the proposed modulation strategies. Finally, we extend our analysis to the vector modulation based on lattice quantization. The reasons to consider the lattices are twofold. First, high dimensional lattices such as the Leech lattice [14] are well-known for their low distortion quantization due to the sphere packing properties. Secondly, due to the inherent structure of lattices, there exist several low-complexity decoding methods [15]. The goal of this paper is to evaluate the performance of aCFP under different modulation strategies to achieve better robustness and a potentially faster search than the conventional pCFP. In addition, we aim at comparing the aCFP with the DWM under the same distortion constraints. This paper is a further extension and generalization of conference work [1] that extends a systematic study of modulation functions, and an analysis of performance combining the results with the pCFP and DWM. The results of computer simulation compare both scalar and vector modulations on the databases of synthetic data and real images.

The paper is organized as follows. Section II contains the overview of aCFP versus pCFP and DWM. Section III briefly introduces the notions of lattices, lattice quantization and decoding, and then considers a common basis for the transform domain pCFP and DWM. Section IV formulates the conventional pCFP and provides the analysis of its performance in terms of the probability of bit error of binary fingerprints and also an anlysis of identification system performance in terms of identification accuracy and complexity. Section V introduces a generic framework of aCFP. A unidimensional aCFP, based on several modulation strategies, is compared with the conventional pCFP and DWM in Section VI . Section VII presents a multidimensional aCFP based on lattice quantization and decoding. The results of numerical modeling are presented in Section VIII . Section IX concludes the paper.

Notations: throughout this paper, we adopt the convention that a scalar random variable is denoted by a capital letter X, a specific value it may take is denoted by the lower case letter x, and its alphabet is designated by the script letter $\mathcal{X}$. As for vectors, a boldface capital letter $\mathbf{X}$ with a corresponding superscript will denote an N-dimensional random vector $\mathbf{X} = \{X_i\}_{i=1}^N$, a boldface lower case letter $\mathbf{x}$ will represent its particular realization $\mathbf{x} = \{x_i\}_{i=1}^N$, and the respective superalphabet is the N^{th} Cartesian power of $\mathcal{X}$, i.e., $\mathcal{X}^N$. $\mathbf{x}[i] = x_i$ stands for the i-th element of vector $\mathbf{x}$. $\mathbb{W} = \{\mathbf{w}_i\}_{i=1}^L$ denotes a matrix with columns $\mathbf{w}_i \in \mathcal{R}^N$. The expectation operator is designated by $\mathrm{E}[\cdot]$. $\mathcal{N}(\mu, \sigma^2)$ stands for the Gaussian distribution with mean μ and variance σ. $\mathcal{B}(L, p)$ denotes the Binomial distribution with L number of trials and probability success p. $\|\cdot\|_2$ stands for the Euclidean norm. $Q_L(\cdot)$ and $Q_\Lambda(\cdot)$ designate for a general vector quantization of dimension L and a lattice quantization with the lattice Λ, respectively. $\mathcal{Q}(x) = \int_{-\infty}^{x} \frac{1}{\sqrt{2\pi}} e^{-\frac{x^2}{2}} \mathrm{d}x$ indicates $\mathcal{Q}$-function. The Kullback-Leibler divergence between two distributions $p(x)$ and $q(x)$ on $\mathcal{X}$ is defined as, $\mathcal{D}(p(x)\|q(x)) = \sum_{x\in\mathcal{X}} p(x) \ln \frac{p(x)}{q(x)}$, with the conventions that $0 \ln 0 = 0$, and $p \ln \frac{p}{0} = \infty$ if $p > 0$.

II. aCFP versus pCFP and Digital watermarking

Generally a content identification system consists of two main stages: *content enrollment* or *marking* and *content identification*. In the following, we will elaborate the difference between content identification systems based on pCFP, DWM and aCFP at the afore-mentioned stages.

In a content identification system based on pCFP, shown in Fig.2a, a content owner provides the content $\mathbf{x} \in \mathcal{X}^N$ and the assigned ID number, ID $\in \mathcal{M}$, $\mathcal{M} = \{1, 2, \cdots, |\mathcal{M}|\}$. The content owner also possesses a secret key $k \in \mathcal{K}$, $\mathcal{K} = \{1, \cdots, |\mathcal{K}|\}$. At the enrollment stage, for a given content $\mathbf{x}$, the fingerprinting block (FP) generates a fingerprint $\mathbf{f}_\mathbf{x} \in \mathcal{F}^L$ for every input $\mathbf{x}$ and k. The generated fingerprint $\mathbf{f}_\mathbf{x}$ and the assigned ID are stored as a database entry related to the user identified by the key k. At the *identification stage*, for a given query $\mathbf{y}$, which might result either from the enrolled content $\mathbf{x}$ or unrelated one $\mathbf{x}' \in \mathcal{X}^N$, a digital fingerprint is extracted following the same approach as at the enrollment phase. Then, the decoder that has the access to the database determines whether or not $\mathbf{y}$ is related to any entry, and if so, to which one.

In a content identification system based on DWM, shown in Fig.2b, a content owner not only can assign an ID number to the content, but also can mark every individual copy of the content $\mathbf{x}$ by embedding a message or mark $\mathbf{m}$. For a given content $\mathbf{x}$, at the enrollment stage, the L character message $\mathbf{m} \in \mathcal{F}^L$ and the assigned ID are stored in the database. Prior to the embedding, the message $\mathbf{m}$ might be encoded

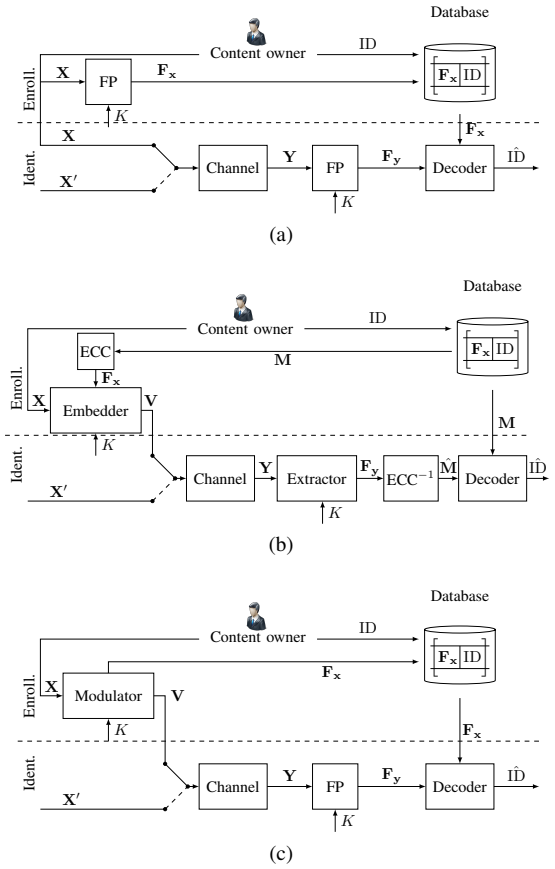


Fig. 1. Generalized models of (a) passive content fingerprinting, (b) digital watermarking and (c) active content fingerprinting.

into the codeword $\mathbf{f_x} \in \mathcal{F}^J, J > L$ using an error correction code (ECC). The codeword $\mathbf{f_x}$ is embedded into the content $\mathbf{x}$, a.k.a. a host, resulting into the marked copy $\mathbf{v}$ based on a secure key k. Obviously, the two operations of both message encoding and embedding can be combined together. However, to highlight the similarity with the pCFP and to reflect the way that most practical digital watermarking methods are designed, we separate these stages. At the identification stage, the extractor produces $\mathbf{f_y} \in \mathcal{F}^J$ based on $\mathbf{y}$ and k. The goal of the next stage is to decode the message $\hat{\mathbf{m}}$ from $\mathbf{f_x}$ using ECC^{-1}. Finally, the decoder of the identification system based on the estimated message $\hat{\mathbf{m}}$ and the database decides that the query $\mathbf{y}$ is related to any entry, and if yes, to which one.

The content identification system based on aCFP is shown in Fig.1c. The aCFP essentially obeys the structure of pCFP with the only difference being at the enrollment stage, where for a given content $\mathbf{x}$, the modulator generates the fingerprint $\mathbf{f_x}$ and the modified content $\mathbf{v}$ for a given key k and defined distortion between $\mathbf{x}$ and $\mathbf{v}$. The identification is performed exactly in exactly the same way as the content identification based on pCFP.

Summarizing the consideration of content identification techniques, one can conclude that:

(a) the main advantage of pCFP lies in its non-invasive type of identification while the principal disadvantage is an inability of pCFP to distinguish different copies of the same content, which, in fact, is not needed for all applications. An additional disadvantage of pCFP consists in high complexity of identification where the fingerprints resemble an analogy with random codes;

(b) the main advantage of DWM lies in its ability to mark each copy of content and the relatively low complexity identification that it achieves, thanks to the usage of structured ECC used for the encoding of embedded information and used modulation/demodulation techniques. The main disadvantage of DWM is the invasive character of embedding.

Leaving security aspects of both techniques aside and considering non-secure applications, a.k.a. assisting functionality, of both techniques used in the 2nd screen and physical object interaction, it is worth mentioning the *granularity* requirement. Granularity is often considered to be the minimum patch or block of data needed to perform a reliable content or object identification [8]. This requirement comes from the on-line identification and synchronization in the 2nd screen applications and object identification based on photos taken by mobile phones under macro-mode that have a limited field of view. The question of superiority of pCFP or DWM with respect to granularity in particular and performance in general is still an open problem [16] and some insights on the comparative study of both techniques would be of great practical importance.

In this respect, the aCFP might represent an interesting alternative to the pCFP and DWM in terms of embedding distortion and granularity needed to reliably identify the desired number of contents or objects with respect to the DWM or the complexity of identification with respect to the pCFP. In fact, some interesting insights can be drawn considering new design principles behind the aCFP.

Finally, aCFP can be considered as a joint source–channel coding problem, when a content is encoded in such a way as to generate the best channel code under given distortion constraints. In this case, the aCFP can be considered to be a sort of compression technique with the centroids representing codewords of channel code. This provides a close link to lattice codes [17] that will be considered in this paper.

III. Definitions and Preliminaries

In this section, we briefly review some basic principles behind the multidimensional quantization based on lattices and transform domain pCFP and DWM. This basis will be needed for the common analysis of both techniques and introduction of aCFP in unidimensional and multidimensional versions.

A. Lattices, Quantization, Lattice decoding

A lattice Λ is a discrete subgroup of the Euclidean space $\mathcal{R}^N$ with the ordinary vector addition operation. Thus, if λ_1 and λ_2 are in Λ, it follows that their sum and difference are also in Λ. A lattice Λ may be specified in terms of a generating matrix. Thus, an $N \times N$ real-valued matrix $\mathbb{G}$ defines a lattice Λ by:

$$\Lambda(\mathbb{G}) = \{\lambda : \lambda = \mathbb{G}\mathbf{n}, \mathbf{n} \in \mathcal{Z}^N\}. \quad (1)$$

That is, the lattice is generated by taking all integer linear combinations of the basis vectors.

A fundamental cell of a lattice Λ denoted by $\mathcal{P}$ is a region in $\mathcal{R}^N$ such that any $\mathbf{x} \in \mathcal{R}^N$ can be written as $\mathbf{x} = \mathbf{x}_0 + \mathbb{G}\mathbf{n}$ for a unique $\mathbf{x}_0 \in \mathcal{P}$ and $\mathbf{n} \in \mathcal{Z}^N$.

The *Voronoi region* of a lattice point $\lambda_0 \in \Lambda(\mathbb{G}) \subset \mathcal{R}^N$, denoted by $\mathcal{V}(\lambda_0)$, is the set of points that are nearer (with respect to Euclidean distance) to that point than to any other lattice point. That is:

$$\mathcal{V}(\lambda_0) = \left\{\mathbf{x} : \|\mathbf{x} - \lambda_0\|_2^2 \leq \|\mathbf{x} - \mathbb{G}\mathbf{n}\|_2^2, \forall \mathbf{n} \in \mathcal{Z}^N\right\}. \quad (2)$$

Due to symmetry of the lattice, the Voronoi region of the lattice Λ, will be denoted by $\mathcal{V}$ for convenience.

A lattice quantizer $Q_\Lambda(\cdot)$ with the lattice $\Lambda(\mathbb{G})$ and quantization basic cell $\mathcal{P}_0$ is a nonlinear mapping from $\mathcal{R}^N$ to $\Lambda(\mathbb{G})$ as given by the relation:

$$Q_\Lambda(\mathbf{x}) = \mathbb{G}\mathbf{n}, \quad (3)$$

where $\mathbf{n}$ is the unique vector satisfying:

$$\mathbf{x} = \mathbf{x}_0 + \mathbb{G}\mathbf{n}, \ \mathbf{x}_0 \in \mathcal{P}_0. \quad (4)$$

The normalized second moment of Λ and basic cell $\mathcal{P}$ is defined as:

$$G(\Lambda, \mathcal{P}) = \frac{1}{N} \frac{\int_{\mathcal{P}} \|\mathbf{x}\|^2 d\mathbf{x}}{V^{1+2/N}} \quad (5)$$

where $V = \det(\mathbb{G})$ is the volume of basic cell $\mathcal{P}$.

Let $\mathbf{U}$ be a random vector statistically independent of the input vector $\mathbf{x}$. Adding this so-called dither before quantization and subtracting it after, we have the subtractive dithered lattice quantization. The error vector is $\boldsymbol{\zeta} = \mathbf{x} - (Q_\Lambda(\mathbf{x}+\mathbf{u}) - \mathbf{u}) = \mathbf{x}+\mathbf{u}-Q_\Lambda(\mathbf{x}+\mathbf{u})$. In the subtractive quantization scheme, the error vector $\boldsymbol{\zeta}$ is statistically independent of the input vector $\mathbf{x}$ and uniformly distributed in $\mathcal{P}_0$ if and only if the dither $\mathbf{u}$ is Nyquist-$\mathbb{G}$, that is $\Phi_{\mathbf{U}}(\mathbb{H}\mathbf{n}) = \delta(\mathbf{n})$, where $\mathbb{H} = 2\pi\mathbb{G}^{-T}$ and $\Phi_{\mathbf{U}}(\cdot)$ is the characteristic function of random vector $\mathbf{U}$ [18].

In the subtractive dithered lattice quantization, we have:

$$\mathrm{E}\left[\|\boldsymbol{\zeta}\|_2^2\right] = NG(\Lambda, \mathcal{P}_0)V^{2/N}, \quad (6)$$

due to the fact that the quantization error $\boldsymbol{\zeta}$ is uniformly distributed over $\mathcal{P}_0$.

B. A Common Basis for Secret Transform Domain Systems

Since most fingerprinting and digital watermarking systems operate in some transform domain, we define the direct and inverse transforms as:

$$\begin{cases} \widetilde{\mathbf{x}} = \mathbb{W}\mathbf{x}, \\ \mathbf{x} = \mathbb{W}^{-1}\widetilde{\mathbf{x}}, \end{cases} \quad (7)$$

for the orthonormal matrix $\mathbb{W}^{-1} = \mathbb{W}^T$.

We will assume that the matrix $\mathbb{W} \in \mathcal{R}^{N\times N}$, $\mathbb{W} = \{\mathbf{w}_i\}_{i=1}^N$ consists of a set of basis vectors $\mathbf{w}_i \in \mathcal{R}^N, 1 \leq i \leq N$. In the part of theoretical analysis, we will assume that this transform is based on any randomized orthogonal matrix $\mathbb{W}$, a.k.a. random projection (RP) transform whose elements are equal likely $w_{i,j} = \{\pm\frac{1}{\sqrt{N}}\}$ based on the secret key k. Such a matrix can be considered as an approximate *orthoprojector*, for which $\mathbb{W}\mathbb{W}^T \approx \mathbb{I}_N$, where $\mathbb{I}_N$ is the unit matrix of size N, and the basis vectors are of a unit norm [6]. The matrix $\mathbb{W}$ can also represent some known types of transforms such as DFT, DCT or DWT.

In this paper, we will use an alternative representation of direct and inverse transforms (7):

$$\begin{cases} \tilde{x}_i = \mathbf{w}_i^T\mathbf{x}, \ 1 \leq i \leq N, \\ \mathbf{x} = \sum_{i=1}^N \tilde{x}_i\mathbf{w}_i = \sum_{i\in\mathcal{K}} \tilde{x}_i\mathbf{w}_i + \sum_{i\notin\mathcal{K}} \tilde{x}_i\mathbf{w}_i, \end{cases} \quad (8)$$

where a set $\mathcal{K} = \{i_1, i_2, \cdots, i_L\}$ represents a set of indices defined by the secret key k. Here, we define the input fingerprint vector $\widetilde{\mathbf{x}}^* = \{\tilde{x}_i\}_{i\in\mathcal{K}}$ as the collection of the selected projection outputs, which will be used further in the fingerprint extraction procedure. This representation corresponds to the direct generation of a set of secret basis vectors or carriers $\mathbf{w}_i$, $i \in \mathcal{K}$, where the fingerprint is computed. This is also closely related to DWM techniques based on *spread spectrum* (SS), *spread transform* (ST) watermarking [19] and *subspace projections* (SSP) [20].

The selection of secret carriers is very important for both the robustness of the scheme and its security as well as statistics of projected coefficients. Therefore, without loss of generality, in our further analysis we will consider only carriers with the disjoint supports due to their ability to generate independent projected coefficients and wide usage in practical fingerprinting and watermarking methods. It should be pointed out that the selection of random basis vectors will also lead to the decorrelation of signal components and results in Gaussian statistics of projected coefficients for certain classes of image models as shown in [6].

IV. Conventional passive content fingerprinting

A. Model of Passive Content Fingerprinting

In case of the conventional pCFP, the original content or its extracted feature vector is not modified and the fingerprint is computed directly from $\widetilde{\mathbf{x}}^*$. The fingerprint extraction procedure, considered in this paper, consists of two steps: mapping $\mathbf{x}$ to the random space applying RP [2] and constructing the input fingerprint vector $\widetilde{\mathbf{x}}^*$ by (8) and finally quantizing $\widetilde{\mathbf{x}}^*$ using a quantizer $Q_L(.)$ that results into a fingerprint $\mathbf{f}_\mathbf{x} = Q_L(\widetilde{\mathbf{x}}^*)$. In the unidimensional case, the quantizer $Q_L(.)$ is represented by a product of scalar quantizers that are often reduced to one-level scalar quantization:

$$\mathbf{f}_\mathbf{x} = Q_1(\tilde{x}_1) \circ Q_1(\tilde{x}_2) \circ \cdots \circ Q_1(\tilde{x}_L), \quad (9)$$

where $Q_1(\tilde{x}) = \mathrm{sign}(\tilde{x})$, "$\circ$" denotes Cartesian product, $\tilde{x}_i = \mathbf{w}_i^T\mathbf{x}$, $i \in \mathcal{K}$ and $\mathrm{sign}(\tilde{x}) = +1$ for $\tilde{x} \geq 0$ and -1, otherwise. The multidimensional case corresponds to the vector version of $Q_L(\cdot)$ and more practically to lattice quantization $Q_\Lambda(\cdot)$ that will be further considered in our analysis.

Assuming an additive noise observation channel [3]:

$$\mathbf{y} = \mathbf{x} + \mathbf{z}, \quad (10)$$

one is interested to estimate the level of degradation to the fingerprint extracted from the degraded content $\mathbf{y}$. For the

[2]Although, in this paper we do not consider geometrical invariance, one can extend the scheme to the case where RP is applied to a set of invariant features or patches around invariant robust points.

[3]The additive model of distortions can be also assumed for some robust feature extraction domains including SIFT descriptors, where the Euclidean metric or the Mahalanobis distance, which are the ML counterparts for the i.i.d. additive and correlated Gaussian noise, respectively, are used for the descriptor matching [21].

purpose of comparison, we will consider the performance of all methods under the additive white Gaussian noise.

To quantify the level of signal distortions we will use the *document-to-noise ratio* (DNR) defined as:

$$\text{DNR} = 10\log_{10}\left(\frac{\frac{1}{N}\mathrm{E}[\|\mathbf{X}\|_2^2]}{\frac{1}{N}\mathrm{E}[\|\mathbf{Z}\|_2^2]}\right) = 10\log_{10}\left(\frac{\sigma_X^2}{\sigma_Z^2}\right), \quad (11)$$

where we assumed that all signals are zero-mean Gaussian vectors, i.e., $\mathbf{X} \sim \mathcal{N}(\mathbf{0}, \sigma_X^2 \mathbb{I}_N)$ with the variance σ_X^2 and noise is also zero-mean Gaussian, $\mathbf{Z} \sim \mathcal{N}(\mathbf{0}, \sigma_Z^2 \mathbb{I}_N)$ with the variance σ_Z^2.

Given a secret key k and the corresponding set of secret carriers, the query fingerprint extraction in the scalar case is performed as[4]:

$$f_{y_i} = \psi(\mathbf{w}_i^T \mathbf{y}) = \text{sign}(\mathbf{w}_i^T \mathbf{y}) = \text{sign}(\tilde{x}_i + \tilde{z}_i),\ i \in \mathcal{K}, \quad (12)$$

where $\Psi(\cdot)$ denotes the fingerprint extraction function, and $\tilde{z}_i = \mathbf{z}^T \mathbf{w}_i$ for $i \in \mathcal{K}$. The resulting query fingerprint is $\mathbf{f_y} = \{f_{y_i}\}_{i=1}^L$. The projected original content, noise and distorted coefficients are distributed i.i.d. as $\tilde{X} \sim \mathcal{N}(0, \sigma_X^2)$, $\tilde{Z} \sim \mathcal{N}(0, \sigma_Z^2)$ and $\tilde{Y} \sim \mathcal{N}(0, \sigma_X^2 + \sigma_Z^2)$.

The performance of pCFP in terms of probability of bit error rate (BER) is measured as [23]:

$$P_{b-\text{pCFP}} = \frac{1}{L}\sum_{i=1}^{L} \Pr\{F_{x_i} \neq F_{y_i}\} = \Pr\{F_{\mathbf{x}} \neq F_{\mathbf{y}}\} \quad (13)$$

$$= \mathrm{E}_{p(|\tilde{x}|)}\left[P_{b||\tilde{x}|}\right] = \frac{1}{\pi}\arctan\left(\frac{\sigma_Z}{\sigma_X}\right), \quad (14)$$

where $P_{b||\tilde{x}|} = \mathcal{Q}\left(\frac{|\tilde{X}|}{\sigma_Z}\right)$ stands for the BER for the coefficient with the magnitude $|\tilde{x}|$.

Remark 1. The probability of bit error $P_{b||\tilde{x}|}$ depends on the magnitude of $|\tilde{x}|$. Therefore, the components with the higher magnitude have lower probability of bit error. Thus, the channel between $F_{\mathbf{x}}$ and $F_{\mathbf{y}}$ is a parametric one or channel with the state defined by $|\tilde{x}|$ [23], [24]. In the case of the decoder neglecting the information about the channel state, the average probability of error (14) is used.

Bit reliability is of great importance for this paper and therefore, we schematically show this concept in Figure 2. The magnitude $|\tilde{x}_i|$ depends on the angle between the vectors $\mathbf{x}$ and $\mathbf{w}_i$. The reliable bit corresponds to almost collinear vectors, while an unreliable one is typical for almost orthogonal vectors. In the case of an unreliable bit, even small perturbation $\mathbf{z}$ to $\mathbf{x}$ might change the sign of $\tilde{x}_i$ resulting in the bit error, i.e., $\text{sign}(\tilde{x}_i) \neq \text{sign}(\tilde{y}_i)$. A reliable bit demonstrates a higher robustness to distortions satisfying the condition $\text{sign}(\tilde{x}_i) = \text{sign}(\tilde{y}_i)$.

The content identification based on binary fingerprints essentially resembles the decoding of random codewords. Unfortunately, the fingerprints do not obey any structure that can be used for efficient decoding. Therefore, generally brute force decoding is used based on the computation of Hamming distances between the probe fingerprint $\mathbf{f_y}$ and the codebook composed of $\mathbf{f_x}(m)$, $1 \leq m \leq |\mathcal{M}|$ [6], [25]. In this case, it can be demonstrated that the number of uniquely recognizable items is bounded as $|\mathcal{M}| \leq 2^{LC_{id}}$, where $C_{id} = \frac{1}{L} I(\mathbf{F_x}; \mathbf{F_y}) = \frac{1}{L}(1 - H_2(P_{b-\text{pCFP}}))$, where C_{id} denotes the identification capacity, $I(\cdot;\cdot)$ stands for the mutual information and $H_2(p) = -p\log_2 p - (1-p)\log_2(1-p)$ represents binary entropy [6], [26].

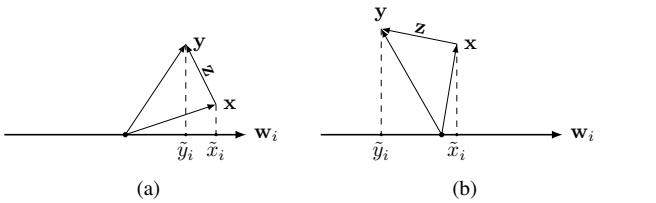


Fig. 2. The bit reliability for the model $\mathbf{y} = \mathbf{x} + \mathbf{z}$: (a) reliable bit and (b) unreliable bit.

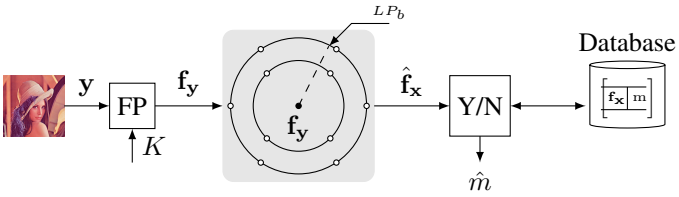


Fig. 3. Identification based on BDD: the estimate of fingerprint is generated in the Hamming sphere around the probe and sequentially tested versus the ordered database. Once the exact match is found, testing is terminated and a match is declared as a decoded index. Otherwise, "no match" is declared when all codewords within the Hamming sphere are tested.

An essential reduction of decoding complexity can be achieved for the low-values of $P_{b-\text{pCFP}}$, i.e., in the high DNR regime, using the BDD [23]. The identification based on the BDD is shown in Figure 3. It was shown that the estimate of true codeword $\hat{\mathbf{f}}_{\mathbf{x}} = \mathbf{f_x}(m)$ based on $\mathbf{f_y}$ can be found in about $\mathcal{O}(2^{LH_2(P_{b-\text{pCFP}})})$ trials. This corresponds to the number of codewords in the Hamming sphere defined by $2^{LH_2(F_x|F_y)}$. The complexity of this decoder is still exponential. However, its complexity is essentially lower than those of brute force decoder $\mathcal{O}(2^{L(1-H_2(P_{b-\text{pCFP}}))})$, if $P_b \leq 0.11$. It is also possible to further reduce the complexity of BDD to $\mathcal{O}(2^{LP_{b-\text{pCFP}}})$, if the BDD will use soft information about the bit reliability. The use of soft information also leads to a new estimate of identification capacity $C_{id} = \frac{1}{L} I(\mathbf{F_x}; \tilde{\mathbf{Y}}^*) \geq \frac{1}{L} I(\mathbf{F_x}; \mathbf{F_y})$ [23], [24].

Remark 2. The complexity of the BDD depends on the probability of bit error. Thus, the set of secret carriers $\{\mathbf{w}_i\}$, $i \in \mathcal{K}$ might be optimized to reduce overall $P_{b-\text{pCFP}}$. However, such an optimization should be performed over all $\mathbf{x}(m), 1 \leq m \leq |\mathcal{M}|$ which is quite a complex problem. Otherwise, the optimal $\{\mathbf{w}_i\}, i \in \mathcal{K}$ should be communicated as side information or tested exhaustively at the decoder. At the same time, the observation that the components with higher magnitude $|\tilde{x}_i|$ have lower $P_{b||\tilde{x}_i|}$ motivates us to consider an "active" counterpart where the coefficients are modulated within the range of allowable distortions to achieve the desirable decrease of bit error rate by increasing the magnitudes of coefficients. In turn, this will also lead to lower decoding complexity.

To underline the importance of $P_{b-\text{pCFP}}$ on both identification system performance and complexity, we will consider the performance of the identification system in terms of the

[4]In this work we do not consider *soft fingerprinting* where the extracted fingerprint also contains additional information about bit reliabilities [22], [23], [24].

probability of correct identification P_{ci} and probability of false acceptance P_{fa} for the BDD.

The probability of correct identification P_{ci} is defined as:

$$P_{ci} = 1 - P_m = \sum_{m=1}^{|\mathcal{M}|} \Pr\{d(\mathbf{F_y}, \mathbf{f_x}(m)) \leq \theta L \cap \bigcap_{m' \neq m} d(\mathbf{F_y}, \mathbf{f_x}(m')) > \theta L | \mathcal{H}_m\} \Pr\{\mathcal{H}_m\}, \quad (15)$$

where P_m denotes the probability of miss, $\theta \geq 0$ is a threshold and L is the length of fingerprint, $\mathcal{H}_m$ corresponds to the case that $\mathbf{F_y}$ is related to the m^{th} entry of the database, $d(\cdot,\cdot)$ denotes a decoding metric, and we assume that all entries will be queried with the same probability, i.e., $\forall m, \Pr\{\mathcal{H}_m\} = \frac{1}{M}$.

The probability of false acceptance is defined as:

$$P_{fa} = \sum_{m=1}^{|\mathcal{M}|} \Pr\{d(\mathbf{F_y}, \mathbf{f_x}(m)) \leq \theta L | \mathcal{H}_0\},$$

where $\mathcal{H}_0$ corresponds to the case when $\mathbf{F_y}$ is unrelated to any entry in the database.

In the case of binary pCFP, the decoding metric $d(\cdot,\cdot)$ corresponds to the Hamming distance $d_H(\cdot,\cdot)$ [6] and:

$$P_{ci-\text{pCFP}} = \frac{1}{|\mathcal{M}|} \sum_{m=1}^{|\mathcal{M}|} \Pr\left\{ d_H(\mathbf{f_x}(m), \mathbf{F_y}) \leq \theta L \cap \bigcap_{m' \neq m} d_H(\mathbf{f_x}(m'), \mathbf{F_y}) > \theta L \middle| \mathcal{H}_m \right\} \overset{(a)}{=} \sum_{d=0}^{\theta L} \binom{L}{d} \epsilon^d (1-\epsilon)^{L-d} \left[\frac{1}{2^L} \sum_{j=d+1}^{L} \binom{L}{j} \right]^{|\mathcal{M}|-1}, \quad (16)$$

where $\epsilon = P_{b-\text{pCFP}}$ is the cross-over probability of the binary symmetric channel between $\mathbf{f_x}(m)$ and $\mathbf{f_y}$ under $\mathcal{H}_m$, (a) follows from the fact that d_H follows $\mathcal{B}(L,\epsilon)$ under $\mathcal{H}_m$, and follows $\mathcal{B}(L, 0.5)$ for other contents.

The probability of false acceptance yields [6]:

$$P_{fa} = \Pr\left\{ \bigcup_{m=1}^{|\mathcal{M}|} d_H(\mathbf{f_x}(m), \mathbf{F_y}) \leq \theta L \middle| \mathcal{H}_0 \right\} = 1 - \left[1 - \left(\tfrac{1}{2}\right)^L \sum_{x=0}^{\theta L} \binom{L}{x} \right]^{|\mathcal{M}|}, \quad (17)$$

where (a) follows from the fact that d_H follows $\mathcal{B}(L, 0.5)$ under $\mathcal{H}_0$.

Using Chernoff bound, one can bound the probabilities of miss P_m and false acceptance for $|\mathcal{M}| = 2^{LR}$ and any $\epsilon \leq \theta \leq \frac{1}{2}$ as follows [6]:

$$P_m \leq \Pr\{d_H(\mathbf{f_x}(m), \mathbf{F_y}) > \theta L\} + \Pr\left\{ \bigcup_{m' \neq m} d_H(\mathbf{f_x}(m'), \mathbf{F_y}) \leq \theta L \right\} \leq 2^{-L\mathcal{D}(\theta \| \epsilon)} + 2^{-L[\mathcal{D}(\theta \| 0.5) - R]},$$

and,

$$P_{fa} \leq 2^{-L[\mathcal{D}(\theta \| 0.5) - R]}.$$

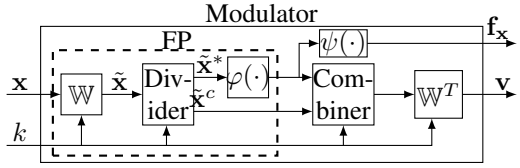


Fig. 4. Modulator and Fingerprinting (FP) block, where $\widetilde{\mathbf{x}}^c = \{\tilde{x}_i\}_{i \notin \mathcal{K}}$ and $\widetilde{\mathbf{x}}^* = \{\tilde{x}_i\}_{i \in \mathcal{K}}$.

Remark 3. The performance of a content identification system is determined by the probability of bit error ϵ, i.e., $P_{b-\text{pCFP}}$ in the case of pCFP. Therefore, similarly to the complexity based on the BDD, the reduction of probability of error is an essential element of reliable system performance.

Finally, the pCFP in the multidimensional case corresponds to the replacement of the scalar quantizer by its vector counterpart $Q_\Lambda(\widetilde{\mathbf{x}}^*)$.

In the following section, we introduce a framework for the reduction of probability of bit error using content modulation in the domain of secret carriers, which we refer to as aCFP.

V. Active content fingerprinting

In the aCFP, we will consider the modification of the original content in the space of secret carriers defined by the key k with the overall goal to improve the performance of identification in terms of the probabilities of correct identification and false acceptance and reducing complexity. For this purpose, we define a general form of aCFP modulator, shown in Fig. 4:

$$\mathbf{v} = \sum_{i \in \mathcal{K}} \varphi(\widetilde{\mathbf{x}}^*)[i] \mathbf{w}_i + \sum_{i \notin \mathcal{K}} \tilde{x}_i \mathbf{w}_i, \quad (18)$$

where $\mathbf{v}$ is the modified content using the inverse transform, $\widetilde{\mathbf{x}}^* = \{\tilde{x}_i\}_{i \in \mathcal{K}}$ denotes the vector of projected coefficients in the secret subspace defined by the secret key k, $\varphi(\widetilde{\mathbf{x}}^*)$ is the modulation function applied to the vector $\widetilde{\mathbf{x}}^*$ and $\varphi(\widetilde{\mathbf{x}}^*)[i]$ implies the i-th element of the modulation output [5]. As illustrated in Fig. 4, the fingerprint extraction procedure is similar to pCFP. The introduction of modulation leads to the distortion. Therefore, we define the measure of distortion as follows.

Definition: D is the *distortion measure per dimension* between sequences $\mathbf{x}$ and $\mathbf{v}$, defined by:

$$D = \tfrac{1}{N} \mathrm{E}\left[\|\mathbf{V} - \mathbf{X}\|_2^2 \right] = \tfrac{1}{N} \mathrm{E}\left[\|\mathbf{S}\|_2^2 \right], \quad (19)$$

where $\mathbf{S} = \sum_{i \in \mathcal{K}} \left(\varphi(\widetilde{\mathbf{X}}^*)[i] - \tilde{X}_i \right) \mathbf{w}_i$ denotes the modulation signal that can be considered as a kind of watermark by analogy to digital watermarking. In the case of pCFP, $D = 0$ while aCFP will be characterized by the distortion determined by the modulation function $\varphi(\cdot)$.

Also, for coherence with digital watermarking, we will define the *document-to-watermark ratio* (DWR) as:

$$\text{DWR} = 10\log_{10}\left(\frac{\frac{1}{N}\mathrm{E}[\|\mathbf{X}\|_2^2]}{\frac{1}{N}\mathrm{E}[\|\mathbf{S}\|_2^2]} \right) = 10\log_{10}\left(\frac{\sigma_X^2}{D} \right), \quad (20)$$

[5] We do not consider here the key-dependent modulation $\varphi(\widetilde{\mathbf{x}}^*, k)$ assuming that the key k is taken into account in the selection of secret carriers. Additionally, the secret key k can be taken into account in the generalization of modulation functions in the form of a secret dither.

which should reflect the existence of content modification by the embedded "watermark" or actually the modulation signal $\mathbf{S}$. Obviously, there is an important difference between the watermark, which bears the information about the content owner, and the modulation signal, which is solely used for the BER reduction. Therefore, it would be more correct to use the term *document-to-modulation signal ratio*. However, to make the comparison with the digital watermarking consistent, we will use the DWR, assuming that the reader can clearly identify the difference between both techniques.

Similarly to the pCFP, we consider the additive noise observation channel, in the context of theoretical performance analysis:

$$\mathbf{Y} = \mathbf{V} + \mathbf{Z}, \tag{21}$$

where $\mathbf{Z}$ denotes the channel distortion, when an estimate is desired of the level of degradations to the fingerprint extracted from the degraded content $\mathbf{Y}$.

Throughout the theoretical analysis, we assume that a content can be modeled as a Gauss-Markov process with variance σ_X^2 [27] and evaluate the performance of all methods under the additive white Gaussian noise (AWGN) with variance σ_Z^2.

Given a secret key k and the corresponding set of secret carriers, the query fingerprint extraction is performed as:

$$\mathbf{f}_{\mathbf{y}} = \psi(\widetilde{\mathbf{y}}^*) = \psi(\varphi(\widetilde{\mathbf{x}}^*) + \widetilde{\mathbf{z}}^*), \tag{22}$$

where $\widetilde{\mathbf{z}}^*$ is a collection of projected coefficients of $\widetilde{\mathbf{z}}$ in a secret subspace defined by k. As shown in [6], applying random projections defined in (8) on the data, which can be modeled by a Gauss-Markov process, makes projected coefficients uncorrelated. Consequently, one can assume that projected coefficients in the secret subspace follow i.i.d. Gaussian distributions, i.e., $\widetilde{\mathbf{Z}}^* \sim \mathcal{N}(\mathbf{0}, \sigma_Z^2 \mathbb{I}_L)$ and $\widetilde{\mathbf{X}}^* \sim \mathcal{N}(\mathbf{0}, \sigma_X^2 \mathbb{I}_L)$.

VI. Unidimensional case of aCFP

In the following discussion, we consider several modulation strategies in terms of their performance and distortions. To establish the link to the DWM, we will consider additive and quantization aCFPs as counterparts of spread spectrum and QIM DWM methods. Then we will introduce an efficient modulation strategy, referred to as "shrinkage based aCFP", that minimizes the embedding distortion. Finally, all techniques will be compared in terms of probabilities of bit errors.

A. Additive active content fingerprinting

Definition: *additive active content fingerprinting* (AddaCFP) is defined by the scalar modulation function of the form:

$$\varphi_A(\tilde{x}_i) = \tilde{x}_i + \alpha \operatorname{sign}(\tilde{x}_i), \tag{23}$$

for $i \in \mathcal{K}$, where $\alpha > 0$ stands for the strength of aCFP. Substituting (47) into (18) yields:

$$\begin{aligned}\mathbf{v} &= \sum_{i\in\mathcal{K}} \left(\tilde{x}_i + \alpha \operatorname{sign}(\tilde{x}_i)\right)\mathbf{w}_i + \sum_{i\notin\mathcal{K}} \tilde{x}_i \mathbf{w}_i \\ &= \mathbf{x} + \alpha \sum_{i\in\mathcal{K}} \operatorname{sign}(\tilde{x}_i)\mathbf{w}_i = \mathbf{x} + \mathbf{s}_A,\end{aligned} \tag{24}$$

where $\mathbf{s}_A = \alpha \sum_{i\in\mathcal{K}} \operatorname{sign}(\tilde{x}_i)\mathbf{w}_i$.

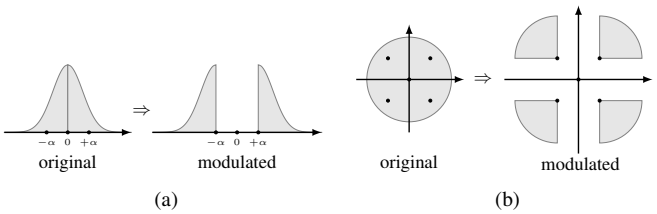


Fig. 5. AddaCFP modulation: impact on the pdf of projected coefficients: (a) one dimensional and (b) two dimensional modulation.

The AddaCFP modulation is shown in Fig. 5 for unidimensional and two-dimensional cases. The distortion of AddaCFP per content sample is:

$$D_A = \tfrac{1}{N}\mathrm{E}\left[\|\mathbf{S}_A\|_2^2\right] \overset{(a)}{=} \tfrac{L}{N}\alpha^2, \tag{25}$$

where (a) follows from the fact that $\tilde{X}_i$ are i.i.d.. Consequentially, the DWR under the AddaCFP is $\mathrm{DWR}_{\mathrm{Add}} = 10\log_{10}\left(\frac{N}{L}\frac{\sigma_X^2}{\alpha^2}\right)$.

The fingerprint extraction for the enrollment can be performed based on $\mathbf{v}$ as:

$$f_{v_i} = \psi(\mathbf{w}_i^T\mathbf{v}) = \operatorname{sign}(\mathbf{w}_i^T\mathbf{v}) = \operatorname{sign}(\tilde{x}_i + \alpha\operatorname{sign}(\tilde{x}_i)),$$

where $i \in \mathcal{K}$. The fingerprint computed at the verification stage is:

$$f_{y_i} = \psi(\mathbf{w}_i^T\mathbf{y}) = \operatorname{sign}(\mathbf{w}_i^T\mathbf{y}) = \operatorname{sign}(\tilde{x}_i + \alpha\operatorname{sign}(\tilde{x}_i) + \tilde{z}_i),$$

where $i \in \mathcal{K}$ and $\mathbf{y} = \mathbf{v} + \mathbf{z}$ (Fig. 1c).

The performance of AddaCFP in terms of BER is given by:

$$\begin{aligned}P_{b-\text{AddaCFP}} &= \mathrm{E}_{p(\varphi_A(\tilde{X}))}\left[\mathcal{Q}\left(\tfrac{|\varphi_A(\tilde{X})|}{\sigma_Z}\right)\right] \overset{(a)}{=} 2\int_0^\infty \mathcal{Q}\left(\tfrac{\tilde{x}+\alpha}{\sigma_Z}\right) \\ &\tfrac{1}{\sqrt{2\pi\sigma_X^2}}\exp\left(\tfrac{-\tilde{x}^2}{2\sigma_X^2}\right)\mathrm{d}\tilde{x} \overset{(b)}{\leq} \exp\left(-\tfrac{\alpha^2}{2\sigma_Z^2}\right) P_{b-\mathrm{pCFP}},\end{aligned} \tag{26}$$

where (a) results from $\varphi_A(\tilde{X})$ which follows the following pdf:

$$p(\varphi_A(\tilde{x})) = \begin{cases}\mathcal{N}(\operatorname{sign}(\tilde{x})|\alpha|, \sigma_X^2), & |\tilde{x}| \geq \alpha, \\ 0, & |\tilde{x}| < \alpha,\end{cases} \tag{27}$$

and (b) follows from the inequality $\mathcal{Q}(x+t) \leq \exp(-\frac{t^2}{2})\mathcal{Q}(x)$ for $x, t \geq 0$.

Remark 4. The difference between the BER of pCFP (14) and AddaCFP is in the positive bias α introduced by the additive modulation that reduces BER by at least the factor of $\exp(-\frac{\alpha^2}{2\sigma_Z^2})$ achieved at the cost of embedding distortion $D_A = \frac{L}{N}\alpha^2$.

Remark 5 (Link to improved spread spectrum (ISS) watermarking)**.** The ISS is a watermarking method aiming at the host interference cancellation in the projected domain with the embedding rate $R_{\mathrm{DW}} = \frac{L}{N}$ bits [28][6]:

$$\mathbf{v} = \mathbf{x} + \nu \sum_{i\in\mathcal{K}} f_{x_i}\mathbf{w}_i - \lambda \sum_{i\in\mathcal{K}} \tilde{x}_i \mathbf{w}_i, \tag{28}$$

where ν and λ control the strength of the watermark and host cancellation, respectively, $f_{x_i} = (-1)^{m_i}$ is considered as a

[6]The presented multi-bit formulation of ISS has two differences with the originally proposed one-bit ISS: the third term is not normalized by $\|\mathbf{w}\|^2$ and L-bit embedding is considered.

modulation feature vector of a watermark and $m_i \in \{0,1\}$. The notation f_{x_i} is introduced by purpose to reflect the link with the extracted bits in the unidimensional pCFP. Under the assumption of unit norm basis vectors, the embedding distortion is:

$$D{=}\tfrac{L}{N}\left(\nu^2+\lambda^2\sigma_X^2\right). \tag{29}$$

It is not difficult to trace the link between the AddaCFP and ISS despite the different objectives behind both techniques. Since the AddaCFP does not target any data hiding, one can disregard the watermark embedding component by setting $\nu = 0$. In the ISS (28), the goal of the term $\lambda\sum_{i\in\mathcal{K}}\tilde{x}_i\mathbf{w}_i$ is to suppress the interference with the host components in the projected domain. The ISS-based modulation is only efficient for large values of $\tilde{x}_i$, since they represent the highest interference to the watermark. However, in the pCFP small-value coefficients represent the main source of bit errors due to the sign flipping. That is why, contrary to the host interference cancellation in the digital watermarking, the AddaCFP increases the small-value coefficients as in (24).

Finally, the BER of ISS corresponds to the mismatch of embedded and extracted bits and is defined as [28]:

$$P_{b-\text{ISS}} = Q\left(\sqrt{\tfrac{(\alpha^2-\lambda^2\sigma_X^2)}{\sigma_Z^2+(1-\lambda^2)\sigma_X^2}}\right), \tag{30}$$

with the optimal $\lambda_{\text{opt}} = \frac{1}{2}\left(1+\frac{\sigma_Z^2}{\sigma_X^2}+\frac{\alpha^2}{\sigma_X^2}\right) - \frac{1}{2}\sqrt{\left(1+\frac{\sigma_Z^2}{\sigma_X^2}+\frac{\alpha^2}{\sigma_X^2}\right)^2-4\frac{\alpha^2}{\sigma_X^2}}$ which minimizes the above BER for the embedding distortion $D_{\text{ISS}} = D_{\text{A}}$.

For the purpose of comparison, we also consider SS-based watermarking which suffers from host interference and can be obtained from the ISS by assigning $\lambda = 0$ that results in:

$$P_{b-\text{SS}} = Q\left(\sqrt{\tfrac{\alpha^2}{\sigma_Z^2+\sigma_X^2}}\right). \tag{31}$$

In summary, one can note that the embedding distortion of the ISS is the necessary cost for both host interference cancellation and watermark embedding itself. The essential amount of the embedding distortion is spent for the host interference cancellation of large magnitude coefficients and it does not contribute to the energy of the information carrier, i.e., watermark. Contrarily, the AddaCFP only increases the magnitudes of coefficients that combinedly work for the decrease of overall BER.

B. Quantization based active content fingerprinting

Definition: *quantization-based active content fingerprinting* (QbaCFP) is defined by the modulation function of form:

$$\begin{aligned}\varphi_Q(\tilde{x}_i) &= c\,\text{sign}(\tilde{x}_i) = \tilde{x}_i + c\,\text{sign}(\tilde{x}_i) - \tilde{x}_i \\ &= \tilde{x}_i + (c-|\tilde{x}_i|)\text{sign}(\tilde{x}_i).\end{aligned} \tag{32}$$

Substituting (32) into (18) yields:

$$\begin{aligned}\mathbf{v} &= \sum_{i\in\mathcal{K}}(\tilde{x}_i + (c-|\tilde{x}_i|)\text{sign}(\tilde{x}_i))\mathbf{w}_i + \sum_{i\notin\mathcal{K}}\tilde{x}_i\mathbf{w}_i \\ &= \mathbf{x} + \sum_{i\in\mathcal{K}}(c-|\tilde{x}_i|)\text{sign}(\tilde{x}_i)\mathbf{w}_i = \mathbf{x}+\mathbf{s}_Q,\end{aligned} \tag{33}$$

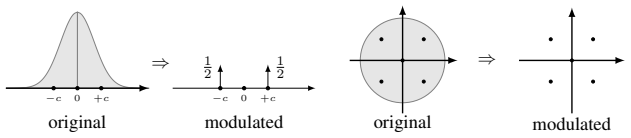


Fig. 6. QbaCFP modulation: impact on the pdf of projected coefficients (a) unidimensional and (b) two-dimensional modulation.

where $\mathbf{s}_Q = \sum_{i\in\mathcal{K}}(c-|\tilde{x}_i|)\text{sign}(\tilde{x}_i)\mathbf{w}_i$.

The QbaCFP modulation is shown in Fig. 6 for unidimensional and two-dimensional cases.

The distortion of QbaCFP is:

$$\begin{aligned}D_Q &= \tfrac{1}{N}\text{E}_{p(\tilde{x}_i)}\left[\|\mathbf{S}_Q\|_2^2\right] = \tfrac{L}{N}\text{E}_{p(|\tilde{x}_i|)}\left[\left(|\tilde{X}|-c\right)^2\right] \\ &\overset{(a)}{=} \tfrac{L}{N}\left(\sigma_X^2 - 2c\sigma_X\sqrt{\tfrac{2}{\pi}}+c^2\right),\end{aligned} \tag{34}$$

where (a) follows from $\text{E}_{p(|\tilde{x}|)}\left[|\tilde{X}|\right] = \sigma_X\sqrt{\frac{2}{\pi}}$, and $|\tilde{X}|$ follows the half-normal distribution.

Finally, the BER of QbaCFP is given by:

$$P_{b-\text{QbaCFP}} = \text{E}_{p(\varphi_Q(\tilde{X}))}\left[\mathcal{Q}\left(\tfrac{|\varphi_Q(\tilde{X})|}{\sigma_Z}\right)\right] \overset{(a)}{=} \mathcal{Q}\left(\tfrac{c}{\sigma_Z}\right), \tag{35}$$

where (a) follows from:

$$p(\varphi_Q(\tilde{x})) = \begin{cases}\frac{1}{2}\delta(\tilde{x}-c),\ \tilde{x}\geq 0,\\ \frac{1}{2}\delta(\tilde{x}+c),\ \tilde{x}<0.\end{cases} \tag{36}$$

For the simple comparison of the QbaCFP with the AddaCFP, ISS and SS, we will make their average distortions equal by forcing $D_Q = D_A = \frac{L}{N}\alpha^2$ that results in:

$$P_{b-\text{QbaCFP}} = \mathcal{Q}\left(\tfrac{\eta}{\sigma_Z}\right), \tag{37}$$

where $\eta = \sigma_X\sqrt{\frac{2}{\pi}}\left(1+\sqrt{1+\frac{\pi}{2}\left(\frac{\alpha^2}{\sigma_X^2}-1\right)}\right)$.

It is also well known in the rate-distortion theory that the reconstruction level c that minimizes the distortion of a one-bit scalar quantizer, which corresponds to the modulation function of QbaCFP, is equal to the mean value of the region, i.e., $c = \text{E}\left[|\tilde{X}|\right] = \sigma_X\sqrt{\frac{2}{\pi}}$ [29] that yields:

$$D_Q = \sigma_X^2\tfrac{L}{N}\left(1-\tfrac{2}{\pi}\right), \tag{38}$$

$$P_{b-\text{QbaCFP}} = Q\left(\sqrt{\tfrac{2}{\pi}}\tfrac{\sigma_X}{\sigma_Z}\right). \tag{39}$$

Finally, the maximum achievable DWR under the QbaCFP, which corresponds to the minimum distortion (38), is $\text{DWR}_Q \leq 10\log_{10}\frac{N}{L\left(1-\frac{2}{\pi}\right)}$.

Remark 6 (Link to spread-transform dither modulation (ST-DM) watermarking)**.** The ST-DM is a hybrid watermarking method aiming at the host interference cancellation in the projected domain, which combines a quantization (binning) strategy with spread transform with the embedding rate $R_{\text{DW}} = \frac{L}{N}$ bits [19]:

$$\mathbf{v}{=}\mathbf{x} + \mu\sum_{i\in\mathcal{K}}\left(Q_{m_i}(\tilde{x}_i)-\tilde{x}_i\right)\mathbf{w}_i, \tag{40}$$

where $0 \leq \mu \leq 1$ is a distortion compensation parameter, and $Q_{m_i}(\cdot)$ is a scalar quantizer, which is defined by the bit m_i with the centroids defined by[7]:

$$c_i{=}\Delta\mathcal{Z} + (-1)^{m_i}\tfrac{\Delta}{4}, \text{for } m_i = \{0,1\}, \tag{41}$$

where Δ is the quantization interval.

Rewriting (33) in the form of (40) by introducing the distortion compensation parameter α, one obtains:

$$\mathbf{v}{=}\mathbf{x} + \alpha\sum_{i\in\mathcal{K}}(c\text{sign}(\tilde{x}_i) - \tilde{x}_i)\mathbf{w}_i, \tag{42}$$

with the remarkable correspondence in part of quantizations $Q_{m_i}(\cdot)$ and $c\text{sign}(\cdot)$. The fundamental difference between the ST-DM and QbaCFP consists in the absence of periodical structure of ST-DM quantizer depending on the message bit m_i, which should compensate for the interference with the host signal. In the case of the QbaCFP similarly to the AddaCFP, one is only interested in increasing the magnitudes of small components in the set $\mathcal{K}$ that is simply achieved by the quantization. Obviously, the cost for this simplicity are the distortions of all components whose values are larger than the centroid c and their decrease causes the increase in the probability of bit error. Therefore, one might imagine more advanced modulation strategies that quantize the low-magnitude coefficients to some predefined levels while preserving the large-magnitude coefficients. This should provide an additional gain in the introduced distortion for the same robustness to the degradations. One can also assume multilevel quantization $Q(\tilde{x}_i)$ instead of one-bit scalar quantizer $c\text{sign}(\tilde{x}_i)$. It should be noted that $Q(\tilde{x}_i)$ does not depend on m_i as in (40). We leave this line of research out of the scope of this paper, targeting here only the introduction of basic principles and advantages of basic aCFP methods in light of existing pCFP and digital watermarking.

The BER of the ST-DM was computed in [12] as $P_{b-\text{ST-DM}} = 2\sum_{k=0}^{\infty}\left\{\mathcal{Q}\left(\frac{(4k+1)\Delta}{2\sigma_Z}\right) - \mathcal{Q}\left(\frac{(4k+3)\Delta}{2\sigma_Z}\right)\right\}$. There also exist other modifications of this basic scheme [20]. However, instead of considering all of them, which is obviously beyond the scope of this paper, we will provide a lower bound on the performance of both spread transform and quantization methods, with the assumption that there will be no host interference and binary modulation. This bound should serve us as a basis for comparison with the aCFP methods.

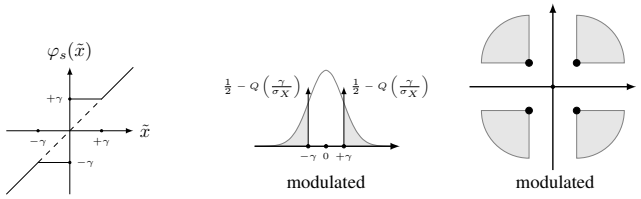
Remark 7 (BER lower bound on all watermarking methods)**.** The BER lower bound on all watermarking techniques that presume no host interference and binary embedding is defined as:

$$P_{b-\text{LB}} = \mathcal{Q}\left(\tfrac{\alpha}{\sigma_Z}\right), \tag{43}$$

which is closely achievable by the binning techniques in the high DNR regime [12].

C. Shrinkage based active content fingerprinting

The goal of this section is to introduce an alternative aCFP modulation strategy that minimizes the modulation distortion and leads to increased performance. To achieve this goal one has to find a trade-off between two conflicting requirements of projected coefficient magnitude increase for the BER reduction and modulation distortion. The basic idea behind the solution of this trade-off consists in the observation that the largest values of $P_{b||\tilde{x}_i|}$ correspond to the low-magnitude coefficients $|\tilde{x}_i|$, $i \in \mathcal{K}$. At the same time, any modulation of these coefficients, and even their complete removal, represents the lowest distortion that is often used in the lossy source coding applications [8]. Additionally, the low-magnitude coefficients concentrated near zero in the host pdf are the most likely ones and represent an essential part of host coefficients. Therefore, it is natural to increase these coefficients to the level of admissible distortion, thus decreasing the $P_{b||\tilde{x}_i|}$.



(a) (b) (c)

Fig. 7. Shrinkage based aCFP: (a) modulation function, (b) unidimensional and (c) two-dimensional modulated pdfs.

Definition: *Shrinkage based active content fingerprinting* (SbaCFP) is defined by the scalar modulation function of the form:

$$\varphi_s(\tilde{x}_i) = \begin{cases} \gamma\,\text{sign}(\tilde{x}_i), |\tilde{x}_i| \leq \gamma, \\ \tilde{x}_i \qquad\quad , |\tilde{x}_i| > \gamma, \end{cases} \tag{44}$$

where $\tilde{x}_i$, $i \in \mathcal{K}$ is the i-th element of $\widetilde{\mathbf{x}}^*$, and $\gamma \geq 0$. The modulation function, unidimensional and two-dimensional modulated pdfs are shown in Fig. 7.

Substituting (44) into (18) yields:

$$\mathbf{v} = \sum_{i\in\mathcal{K},|\tilde{x}_i|\leq\gamma}\gamma\,\text{sign}(\tilde{x}_i)\mathbf{w}_i + \sum_{i\in\mathcal{K},|\tilde{x}_i|>\gamma}\gamma\,\text{sign}(|\tilde{x}_i|)\mathbf{w}_i$$
$$+\sum_{i\notin\mathcal{K}}\tilde{x}_i\mathbf{w}_i = \mathbf{x} + \sum_{i\in\mathcal{K},|\tilde{x}_i|\leq\gamma}(\gamma\,\text{sign}(\tilde{x}_i) - \tilde{x}_i)\mathbf{w}_i = \mathbf{x} + \mathbf{s}_S,$$

where $\mathbf{s}_S = \sum_{i\in\mathcal{K},|\tilde{x}_i|\leq\gamma}(\gamma\,\text{sign}(\tilde{x}_i) - \tilde{x}_i)\mathbf{w}_i$. The distortion of SbaCFP per content sample is:

$$D_s = \tfrac{1}{N}\text{E}_{p(\tilde{x}_i)}\left[\|\mathbf{S}_S\|_2^2\right] = 2\tfrac{L}{N}\int_0^{\gamma}(\gamma - t)^2 p_{\tilde{X}}(t)\text{d}t, \tag{45}$$

where $p_{\tilde{X}} = \mathcal{N}(0,\sigma_X^2)$. In this scheme, the fingerprint extractor function is just the sign function, i.e., $\psi(t) = \text{sign}(t)$. The fingerprint is then calculated during enrollment as:

$$f_{x_i} = \psi(\varphi_s(\tilde{x}_i)) = \text{sign}(\varphi_s(\tilde{x}_i)),\ i \in \mathcal{K}.$$

The fingerprint computed at the identification stage is:

$$f_{y_i} = \psi(\tilde{y}_i) = \text{sign}(\tilde{y}_i) = \text{sign}(\varphi_s(\tilde{y}_i) + \tilde{z}_i),\ i \in \mathcal{K},$$

where $\tilde{y}_i$ similar to $\tilde{x}_i$ is a projection coefficient in the secret subspace defined by k. The performance of SbaCFP

[7]We consider only binary embedding here.

[8]The shrinkage function (Fig. 7) is very close to the dead–zone scalar quantizer that is often used in lossy image compression. Therefore, the aCFP based on shrinkage can be considered as a joint source–channel coding problem.

is determined by the BER and is given by:

$$P_{b-\text{SbaCFP}} = \mathrm{E}\left[\mathcal{Q}\left(\frac{|\varphi_s(\tilde{X})|}{\sigma_Z}\right)\right] = 2\int_0^{\gamma} \mathcal{Q}\left(\frac{\gamma}{\sigma_Z}\right) p_{\tilde{X}}(t)\,\mathrm{d}t + 2\int_{\gamma}^{\infty} \mathcal{Q}\left(\frac{t}{\sigma_Z}\right) p_{\tilde{X}}(t)\,\mathrm{d}t = 2\mathcal{Q}\left(\frac{\gamma}{\sigma_Z}\right)\left[\frac{1}{2} - \mathcal{Q}\left(\frac{\gamma}{\sigma_x}\right)\right] + 2\int_{\gamma}^{\infty} \mathcal{Q}\left(\frac{t}{\sigma_Z}\right) p_{\tilde{X}}(t)\,\mathrm{d}t. \quad (46)$$

D. Comparison of DWM, pCFP and ACFP methods

In this section, we compare the performance of DWM, pCFP and aCFP in terms of BER. Fig. 8 shows the BERs for the DWM based on SS, ISS, ST-DM and LB for the DWM as well as pCFP, AddaCFP, QbaCFP and SbaCFP for $\frac{L}{N}$ equals to 0.01 and 0.001, and DWRs equal to 20dB and 24dB.

Remark 8. Surprisingly, the performance of pCFP and SS-based DWM is very close and for $\frac{L}{N} = 0.01$ pCFP even outperforms both SS and ISS for DWR=24dB. ST-DM DWM closely approaches the lower bound for all DWRs and $\frac{L}{N}$ ratios.

Remark 9. All three aCFP techniques considerably outperform DWM and pCFP methods. Overall, SbaCFP offers the best performance providing very low BER. However, from Fig. 8 and 7 one can conclude that the performance of SbaCFP and QbaCFP can be very close.

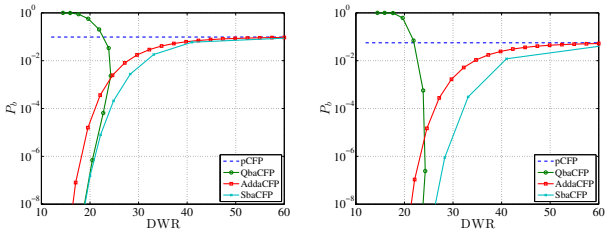
Remark 10. Since the BER of an aCFP is considerably lower than a pCFP, one can envision even the replacement of BDD by the exact fingerprint matching similar to the cryptographic hashing. Indeed it is known [6] that the capacity achieving threshold θ of pCFP should be close to ϵ. If $\epsilon \to 0$, θL in the BDD also converges to zero, which justifies the exact fingerprint matching. Finally, the probability of correct identification (see (16)) is given by $P_{ci} = (1-\epsilon)^L$, if θ is assumed to be equal to 0 thus indicating a high probability of obtaining the match $\mathbf{f_y} = \mathbf{f_x}(m)$ directly without any search.

Therefore, the aCFP is capable of enhancing the performance of both the pCFP and the DWM, by providing BERs which are not achievable by these techniques and by considerably reducing the search complexity that ultimately leads to identification based on exact matching of fingerprints.

On the other hand, it should be pointed out that fingerprints generated by both pCFP and aCFP are essentially composed of random bits and do not possess any structure. In addition, the extracted fingerprints are binary, i.e., $f_{x_i} \in \{0,1\}$, which limits the identification rate as $R_{id} \leq 1 - H_2(\epsilon)$. Furthermore, as previously mentioned, BDD is only efficient, in terms of complexity, when $\epsilon \leq 0.11$. Consequently, to investigate a potential gain that can be achieved based on structured modulation techniques, we consider multidimensional modulation based on lattices.

VII. Multidimensional case of aCFP

To achieve lower decoding complexity and distortion, we extend our analysis to lattice based quantization as a vector modulation function. We will refer to this sort of aCFP as lattice based aCFP.



(a) DNR=10dB (b) DNR=15dB

Fig. 9. Comparison of different aCFP scheme for fixed $\frac{L}{N} = 0.01$.

Definition: *Lattice based active content fingerprinting* (LbaCFP) is defined by the vector modulation function of the form:

$$\varphi_\Lambda(\widetilde{\mathbf{x}}^*) = Q_\Lambda(\widetilde{\mathbf{x}}^* + \mathbf{u}) - \mathbf{u}, \quad (47)$$

where $Q_\Lambda(\cdot)$ is the lattice based quantization using the lattice $\Lambda \subset \mathcal{R}^L$ with the generator matrix $\mathbb{G}$ and the fundamental Voronoi region $\mathcal{V}$, and $\mathbf{u}$ is the dither [9] as defined in Section III. The distortion of LbaCFP per content sample is:

$$D_\Lambda = \tfrac{1}{N}\mathrm{E}\left[\|\mathbf{S}_\Lambda\|_2^2\right] = \tfrac{1}{N}\mathrm{E}\left[\sum_{i\in\mathcal{K}}(\varphi_\Lambda(\widetilde{\mathbf{X}}^*)[i] - \tilde{X}_i)^2\right] = \tfrac{1}{N}\mathrm{E}\left[\sum_{i\in\mathcal{K}}\zeta_i^2\right] = \tfrac{1}{N}\mathrm{E}\left[\|\boldsymbol{\zeta}\|^2\right] \overset{(a)}{=} \tfrac{L}{N}G(\Lambda,\mathcal{V})V^{2/L},$$

where (a) follows from the fact that the error vector $\boldsymbol{\zeta}$ is uniformly distributed over $\mathcal{V}$ [18] and $G(\Lambda,\mathcal{V})$ is the normalized second moment of the lattice Λ with the Voronoi region $\mathcal{V}$ and $V = \det(\mathbb{G})$ is the volume of $\mathcal{V}$.

In this scheme, the fingerprint extractor function is the identity function, i.e., $\psi(x) = x$. The fingerprint at the enrollment stage is:

$$\mathbf{f_x} = \psi(\varphi_\Lambda(\widetilde{\mathbf{x}}^*)) = \varphi_\Lambda(\widetilde{\mathbf{x}}^*).$$

The fingerprint computed at the identification stage is:

$$\mathbf{f_y} = \psi(\varphi_\Lambda(\widetilde{\mathbf{y}}^*)) = \varphi_\Lambda(\widetilde{\mathbf{y}}^*) = \varphi_\Lambda(\mathbf{f_x} + \tilde{\mathbf{z}}^*),$$

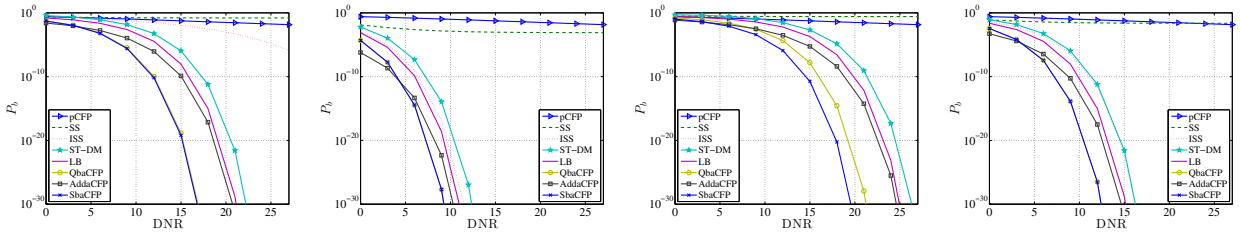
where $\tilde{\mathbf{z}}^*$ is a collection of projected coefficients of $\tilde{\mathbf{z}}$ in a secret subspace defined by k and $\widetilde{\mathbf{Z}}^* \sim \mathcal{N}(\mathbf{0}, \sigma_Z^2\mathbb{I}_L)$ [6]. Consequently, one can consider the lattice quantizer as a decoder under AWGN.

The performance of the content identification using LbaCFP can be upper bounded for P_{ci} as follows, for any $\gamma_{\text{eff}} \leq \sigma_Z^2$:

$$\begin{aligned} P_{ci} &= \sum_{m=1}^{|\mathcal{M}|} \Pr\left\{\sum_{\widetilde{\mathbf{y}}^*\in\mathcal{V}(\mathbf{f_x}(m))} p(\widetilde{\mathbf{y}}^*|\mathbf{f_x}(m))\big|\mathcal{H}_m\right\}\Pr\{\mathcal{H}_m\} \\ &\overset{(a)}{\leq} \tfrac{1}{|\mathcal{M}|}\sum_{m=1}^{|\mathcal{M}|}\Pr\{\|\mathbf{f_x}(m) - \widetilde{\mathbf{Y}}^*\|_2^2 \leq \gamma_{\text{eff}}L|\mathcal{H}_m\} \\ &\overset{(b)}{=} \Pr\{\|\mathbf{f_x}(1) - \mathbf{F_y}\|_2^2 \leq \gamma_{\text{eff}}L|\mathcal{H}_1\} \\ &\overset{(c)}{\leq} \exp\left(-L\left(\tfrac{\gamma_{\text{eff}}}{2\sigma_Z^2} - 0.5\ln\left(\tfrac{\gamma_{\text{eff}}e}{\sigma_Z^2}\right)\right)\right), \end{aligned} \quad (48)$$

where $\gamma_{\text{eff}}L$ is the radius of a sphere that has the same volume as the Voronoi region $\mathcal{V}$, (a) holds since the value of the probability density of the AWGN at any point of a sphere of

[9] The dither can be key-dependent.



(a) DWR=20dB, $\frac{L}{N} = 0.01$ (b) DWR=20dB, $\frac{L}{N} = 0.001$ (c) DWR=24dB, $\frac{L}{N} = 0.01$ (d) DWR=24dB, $\frac{L}{N} = 0.001$

Fig. 8. Comparison of pCFP, different aCFP and DWM schemes and Lower Bound (LB) in DWM.

radius r is larger than it is at any point outside of the sphere, (b) follows due to the assumption that $\mathbf{f}_{\mathbf{x}}(m) \neq \mathbf{f}_{\mathbf{x}}(m')$ for all $m \neq m'$ and all the Veronoi regions have the same geometry, and (c) follows from the Cramer-Chernoff bound [30].

As mentioned in Section III, the projected coefficients in the secret subspace follow i.i.d. Gaussian pdf, i.e., $\widetilde{\mathbf{X}}^* \sim \mathcal{N}(\mathbf{0}, \sigma_X^2 \mathbb{I}_L)$. For large subspace dimension L, this is equivalent to a uniform distribution over a sphere of radius $\sqrt{L\sigma_X^2}$ [17]. Consequently, the probability of false acceptance can be approximately evaluated as the volume of the sphere of radius $\sqrt{L\sigma_X^2}$ over the volume of Voronoi region $\mathcal{V}$ of the lattice Λ times by the number of items $|\mathcal{M}|$, i.e., $P_{fa} \approx \frac{C_L (L\sigma_X^2)^{\frac{L}{2}}}{V|\mathcal{M}|}$, where $C_L = \frac{\pi^{\frac{L}{2}}}{\Gamma(\frac{L}{2}+1)}$.

VIII. Numerical evaluation

A. Numerical results using Synthetic Database

In this Section, we perform the analysis of theoretical results presented in Sections VI and VII based on a synthetic database of size $|\mathcal{M}| = 1024$ and $N = 1024$ generated from the Gauss-Markov process[10] with the normalized correlation coefficients $\rho = 0, 0.5$ and 0.75 and $\sigma_X^2 = 1$ to have fair comparison with our previous results in [6]. As in the enrollment stage, we first modify the generated synthetic contents based on the embedding schemes introduced in Section VI over the secret subspace of dimension $L = 32$ and DWR=22dB, and then the corresponding fingerprints will be extracted. Fig. 10 shows BER's of different modulation schemes, empirical evaluation versus theoretical ones. These results demonstrate a very good correspondence between the theoretical and simulation results on the synthetic data for all ρ's. Fig. 11 shows ROC (Receiver operating characteristic) curves, which is for θ varying from 0 to 1, of the database with $\rho = 0.75$ under AWGN distortions with DNR's in the range of 0 to 10 dB. Following results indicated in Fig 11, one can conclude that an aCFP can bring considerable improvement to content identification performance, even for very low embedding distortion. It should be noted that QbaCFP performs very closely to SbaCFP.

B. Numerical results using Image Database

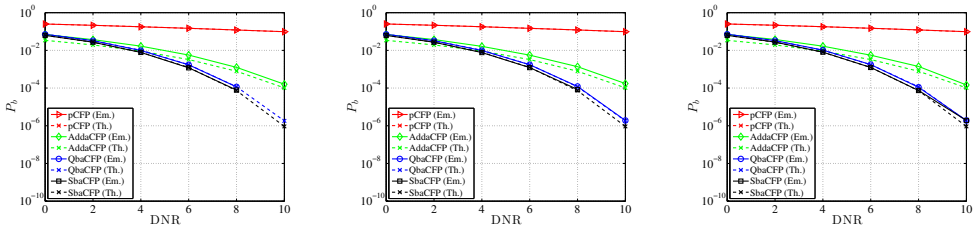
In this section, we evaluate the performance of content identification using a gray scaled version of the UCID database [31], consisting of 1338 images of size 384 by 512. To extract a feature vector from each gray scaled image, an image is divided into 16 by 16 blocks and the 2D DCT of each block is computed. The feature vector is constructed by concatenating the DCT coefficients at the coordinates $(1, 2)$ inside each block [6] resulting in a vector of length $N = 768$. Finally, the fingerprint of length L from each feature vector is extracted by using RP followed by different modulation approaches considered in Sections VI and VII. Table I shows the parameters of the modulation schemes and their average embedding distortion over all images in UCID, based on Peak Signal to Distortion Ratio (PSDR$= 10\log_{10}(\frac{255^2}{D})$) and the average of the mean structural similarity index (MSSIM) [32] to quantitatively evaluate the imperceptibility of the modulations.

Since QbaCFP and SbaCFP have shown the close performance according to Fig. 8, 9 and 11, we limit our performance analysis to the case of SbaCFP. Fig. 12 shows the ROC curves obtained for θ varying from 0 to 1, of the content identification system using AddaCFP and SbaCFP with $L = 32$ and the parameters indicated in Table I under AWGN with different Peak Signal to Noise Ratio (PSNR$= 10\log_{10}(\frac{255^2}{\sigma_Z^2})$), JPEG compression with different Quality Factors (QF) and histogram equalization (Histeq) distortions. It should be pointed out that the performance of AddaCFP and SbaCFP in terms of the probability of correct identification P_{ci} improves with increasing θ. However, it requires higher decoding complexity.

Table II shows the performance of content identification for AddaCFP and SbaCFP using BDD with $\theta L = 2$ that ensures $P_{fa} = 0$ under the afore-mentioned attacks. The closeness of theoretical probability bit errors P_b to empirical ones $\hat{P}_b$ confirms the accuracy of the models introduced in Section VI. Moreover, P_b's of aCFPs show remarkable improvements w.r.t. pCFP that leads to high P_{ci}. For example, under JPEG compression with QF=1, the BERs in SbaCFP, AddaCFP and pCFP are respectively, 0.03, 0.05 and 0.11. This improves P_{ci} from 0.33 in pCFP to 0.79 and 0.92 in AddaCFP and SbaCFP, respectively. Finally, the results from Fig. 12 and Table II show that SbaCFP outperforms AddaCFP.

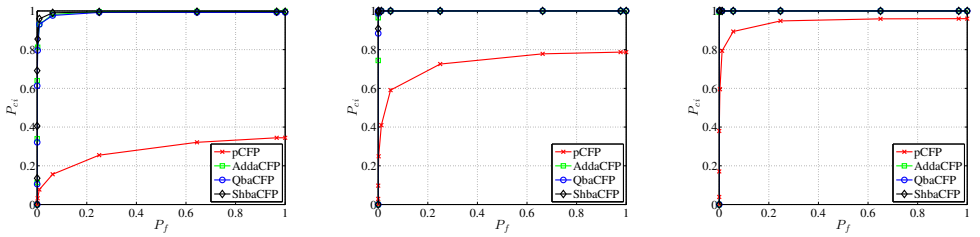
In this last section, we compare the performance of content identification using AddaCFP and SbaCFP in unidimension versus LbaCFP in multidimension cases. In order to have a fair comparison between the performance of the modulators in terms of memory storage, embedding rate, and complexity, we investigate their performance for $L = 24$ in LbaCFP (Leech Lattice dimension) and corresponding $L = 192$ in AddaCFP and SbaCFP, and we set $\theta = \frac{2}{192}$ for BDD used in a unidimension case to have roughly the same decoding com-

[10]the Gauss-Markov process is a Gaussian first order autoregressive process $X_n = \rho X_{n-1} + \Xi_n$, where ρ is the normalized correlation coefficient and $\Xi_n \sim \mathcal{N}(0, \sigma_\Xi^2)$. The Gauss-Markov process model is used for the modeling of correlation in real images [27] and it is also used for the evaluation of performance in digital fingerprinting and watermarking.



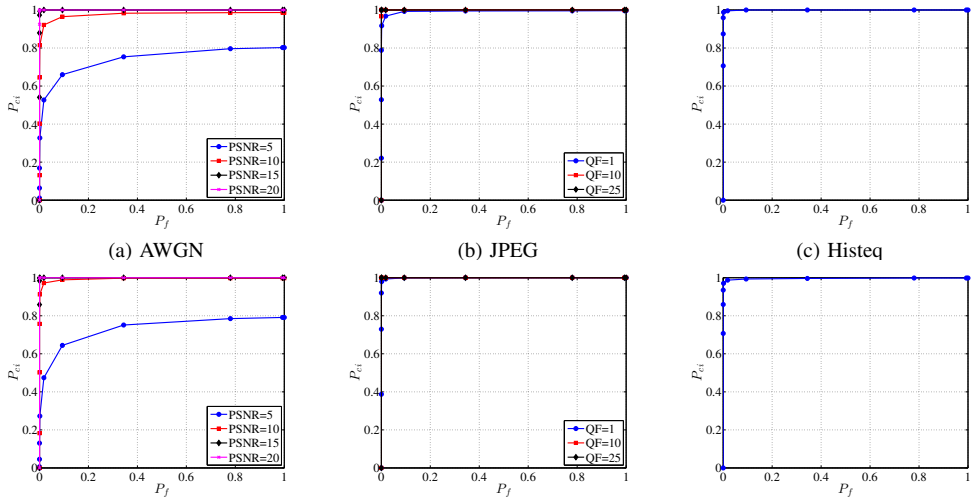
(a) $\rho = 0$ (b) $\rho = 0.5$ (c) $\rho = 0.75$

Fig. 10. Comparison of BER's for different aCFP schemes (theoretical (Th.) vs emperical (Em.)) over the synthetic database of size $|\mathcal{M}| = 1024$ and $N = 1024$ generated by Gauss-Markov process with ρ's equal to 0, 0.5 and 0.75, secret subspace of dimension $L = 32$ and DWR=22dB.



(a) DNR=0 dB (b) DNR=5 dB (c) DNR=10 dB

Fig. 11. ROC curves of identification based on pCFP, AddaCFP, QbaCFP and ShbaCFP with $|\mathcal{M}| = 1024$, $N = 4096$, $\rho = 0.75$, $L = 32$ and DWR= 22 dB under AWGN.



(a) AWGN (b) JPEG (c) Histeq

(d) AWGN (e) JPEG (f) Histeq

Fig. 12. ROC curves of real image identification based on (First row) AddaCFP with $N = 768$, $L = 32$ and $\alpha = 45$, and (Second row) SbaCFP with $N = 768$, $L = 32$ and $\gamma = 110$ under: AWGN, JPEG and Histeq distortions.

TABLE I
DIFFERENT MODULATION SCHEMES.

Modulators	Parameters	PSDR	MSSIM
AddaCFP	$L = 32$, $\alpha = 45$	53 dB	0.999
	$L = 192$, $\alpha = 18$	53 dB	0.999
SbaCFP	$L = 32$, $\gamma = 110$	53 dB	0.999
	$L = 192$, $\gamma = 60$	53 dB	0.999
LbaCFP	$L = 24$, scale= 70	53 dB	0.999

plexity w.r.t. Leech Lattice BDD with the complexity of 519 operations [15]. We evaluate the ability of the identification system to correctly identify an image after it has undergone the malicious attacks listed in Table III. From the results in Table III, one can conclude that, with the exception of Histeq distortion, LbaCFP generally outperforms modulation schemes in a unidimensional case.

IX. CONCLUSION

In this paper we introduced the concept of aCFP and considered its practical implementations based on additive, quantization and shrinkage methods in unidimensional cases, and a multidimensional case based on Leech–lattice. Simulation and practical results show that SbaCFP outperforms other modulation schemes in the unidimensional case. However, LbaCFP in a multidimensional case outperforms all methods

TABLE III
COMPARISON OF PROBABILITY OF CORRECT IDENTIFICATION P_{ci} USING DIFFERENT MODULATION SCHEMES WITH APPROXIMATELY CLOSE DECODING COMPLEXITY.

Attack	Parameters		P_{ci}		
	im. domain	proj. domain (DNR)	SbaCFP $L = 192$ $\theta = 2/L$	AddaCFP $L = 192$ $\theta = 2/L$	LbaCFP $L = 24$
AWGN	PSNR=20 dB	18 dB	0.96	0.34	1
	15 dB	13 dB	0.07	0.02	0.83
	10 dB	8 dB	0	0	0.01
	5 dB	3 dB	0	0	0
JPEG	QF=25	27 dB	1	1	1
	10	19 dB	0.998	0.56	1
	1	10 dB	0	0	0.14
Histeq		6 dB (LbaCFP 3 dB)	0.3	0.14	0.11

TABLE II
LIST OF ATTACKS TESTED AND THE CORRESPONDING PROBABILITY OF CORRECT IDENTIFICATION, REAL $\hat{P}_b$ AND PREDICTED P_b BERS IN UNIDIMENSIONAL ACFP.

Modulators	Attack	Parameters		Performance					
		im. domain	proj. domain (DNR)	P_{ci}	P_{fa}	$\hat{P}_b$	P_b	$P_{b-\text{pCFP}}$	P_{ci}(pCFP)
SbaCFP $L=32$ $\theta=2/L$ $\gamma=110$	AWGN	PSNR=20 dB	18 dB	1	0	0	0	0.05	0.80
		15 dB	13 dB	0.998	0	0.004	0.003	0.08	0.55
		10 dB	8 dB	0.76	0	0.05	0.05	0.13	0.21
		5 dB	3 dB	0.15	0	0.15	0.14	0.21	0.05
	JPEG	QF=25	27dB	1	0	0	0	0.01	0.98
		10	20dB	1	0	0	0	0.04	0.86
		1	10dB	0.92	0	0.03	0.02	0.11	0.33
	Histeq		4dB	0.94	0	0.02	0.12	0.14	0.43
AddaCFP $L=32$ $\theta=2/L$ $\alpha=45$	AWGN	PSNR=20 dB	19 dB	1	0	0.002	0.001	0.05	0.80
		15 dB	14 dB	0.97	0	0.019	0.013	0.08	0.55
		10 dB	9 dB	0.63	0	0.07	0.05	0.13	0.21
		5 dB	4 dB	0.16	0	0.15	0.12	0.21	0.05
	JPEG	QF=25	28dB	1	0	0	0	0.01	0.98
		10	21dB	1	0	0.001	0.001	0.04	0.86
		1	11dB	0.79	0	0.05	0.03	0.11	0.33
	Histeq		5dB	0.96	0	0.01	0.1	0.14	0.43

in unidimension due to its normalized second moment property and very low decoding complexity.

It should be pointed out that, besides its remarkable performance, the aCFP does not intend to replace DWM in the copyright protection applications. However, it could be considered as a reasonable alternative in those applications that require content related management, tracking, tracing and monitoring.

In the future, we will extend our analysis to consider the security of aCFP and enhanced search strategies. Moreover, we intend to extend the results to geometrical invariant features. It is also interesting to establish the link between aCFP and uncoded transition. Finally, the joint compression and modulation within the aCFP framework looks very attractive.

REFERENCES

[1] S. Voloshynovskiy, F. Farhadzadeh, O. Koval, and T. Holotyak, "Active content fingerprinting: a marriage of digital watermarking and content fingerprinting," in *IEEE WIFS*, Tenerife, Spain, Dec. 2012.

[2] Wikipedia. Second screen – wikipedia, the free encyclopedia, 2013. [Online; accessed June-16-2013].

[3] Digimarc, http://www.digimarc.com/discover/magazines, [Online; accessed June-16-2013].

[4] V. Lohweg, J. L. Hoffmann, H. Drksen, R. Hildebrand, E. Gillich, J. Hofmann, and J. Schaede, "Banknote authentication with mobile devices," in *SPIE*, California, USA, Feb. 2013.

[5] S. Voloshynovskiy, O. Koval, F. Beekhof, and T. Pun, "Unclonable identification and authentication based on reference list decoding," in *Conf. on SCSI*, Germany, Mar. 2008.

[6] F. Farhadzadeh, S. Voloshynovskiy, and O. Koval, "Performance analysis of content-based identification using constrained list-based decoding," *IEEE Trans. on Info. Forensics and Sec.*, vol. 7, no. 5, pp. 1652 –1667, oct. 2012.

[7] J. Fridrich, "Robust bit extraction from images," in *IEEE Conf. on MCS*, vol. 2, july 1999, pp. 536 –540.

[8] J. Haitsma and T. Kalker, "A highly robust audio fingerprinting system." in *ICM Info. Retrieval,*, 2002.

[9] I. J. Cox, M. L. Miller, and J. A. Bloom, *Digital Watermarking*. Morgan Kaufmann Publishers (Academic Press), 2002.

[10] F. Perez-Gonzalez, F. Balado, and J. Martin, "Performance analysis of existing and new methods for data hiding with known-host information in additive channels," *IEEE Trans. on Signal Processing*, vol. 51, no. 4, pp. 960 – 980, apr 2003.

[11] L. Pérez-Freire and F. Pérez-González, "Spread spectrum watermarking security," *IEEE Trans. on Info. Forensics and Sec.*, vol. 4, no. 2–24, pp. 969–978, Mar. 2009.

[12] F. Pérez-González, F. Balado, and J. R. Hernández, "Performance analysis of existing and new methods for data hiding with known-host information in additive channels," *IEEE Trans. on Signal Processing*, vol. 51, no. 4, Apr. 2003.

[13] O. Koval, S. Voloshynovskiy, P. Bas, and F. Cayre, "On security threats for robust perceptual hashing," in *SPIE*, San Jose, USA, 2009.

[14] J. Leech, "Notes on sphere packings," *Canadian Journal of Mathematics*, 1967.

[15] O. Amrani, Y. Be'ery, A. Vardy, F.-W. Sun, and H. van Tilborg, "The leech lattice and the golay code: bounded-distance decoding and multilevel constructions," *IEEE Trans. on Info. The.*, vol. 40, no. 4, pp. 1030 –1043, jul. 1994.

[16] C. B. Zid, S. Baudry, B. Chupeau, and G. Doërr, "A sneak peek into the camcorder path," in *SPIE*, 2013.

[17] U. Erez and R. Zamir, "Achieving 1/2 log (1+snr) on the awgn channel with lattice encoding and decoding," *IEEE Trans. on Info. Th.*, vol. 50, no. 10, pp. 2293–2314, 2004.

[18] A. Kirac and P. Vaidyanathan, "Results on lattice vector quantization with dithering," *IEEE Trans. on Circuits and Systems*, vol. 43, no. 12, pp. 811 –826, dec. 1996.

[19] B. Chen and G. W. Wornell, "Quantization index modulation: A class of provably good methods for digital watermarking and information embedding," *IEEE Trans. on Info. The.*, vol. 47, pp. 1423–1443, 2001.

[20] R. F. H. Fischer and R. Bäuml, "Lattice costa schemes using subspace projection for digital watermarking," *European Trans. Telecom.*, vol. 15, pp. 351–362, 2004.

[21] D. Lowe, "Object recognition from local scale-invariant features," in *IEEE Conf. on CV*, vol. 2, 1999, pp. 1150 –1157.

[22] E. McCarthy, F. Balado, G. Silverstreand, and N. Hurley, "A framework for soft hashing and its application to robust image hashing," in *IEEE ICIP*, Singapore, Oct. 2004.

[23] S. Voloshynovskiy, O. Koval, F. Beekhof, F. Farhadzadeh, and T. Holotyak, "Information-theoretical analysis of private content identification," in *IEEE ITW*, Dublin, Ireland, Sep. 2010.

[24] S. Voloshynovskiy, T. Holotyak, O. Koval, F. Beekhof, and F. Farhadzadeh, "Sign-magnitude decomposition of mutual information with polarization effect in digital identification," in *IEEE ITW*, Brazil, Oct. 2011.

[25] A. L. Varna, A. Swaminathan, and M. Wu, "A decision theoretic framework for analyzing hash-based content identification systems," in *ACM*, Oct. 2008, pp. 67–76.

[26] F. Willems, "On the capacity of a biometrical identification system," in *IEEE ISIT*, 2003, pp. 8–2.

[27] A. K. Jain, *Fundamentals of digital image processing*. Upper Saddle River, NJ, USA: Prentice-Hall, Inc., 1989.

[28] H. S. Malvar and D. A. F. Florencio, "An improved spread spectrum: A new modulation technique for robust watermarking," *IEEE Trans. on Signal Processing*, vol. 51, no. 4, pp. 898–905, Apr. 2003.

[29] T. M. Cover and J. A. Thomas, *Elements of information theory*. New York, NY, USA: Wiley-Interscience, 1991.

[30] G. Poltyrev, "Bounds on the decoding error probability of binary linear codes via their spectra," *IEEE Trans. on Info. The.*, vol. 40, no. 4, pp. 1284 –1292, jul. 1994.

[31] G. Schaefer and M. Stich, "Ucid - an uncompressed colour image database," in *Storage and Retrieval Methods and Applications for Multimedia*, ser. SPIE, 2004.

[32] Z. Wang, A. Bovik, H. Sheikh, and E. Simoncelli, "Image quality assessment: from error visibility to structural similarity," *IEEE Trans. on Image Processing*, vol. 13, no. 4, pp. 600 –612, apr. 2004.