A Robust Optimization Solution to the Data Hiding Problem using Distributed Source Coding Principles

Jim Chou, Sandeep Pradhan, Laurent El Ghaoui and Kannan Ramchandran

269 Cory Hall
University of California - Berkeley,
Berkeley, CA 94708
E-mail: jimchou, pradhan5, elghaoui,kannanr@eecs.berkeley.edu

ABSTRACT

Inspired by a recently proposed constructive framework for the distributed source coding problem,[1] we propose a powerful constructive approach to the watermarking problem, emphasizing the dual roles of "source codes" and "channel codes." In our framework, we explore various source and channel codes to achieve watermarks that are robust to attackers in terms of maximizing the distortion between the corrupted coded-source signal and the original signal while holding the distortion between the coded-source signal and the original signal constant. We solve the resulting combinatorial optimization problem using an original technique based on robust optimization and convex programming.

Keywords: Data Hiding, Digital Watermarking, Multimedia, Convex Optimization, Robustness

1. INTRODUCTION

Digital watermarking (data hiding) is an emerging research area that has received a considerable amount of attention in recent years. The basic idea behind digital watermarking is to embed information into a signal such that the amount of distortion incurred on the signal is minimal. The motivation behind embedding information into a signal is that if the embedded information can be reliably recovered, then this information can specify the affiliation between the signal and its original owner; thus the information must be embedded in a manner that will preclude others from destroying it easily. Methods for embedding watermarks are wide and varied; popular methods range from modulating the information onto the least significant bits[2] to using the information as a key for indexing pseudo-random noise sequences which are additively combined with the signal.[3] Irrespective of the method, the main goal that each method has in common is to embed the maximum amount of information possible given a fixed distortion constraint between the signal and the watermarked signal, while allowing for reliable recovery of the embedded information subject to a fixed-distortion attack. Recent research has focused on the achievable capacity of various watermarking systems.[4–6] In particular, Chen et. al.[4] made the connection between the capacity of the watermarking problem and the capacity of the problems considered in other works.[7–9] In doing so, a practical system was devised by Chen et. al.[4] that achieves a higher capacity than that of systems based on low-bits-modulation and spread spectrum techniques.

In this paper, we consider the problem of source coding with side information at the decoder and relate it to the "dual" problem of channel coding with side information at the encoder. The key insight is that data hiding can be posed as a problem of channel coding with side information at the encoder and is therefore amenable to solutions which are in many senses the dual of the solution to the distributed source coding problem. We thus construct a solution to the watermarking problem based on distributed source coding principles and the insightful theoretical results presented by Costa et. al and Pinsker et al and Moulin et al.[8,7,6] Furthermore, we apply optimization techniques to approximate the constraints that are necessary for our solution to be optimal in a rate-distortion sense. We will provide experimental results based on gaussian sources and real images with additive white gaussian noise (AWGN) attacks to demonstrate the efficacy of this approach.

2. DUALITY OF DISCUS AND WATERMARKING

In this section we explore the duality between the distributed source coding problem and the watermarking problem.

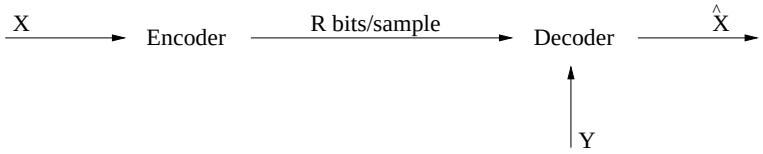

Figure 1. Distributed source coding: only decoder has access to the side information Y.

2.1. Distributed source coding

Source coding with side information is shown schematically in Fig. 1. In this problem, the source X is to be encoded and transmitted to a receiver which has access to some side information Y. Even though the encoder does not have access to Y, using joint statistics of X and Y, the encoder can perform as well as the case when both the encoder and decoder have access to Y. This is known as the Slepian-Wolf theorem in information theory.[10] Let us consider the following illustrative example.

Example 1: Discrete Case: Consider X and Y to be equiprobable 3-bit data sets which are correlated in the following way: $d_H(X,Y) \leq 1$, where $d_H(.,.)$ denotes Hamming distance. When Y is known both at the encoder and decoder, we can compress X to 2 bits, conveying the information about the uncertainty of X given Y (i.e., the modulo-two sum of X and Y given by: (000),(001),(010) and (100)). Now if Y is known only at the decoder, we can surprisingly still compress X to 2 bits. The method of construction stems from the following argument: if the decoder knows that X=000 or X=111, then it is wasteful to spend any bits to differentiate between the two. In fact, we can group X=000 and X=111 into one coset (it is exactly the principal coset of the length-3 repetition code). In a similar fashion, we can partition the remaining space of 3-bit binary codewords into 3 more different cosets with each coset containing the original codewords offset by a unique and "correctable" error pattern. Since there are 4 cosets, we need to spend only 2 bits to specify the coset in which X belongs. The four cosets are given as

$$\begin{aligned} \text{coset-1} &= (000, 111), \quad \text{coset-2} = (001, 110), \\ \text{coset-3} &= (010, 101), \quad \text{coset-4} = (011, 110) \end{aligned}$$

The decoder can recover X perfectly by decoding Y to the closest codeword (in Hamming distance) to the entries in the coset specified by the encoder. Thus the encoder does not need to know Y for optimum encoding.

The above concepts can be generalized under certain conditions to lossy compression to include the encoding/decoding of continuous-valued random variables, where the decoder uses Y to reconstruct the source based on a fidelity criterion. The minimum rate of encoding[11] for a given fidelity criterion D, is

$$R(D) = min_{\hat{X}=f(U,Y),p(U|X)}[I(U;X) - I(U;Y)] \tag{1}$$

where U is the set of codewords representing X, $I(U;X)$ is the Shannon mutual information,[12] and the minimization is carried out over all conditional probability density functions $p(U|X)$ and a function $f(U;Y)$ such that $E(X-\hat{X})^2 \leq D$. The main idea of encoding is as follows: (1) build a source code to represent X using about $2^{nI(U;X)}$ codewords, (2) partition this set into $2^{nR(D)}$ cosets with each coset containing $2^{nI(U;Y)}$ codewords (the set of codewords acts as a channel code for the fictitious channel between U and Y), and (3) find the optimal reconstruction, which is obtained as a function $f(U,Y)$.

To elucidate the encoding construction, consider the case where X is an independent identically distributed (*i.i.d.*) gaussian random variable with side information Y given by $Y = X + N$ (N is an *i.i.d.* Gaussian random variable independent of X). The side information is therefore a noisy version of X and as in the discrete case, it has been shown[11] that an encoder can be designed to represent X matching the compression efficiency of the case where the encoder also has access to Y. The practical method proposed by Pradhan et. al.[1] is to design a source code and partition it into a bank of cosets of channel codes. The source code is designed for the optimal representation of X, and is partitioned into cosets which have good distance properties. The source X is quantized to a codeword (referred to as the active codeword) and the index of the coset containing this codeword is sent to the decoder. With

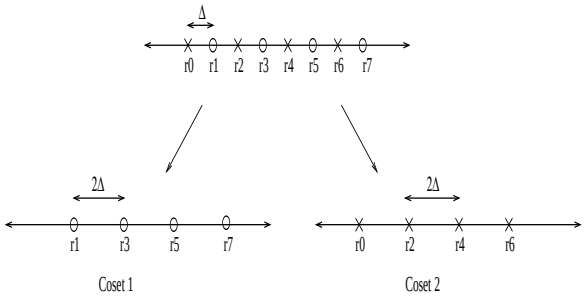

Figure 2. An 8-level quantizer partitioned into 2 cosets containing 4 levels each.

the help of Y, the decoder finds the active codeword in the coset whose index is given by the encoder. A practical example is as follows:

Example 2: Continuous Case: Consider an 8-level fixed-length scalar quantizer as the source coder (see Fig. 2), with the rate of transmission fixed at 1 bit/source sample. To partition the source code, we partition the reconstruction levels into two cosets. The reconstruction levels in coset 1 are denoted as (r_0, r_2, r_4, r_6) and the reconstruction levels in coset 2 are (r_1, r_3, r_5, r_7) (see Fig. 2). The partition is constructed in such a way that the cosets are symmetric and the minimum distance between any two elements within a coset is kept as large as possible. Upon encoding, X is quantized to a codeword in the composite 8-level quantizer and the index of the coset containing this codeword is sent to the decoder. The decoder finds this codeword in the coset whose index is sent by the encoder, as the one which is closest (in the appropriate distance measure) to Y. In doing so, the decoder can reconstruct the value of X to a codeword which is "close" in terms of squared-error distortion.

In the above example, a simple 8-level scalar quantizer was considered for the sake of simplicity. Encoding constructions, however, can be generalized to more sophisticated source coders. For example, the partition can be done in higher-dimensional spaces using trellis codes. This was proposed by Pradhan et. al.[1] where the n-dimensional product space of scalar quantized reconstruction levels is partitioned into cosets of trellis coded quantizer codebooks.

2.2. Watermarking problem

The digital watermarking problem can be formulated as follows. The encoder has access to two signals; the information (an index set), M, to be embedded, and the signal that the information is to be embedded in. The output of the encoder is the watermarked signal. The attacker will attempt to degrade the signal so that the decoder (who wants to authenticate the watermarked signal) will fail to decode the watermark. The attacker has a distortion constraint on the amount of degradation that he can inflict on the signal. The encoder has to be designed such that the attacker needs to inflict an amount of distortion (to destroy the watermark) that will render the signal to be useless. Mathematically, the goal is to solve the following constrained minimization problem:

$$min_{||W-S||^2 \leq D_1, ||Y-W||^2 \leq D_2} P_e(\hat{M}) \tag{2}$$

where $P_e(\hat{M})$ represents the probability of decoding error.

It was shown[4,6] that this problem can be viewed as channel coding with side information (about the channel) at the encoder (see fig. 3). We will consider the case when the side information is jointly gaussian with the noise source. It is obvious that this problem becomes the regular AWGN communications problem when both the encoder and decoder have access to the side information. Surprisingly, in the case when only the encoder has access to the side information, it has been shown that this system can *perform as well as when both the encoder and decoder have access to the side information*[8]. In fact, this is the key insight which motivates our analysis for the digital watermarking system through exploitation of its dual nature with the distributed source coding problem. Before formulating the details for constructing an encoder/decoder for the watermarking problem, we first look at an analogous problem that will illustrate key concepts.

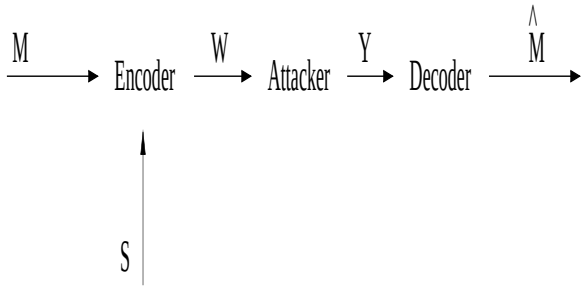

Figure 3. The watermarking problem: the encoder has access to a signal S. M represents the information to be embedded and W represents the watermarked signal.

The problem of writing on defective memories[9] is very similar to the watermarking problem. To illustrate this problem, consider the case where the encoder attempts to write data onto faulty registers with one of the registers always being "stuck-at" 0 or 1. The encoder, however, has full knowledge of which register is always "stuck". If the decoder also has this knowledge, then $n-1$ bits can be stored reliably in n registers with one "stuck-at" fault. It was shown[9] that the encoder can be designed to perform as well as the case when both the encoder and the decoder have access to the "stuck-at" position. We illustrate the method for constructing such an encoder in the following example*.

Example3: Faulty Registers: Consider a 3-bit memory device, which has one "stuck-at" in any position with a uniform probability of the "stuck-at" being either a one or a zero. When both the encoder and the decoder know the value and the position of the "stuck-at", the encoder can write 2 bits reliably. Now consider the case when only the encoder has access to the "stuck-at". Let A denote the set of binary three tuples: $\{0,1\}^3$. We partition A into cosets of codewords which are compatible with any type of "stuck-at". This partition is the same as the one considered in Example 1. If the encoder wants to write the first message, he chooses a codeword from the first coset which is compatible with the "stuck-at". If the first bit has a "stuck-at" fault of 1, then he chooses [1 1 1] as the codeword to be written on the memory. In this fashion, we can again reliably write two bits (corresponding to the index of the coset), even though the decoder does not have access to the "stuck-at".

The above problem can be generalized into the problem of channel coding with side information, for which the capacity can be calculated. The capacity[9,7,8] of such systems is given by

$$C = max_{p(U,S|X)}[I(U;Y) - I(U;S)] \tag{3}$$

where the maximization is over all conditional probability density functions $p(U,X|S)$. The signal S is the side information about the channel and U represents the codeword space. The main steps for encoding should be as follows: (1) build a channel code over the space of U, with the number of codewords being nearly equal to $2^{nI(U;Y)}$ (where n is the block length of encoding), (2) partition this channel codeword space into cosets of source codes with each coset containing nearly $2^{nI(U;S)}$ codewords, and (3) choose a codeword, U, to represent S from the coset which has an index equal to the message. The size of the index set should be nearly equal to 2^{nC}, where C is given in (3). The signal, X, which is transmitted over the channel is a function $f(U,S)$.

Returning to the watermarking problem, we consider a related problem for which previous work has already been done. In particular, we consider the case where the channel is AWGN (see Fig. 3) and the signal (side information, S) is *i.i.d.* gaussian. In this case, it was shown[8] that the capacity (3) is given as

$$C = \frac{1}{2} log\left\{\frac{P}{N} + 1\right\} \tag{4}$$

where P and N represent the transmitter power constraint and the variance of the channel noise respectively. To achieve capacity,[8] showed that the space of codewords must be of the form $U = X + \alpha S$, where $\alpha = \frac{P}{P+N}$ and N is

*We use the term "stuck-at" loosely to refer to a faulty memory register which is constrained to always be a 1 or a 0.

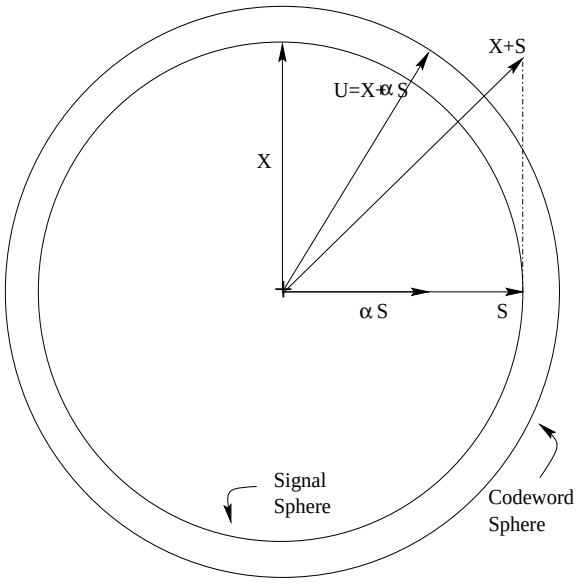

Figure 4. Geometric interpretation of the optimal encoding: U and S denote the codeword chosen for encoding and the signal to be watermarked respectively. The watermark added onto the signal S is X.

the variance of the attacker's noise. The codeword U should be chosen in a way that will ensure that the projection of U along the direction orthogonal to S has power nP, where n is the signal-dimension.

The watermarking problem can be posed as a version of the above problem, where the signal (i.e., image, audio, etc.) is given by S, the attack is AWGN[†] (with a distortion constraint of N) and the distortion constraint between the signal and the watermarked signal is given as P. To solve this problem requires searching through the set of all codewords in U and finding the one which has the "correct" projection onto the signal S and the "correct" projection onto the direction orthogonal to S. By "correct" we mean that given some probability distribution on X and S, we should be able to determine the expected magnitude of the projection onto each component. For example, let us consider the case when S (zero mean, variance Q) and X (zero mean, variance P) are *i.i.d.* Gaussian random variables, and both X and S are n-dimensional vectors. In this case, when n is large, S will be distributed on a hyper-sphere with radius $\sqrt{nQ}$ and U will be distributed on a hyper-sphere with radius $\sqrt{P+\alpha^2 Q}$ (see Fig. 4). The codeword U which is optimal in the sense of Costa's work[8] should then have a projection onto the signal S which has power equal to αnQ and a projection onto the direction orthogonal to S with power equal to nP where n represents the dimension of the signal S and the watermark X. A naive approach to finding such a codeword would be to perform an exhaustive search over the codebook specified by U. Depending upon the codebook, this can be an insurmountable task. Instead, we propose the following optimization problem to find the proper codeword:

$$\begin{gathered} min\ (S^T U - \alpha nQ)^2 \ \ subject\ \ to \\ (U-\alpha S)^T (U-\alpha S) = nP \\ U \in L \end{gathered} \tag{5}$$

where $L = [-N\Delta, ..., N\Delta]^n$ is a lattice in R^n and $\Delta > 0$. Notice here that we choose as our codebook a one-dimensional lattice which can also be thought of as a uniform scalar quantizer. The equality constraint in (5) is often impossible to satisfy, but we can relax it into an inequality constraint and pose an approximation to the above problem:

$$min(S^T U - \alpha nQ)^2 \ \ subject\ to$$

[†]It has been shown[6] that AWGN is the optimal attack when the signal S is gaussian.

$$
\begin{aligned}
U^T U &\le n(P + \alpha^2 Q) \\
(U - \alpha S)^T (U - \alpha S) &\le nP \\
(U - \alpha S)^T S &\le \epsilon \\
U &\in L
\end{aligned} \tag{6}
$$

The new problem ensures that X is approximately orthogonal to S, that both X and U satisfy the approximate power constraints and that the projection of U onto S is as close to its expected magnitude as possible.

3. SOLVING THE OPTIMIZATION PROBLEM

In this section, we address the problem of solving the optimization program (7). The problem involves a quadratic convex objective and constraints, except for the lattice constraint $U \in L$. The latter constraint makes the problem combinatorial (NP-hard), which means it is impossible to solve it using exhaustive search.

To address this hard problem, we use a relaxation technique for integer programming that is based on robust optimization, first introduced in.[13] To illustrate the method, consider a quadratic constraint

$$f(U) \le 0, \;\; U \in L$$

where f is some scalar-valued function of vector U, and L is a lattice. Let $\rho \ge 0$ be the smallest number $\rho \ge 0$ such that for every U, the ball $\mathcal{B}_\rho(U) := \{\xi \mid \|\xi - U\|_\infty \le \rho\}$ contains the vector in the lattice nearest to U (with distance measured by the maximum norm $\|\cdot\|_\infty$). (In our case, an appropriate choice for ρ is $\rho = \Delta/2$.)

Suppose we find a *real* vector U such that the robustness constraint

$$f(U + \delta U) \le 0 \text{ for every } \delta U, \;\; \|\delta U\|_\infty \le \rho \tag{7}$$

is enforced. Let us now round U to a nearest vector in the lattice $U_{\text{lat}} \in L$, "nearest" being taken in the sense of the maximum norm $\|\cdot\|_\infty$. The lattice vector U_{lat} can be written as $U_{\text{lat}} = U + \delta$ for some δ, $\|\delta\| \le \rho$. From (7), we conclude that

$$f(U_{\text{lat}}) \le 0, \;\; U_{\text{lat}} \in L.$$

Thus, by enforcing a robustness constraint, we are computing a real vector such that its nearest lattice approximation is guaranteed to be feasible. The above method can be called a robustness relaxation to lattice programming. We note that the method hinges on our ability to solve the robustness problem (7).

Since the function f is quadratic, we can directly use the robust optimization technique introduced in,[13] to find a sufficient condition for the robust quadratic constraint (7) to hold, in the form of a tractable, convex problem, with linear objective and positive semidefiniteness constraints. This type of problem, called semidefinite program (SDP), can be very efficiently solved in polynomial-time using interior-point techniques.[14]

Applying this technique to our problem results in the following SDP:

$$\textit{min } t \quad \textit{subject} \quad \textit{to}$$

$$\begin{bmatrix} \mathbf{C}_\alpha - \mathbf{I} & -U & \mathbf{0} \\ -U^T & \eta^2 - \rho^2 Tr(\mathbf{C}_\alpha) & U^T \\ \mathbf{0^T} & U & \mathbf{I} \end{bmatrix} \succeq 0,$$

$$\mathbf{C}_\alpha \textit{ diagonal}$$

$$\begin{bmatrix} \mathbf{C}_\beta - SS^T & (S^T U - \gamma)S & \mathbf{0} \\ (S^T U - \gamma)S^T & t - \rho^2 Tr(\mathbf{C}_\beta) & S^T U - \gamma \\ \mathbf{0^T} & S^T U - \gamma & 1 \end{bmatrix} \succeq 0,$$

$$\mathbf{C}_\beta \textit{ diagonal}$$

$$\begin{bmatrix} \mathbf{C}_\mu - \mathbf{I} & \alpha S - U & \mathbf{0} \\ (\alpha S - U)^T & nP - \rho^2 Tr(C_\mu) & (U - \alpha S)^T \\ \mathbf{0^T} & U - \alpha S & \mathbf{I} \end{bmatrix} \succeq 0,$$

$$\mathbf{C}_\mu \ \textit{diagonal}$$

$$U^T S \leq \epsilon + \alpha S^T S - \sum_{i=1}^{n} |S_i|$$

where $\eta = \sqrt{n(P+\alpha^2 Q)}$, $\gamma = \alpha n Q$, $\rho = \frac{\Delta}{2}$ (Δ is the quantizer step-size) and $\mathbf{C}_\alpha, \mathbf{C}_\beta$ are diagonal, positive semi-definite matrices determined by the parameters used for the Lagrange relaxation. The above problem is now clearly convex and can be solved using a variety of generic methods.[14] Details on the above result can be found in the Appendix. Upon solving the above SDP, it can be shown (see Appendix) that for any feasible U and t, the lattice point which is closest to U will also satisfy the constraints defined in (6).

4. SUMMARY

As a result, we have a constructive approach for encoding watermarks into a given signal (which is Gaussian). The encoding process is summarized as follows: (1) pick a channel code (i.e., a lattice L) over the space of U, (2) partition the channel code into cosets of source codewords with each codeword separated by some minimum distance ρ and (3) choose the index of the coset in which the codeword needs to be found to calculate the watermark that is to be encoded into the signal.

Upon decoding, the decoder finds a codeword in the composite channel code (containing nearly $2^{nI(U;Y)}$) which is closest to the received vector (in some sense). The coset containing the decoded codeword is declared as the decoded watermark message. It can be shown[8] that the scaling which is done at the encoder provides the required distance property to tolerate the attacker's noise at the decoder.

5. DESIGN AND CONSTRUCTION: SIMULATION RESULTS

To test the effectiveness of our above encoding and decoding approach, we have devised three practical solutions based on different channel codebooks and source partitions. In our first approach, we use a 16-level uniform lattice (with $\Delta = 0.2$) as our channel codebook and partition it into scalar-quantized cosets; we refer to this method as SQ-SQ. The next approach entails using an 8-level uniform lattice (with $\Delta = 0.4$) as our channel codebook and partition it into trellis-coded-quantized cosets; we will refer to this method as TCM-SQ. The final method is based on a TCM codebook which is partitioned into trellis-coded-quantized (TCQ) cosets; we refer to this method as TCM-TCQ. In each of the above three methods we determine the watermarked signal based on the channel codebook and source partition using the optimization techneques described in the previous section.

The first set of simulations that we conducted, assumed that the signal was gaussian and that the attack was AWGN. We fixed the probability of decoding error to be less than 10^{-5} and determined the amount of distortion that was necessary to ensure this probability of error given a fixed-distortion attack. The performance curves representing the signal-to-distortion (i.e., signal power vs. watermark power) ratio vs. signal-to-noise (i.e., signal power vs. noise power) at $P_e(\hat{M}) = 10^{-5}$ and a rate of 1 embedded-bit/sample is given in Fig. 5. It can be seen from the figure that TCM-TCQ outperforms both TCM-SQ and SQ-SQ by a wide margin. The reason for this performance discrepancy is that TCM-TCQ uses a better source codebook than either TCM-SQ and SQ-SQ.

The next simulation that we conducted is of a more practical use. We consider the case where S is the 'Lena' image and the attack is confined to JPEG compression. We again tested all three schemes. At an embedding rate of $\frac{1}{64}$ bits/sample, SQ-SQ incurred a Peak-Signal-to-Distortion-Ratio (PSDR) of 40.5dB in order to withstand a JPEG compression factor of 75%. TCM-SQ incurred a Peak-Signal-to-Distortion-Ratio (PSDR) of 41.23dB in order to withstand a JPEG compression factor of 40%. On the other hand, TCM-TCQ incurred a PSDR of 42.5dB in order to withstand a JPEG compression quality factor of 25%. Again TCM-TCQ outperforms both SQ-SQ and TCM-SQ. In Fig. 6 we show the degraded JPEG (Q=25%) image. The probability of decoding error was less than 10^{-5}.

6. CONCLUSION

In conclusion, we have shown that the watermarking problem and the distributed source coding problem can be viewed as the duals of each other, where the former is channel coding with side information at the transmitter and the latter is source coding with side information at the receiver. In the first case, the channel code is partitioned into a bank of cosets of source codes, while in the latter the source code is partitioned into a bank of cosets of channel

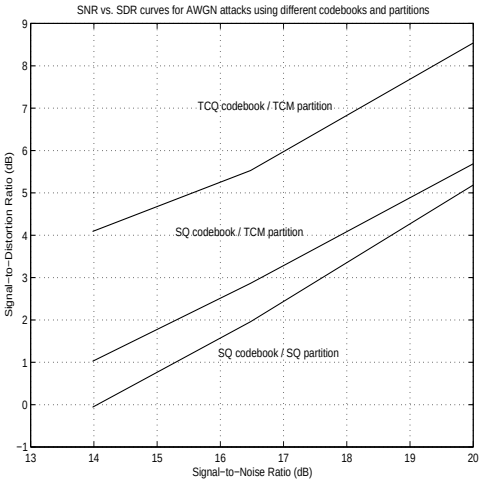

Figure 5. Performance curves of SQ-SQ TCM-SQ and TCM-TCQ given fixed distortion attacks.

Figure 6. Degraded Lena image (JPEG compressed, Q=25%). With an embedding rate of 1/64 bps, all bits were recovered successfully.

codes. The 3-bit binary data example clearly demonstrates the dual nature of the two problems. We then formulated a practical solution for the watermarking problem based on robust optimization techniques. Through simulation results, it can be seen that this method performs fairly well and is amenable to codes built for higher-dimensions.[1]

Appendix

The robust optimization approach for solving the quadratic problem (6) with lattice constraints leads to the alternative problem:

$$\begin{gathered} min_{U,t}\ t \\ subject \quad to \\ (S^T(U+\delta U)-\gamma)^2 \le t \quad \forall \delta U,\ \delta U_i^2 \le \rho^2 \quad i=1,...,n \\ ||U+\delta U||_2^2 \le \eta^2 \quad \forall \delta U,\ \delta U_i^2 \le \rho^2 \quad i=1,...,n \\ ||U+\delta U-\alpha S||_2^2 \le \eta^2 \quad \forall \delta U,\ \delta U_i^2 \le \rho^2 \quad i=1,...,n \\ (U+\delta U-\alpha S)^T S \le \epsilon \quad \forall \delta U,\ \delta U_i^2 \le \rho^2 \quad i=1,...,n \end{gathered} \tag{8}$$

where $\gamma = \alpha n Q$, $\eta = \sqrt{n(P+\alpha^2 Q)}$ and it is assumed that for any $Y \in R^n$ there exists a point in the lattice which has a distance that is no larger than $\rho = \frac{\Delta}{2}$ in each dimension from Y, where Δ represents the distance (in each dimension) between codewords in the lattice. Thus, solving (7) will give us a result that satisfies the constraints posed in (6) and the point in the lattice which is closest to this result will also satisfy the constraints of (6).

We now consider each robustness constraint in the above problem one by one, and for each find a sufficient condition that ensures its validity. Let us start with the first robustness constraint:

$$||U+\delta U||_2^2 \le \eta^2 \quad \forall \delta U,\ \delta U_i^2 \le \rho^2 \quad i=1,...,n. \tag{9}$$

Using Lagrange relaxation,[13] we find that a sufficient condition for is that there exist non-negative numbers λ_i, $i=1,...,n$, such that:

$$\forall \delta U \in R^n,$$

$$\begin{bmatrix} \delta U \\ 1 \end{bmatrix}^T \begin{bmatrix} -\mathbf{I} & -U \\ -U^T & \eta^2 - U^T U \end{bmatrix} \begin{bmatrix} \delta U \\ 1 \end{bmatrix} \ge$$

$$\sum_{i=1}^{n} \lambda_i (\rho^2 - \delta U_i^2)$$

which can be rewritten as a matrix inequality:

$$\begin{bmatrix} \mathbf{C}_\alpha - \mathbf{I} & -U \\ -U^T & \eta^2 - U^T U - \rho^2 Tr(\mathbf{C}_\alpha) \end{bmatrix} \succeq 0$$

where $\mathbf{C}_\alpha = \mathbf{diag}(\lambda_1, ..., \lambda_n)$. The above matrix inequality can be written as a "linear matrix inequality" (LMI, see[14]) using Schur complements:

$$\begin{bmatrix} \mathbf{C}_\alpha - \mathbf{I} & -U & \mathbf{0} \\ -U^T & \eta^2 - \rho^2 Tr(\mathbf{C}_\alpha) & U^T \\ \mathbf{0^T} & U & \mathbf{I} \end{bmatrix} \succeq 0,\ C_\alpha\ diagonal$$

In an identical manner, the constraint

$$||U+\delta U-\alpha S||_2^2 \le \eta^2 \quad \forall \delta U,\ \delta U_i^2 \le \rho^2 \quad i=1,...,n \tag{10}$$

is satisfied if there exists a diagonal matrix $\mathbf{C}_\mu$ such that the LMI in $U, \mathbf{C}_\mu$

$$\begin{bmatrix} \mathbf{C}_\mu - \mathbf{I} & \alpha S - U & \mathbf{0} \\ (\alpha S - U)^T & nP - \rho^2 Tr(C_\mu) & (U-\alpha S)^T \\ \mathbf{0^T} & U - \alpha S & \mathbf{I} \end{bmatrix} \succeq 0$$

holds.

Likewise, for every t, the constraint

$$(S^T(U+\delta U)-\gamma)^2 \le t \quad \forall \delta U,\ \delta U_i^2 \le \rho^2 \quad i=1,...,n \tag{11}$$

is satisfied if there exists a diagonal matrix $\mathbf{C}_\beta$ such that

$$\left[\begin{array}{cc} \mathbf{C}_\beta - SS^T & (S^TU-\gamma)S \\ (S^TU-\gamma)S^T & t-(S^TU-\gamma)^2-\rho^2 Tr(\mathbf{C}_\beta) \end{array} \right] \succeq 0.$$

Again, the above matrix inequality can be written as a LMI in $\mathbf{C}\beta$ and U through the use of Schur complements:

$$\left[\begin{array}{ccc} \mathbf{C}_\beta - SS^T & (S^TU-\gamma)S & \mathbf{0} \\ (S^TU-\gamma)S^T & t-\rho^2 Tr(\mathbf{C}_\beta) & S^TU-\gamma \\ \mathbf{0^T} & S^TU-\gamma & 1 \end{array} \right] \succeq 0.$$

Finally, the "orthogonality" constraint

$$(U+\delta U-\alpha S)^T S \le \epsilon \quad \forall \delta U,\ \delta U_i^2 \le \rho^2 \quad i=1,...,n$$

is easily seen to be equivalent to the linear inequality

$$U^TS \le \epsilon + \alpha S^TS - \sum_{i=1}^{n} |S_i|.$$

Together, the above constraints are jointly convex in $t, U, \mathbf{C}_\alpha, \mathbf{C}_\beta, \mathbf{C}_\mu$, and the problem of minimizing t under these constraints is a semidefinite program, as claimed.

REFERENCES

1. S. S. Pradhan and K. Ramchandran, "Distributed source coding using syndromes: Design and construction," *Proceedings of the Data Comression Conference (DCC)* , March 1999.
2. J. M. Barton, "Method and apparatus for embedding authentication information within digital data," *United States Patent #5,646,997* , Issued July 8 1997.
3. I. Cox, J. Killian, T. Leighton, and T. Shamoon, "Secure spread spectrum watermarking for multimedia," *IEEE Trans. on Image Processing* **6**, pp. 1673–1687, December 1997.
4. B. Chen and G. W. Wornell, "An information-theoretic approach to the design of robust digital watermarking systems," *Proc. Int. Conf. on Acoustics, Speech and Signal Processing (ICASSP)* , March 1999.
5. C. Cachin, "An information-theoretic model for steganography," *Proc 1998 Workshop on Information Hiding* , 1998.
6. P. Moulin and J. O'Sullivan, "Information-theoretic analysis of information hiding," *Preprint* .
7. S. Gel'fand and M. Pinsker, "Coding for channel with random parameters," *Problems of Control and Information Theory* **9**, pp. 19–31, 1980.
8. M. Costa, "Writing on dirty paper," *IEEE Trans. on Information Theory* **29**, pp. 439–441, May 1983.
9. C. Heegard and A. El Gamal, "On the capacity of computer memory with defects," *IEEE Trans. on Information Theory* **29**, pp. 731–739, September 1983.
10. D. Slepian and J. K. Wolf, "Noiseless encoding of correlated information sources," *IEEE Trans. on Inform. Theory* **IT-19**, pp. 471–480, July 1973.
11. A. D. Wyner and J. Ziv, "The rate-distortion function for source coding with side information at the decoder," *IEEE Trans. on Inform. Theory* **IT-22**, pp. 1–10, January 1976.
12. T. M. Cover and J. A. Thomas, *Elements of Information theory*, Wiley, New York, 1991.
13. L. El Ghaoui, F. Oustry, and H. Lebret, "Robust solutions to uncertain semidefinite programs," *SIAM J. Optimization* **9**(1), pp. 33–52, 1998.
14. S. Boyd and L. Vanderberghe, *Introduction to convex optimization with engineering applications. Lecture notes for EE364, Stanford University*, 1996.