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

Husrev T. Sencar a,*, Mahalingam Ramkumar b, Ali N. Akansu c, Amol Sukkerkar 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-
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 iden-
tically distributed (iid) assumption on the random
variables.

Table 1

<table>
<thead>
<tr>
<th>Notation used in the paper</th>
</tr>
</thead>
<tbody>
<tr>
<td>C</td>
</tr>
<tr>
<td>$X_m$</td>
</tr>
<tr>
<td>Z</td>
</tr>
<tr>
<td>$W_m$</td>
</tr>
<tr>
<td>$W^{\hat{m}}_{m}$</td>
</tr>
<tr>
<td>$\rho_{m,j}$</td>
</tr>
<tr>
<td>$\hat{W}^m$</td>
</tr>
<tr>
<td>$\hat{W}^m_{T_i}$</td>
</tr>
<tr>
<td>$\rho_{m,j}^v$</td>
</tr>
<tr>
<td>$P_E$</td>
</tr>
<tr>
<td>W.N.R</td>
</tr>
</tbody>
</table>

In the next section, Section 2, we briefly describe
the characteristics of scalar quantization based embed-
ding/detection techniques. We introduce the
multiple codebook data hiding technique in Section 3. The performance analysis methodology for sin-
gle and multiple codebook data hiding cases, and
the performance results are presented in the follow-
ing 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 tech-
niques 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 quan-
tization followed by a post-processing like thresh-
olding [13], distortion compensation (DC) [14,15],
or Gaussian mapping (GM) [16]. That is, the cov-
ersignal coefficients are first quantized with respect
to the watermark samples where each sample takes

![Fig. 1. Channel model for embedding and detection operations.](image)
on discrete values from a finite set and each value is
associated with a different quantizer. Then, the res-
tulting 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 deci-
sions [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 embed-
ding/detection techniques is displayed in Figure 1. In
this model, $m$ is the message to be hidden, $C$ is the
cover-signal, $W$ is the watermark, $X$ is the distor-
tion introduced to $C$ due to quantization, $X_t$ is the
processing distortion due to post-processing, $S$ is the
embedded-signal, $Z$ is the additive channel noise, $Y$
is the distorted embedded-signal, $W$ is an estimate
of $W$, and $\hat{m}$ is the detected message. It should be
noted that $X_t$ is generated as a function of $X$ de-
pending on the expected noise level in the channel.
The embedded-signal $S$ can be defined in terms of the
codeword, $X_n = X - X_t$, as $S = C + X_n$ where
the collection of $X_n$ corresponding to all messages
constitute a codebook. Consequently, the per-sample
distortion $P$ introduced to cover-signal due to em-
bedding can be expressed as $P = \frac{1}{N}||X_n||^2$. In the
model, the signal $C + X$ refers to a quantized signal
and the watermark $W$ can be perfectly recovered
from this signal. Since detector is not aware of the
post-processing at the embedder, the corresponding
processing distortion $X_t$, introduced at the embed-
der, can be considered as another source of noise.
Therefore, the effective noise at the detector, that
distorts the embedded watermark $W$, can be defined
as $Z_f = Z - X_t$. Correspondingly, the performance
analysis of an embedding/detection scheme can be
conducted in terms of the statistics of $X, X_t$, and $Z$.

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

$$S = Q_{\Delta}(C + W) - W \quad \text{(1)}$$
where $Q_{\Delta}(\cdot)$ can be considered to be a product quantizer generated by a Cartesian product of $N$ uniform scalar quantizers, $q_{\Delta}$, with quantization step size $\Delta$. Therefore, embedding can be viewed as $N$ successive scalar quantization, of the coefficients of $\mathbf{C} = (C_1, \ldots, C_N)$, dithered with the watermark vector $\mathbf{W}_m = (W_1, \ldots, W_N)$. Each distinct value of the watermark (dither) signal is associated with a quantizer that is generated by properly shifting the reconstruction points of $q_{\Delta}$. 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

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

where $\mathbf{W}_m$ is the watermark associated with the message index $m$. The presence of watermarks $\mathbf{W}_1, \ldots, \mathbf{W}_M$ 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 $\mathbf{W}$ 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

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

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 $\hat{W} = (\hat{W}_1, \ldots, \hat{W}_N)$ and then matching $\hat{W}$ to one of $\mathbf{W}_1, \ldots, \mathbf{W}_M$.

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 $\Delta$ 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 $\Delta$ 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 $\Delta$), 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 $\mathcal{N}(\sigma_{\hat{X}_n}^2, \sigma_P^2/N)$ [11] where

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

Accordingly, when $N$ is large, the distortion $P$ introduced to the host signal becomes $P = \sigma_{\hat{X}_n}^2$. How-
ever, with decreasing $N$, $P$ varies more significantly around $\sigma_X^2$, depending on the distribution of $X$. Typically, embedding/detection parameters are optimized to maximize the performance at the permitted embedding distortion, $P_E = \sigma_X^2$, and the given channel noise level, $\sigma_Z^2$. 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 $C$ 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 $c = (c_1, c_2)$ using a a two-dimensional lattice. The lattice points associated with each binary sample is marked by $\times$ and $o$ symbols and embedding is performed by translating vector $c$ to the nearest centroid associated with the symbol to be embedded. In the considered case, the binary symbol corresponding to $\times$ is embedded into $c$ and into two of its transformed (rotated) versions $c_2$ and $c_3$. The embedding distortions between the signal pairs $(c, c)$, $(c_2, c_2)$, and $(c_3, c_3)$ are measured, in terms of Euclidean distance, as $d_1$, $d_2$, and $d_3$, respectively. When $c_2$ and $c_3$ are inverse transformed, one can observe that the distortions introduced to $c$ due to three embedding operations are not the same, and $c_2$ (inverse transformed $\hat{c}_2$) 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 $c_2$ and $c_3$ remain same after inverse transformation.) Thus, with the added complexity of transformations, a binary symbol can be embedded into $c$ at a lower embedding distortion level.

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, $\sigma_X^2$, at the given separation of reconstruction points, $\Delta$, or to further increase the $\Delta$ at the fixed $\sigma_X^2$, while satisfying $P = P_E$. Both actions lead to an improvement in the detection performance. However, when $N \rightarrow \infty$, for any $C$, the embedding distortion converges to the expected value, $P = P_E = \sigma_X^2$, 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 $C$ and $W$, the goal is to generate many realizations of $P$ by embedding $W$ into many variants of $C$ 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 $C$ 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 conforms 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 $\Delta$ and the threshold $0 < \beta < \Delta$. Correspondingly, the expressions for the processing distortion $X_i$ and the codeword $X_n$ are obtained as $X_i = \max(0, |X| - \frac{\Delta}{2}) \text{sign}(X)$ and $X_n = \min(|X|, \frac{\Delta}{2}) \text{sign}(X)$, 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 $W_1, \ldots, W_M \in \mathbb{R}^N$ that are associated with $M$ distinct messages and the other is the set of $L$, $N \times N$, unitary transform bases, i.e., $I = T_i^T T_i$ for $1 \leq i \leq L$ where $I$ is the $N \times N$ identity matrix and $T^T$ denotes the matrix transpose operation. The overall data hiding system can be outlined in an additive model as

$$W : m \rightarrow W_m, \quad S_k = E(T_k, C, W_m), \quad 1 \leq k \leq L,$$
$$S_k = T_i^T S_k,$$
$$Y = S_k + Z = C + X_n + Z,$$
$$W_m = D(T_i, Y), \quad i = 1, \ldots, L,$$
$$W_m^{-1} : \hat{W}_m \rightarrow \hat{m}.$$  

In the model, $C$ is the iid Gaussian distributed cover-signal with the marginal $C \sim \mathcal{N}(0, \sigma^2_C)$, $X_n = X_{n_k}$ is the codeword and $Z$ is the AWGN vector where $Z \sim \mathcal{N}(0, \sigma^2_Z)$. Figure 3 displays the block diagram of an L-codebook embedding and detection scheme.

With the use of multiple codebooks, the choice of $T_k$ determines the codeword $X_{n_k}$ among codewords $\{X_{n_1}, \ldots, X_{n_k}\}$.

The most crucial step of multiple codebook data hiding is the selection of the transformation basis $T_k$, $1 \leq k \leq L$, which yields the codeword that adapts to $C$ best at the permitted embedding distortion $P_E$. For this, the watermark $W_m$ is embedded into $L$ transformations of the cover-signal, $C_i = T_i C$ 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 $X_i$, the signal $W_m$ embedded into $C_i$ will differ from the corresponding extraction $W_m$, i.e., $D(E(C, W_m)) \neq W_m$. Therefore, embedder can decide on the transformation basis by measuring the similarity between $W_m$ embedded into all transformations of $C$ and the corresponding extractions $\hat{W}_m$ through computing and comparing normalized correlations, $\tilde{\rho}_{m,m}$. The value of $i$ index that yields the highest correlation $\tilde{\rho}_{m,m}$, is chosen as the index of the transformation basis $T_k$, $k = \arg \max_i (\tilde{\rho}_{m,m})$ for $\tilde{\rho}_{m,m} = \frac{W_m^T \hat{W}_m}{||W_m|| ||\hat{W}_m||}$. Then, the embedded-signal in the transform domain, $\hat{S}_k$, is inverse transformed to signal domain, $S_k$.

At the receiver side, on the other hand, the sent message is detected from the received signal $Y$ without knowing which of the $L$ transformation bases is used for embedding. Hence, the extractor tries all transformations of $Y$ and extracts signals $W_{m}^{i} = D(T_i, Y)$. Then, the set of extracted signals $\{W_{m_1}, \ldots, W_{m_L}\}$, of which only $W_{m_k}$ is a valid extraction, is compared with the set of watermarks $\{W_1, \ldots, W_M\}$ by computing the normalized correlations $\rho_{m,j}$, where $i = 1, \ldots, L$ and $j = 1, \ldots, M$. Among all $(i,j)$ index pairs, the $j$ index of the pair that maximizes $\rho_{m,j}$ is the index of the detected message $\hat{m}$, $\hat{m} = \arg \max_{m,j} (\rho_{m,j})$.

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 \( W_m^n = [W_{m1}, \ldots, W_{mn}] \) be a length \( N \) iid zero mean binary random vector corresponding to message \( m \) and \( \tilde{W}_m^n = [\tilde{W}_{m1}, \ldots, \tilde{W}_{mn}] \) 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, \( W_m \) is an \( iid \) zero mean random vector, and a detection error is due to \( \tilde{W}_m \) having the highest correlation with any of \( \{W_1, \ldots, W_M\} \) other than \( W_m \). Then, an event \( E_j \) that the detector will pick \( \hat{m} \) as the detected message instead of \( m \) is denoted as

\[ E_j = \{ \rho_{m,j} \geq \rho_{m,m} \}, \quad 1 \leq j \leq M, \; j \neq m. \]  

(7)

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

\[ P_{\text{one}} = \Pr(E_{\text{one}}) \leq \sum_{j=1, j \neq m}^M \Pr(\rho_{m,j} \geq \rho_{m,m}). \]  

(8)

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

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

(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

\[ P_{\text{one}} \leq \sum_{j=1, j \neq m}^M \int_{-\infty}^{\infty} \int_{-\infty}^{\infty} f_{\rho_j}(\rho_j) f_{\rho_m}(\rho_m) d\rho_j d\rho_m. \]  

(10)

The inner integral in (10) can be expressed in terms of Gaussian Q-function, i.e., \( Q(x) = \frac{1}{\sqrt{\pi}} \int_{-x}^{\infty} e^{-t^2} dt \). Since statistics of \( \rho_j \) 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 \( P_{\text{one}} \) simplifies to

\[ P_{\text{one}} \leq (M - 1) \int_{-\infty}^{\infty} Q(\rho_m \sqrt{N}) f_{\rho_m}(\rho_m) d\rho_m. \]  

(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 \( \rho_{m,m} \), \( \forall i \in [1, \ldots, L] \), measured between \( W_m \) embedded into \( L \) transformations of \( C \) and the corresponding extractions \( W_i^m \). Due to channel noise \( Z \), the dependency between the embedded watermark and the extracted signal at the detector reduces. Therefore, the correlation \( \rho_{m,m} \), between \( W_m \) and its extracted version from \( Y \), would be less than \( \rho_{m,m} \). 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., \( \arg\max (\rho_{m,m}) = \arg\max (\rho_{m,m}) \).

Let the maximum of \( \rho_{m,m} \) be denoted by \( \rho_{\text{max}} \) with the pdf given as

\[ \rho_{\text{max}} \sim \max (\rho_{m,m}, \ldots, \rho_{m,m}^L) \]  

(12)

where \( \rho_{m,m} \) are independent random variables with \( \rho_{m,m} \sim \rho_{\text{dep}} \) (Section D of Appendix). With multiple codebook data hiding, detection errors are due to any of the normalized correlation values \( \rho_{m,j}, J \neq m \), being greater than the correlation value \( \rho_{\text{max}} \).

Assuming \( T_k \) is the transformation basis used for embedding in all cases, an event \( E_j \) that the detector will pick \( \hat{m} \) instead of \( m \) is denoted as

\[ E_j = \{ \rho_{m,j} \geq \rho_{\text{max}} \}, \quad 1 \leq i \leq L, \; 1 \leq j \leq M, \; j \neq m. \]  

(13)

The event \( E_{\text{mul}} \) that the detector makes an error is

\[ E_{\text{mul}} = \bigcup_{i=1}^L \bigcup_{j=1, j \neq m}^M E_j. \]  

Hence, the union bound on probability of detecting a wrong message for multiple codebook hiding, \( P_{\text{mul}} \), is obtained as

\[ P_{\text{mul}} = \Pr(E_{\text{mul}}) \leq \sum_{i=1}^L \sum_{j=1, j \neq m}^M \Pr(\rho_{m,j} \geq \rho_{\text{max}}). \]  

(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 \( \rho_{\text{max}} \). The distribution of \( \rho_{m,j} \) can be generalized as

\[ \rho_{m,j} \sim \begin{cases} N(0, \frac{1}{N}) & \text{if } i \neq k, \text{ or } i = k \text{ and } j \neq m, \\ \rho_{\text{dep}} & \text{if } i = k \text{ and } j = m. \end{cases} \]  

(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

\[ P_{\text{mul}} \leq \sum_{i=1}^L \sum_{j=1, j \neq m}^M \int_{-\infty}^{\infty} \int_{-\infty}^{\infty} f_{\rho_i}(\rho_i) f_{\rho_{\text{max}}}(\rho) d\rho_i d\rho_{\text{max}}. \]  

(16)

where the first subscript referring to the transmitted message \( m \) is dropped. Since the inner integral in
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= $10 \log_{10}(\frac{Z}{N})$. 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 $\tilde{\rho}_{m,m}^1, \cdots, \tilde{\rho}_{m,m}^M$ 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 i.i.d 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 $P_e^{\text{mul}}$ by the factor of $L$, compared to $P_e^{\text{single}}$, 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 $\rho_{dep}$ to $\rho_{max}$. 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 $\rho_{max}$ rather than of $\rho_{dep}$.

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 \( \mathbb{R}^N \) 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 \( \Delta \) are derived for the considered WNRs. The parameter \( \beta \), 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 \( 64 \times 32, 128 \times 64, 256 \times 128 \) 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., \( \frac{\pi}{2L} \). That is, the distance between the codewords of consecutive codebooks depends on \( \frac{\pi}{2L} \) (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 setting, the cover-signal is a \( 256 \times 256 \) 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 \( \Delta \) used for embedding and detection is selected to maximize robustness against lossy JPEG compression at the given quality factor (QF) and the parameter \( \beta \) 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-
embedding 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 \( 8 \times 8 \) 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 \( N \).

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 provides 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 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 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 \( \rho_{\text{ind}} \)

If \( W_m \) and \( \hat{W}_m \) have a zero covariance matrix, because of distortion by channel noise, the normalized correlation \( \rho_{\text{ind}} \) between \( W_m \) and \( \hat{W}_m \) is defined as

\[
\rho_{\text{ind}} = \frac{W_m^T \hat{W}_m}{||W_m|| \cdot ||\hat{W}_m||} = \sum_{l=1}^{l=N} \frac{W_{ml} \hat{W}_{ml}}{||W_m|| \cdot ||\hat{W}_m||}. \tag{A.1}
\]

Due to iid assumption, the normalized rv’s \( \frac{W_{ml}}{||W_m||} \) and \( \frac{\hat{W}_{ml}}{||\hat{W}_m||} \) are both zero mean with variance \( \frac{1}{N} \). Consequently, through central limit theorem, the normalized correlation \( \rho_{\text{ind}} \) approximates Gaussian distribution with mean \( m_{\rho_{\text{ind}}} = 0 \) and variance \( \sigma_{\rho_{\text{ind}}}^2 = \frac{1}{N} \), \( \rho_{\text{ind}} \sim \mathcal{N}(0, \frac{1}{N}) \). Similarly, when \( \hat{W}_m \) and \( W_j \) are independent iid random vectors, the normalized correlation \( \rho_{m,j} \sim \rho_{\text{ind}} \).
Appendix B. Distribution of \( \rho_{dep} \)

When \( W_m \) and \( \hat{W}_m \) are dependent, the samples \( W_{m1} \) and \( \hat{W}_{m1} \), \( 1 \leq l \leq N \), are somewhat correlated, and the normalized correlation \( \rho_{dep} \), defined between \( W_m \) and \( \hat{W}_m \), is as given in (A.1).

Since \( Z_f \) is the noise that distorts the embedded \( W_m \), the signal \( W_m \), extracted from \( Y \), can be expressed in terms of \( Z_f \) and \( \hat{W}_m \). Hence, for a binary distributed watermark sample \( W \), of the iid vector \( W_m \), 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

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

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

Due to memoryless embedding/detection and attack schemes, generation of the vector \( W_m \) from \( W \) is straightforward. The pdf of \( Z_f \) can be found in terms of the pdf’s of \( Z_f \) and \( X \). [11] Ultimately, \( \rho_{dep} \) 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 is due to the embedding distortion \( P \). When \( N \) is not large enough, \( P \) deviates from \( P_E = \sigma_X^2 \). 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 \( W_m \) and \( \hat{W}_m \) is actually a rv conditioned on \( P \), \( \rho_{dep|P} \), with the mean \( m_{\rho_{\ast}} \), and variance \( \sigma_{\rho_{\ast}}^2 \) as given in Section C. Since covariance matrix of \( W_m \) and \( \hat{W}_m \) is diagonal, distribution of \( \rho_{dep|P} \) approximates Gaussian distribution, \( \rho_{dep|P} \sim N(m_{\rho_{\ast}}, \sigma_{\rho_{\ast}}^2) \), and pdf of \( \rho_{dep} \) can be obtained as

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

where \( P \sim N(\sigma^2_X, \sigma^2_P) \).

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

The mean \( m_{\rho_{\ast}} \) of \( \rho_{dep|P} \) 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 \( f(w) = \frac{1}{2} \delta(w - \frac{\Delta}{4}) + \frac{1}{2} \delta(w + \frac{\Delta}{4}) \). Hence, the \( pq \)-th joint moment of \( W \) and \( \hat{W} \) is defined as

\[
E[W^p \hat{W}^q] = \int_{-\infty}^{\infty} \int_{-\infty}^{\infty} w^p \hat{w}^q f_W(w) f_{\hat{W}}(\hat{w}) dwd\hat{w}.
\] (C.1)

The joint pdf in the above integral can be expressed in terms of marginal and conditional pdf’s, \( f_W(w, \hat{w}) = f_W(w) f_{\hat{W}}(\hat{w}) \). 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, \( E[W] = \int_{-\infty}^{\infty} \hat{w}(z_f) f_{Z_f}(z_f) dz_f \), and since all pdf’s are assumed to be symmetric, (C.1) can be written as

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

where \( R(q) \) is defined as

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

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

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

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

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

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

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

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

\[
E[\hat{W}^p] = \left\{\begin{array}{ll}
0, & \text{if } p \text{ is odd} \\
R(p), & \text{if } p \text{ is even}.
\end{array}\right.
\] (C.7)

Based on (C.1)-(C.7), \( m_{\rho_{\ast}} \) is derived as

\[
m_{\rho_{\ast}} = \frac{E[\hat{W}^2] - E[\hat{W}^2]^2}{E[\hat{W}^2]} = \frac{\Delta R(1)}{\sqrt{\Delta^2 R(2)}} = \frac{R(1)}{\sqrt{R(2)}}.
\] (C.8)

The variance \( \sigma^2_{\rho_{\ast}} \) is the variation of the correlation coefficient around its mean \( m_{\rho_{\ast}} \). 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
\( \sigma_{r^*} \) is not straightforward. Therefore, Monte-Carlo simulations are performed to obtain the \( \sigma_{r^*} \) values for the considered \( N \) by computing the correlations between the embedded \( W_m \) and extracted \( \hat{W}_m \) at the assumed \( WNR \) and, then, by measuring the deviation from \( m_{r^*} \).

Appendix D. Distribution of \( \rho_{\text{max}} \) and \( \rho_{m,j}^{\text{ind}} \)

The rv \( \rho_{\text{max}} \) is the maximum of \( L \) random variables, (12), that are all distributed according to pdf of \( \rho_{d_{\text{dep}}} \). Correspondingly, the pdf of \( \rho_{\text{max}} \) can be expressed in terms of the pdf of \( \rho_{d_{\text{dep}}} \) as

\[
\rho_{\text{max}} = \text{max} \left( \rho_{d_{\text{dep}}} \right) \Rightarrow \rho_{\text{max}}(x) = \int_{-\infty}^{\infty} f_X(x) \, dx.
\]

On the other hand, the pdf of the rv's \( \rho_{m,j}^{\text{ind}} \) 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 \( \rho_{m,j}^{\text{ind}} \), \( 1 \leq j \leq M \), is equivalent to random variables \( \rho_{\text{dep}} \) and \( \rho_{\text{ind}} \) 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 \( T_i \) transformation of the received signal does not provide meaningful information about \( W_m \) since embedding transformation was \( T_k \). Consequently, the binary distributed \( W_m \) with values in \( \{-\frac{\Delta}{2}, \frac{\Delta}{2}\} \) is extracted, \( \hat{W}_m^{\text{ind}} \), as a uniformly distributed sample sequence in the range \( [-\frac{\Delta}{2}, \frac{\Delta}{2}] \) which is independent from \( W_m \). Therefore, the normalized correlation \( \rho_{m,j}^{\text{ind}} \), \( i \neq k \) and \( \forall j \), has the same statistics as the \( \rho \) in its statistics respectively for \( j = m \) and \( j \neq m \).

References


