ISIT 2000, Sorrento, Italy, June 25-30,2000

Quantization Index Modulation: A Class of Provably Good Methods for Digital Watermarking and Information Embedding

Brian Chen and Gregory W. Wornell
Research Laboratory of Electronics and Department of Electrical Engineering and Computer Science
Massachusetts Institute of Technology, Room 36-631
Cambridge, MA 02139
Email: `bchen@alum.mit.edu` and `gww@allegro.mit.edu`

Abstract — **We consider the problem of embedding one signal (*e.g.*, a digital watermark), within another "host" signal to form a third, "composite" signal. The goal is to achieve efficient rate-distortion-robustness trade-offs. We introduce a new class of embedding methods called distortion-compensated quantization index modulation. In several different contexts involving both intentional and unintentional attacks, capacity-achieving methods exist within this class, while in other contexts these methods achieve provably better rate-distortion-robustness performance than previously proposed spread-spectrum and generalized low-bit(s) modulation methods.**

I. INTRODUCTION

Digital watermarking and information embedding systems embed information in a host signal, which is typically an image, audio signal, or video signal. The host signal is not degraded unacceptably in the process, and one can recover the watermark even if the composite host and watermark signal undergo a variety of attacks as long as these corruptions do not unacceptably degrade the host signal. These systems play an important role at least three major application areas: (1) copyright protection of multimedia content, (2) authentication and tamper-detection, and (3) backwards-compatible upgrading of existing legacy communication networks [1].

II. PROBLEM MODEL

We wish to embed a message $m \in \{1, 2, \ldots, 2^{NR_m}\}$, sometimes called a digital watermark, in some host signal vector $\mathbf{x} \in \Re^N$, where R_m is the embedding *rate* in bits per host signal sample. Specifically, m and $\mathbf{x}$ are mapped onto a composite signal vector $\mathbf{s} \in \Re^N$ using some embedding function $\mathbf{s}(\mathbf{x}, m)$, and we define a *distortion* measure between $\mathbf{x}$ and $\mathbf{s}$. Equivalently, we can define a host-dependent distortion signal $\mathbf{e}(\mathbf{x}, m)$ that is added to $\mathbf{x}$ to obtain $\mathbf{s}$. The composite signal $\mathbf{s}$ is subjected to unintentional attacks and possibly to intentional attacks inside some channel, which produces an output vector $\mathbf{y} \in \Re^N$. A decoder generates an estimate $\hat{m}$ of m after observing $\mathbf{y}$, *i.e.*, we consider the "host-blind" case, where $\mathbf{x}$ is not available to the decoder. Ideally, the decoder can reliably recover the embedded information as long as the channel degradations are not too severe. Thus, the tolerable severity of the degradations is a measure of the *robustness* of the system. The goodness of $\mathbf{s}(\mathbf{x}, m)$ and its corresponding decoder is measured by the achievable rate-distortion-robustness trade-offs.

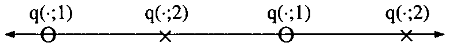

Fig. 1: Quantization index modulation information embedding.

III. DISTORTION-COMPENSATED QUANTIZATION INDEX MODULATION

Quantization index modulation (QIM) embedding functions arise by defining an ensemble of quantizers $\mathbf{q}(\cdot; m)$, one quantizer in the ensemble for each possible value of m. Then, $\mathbf{s}(\mathbf{x}, m) = \mathbf{q}(\mathbf{x}; m)$. An example is shown in Fig. 1 for the case where $N = 1$, $R_m = 1$, and the quantizers are uniform, scalar quantizers. One can decode, for example, by determining whether $\mathbf{y}$ is closer to a $\circ$ point ($\hat{m} = 1$) or to a $\times$ point ($\hat{m} = 2$). Thus, the $\times$ and $\circ$ points represent both source codewords for representing $\mathbf{x}$ and channel codewords for communicating m. QIM systems reject interference from the host signal since $\mathbf{x}$ determines which $\circ$ or $\times$ point is chosen but does not deflect $\mathbf{s}$ or $\mathbf{y}$ away from these points. Distortion-compensated QIM (DC-QIM) systems add back some fraction $1 - \alpha$ of the quantization error, $\mathbf{s}(\mathbf{x}, m) = \mathbf{q}(\mathbf{x}; m) + (1 - \alpha)[\mathbf{x} - \mathbf{q}(\mathbf{x}; m)]$, which can be shown [1] to improve rate-distortion-robustness performance with the proper choice of α.

IV. PERFORMANCE AGAINST ATTACKS

In fact, one can derive sufficient conditions under which capacity-achieving DC-QIM systems exist [1]. These conditions are satisfied in at least three cases: (1) the additive Gaussian noise channel and Gaussian host signal scenario of [2], (2) the case of squared error distortion-constrained attacks and a Gaussian host signal described in [3], and (3) the case of squared error distortion-constrained attacks, a non-Gaussian host signal, asymptotically small embedding-induced distortion, and asymptotically small attacker's distortion described in [3].

In a number of other contexts where the capacity is unknown, DC-QIM methods achieve provably better performance than previously proposed additive spread-spectrum methods, which do not reject interference from the host signal, and generalized low-bit(s) modulation methods. These cases are discussed in [1], along with practical implementations of DC-QIM and QIM systems.

REFERENCES

[1] B. Chen, *Design and Analysis of Digital Watermarking, Information Embedding, and Data Hiding Systems.* PhD thesis, MIT, Cambridge, MA, June 2000.

[2] M. H. M. Costa, "Writing on dirty paper," *IEEE Trans. on Info. Thy.*, vol. 29, pp. 439–441, May 1983.

[3] P. Moulin and J. A. O'Sullivan, "Information-theoretic analysis of information hiding," Preprint, Oct. 1999.

This work has been supported in part by the Office of Naval Research under Grant No. N00014-96-1-0930, by the Air Force Office of Scientific Research under Grant No. F49620-96-1-0072, by the MIT Lincoln Laboratory Advanced Concepts Committee, and by a National Defense Science and Engineering Graduate Fellowship.

0-7803-5857-0/00/$10.00 ©2000 IEEE.