

Improved Utilization of Embedding Distortion in Scalar Quantization Based Data Hiding Techniques

Husrev T. Sencar^{a,*}, Mahalingam Ramkumar^b, Ali N. Akansu^c, Amol Sukerkar^c

^a*Polytechnic University, Department of Computer and Information Science, Brooklyn, NY 11201, USA*

^b*Mississippi State University, Department of Computer Science & Engineering, Mississippi State, MS 39762, USA*

^c*New Jersey Institute of Technology, Department of Electrical and Computer Engineering, Newark, NJ 07102, USA*

Abstract

In this paper, we analyze the performance of scalar quantization based data hiding techniques with decreasing cover-signal sizes under mean squared error distortion measure. We introduce a new scheme, called multiple codebook data hiding, that enables conventional embedding/detection techniques to utilize the permitted embedding distortion more efficiently. The proposed method treats the embedding distortion introduced to a cover-signal as a random variable, and utilizes the fact that decreasing cover-signal size increases the deviation of the embedding distortion from its expected value. This is exploited by embedding the watermark into a variant of the cover-signal that yields a lower embedding distortion. In the proposed method, variants of the cover-signal are obtained by deploying a set of real unitary transformations known to both embedder and detector. For the given cover-signal, the embedder chooses a transformation basis and embeds the message in the transformed cover-signal, whereas the detector has to search all transformations of the received signal for the embedded message. We evaluate the performance improvement due to multiple codebook data hiding and compare it with the conventional (single codebook) approaches, under additive white Gaussian noise attacks, in terms of the bound on the probability of detection error. Performance results obtained from simulation and by applying the technique to image watermarking problem under JPEG compression attack are also presented.

Key words: Data hiding; Embedder/Detector; Watermark; Post-processing; Embedding distortion; Codebooks.

1. Introduction

In all digital communication systems, a general objective is the efficient use of the available resources, *i.e.*, bandwidth, power, and affordable complexity, to achieve a specified performance goal expressed in terms of error probability or reconstruction quality. The design of a communication system very often requires tradeoffs among these

resources depending on the channel description which characterizes the power limitations, accessible bandwidth, and the nature of the channel noise and its statistics. In many applications, one of the two primary communications resources, power or bandwidth, is more scarcer than the other. This limitation on the communication system is fundamental to the choice of a modulation scheme.

Data hiding is a form of communication where an information signal is transmitted by embedding it in a cover-signal in an imperceptible and unobtrusive manner. Accordingly, the notion of channel in a communications scenario, which is defined as the propagating medium between the transmit-

* Corresponding author.

Email addresses: taha@isis.poly.edu (Husrev T. Sencar), ramkumar@cse.msstate.edu (Mahalingam Ramkumar), ali@oak.njit.edu (Ali N. Akansu), ans9@njit.edu (Amol Sukerkar).

ter and receiver, can be reinterpreted as the cover-signal in the context of data hiding, as it is the message bearing medium between the embedder and detector. Correspondingly, the power constraints and the available transmission bandwidth of a communications channel can be associated with the embedding distortion (maximum amount of distortion that may be introduced to the cover-signal) and the cover-signal size in the data hiding channel model, respectively. In data hiding, the combination of the message signal (watermark) and the cover-signal is called the embedded-signal, and the channel noise is analogous to all forms of disturbances that affect the embedded-signal.

Data hiding techniques are most often evaluated based on three main criteria: robustness, imperceptibility and payload (hiding rate). These goals are conflicting in nature, and the designer is required to make the proper trade-off between these goals depending on the requirements of the specific application. In data hiding, the resource of the communication between the embedder and detector is the distortion introduced to cover-signal during embedding. Therefore, at a given level of robustness, payload increases with the permitted embedding distortion (per cover-signal component) and the size of the cover-signal, and information hider needs to design the embedder and detector that make effective use of these resources.

The duality between the communications and data hiding frameworks has been well studied and incorporated into the design of embedding/detection techniques [1–4]. In this regard, techniques based on the principles of linear (additive and multiplicative) spread-spectrum modulation [5] and schemes based on binning and coset formation (quantization with one- or multi-dimensional lattices) [6] have exploited this connection most effectively. However, in terms of additive noise attacks, quantization based methods provide superior efficacy as compared to linear methods due to their ability to reject cover-signal interference at the detector [7]. In essence, quantization based embedding/detection techniques are designed to achieve the data hiding capacity [8–11] by deploying optimal or near-optimal constructions in exploiting cover-signal information during embedding [2,3,12]. The formulation of quantization based data hiding techniques is often aimed at maximizing the data hiding rate, which is the average amount of information that can be reliably extracted from each embedded-signal sample and it has an asymptotic behavior in

cover-signal size. The solution to the constrained formulation links the embedding distortion per cover-signal sample to hiding rate which approaches exactness at large cover-signal sizes. Most quantization based data hiding techniques are designed and evaluated under this assumption and, therefore, the influence of cover-signal size on the performance is ruled out. In this paper, we study methods that are based on scalar quantization based data hiding techniques and that enable better exploitation of the permitted embedding distortion under varying cover-signal sizes. We assume that the embedding distortion is measured by mean squared difference and the embedded-signal is subjected to additive white Gaussian noise (AWGN) attack.

To make effective use of the embedding distortion with the increasing cover-signal sizes, the general approach has been the incorporation of redundancy coding into watermark generation and embedding. In this regard the most popular approach has been the *spread transforming* (ST) method [6]. On the contrary for smaller cover-signal sizes, better utilization of the embedding distortion has not been adequately addressed and, implicitly, validity of conventional methods is assumed. To fill this gap, we propose the use of *multiple codebook data hiding* method built based on existing embedding/detection techniques. Multiple codebook data hiding method enables generating a set of codewords for each message to be embedded and picks the codeword that adapts to cover-signal best. In multiple codebook data hiding, the detection performance is improved due to ability to embed the watermark at a reduced embedding distortion. The crux of the proposed method lies in taking the minimum of several realizations of a random variable and utilizing the difference with the (expected) mean. In our case, the random variable is the measured mean square error distortion over a finite number of samples, *i.e.*, embedding distortion, and each realization of the random variable is associated with a different codeword. We compare single codebook (*conventional case*) and multiple codebook data hiding methods by analytically computing the bound on the probability of error in detecting the wrong message and by simulations for various cover-signal sizes and number of codebooks. We further examine the performance of the proposed scheme by applying the technique to digital image watermarking problem under JPEG compression attack.

In the text, we denote vectors with boldfaced characters, random variables with capital letters and

their realizations with the corresponding lower case letters, and the matrix variables with ‘blackboard bold’ fonts. Table 1 lists all the notations we use in this paper. For the general case all signals are assumed to be random vectors of size N , however, in some of the derivations individual random variables are used, rather than vector representations, for the sake of simplicity. In such cases, vector extensions are straightforward due to independent and identically distributed (*iid*) assumption on the random variables.

Table 1
Notation used in the paper

\mathbf{C}	Cover-signal
\mathbf{X}_n	Codeword
\mathbf{S}	Information hidden signal
\mathbf{Z}	Channel noise
\mathbf{Y}	Distorted \mathbf{S}
\mathbf{W}_m	Watermark corresponding to message m
$\hat{\mathbf{W}}_m$	Extracted signal when \mathbf{W}_m is embedded
$\rho_{m,j}$	The normalized correlation between $\hat{\mathbf{W}}_m$ and \mathbf{W}_j
$\tilde{\mathbf{W}}_m^i$	Extracted signal from \mathbb{T}_i transformation of \mathbf{S} when \mathbf{W}_m is embedded
$\hat{\tilde{\mathbf{W}}}_m^i$	Extracted signal from \mathbb{T}_i transformation of \mathbf{Y} when \mathbf{W}_m is embedded
$\tilde{\rho}_{m,j}^i$	The normalized correlation between $\tilde{\mathbf{W}}_m^i$ and \mathbf{W}_j
$\rho_{m,j}^i$	The normalized correlation between $\hat{\tilde{\mathbf{W}}}_m^i$ and \mathbf{W}_j
P	Embedding distortion
P_E	Permitted embedding distortion
WNR	Embedding distortion to channel distortion ratio

In the next section, Section 2, we briefly describe the characteristics of scalar quantization based embedding/detection techniques. We introduce the multiple codebook data hiding technique in Section 3. The performance analysis methodology for single and multiple codebook data hiding cases, and the performance results are presented in the following sub-sections. Our remarks and conclusions are given in Section 4.

2. Embedding and Detection

References [13–16] proposed scalar quantization based low-complexity embedding/detection techniques that approach the *data hiding capacity* under mean squared error distortion measure and AWGN attacks [11]. These methods, in common, perform the embedding operation as a form of dithered quantization followed by a post-processing like thresholding [13], distortion compensation (DC) [14,15], or Gaussian mapping (GM) [16]. That is, the cover-signal coefficients are first quantized with respect to the watermark samples where each sample takes

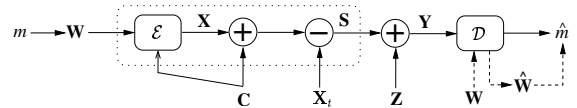


Fig. 1. Channel model for embedding and detection operations.

on discrete values from a finite set and each value is associated with a different quantizer. Then, the resulting quantized cover-signal coefficients undergo a post-processing to generate the embedded-signal. The detection of the sent message is by maximum likelihood decoding through sample-wise hard decisions [15,16] or soft decisions [13,14] based on the availability of the set of watermarks at the detector.

The channel model for quantization based embedding/detection techniques is displayed in Figure 1. In this model, m is the message to be hidden, \mathbf{C} is the cover-signal, \mathbf{W} is the watermark, \mathbf{X} is the distortion introduced to \mathbf{C} due to quantization, \mathbf{X}_t is the processing distortion due to post-processing, \mathbf{S} is the embedded-signal, \mathbf{Z} is the additive channel noise, \mathbf{Y} is the distorted embedded-signal, $\hat{\mathbf{W}}$ is an estimate of \mathbf{W} , and \hat{m} is the detected message. It should be noted that \mathbf{X}_t is generated as a function of \mathbf{X} depending on the expected noise level in the channel. The embedded-signal \mathbf{S} can be defined in terms of the *codeword*, $\mathbf{X}_n = \mathbf{X} - \mathbf{X}_t$, as $\mathbf{S} = \mathbf{C} + \mathbf{X}_n$ where the collection of \mathbf{X}_n corresponding to all messages constitute a *codebook*. Consequently, the per-sample distortion P introduced to cover-signal due to embedding can be expressed as $P = \frac{1}{N} \|\mathbf{X}_n\|^2$. In the model, the signal $\mathbf{C} + \mathbf{X}$ refers to a quantized signal and the watermark \mathbf{W} can be perfectly recovered from this signal. Since detector is not aware of the post-processing at the embedder, the corresponding processing distortion \mathbf{X}_t , introduced at the embedder, can be considered as another source of noise. Therefore, the effective noise at the detector, that distorts the embedded watermark \mathbf{W} , can be defined as $\mathbf{Z}_f = \mathbf{Z} - \mathbf{X}_t$. Correspondingly, the performance analysis of an embedding/detection scheme can be conducted in terms of the statistics of \mathbf{X} , \mathbf{X}_t , and \mathbf{Z} .

Dither modulation (DM) technique is central to the design of scalar quantization based embedding/detection techniques. In DM, each quantizer in the ensemble is generated from a base quantizer by shifting the quantization cells and reconstruction points. The embedded-signal is generated by quantizing the cover-signal with the corresponding dithered quantizer as

$$\mathbf{S} = Q_{\Delta}(\mathbf{C} + \mathbf{W}) - \mathbf{W} \quad (1)$$

where $Q_\Delta(\cdot)$ can be considered to be a product quantizer generated by a Cartesian product of N uniform scalar quantizers, q_Δ , with quantization step size Δ . Therefore, embedding can be viewed as N successive scalar quantization, of the coefficients of $\mathbf{C} = (C_1, \dots, C_N)$, dithered with the watermark vector $\mathbf{W}_m = (W_1, \dots, W_N)$. Each distinct value of the watermark (dither) signal is associated with a quantizer that is generated by properly shifting the reconstruction points of q_Δ . It should be noted that DM delivers the optimal performance when the channel noise is very low or absent. However, with the increasing channel noise level, the performance of DM drops rapidly. This sharp deterioration in the performance is compensated by incorporating the post-processing in watermark embedding. Consequently, the codeword \mathbf{X}_n is defined as

$$\mathbf{X}_n = (Q_\Delta(\mathbf{C} + \mathbf{W}_m) - \mathbf{W}_m) - \mathbf{C} - \mathbf{X}_t. \quad (2)$$

where \mathbf{X}_t is the processing distortion obtained by subjecting the quantization error to the particular *post-processing*.

In scalar quantization based data hiding methods, the extraction of the sent message, from the received signal \mathbf{Y} , can be realized by minimum distance decoding or by maximum correlation rule. With the use of minimum distance decoder, detection is simply the quantization of the received signal \mathbf{Y} by all quantizers in the ensemble [17]. Accordingly, the message index associated with the quantizer that yields the minimum Euclidean distance to received \mathbf{Y} is deemed to be the sent message. The watermark detection can be written, in terms of $\mathbf{Y}_m = \mathbf{Y} + \mathbf{W}_m$, as

$$\hat{m} = \mathcal{D}(\mathbf{Y}) = \arg \min_m \|\mathbf{Y}_m - Q_\Delta(\mathbf{Y}_m)\|, \quad 1 \leq m \leq M. \quad (3)$$

where \mathbf{W}_m is the watermark associated with the message index m . The presence of watermarks $\mathbf{W}_1, \dots, \mathbf{W}_M$ at the detector, leads to an improved detection of the sent message since they can be utilized in detection operation. In this case, detection of each sample is by soft decisions. This can be realized by mapping each coefficient Y_m of \mathbf{Y}_m over a discontinuous (sawtooth) function [11]. The norm of the resulting signal values is the distance between \mathbf{Y} and \mathbf{W}_m . Hence, the watermark that has the minimum distance to \mathbf{Y} is regarded as the embedded signal. Alternatively, when the demodulation scheme is based on maximum correlation rule, an estimate $\hat{\mathbf{W}}$ of embedded \mathbf{W} is extracted from the received signal. Then, the sent message is detected

by matching the estimate of the embedded watermark to one of the watermarks using a correlation based similarity measure as

$$\hat{\mathbf{W}} = \mathcal{D}(\mathbf{Y}), \quad \hat{m} = \arg \max_m \frac{\mathbf{W}_m \hat{\mathbf{W}}}{\|\mathbf{W}_m\| \|\hat{\mathbf{W}}\|}, \quad 1 \leq m \leq M. \quad (4)$$

To avoid hard decisions in watermark extraction, [13] proposed a continuous periodic triangular extraction function. Hence, an estimate of the embedded watermark is obtained by mapping each coefficient of \mathbf{Y} over the periodic triangular function. Message detection is achieved by combining the sample estimates into $\hat{\mathbf{W}} = (\hat{W}_1, \dots, \hat{W}_N)$ and then matching $\hat{\mathbf{W}}$ to one of $\mathbf{W}_1, \dots, \mathbf{W}_M$.

In scalar quantization based methods, the embedding and detection operations are controlled by a pair of parameters. One of the parameters is the quantization step size Δ which designates the distance between the reconstruction points. The other parameter controls the amount of processing distortion introduced to quantized signal by the post-processing, and it is parameterized depending on the type of post-processing employed at the embedder. For successful operation, the parameter Δ needs to be available both at the embedder and detector whereas the post-processing parameter is only known to embedder. The values for the parameters are obtained as a function of the presumed channel noise level and the permitted embedding distortion amount P_E (implicitly assuming embedding signal size N is large).

3. Multiple Codebook Data Hiding

In scalar quantization based data hiding methods, the embedding distortion P introduced to cover-signal \mathbf{C} is computed over all embedded-signal coefficients, *i.e.*, $P = \frac{1}{N} \|\mathbf{X}_n\|^2$. Since quantization based embedding/detection techniques can be made independent of assumptions on the cover-signal statistics (by key dithering or properly selecting Δ), the distortion introduced to each cover-signal sample C has the statistics of X_n . In other words, the distortion P is a random variable (*rv*) and its distribution approximates $\mathcal{N}(\sigma_{X_n}^2, \frac{\sigma_P^2}{N})$ [11] where

$$\frac{\sigma_P^2}{N} = \frac{1}{N} \left(\int_{-\infty}^{\infty} x_n^4 f_{X_n}(x_n) dx_n - (\sigma_{X_n}^2)^2 \right). \quad (5)$$

Accordingly, when N is large, the distortion P introduced to the host signal becomes $P = \sigma_{X_n}^2$. How-

ever, with decreasing N , P varies more significantly around $\sigma_{X_n}^2$ depending on the distribution of X_n . Typically, embedding/detection parameters are optimized to maximize the performance at the permitted embedding distortion, $P_E = \sigma_{X_n}^2$, and the given channel noise level, σ_Z^2 . Hence, implicitly, a very large embedding signal size N is assumed. Embedding and detection with the parameters obtained through an optimization procedure that disregards this aspect of the problem may cause the data hiding method to perform worse than expected.

An obvious approach to this problem is to fine-tune the parameters obtained with the assumption of large N , so that the resulting embedding distortion P is neither above nor below the permitted distortion level P_E . The question now is, can we do better? Can the fact that the embedding distortion has a large variance be utilized to improve the performance of data hiding? Multiple codebook data hiding method exploits this phenomenon (that embedding distortion has a large variance for small N) by choosing a different representation for \mathbf{C} which yields lower embedding distortion. Then, the ability to embed a watermark at a lower embedding distortion, rather than at the permitted distortion level, is translated into more robust embedding of the watermark.

The essence of the method is depicted in Figure 2 where a binary symbol is embedded into a signal vector $\mathbf{c} = (c_1, c_2)$ using a two-dimensional lattice. The lattice points associated with each binary sample is marked by \times and \circ symbols and embedding is performed by translating vector \mathbf{c} to the nearest centroid associated with the symbol to be embedded. In the considered case, the binary symbol corresponding to \times is embedded into \mathbf{c} and into two of its transformed (rotated) versions \mathbf{c}_2 and \mathbf{c}_3 . The embedding distortions between the signal pairs $(\mathbf{c}, \hat{\mathbf{c}})$, $(\mathbf{c}_2, \tilde{\mathbf{c}}_2)$, and $(\mathbf{c}_3, \tilde{\mathbf{c}}_3)$ are measured, in terms of Euclidean distance, as d_1 , d_2 , and d_3 , respectively. When $\tilde{\mathbf{c}}_2$ and $\tilde{\mathbf{c}}_3$ are inverse transformed, one can observe that the distortions introduced to \mathbf{c} due to three embedding operations are not the same, and $\hat{\mathbf{c}}$ (inverse transformed $\tilde{\mathbf{c}}_2$) yields the lowest embedding distortion, d_2 . (It is important to note that since the transformations are assumed to be unitary the embedding distortions introduced to \mathbf{c}_2 and \mathbf{c}_3 remain same after inverse transformation.) Thus, with the added complexity of transformations, a binary symbol can be embedded into \mathbf{c} at a lower embedding distortion level.

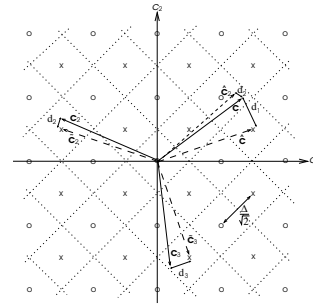


Fig. 2. Depiction of embedding a binary symbol into the cover-signal $\mathbf{c} = (c_1, c_2)$ and into its two transformations using a 2-D lattice.

This idea can be easily generalized to signals of size N by employing $N \times N$ unitary transformations. *Since transformations enable embedding at lower distortion levels, $P < P_E$, the difference between the permitted and actual embedding distortions is utilized by the embedder to either reduce the processing distortion, $\sigma_{X_t}^2$, at the given separation of reconstruction points, Δ , or to further increase the Δ at the fixed $\sigma_{X_t}^2$ while satisfying $P = P_E$.* Both actions lead to an improvement in the detection performance. However, when $N \rightarrow \infty$, for any \mathbf{C} , the embedding distortion converges to the expected value, $P \rightarrow P_E = \sigma_{X_n}^2$, and multiple codebook hiding does not provide any advantage over single codebook hiding.

The underlying idea of our approach is based on the premise that with decreasing N the embedding distortion P assumes a random behavior. Hence, for a given \mathbf{C} and \mathbf{W} , the goal is to generate many realizations of P by embedding \mathbf{W} into many *variants* of \mathbf{C} and to choose the one that yields the lowest P . Use of unitary transformations for this purpose offer two main advantages. First, the rotation of the coordinates, as in Figure 2, ensures that transformations of \mathbf{C} are far enough apart (in the Euclidean sense) in the signal space so that the resulting embedding distortion values are significantly different. Second advantage is due to unitary transformation of coordinates which preserves the signal energy. This guarantees that the embedded-signal in the transformation domain also confirms to distortion constraints when inverse transformed, greatly simplifying the optimization of parameter values.

For a message to be transmitted, the use of multiple codebooks provides the embedder with a freedom in generating a set of codewords and choosing the *best* among them. Correspondingly, the detector has to search over all codebooks for successful extraction of the message. That is, detector should be able to differentiate the correct transformation

from among all transformations of the received signal. Apparently, such a detection of the message is more prone to errors. It is shown in this paper that for AWGN attack, Gaussian distributed cover-signal and square error distortion measure, the increase in probability of error due to use of multiple codebooks is compensated due to embedder's ability to adapt the codeword to the cover-signal. For this, we incorporate the proposed scheme into binary DM with thresholding type of post-processing; however, the concept is applicable to all quantization based embedding/detection techniques. The improvement in the utilization of the permitted embedding distortion is evaluated analytically assuming correlation based detection (rather than minimum distance decoding) because of tractability considerations. However, the effectiveness of the method has been verified by simulation for both types of detection [13,17]. In DM with thresholding, the embedding and detection operations are characterized by the quantization step size Δ and the threshold $0 < \beta < \Delta$. Correspondingly, the expressions for the processing distortion \mathbf{X}_t and the codeword \mathbf{X}_n are obtained as $\mathbf{X}_t = \max(0, |\mathbf{X}| - \frac{\beta}{2})\text{sign}(\mathbf{X})$ and $\mathbf{X}_n = \min(|\mathbf{X}|, \frac{\beta}{2})\text{sign}(\mathbf{X})$, respectively. Detection of the embedded message is as described in (3) or (4).

3.1. Channel Model for Multiple Codebook Data Hiding

In the multiple codebook data hiding scenario, embedder and detector share two sets of information. One is the set of sequences $\mathbf{W}_1, \dots, \mathbf{W}_M \in \mathfrak{R}^N$ that are associated with M distinct messages and the other is the set of L , $N \times N$, unitary transform bases, *i.e.*, $\mathbb{I} = \mathbb{T}_i^T \mathbb{T}_i$ for $\forall i \in [1 \leq i \leq L]$ where \mathbb{I} is the $N \times N$ identity matrix and T denotes the matrix transpose operation. The overall data hiding system can be outlined in an additive model as

$$\begin{aligned}
\mathcal{W} : m &\longrightarrow \mathbf{W}_m, \\
\hat{\mathbf{S}}_k &= \mathcal{E}(\mathbb{T}_k \mathbf{C}, \mathbf{W}_m), \quad 1 \leq k \leq L, \\
\mathbf{S}_k &= \mathbb{T}_k^T \hat{\mathbf{S}}_k, \\
\mathbf{Y} &= \mathbf{S}_k + \mathbf{Z} = \mathbf{C} + \mathbf{X}_{n_k} + \mathbf{Z}, \\
\hat{\mathbf{W}}_m^i &= \mathcal{D}(\mathbb{T}_i \mathbf{Y}), \quad i = 1, \dots, L, \\
\mathcal{W}^{-1} : \hat{\mathbf{W}}_m^i &\longrightarrow \hat{m}.
\end{aligned} \tag{6}$$

In the model, \mathbf{C} is the *iid* Gaussian distributed cover-signal with the marginal $C \sim \mathcal{N}(0, \sigma_C^2)$, $\mathbf{X}_n = \mathbf{X}_{n_k}$ is the codeword and \mathbf{Z} is the AWGN vector where $Z \sim \mathcal{N}(0, \sigma_Z^2)$. Figure 3 displays the block diagram of an L -codebook embedding and detection scheme.

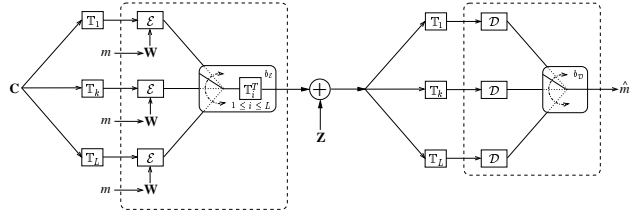


Fig. 3. Multiple codebook embedding and detection.

With the use of multiple codebooks, the choice of \mathbb{T}_k determines the codeword \mathbf{X}_{n_k} among codewords $\{\mathbf{X}_{n_1}, \dots, \mathbf{X}_{n_L}\}$.

The most crucial step of multiple codebook data hiding is the selection of the transformation basis \mathbb{T}_k , $1 \leq k \leq L$, which yields the codeword that adapts to \mathbf{C} best at the permitted embedding distortion P_E . For this, the watermark \mathbf{W}_m is embedded into L transformations of the cover-signal, $\mathbf{C}_i = \mathbb{T}_i \mathbf{C}$ for all i , consecutively. Noting that in scalar quantization based methods embedding and detection functions are not inverses of each other, due to the processing distortion \mathbf{X}_t , the signal \mathbf{W}_m embedded into \mathbf{C}_i will differ from the corresponding extraction $\tilde{\mathbf{W}}_m^i$, *i.e.*, $\mathcal{D}(\mathcal{E}(\mathbf{C}, \mathbf{W}_m)) \neq \mathbf{W}_m$. Therefore, embedder can decide on the transformation basis by measuring the similarity between \mathbf{W}_m embedded into all transformations of \mathbf{C} and the corresponding extractions $\tilde{\mathbf{W}}_m^i$ through computing and comparing normalized correlations, $\tilde{\rho}_{m,m}^i$. The value of i index that yields the highest correlation $\tilde{\rho}_{m,m}^i$, is chosen as the index of the transformation basis \mathbb{T}_k , $k = \arg \max_i (\tilde{\rho}_{m,m}^i)$ for $\tilde{\rho}_{m,m}^i = \frac{\mathbf{W}_m^T \tilde{\mathbf{W}}_m^i}{\|\mathbf{W}_m\| \|\tilde{\mathbf{W}}_m^i\|}$. Then, the embedded-signal in the transform domain, $\hat{\mathbf{S}}_k$, is inverse transformed to signal domain, \mathbf{S}_k .

At the receiver side, on the other hand, the sent message is detected from the received signal \mathbf{Y} without knowing which of the L transformation bases is used for embedding. Hence, the extractor tries all transformations of \mathbf{Y} and extracts signals $\hat{\mathbf{W}}_m^i = \mathcal{D}(\mathbb{T}_i \mathbf{Y})$. Then, the set of extracted signals $\{\hat{\mathbf{W}}_m^1, \dots, \hat{\mathbf{W}}_m^L\}$, of which only $\hat{\mathbf{W}}_m^k$ is a valid extraction, is compared with the set of watermarks $\{\mathbf{W}_1, \dots, \mathbf{W}_M\}$ by computing the normalized correlations $\rho_{m,j}^i$, where $i = 1, \dots, L$ and $j = 1 \dots, M$. Among all (i, j) index pairs, the j index of the pair that maximizes $\rho_{m,j}^i$ is the index of the detected message \hat{m} , $\hat{m} = \arg_j \max_{i,j} (\rho_{m,j}^i)$.

In Sections 3.2 and 3.3, single and multiple codebook data hiding methods are studied and analyzed in terms of their probability of error performances.

3.2. Single Codebook Data Hiding

Let $\mathbf{W}_m^T = [W_{m_1}, \dots, W_{m_N}]$ be a length N *iid* zero mean binary random vector corresponding to message m and $\hat{\mathbf{W}}_m^T = [\hat{W}_{m_1}, \dots, \hat{W}_{m_N}]$ be the extracted real valued signal at the detector. Since the embedding and detection processes are memoryless and both cover-signal and channel noise are white, $\hat{\mathbf{W}}_m$ is an *iid* zero mean random vector, and a detection error is due to $\hat{\mathbf{W}}_m$ having the highest correlation with any of $\{\mathbf{W}_1, \dots, \mathbf{W}_M\}$ other than \mathbf{W}_m . Then, an event E_j that the detector will pick \hat{m} as the detected message instead of m is denoted as

$$E_j = \{\rho_{m,j} \geq \rho_{m,m}\}, \quad 1 \leq j \leq M, j \neq m. \quad (7)$$

The event E^{one} that detector makes a detection error is $E^{one} = \bigcup_{j=1, j \neq m}^M E_j$. Hence, the upper-bound on probability of error for single codebook data hiding, P_e^{one} , can be expressed using (7) as

$$P_e^{one} = Pr(E^{one}) \leq \sum_{j=1, j \neq m}^M Pr(\rho_{m,j} \geq \rho_{m,m}). \quad (8)$$

As detailed in Sections A and B of the Appendix, the pdf of *rv* $\rho_{m,j}$, in (8), can be generalized as

$$\rho_{m,j} \sim \begin{cases} \mathcal{N}(0, \frac{1}{N}), & \text{if } m \neq j \\ \rho_{dep}, & \text{if } m = j. \end{cases} \quad (9)$$

Assuming m is the index of the transmitted message for all the cases, the first subscript, m , of $\rho_{m,j}$ can be dropped for the sake of simplicity. Thus, (8) can be rewritten using (9) as

$$P_e^{one} \leq \sum_{j=1, j \neq m}^M \int_{-\infty}^{\infty} \int_{-\infty}^{\infty} f_{\rho_j}(\rho_j) f_{\rho_m}(\rho_m) d\rho_j d\rho_m. \quad (10)$$

The inner integral in (10) can be expressed in terms of Gaussian Q -function, *i.e.*, $Q(x) = \frac{1}{\sqrt{2\pi}} \int_x^{\infty} e^{-\frac{t^2}{2}} dt$. Since statistics of ρ_j are independent of the index j when $j \neq m$, the sum operator in (10) can be replaced with the factor $M - 1$ and the inequality in P_e^{one} simplifies to

$$P_e^{one} \leq (M - 1) \int_{-\infty}^{\infty} Q(\rho_m \sqrt{N}) f_{\rho_m}(\rho_m) d\rho_m. \quad (11)$$

3.3. Multiple Codebook Data Hiding

In multiple codebook data hiding method, the embedder searches for the transformation basis that yields the least processing distortion. This is done

by choosing the maximum of the correlations $\tilde{\rho}_{m,m}^i$, $\forall i \in [1, \dots, L]$, measured between \mathbf{W}_m embedded into L transformations of \mathbf{C} and the corresponding extractions $\tilde{\mathbf{W}}_m^i$. Due to channel noise \mathbf{Z} , the dependency between the embedded watermark and the extracted signal at the detector reduces. Therefore, the correlation $\rho_{m,m}^i$, between \mathbf{W}_m and its extracted version from \mathbf{Y} , would be less than $\tilde{\rho}_{m,m}^i$. However, unless the noise level is too high, the transformation basis that yields the highest correlation at the embedder will yield the highest correlation at the detector, *i.e.*, $\arg_i \max(\tilde{\rho}_{m,m}^i) = \arg_i \max(\rho_{m,m}^i)$.

Let the maximum of $\rho_{m,m}^i$ be denoted by ρ_{max} with the pdf given as

$$\rho_{max} \sim \max(\rho_{m,m}^1, \dots, \rho_{m,m}^L) \quad (12)$$

where $\rho_{m,m}^i$ are independent random variables with $\rho_{m,m}^i \sim \rho_{dep}$ (Section D of Appendix). With multiple codebook data hiding, detection errors are due to any of the normalized correlation values $\rho_{m,j}^i$, $j \neq m$, being greater than the correlation value ρ_{max} . Assuming \mathbb{T}_k is the transformation basis used for embedding in all cases, an event E_j^i that the detector will pick \hat{m} instead of m is denoted as

$$E_j^i = \{\rho_{m,j}^i \geq \rho_{max}\}, \quad 1 \leq i \leq L, 1 \leq j \leq M, j \neq m. \quad (13)$$

The event E^{mul} that the detector makes an error is $E^{mul} = \bigcup_{i=1}^L \bigcup_{j=1, j \neq m}^M E_j^i$. Hence, the union bound on probability of detecting a wrong message for multiple codebook hiding, P_e^{mul} , is obtained as

$$P_e^{mul} = Pr\{E^{mul}\} \leq \sum_{i=1}^L \sum_{j=1, j \neq m}^M Pr(\rho_{m,j}^i \geq \rho_{max}). \quad (14)$$

Comparing (8) with (14), one sees that the advantage of multiple codebook embedding over single codebook embedding is reflected in the statistics of $\rho_{m,m}$ and ρ_{max} . The distribution of $\rho_{m,j}^i$ can be generalized as

$$\rho_{m,j}^i \sim \begin{cases} \mathcal{N}(0, \frac{1}{N}), & \text{if } i \neq k, \text{ or } i = k \text{ and } j \neq m, \\ \rho_{dep}, & \text{if } i = k \text{ and } j = m, \end{cases} \quad (15)$$

as detailed in Section D of Appendix. The probability of error for multiple codebook data hiding, (14), can be further rewritten using the above results as

$$P_e^{mul} \leq \sum_{i=1}^L \sum_{j=1, j \neq m}^M \int_{-\infty}^{\infty} \int_{-\infty}^{\infty} f_{\rho_j^i}(\rho_j^i) f_{\rho_{max}}(\rho) d\rho_j^i d\rho_{max} \quad (16)$$

where the first subscript referring to the transmitted message m is dropped. Since the inner integral in

(16) is the Gaussian- Q function and does not depend on the index j , (16) can be simplified to

$$P_e^{mul} \leq L(M-1) \int_{-\infty}^{\infty} Q(\rho_{max}\sqrt{N}) f_{\rho_{max}}(\rho_{max}) d\rho_{max}. \quad (17)$$

Note that for $L = 1$, (17) reduces to (11). Essentially, the change in P_e^{mul} with respect to L depends on (12). That is, as long as the increase in ρ_{max} with L can compensate for the additional errors due to search over L codebooks, P_e^{mul} will decrease; otherwise, P_e^{mul} will start increasing.

3.4. Comparisons

Figures 4 a-c display the union bound on the probability of error for thresholding type of post-processing and correlation based detection. The curves are obtained by numerically solving (17) for varying number of codebooks and codebook sizes $M \times N$ and at different WNRs, where $WNR = 10 \log_{10}(\frac{P_E}{\sigma_Z^2})$. In all cases, at a fixed N , as the number of codebooks increases the bound on the probability of error decreases. On the other hand, at a fixed WNR, probability of error for single codebook data hiding also decreases with the increasing signal size N . Intuitively, this is due to the increasing confidence in the detection with N . With reference to the analysis in Section 3.3, as $m_{\rho_{dep}}$ increases and $\sigma_{\rho_{dep}}^2$ decreases, the maximum of the ensemble of random variables $\tilde{\rho}_{m,m}^1, \dots, \tilde{\rho}_{m,m}^L$ is less likely to differ from the rest. Hence, for large values of N this deviation will get smaller and less number of variables will be sufficient to utilize the gap between the permitted and estimated distortions. Therefore, with increasing N the improvement due to use of multiple codebooks becomes more pronounced at lower WNRs where single codebook embedding performs relatively poorly.

Another concern is the probability of false-positives. When a cover-signal is subjected to watermark detection, the extracted signal will be *iid* uniformly distributed and uncorrelated with any of the watermarks. As a result, the normalized correlation measured between the extracted signal and the watermarks will be irrespective of the channel noise level and distributed as $\mathcal{N}(0, \frac{1}{N})$. Considering a fixed threshold for message detection, the false-positive rate within multiple codebook data hiding increases, approximately, with a factor of L compared to single codebook data hiding (as there are so many comparisons that may yield a false

positive). However, noting that the use of multiple codebooks enables embedding a watermark with less processing distortion, the detection performance is improved. The numerical solutions of (17) indicates that the increase in the P_e^{mul} by the factor of L , compared to P_e^{one} , is compensated by embedder's ability to better adapt the codeword to the cover-signal as a result of which detection statistics are improved from those of ρ_{dep} to ρ_{max} . In a similar manner, the increase in false-positive rate with the number of codebooks can be compensated by proper selection of the threshold which relies on the statistics of ρ_{max} rather than of ρ_{dep} .

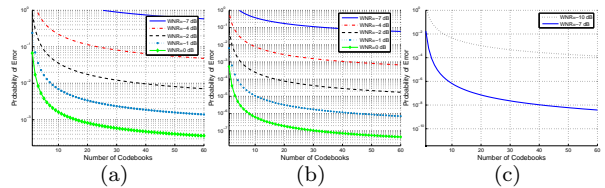


Fig. 4. Probability of error performance for multiple codebook hiding for (a) $M=100$ and $N=50$, (b) $M=200$ and $N=100$ and (c) $M=1000$ and $N=500$.

The computational complexity of the multiple codebook data hiding technique ultimately depends on the number of codebooks employed. In multiple codebook data hiding, the embedder selects the codeword associated with a message from among a set of codewords it generated and the detector performs a search over all codebooks prior to message extraction. Multiple codebook embedding, when compared with single codebook embedding, requires the embedding of the watermark into transformations of the cover-signal and a comparison based on the resulting signals, in order to select the transformation basis. On the other hand, at the detector, extraction should be repeated for each transformation basis. Therefore, the computational complexity increases almost linearly with the number of codebooks, see Figure 3.

3.5. Implementation and Simulation Results

Optimum codeword selection in multiple codebook hiding depends on designing the set of transform bases properly, (*i.e.*, they should be able to generate maximally separated transformations of the cover-signal). One intuitive way of picking such a set of transform bases is by choosing them among rotation matrices so that each transformation of the cover-signal is a rotated version of the others. This can be achieved by designing the transformation

bases using Givens rotations, which provide orthogonal transformations in \mathfrak{R}^N to rotate a vector by a chosen angle [18].

The watermarks are generated as the rows of $N \times N$ Hadamard transform matrix and its negated version, *i.e.*, $M = 2N$, due to its simplicity. The cover-signal and channel noise are *iid* zero mean Gaussian vectors. Prior to embedding, the permitted embedding distortion P_E is fixed, and the optimal values for the embedding parameter Δ are derived for the considered WNRs. The parameter β , however, is properly set in order to ensure an embedding distortion of P_E .

We performed hiding with up to 25 codebooks considering codebook sizes of 64×32 , 128×64 , 256×128 and the WNR range of -12 dB to 0 dB. Figure 5 displays the probability of success rates for $L = \{1, 3\}$ and varying N values. The increase in the embedding signal size N , at a fixed number of codebooks, improves the detection statistics since normalized correlation gives more reliable results with the larger signal sizes. Figure 6-a displays the probability of success for $N = 128$ and $L = \{1, 3, 5, 9, 14, 25\}$ when correlation based detector is used. Similarly, the results obtained for minimum distance decoding are displayed in Figure 6-b. It is observed from these performance simulations that the use of three codebooks for embedding and detection improves the performance substantially, compared to single codebook case, for both of the detection approaches. On the other hand, for larger number of codebooks the improvement is modest but still noticeable. It should also be considered that in the proposed method, each codeword is generated by rotating a fixed watermark vector in the high-dimensional space at equally spaced angles, *e.g.*, $\frac{2\pi}{L}$. That is, the distance between the codewords of consecutive codebooks depends on $\frac{2\pi}{L}$ (Figure 2), and with increasing L this distance gets smaller and smaller. Therefore, when the channel noise drags the signal to its neighbor transformation (*i.e.*, codebook) the resulting signal will still be mapped to the same codeword but of a different codebook, and will not necessarily lead to a detection error.

3.6. Case Study: Image Watermarking under JPEG Compression Attack

To further assess the effectiveness of the technique, we consider digital image watermarking application under JPEG compression attack. In our

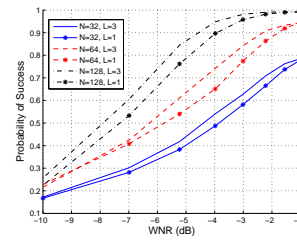


Fig. 5. (Probability of success performance for 3-codebook hiding for various watermark sizes of $N = 32$, $N = 64$ and $N = 128$).

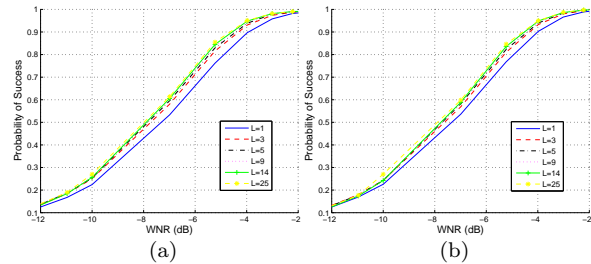


Fig. 6. Probability of success performance for multiple codebook hiding for various number of codebooks $L = \{1, 3, 5, 9, 14, 25\}$ and $N = 128$ using (a) correlation based detection and (b) minimum distance decoding.

setting, the cover-signal is a 256×256 gray-level uncompressed image and the watermark is a random bit sequence which is embedded only in the middle frequency DCT coefficient channels using binary DM with thresholding type of post-processing. In each frequency channel, a number of coefficients, depending on the codebook size, are selected and embedded using varying number of codebooks. Then, the marked-image is JPEG compressed. This is followed by detecting the watermark from the compressed embedded-image through sample-wise soft decisions.

The parameter Δ used for embedding and detection is selected to maximize robustness against lossy JPEG compression at the given quality factor (QF) and the parameter β is fine-tuned by embedder to comply with the permitted embedding distortion. We consider $L \in \{1, 3, 5\}$ codebooks and codeword lengths of $N \in \{256, 512, 1024\}$ for embedding/detection. The transformation bases of size $N \times N$ are obtained as described in Section 3.5. In our experiments, the mean squared embedding distortion introduced to DCT coefficients is permitted to change in the range between $P_E = 5$ and $P_E = 15$ (with PSNR greater than 40 dB and no visual artifacts).

To compare single and multiple codebook em-

bedding cases, the embedding and detection operations are repeated for 500 randomly generated watermarks at each permitted embedding distortion level. Performance results are given in terms of probability of successful watermark detection values, obtained based on the number of correctly detected watermarks, and the averaged correlation values, measured between the embedded and correctly detected watermarks. Figure 7 displays corresponding results when $N = 256$ and the marked image is compressed at QF 75 and 60. These results show that for fixed permitted embedding distortion levels (in DCT domain) the performance improves with the proposed scheme at both compression levels. In a similar manner, Figure 8 show results when the codeword length is increased to 512 and 1024 under JPEG compression with QF 75. In this case, it is observed that the overall performance is relatively insensitive to increase in signal size N , and the use of multiple codebooks still offers an improvement over single codebook case. This can be attributed to the non-*iid* nature of DCT coefficient channels due to high correlation among 8×8 sub-blocks of the cover-image. That is, the embedding distortion, when transformed, continues to exhibit sufficient variation at higher signal lengths, thereby making the scheme viable at larger N .

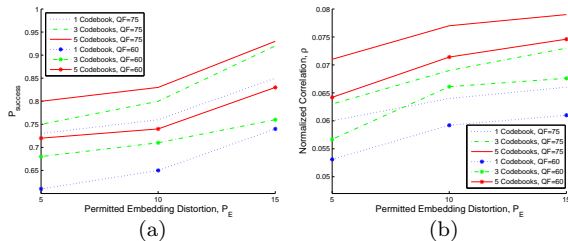


Fig. 7. (a) Probability of successful detection and (b) averaged correlation values between the embedded and successfully extracted watermarks for $N = 256$ under JPEG compression attack at QF 75 and 60.

4. Conclusions

In this paper, we studied multiple codebook data hiding technique to improve the performance of scalar quantization based embedding/detection techniques when the cover-signal size is smaller. The use of multiple codebooks provide the embedder with a codeword that better adapts to the cover-signal. For a given cover-signal and a watermark, this is achieved by embedding the watermark in a transformed version of the cover-signal, which yields a lower embedding distortion, so that the

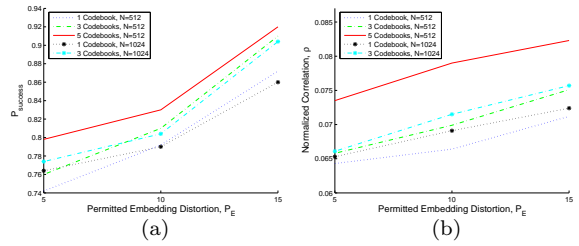


Fig. 8. (a) Probability of successful detection and (b) averaged correlation values between the embedded and successfully extracted watermarks for $N = 512$ and $N = 1024$ under JPEG compression attack at QF 75.

resulting gap with the permitted distortion can be used to improve the robustness. The proposed method does not require any changes in the embedding and detection operations of the underlying scheme. It merely requires the embedding to be performed multiple times in order to choose the codeword corresponding to the message being embedded. Similarly, multiple extractions are performed before making a decision on the received message. Analytical results indicate that the the probability of detection error decreases with the number of codebooks. Simulation results obtained for synthetically generated cover-signals under AWGN attacks show that the use of multiple codebooks for data hiding is indeed superior to single codebook data hiding. Furthermore, the potential of the proposed method is also demonstrated by applying the proposed method to image watermarking problem considering JPEG compression attack.

Appendix A. Distribution of ρ_{ind}

If \mathbf{W}_m and $\hat{\mathbf{W}}_m$ have a zero covariance matrix, because of distortion by channel noise, the normalized correlation ρ_{ind} between \mathbf{W}_m and $\hat{\mathbf{W}}_m$ is defined as

$$\rho_{ind} = \frac{\mathbf{W}_m^T \hat{\mathbf{W}}_m}{\|\mathbf{W}_m\| \|\hat{\mathbf{W}}_m\|} = \sum_{l=1}^{l=N} \frac{W_{m_l} \hat{W}_{m_l}}{\|\mathbf{W}_m\| \|\hat{\mathbf{W}}_m\|}. \quad (\text{A.1})$$

Due to *iid* assumption, the normalized *rv*'s $\frac{W_{m_l}}{\|\mathbf{W}_m\|}$ and $\frac{\hat{W}_{m_l}}{\|\hat{\mathbf{W}}_m\|}$ are both zero mean with variance $\frac{1}{N}$. Consequently, through central limit theorem, the normalized correlation ρ_{ind} approximates Gaussian distribution with mean $m_{\rho_{ind}} = 0$ and variance $\sigma_{\rho_{ind}}^2 = \frac{1}{N}$, $\rho_{ind} \sim \mathcal{N}(0, \frac{1}{N})$. Similarly, when $\hat{\mathbf{W}}_m$ and \mathbf{W}_j are independent *iid* random vectors, the normalized correlation $\rho_{m,j} \sim \rho_{ind}$.

Appendix B. Distribution of ρ_{dep}

When \mathbf{W}_m and $\hat{\mathbf{W}}_m$ are dependent, the samples W_{m_l} and \hat{W}_{m_l} , $1 \leq l \leq N$, are somewhat correlated, and the normalized correlation ρ_{dep} , defined between \mathbf{W}_m and $\hat{\mathbf{W}}_m$, is as given in (A.1).

Since \mathbf{Z}_f is the noise that distorts the embedded \mathbf{W}_m , the signal $\hat{\mathbf{W}}_m$, extracted from \mathbf{Y} , can be expressed in terms of \mathbf{Z}_f and \mathbf{W}_m . Hence, for a binary distributed watermark sample W , of the *iid* vector \mathbf{W}_m , with a value in $\{-\frac{\Delta}{4}, \frac{\Delta}{4}\}$, the extracted sample \hat{W} is expressed in terms of Z_f and W as

$$\hat{W} = \begin{cases} (\frac{(2i+1)\Delta}{4} - Z_f)(-1)^i, & \text{if } W = \frac{\Delta}{4}, \\ (-\frac{(2i+1)\Delta}{4} + Z_f)(-1)^i, & \text{if } W = -\frac{\Delta}{4}. \end{cases} \quad (\text{B.1})$$

where $i\frac{\Delta}{2} < Z_f \leq \frac{(i+1)\Delta}{2}$ and $i \in \mathcal{Z}$.

Due to memoryless embedding/detection and attack schemes, generation of the vector $\hat{\mathbf{W}}_m$ from \hat{W} is straightforward. The pdf of Z_f can be found in terms of the pdf's of Z and X_t [11]. Ultimately, ρ_{dep} can be calculated in terms of embedding/detection parameters, N , and statistics of Z_f and W .

It should be noted that another source of randomness in ρ_{dep} is due to the embedding distortion P . When N is not large enough, P deviates from $P_E = \sigma_{X_n}^2$. This requires a refinement of the embedding parameter values, which are optimized for large N , so that they yield $P = P_E$. Therefore, the correlation of \mathbf{W}_m and $\hat{\mathbf{W}}_m$ is actually a *rv* conditioned on P , $\rho_{dep|P}$, with the mean m_{ρ^*} and variance $\sigma_{\rho^*}^2$ as given in Section C. Since covariance matrix of \mathbf{W}_m and $\hat{\mathbf{W}}_m$ is diagonal, distribution of *rv* $\rho_{dep|P}$ approximates Gaussian distribution, $\rho_{dep|P} \sim \mathcal{N}(m_{\rho^*}, \sigma_{\rho^*}^2)$, and pdf of ρ_{dep} can be obtained as

$$f_{\rho_{dep}}(\rho_{dep}) = \int_{-\infty}^{\infty} f_{\rho_{dep|P}}(\rho_{dep|P})f_P(P)dP, \quad (\text{B.2})$$

where $P \sim \mathcal{N}(\sigma_{X_n}^2, \frac{\sigma_P^2}{N})$.

Appendix C. Statistics of $\rho_{dep}|P$

The mean m_{ρ^*} of *rv* $\rho_{dep|P}$ can be computed through deriving the joint and marginal moments of W and \hat{W} . The *rv* \hat{W} can be expressed in terms of Z_f and W as in (B.1), where W is a *rv* with pdf $f_W(w) = \frac{1}{2}\delta(w - \frac{\Delta}{4}) + \frac{1}{2}\delta(w + \frac{\Delta}{4})$. Hence, the pq -th joint moment of W and \hat{W} is defined as

$$E[W^p \hat{W}^q] = \int_{-\infty}^{\infty} \int_{-\infty}^{\infty} w^p \hat{w}^q f_{W, \hat{W}}(w, \hat{w}) dw d\hat{w}. \quad (\text{C.1})$$

The joint pdf in the above integral can be expressed in terms of marginal and conditional pdf's, $f_{W, \hat{W}}(w, \hat{w}) = f_{\hat{W}}(\hat{w}|w)f_W(w)$. Since the expectation of a function of a random variable can be expressed in terms of the pdf of the random variable itself rather than of the function, $E[\hat{W}] = \int_{-\infty}^{\infty} \hat{w}(z_f)f_{Z_f}(z_f)dz_f$, and since all pdf's are assumed to be symmetric, (C.1) can be written as

$$E[W^p \hat{W}^q] = \left(\frac{1}{2} + \frac{(-1)^{p+q}}{2}\right) \left(\frac{\Delta}{4}\right)^p R(q), \quad (\text{C.2})$$

where $R(q)$ is defined as

$$R(p) = 2 \sum_{i=0}^{i=\infty} \int_{\frac{i\Delta}{2}}^{\frac{(i+1)\Delta}{2}} \left(\left(\frac{(2i+1)\Delta}{4} - z_f\right)(-1)^i\right)^p f_{Z_f}(z_f)dz_f. \quad (\text{C.3})$$

Hence, the joint moment of W and \hat{W} is generalized, based on (C.2), as

$$E[W^p \hat{W}^q] = \begin{cases} \left(\frac{\Delta}{4}\right)^p R(q), & \text{if } p, q \text{ are both even or odd,} \\ 0, & \text{otherwise.} \end{cases} \quad (\text{C.4})$$

Marginal moments of W are derived straightforwardly, due to the binary distribution, as

$$E[W^p] = \begin{cases} 0, & \text{if } p \text{ is odd,} \\ \left(\frac{\Delta}{4}\right)^p, & \text{if } p \text{ is even.} \end{cases} \quad (\text{C.5})$$

The moments of the *rv* \hat{W} can be computed in a manner similar to (C.2) as

$$E[\hat{W}^p] = \left(\frac{1}{2} + \frac{1}{2}(-1)^p\right) R(p). \quad (\text{C.6})$$

Finally, $E[\hat{W}^p]$ can be summarized as

$$E[\hat{W}^p] = \begin{cases} 0, & \text{if } p \text{ is odd,} \\ R(p), & \text{if } p \text{ is even.} \end{cases} \quad (\text{C.7})$$

Based on (C.1)-(C.7), m_{ρ^*} is derived as

$$m_{\rho^*} = \frac{E[W\hat{W}]}{\sqrt{E[W^2]E[\hat{W}^2]}} = \frac{\frac{\Delta}{4}R(1)}{\sqrt{(\frac{\Delta}{4})^2 R(2)}} = \frac{R(1)}{\sqrt{R(2)}}. \quad (\text{C.8})$$

The variance $\sigma_{\rho^*}^2$ is the variation of the correlation coefficient around its mean m_{ρ^*} . When \hat{W} and W are from a bivariate Gaussian distribution, the variance is as given in [19]. However, when the samples are from non-Gaussian distributions, derivation of

σ_{ρ^*} is not straightforward. Therefore, Monte-Carlo simulations are performed to obtain the $\sigma_{\rho^*}^2$ values for the considered N by computing the correlations between the embedded \mathbf{W}_m and extracted $\hat{\mathbf{W}}_m$ at the assumed WNR and, then, by measuring the deviation from m_{ρ^*} .

Appendix D. Distribution of ρ_{max} and $\rho_{m,j}^i$

The *rv* ρ_{max} is the maximum of L random variables, (12), that are all distributed according to pdf of *rv* ρ_{dep} . Correspondingly, the pdf of ρ_{max} can be expressed in terms of the pdf of ρ_{dep} as $f_{\rho_{max}}(\rho_{max}) = LF_{\rho_{dep}}^{L-1}(\rho_{max})f_{\rho_{dep}}(\rho_{max})$ where $F_X(x) = \int_{-\infty}^x f_X(x)dx$.

On the other hand, the pdf of the *rv*'s $\rho_{m,j}^i$ can be found based on the choice of i and j . When detector assumes $i = k$, the transformations used for embedding and detection matches. Therefore, the results of the analysis given in Sections A and B also apply to multiple codebook data hiding. Consequently, the normalized correlation $\rho_{m,j}^k$, $1 \leq j \leq M$, is equivalent to random variables ρ_{dep} and ρ_{ind} in its statistics respectively for $j = m$ and $j \neq m$.

If there is a mismatch between the embedding and detection transformations, $i \neq k$, then an extraction from \mathbb{T}_i transformation of the received signal does not provide meaningful information about \mathbf{W}_m since embedding transformation was \mathbb{T}_k . Consequently, the binary distributed \mathbf{W}_m with values in $\{-\frac{\Delta}{4}, \frac{\Delta}{4}\}$ is extracted, $\hat{\mathbf{W}}_m^i$, as a uniformly distributed sample sequence in the range $[-\frac{\Delta}{4}, \frac{\Delta}{4}]$ which is independent from \mathbf{W}_m . Therefore, the normalized correlation $\rho_{m,j}^i$, $i \neq k$ and $\forall j$, has the same statistics as the *rv* ρ_{ind} , $\rho_{m,j}^i \sim \mathcal{N}(0, \frac{1}{N})$.

References

- [1] J. Chou, S. S. Pradhan, L. E. Ghaoui, K. Ramchandran, On the duality between data hiding and distributed source coding, in: Proc. of 33rd Annual Asilomar conference on Signals, Systems, and Computers, 1999.
- [2] R. J. Barron, B. Chen, G. W. Wornell, The duality between information embedding source coding with side information and its implications—applications, IEEE Transactions on Information Theory 49 (5) (2003) 1159–1180.
- [3] R. Zamir, S. Shamai, U. Erez, Nested linear/lattice codes for structured multiterminal binning, IEEE Transactions on Information Theory 48 (5) (2002) 1250–1276.
- [4] I. J. Cox, M. L. Miller, A. L. McKellips, Watermarking as communication with side information, Proc. of IEEE 87 (1999) 1127–1141.
- [5] I. J. Cox, J. Kilian, T. Leighton, T. Shamoon, Secure spread spectrum watermarking for multimedia, IEEE Transactions on Image Processing 6 (12) (1997) 1673–1687.
- [6] B. Chen, G. W. Wornell, Quantization index modulation: A class of provably good methods for digital watermarking and information embedding, IEEE Transactions on Information Theory 47 (4) (2001) 1423–1443.
- [7] L. Perez-Freire, F. Perez-Gonzalez, Spread-spectrum vs. quantization-based data hiding: Misconceptions and implications, in: SPIE Conference on Security, Steganography, and Watermarking of Multimedia Contents VII, Vol. 5681, 2005.
- [8] M. Costa, Writing on dirty paper, IEEE Transactions on Information Theory 29 (1983) 439–441.
- [9] P. Moulin, J. A. O’Sullivan, Information-theoretic analysis of information hiding, IEEE Transactions on Information Theory 49 (2003) 563–593.
- [10] A. S. Cohen, A. Lapidot, The gaussian watermarking game, IEEE Transactions on Information Theory 48 (2002) 1639–1667.
- [11] H. T. Sencar, M. Ramkumar, A. N. Akansu, An overview of scalar quantization based data hiding methods, Signal Processing 86 (3) (2006) 893–914.
- [12] P. Moulin, R. Koetter, Data hiding codes, Proc. of IEEE 93 (2005) 2083–2126.
- [13] M. Ramkumar, A. N. Akansu, Self-noise suppression schemes for blind image steganography, in: Proc SPIE International Workshop on Voice, Video and Data Communication, Multimedia Applications, Vol. 3845, 1999.
- [14] B. Chen, G. Wornell, Preprocessed and postprocessed quantization index modulation methods for digital watermarking, in: Proc SPIE: Security and Watermarking of Multimedia Contents II, Vol. 3971, 2000, pp. 48–59.
- [15] J. J. Eggers, R. Bauml, R. Tzschoppe, B. Girod, Scalar Costa scheme for information embedding, IEEE Transactions on Signal Processing 51 (4) (2003) 1003–1019.
- [16] F. Perez-Gonzalez, F. Balado, J. R. Hernandez Martin, Performance analysis of existing and new methods for data hiding with known-host information in additive channels, IEEE Transactions on Signal Processing 51 (4) (2003) 960–980.
- [17] B. Chen, G. W. Wornell, Dither modulation: A new approach to digital watermarking and information embedding, in: Proc. of SPIE: Security and Watermarking of Multimedia Contents, Vol. 3657, 1999, pp. 342–353.
- [18] D. S. Watkins, Fundamentals of Matrix Computations, New York: John Wiley & Sons, 1991.
- [19] R. A. Fisher, Statistical Methods for Research Workers, New York: Hafner Press, 1970.