MULTIPLE WATERMARKING: IS POWER SHARING BETTER THAN TIME SHARING?

Yi-Wen Liu and Julius O. Smith

Center for Computer Research in Music and Acoustics, Stanford, CA 94305, USA

ABSTRACT

Writing on dirty paper, a mathematical model for watermarking, is examined and extended to allow simultaneous watermarking at multiple rates. Based on the extended model, it is argued that the optimal watermarking strategy is to share the power. The argument is further supported by simulations of joint spread spectrum and quantization watermarking.

1. INTRODUCTION AND DEFINITIONS OF TERMINOLOGY

Watermarking is defined here as embedding information-carrying signals in a host signal. The information-carrying signals, called the *watermarks*, should be embedded without introducing perceptible distortion to the host signal. This is often called the *transparency* requirement. The amount of information a watermark carries is called the *payload*. Originally, watermarking was invented for multimedia copyright protection [1]. To serve as an evidence of copyright, a transparent watermark can carry a payload of one bit, which is expected to be detectable with very low probability of false alarm. Alternatively, the watermark can be modulated by a payload of multiple bits that represents a copyright message, and the message is expected to be decodable without ambiguity. A rate of payload that can be successfully decoded is called an *achievable data hiding rate*, often in the unit of b/sec (audio) or b/pixel (image). The supremum of all achievable rates is called the *data hiding capacity*.

In addition to transparency, a watermark is typically required to be robust and secure. Here, *robustness* refers to a watermark's ability to be detected or decoded after some "standard" signal processing procedures, with or without intentions to remove the watermark. In this paper, such procedures are called *blind attacks*. On the contrary, *Security* refers to a watermark's ability to survive attacks from an adversary processing the watermarked signal with an intention to remove the watermark or, if the removal is impossible, to fail the detection of it. Such attacks are often called *malicious attacks*. This paper deals with blind attacks and models them as additive noise.

Beside copyright protection, newer watermark applications have emerged recently, such as copy control, transaction tracking, broadcast monitoring, and tamper detection (see e.g. [1] for a general discussion). Each application has its own robustness requirements, and each aims at a different data hiding rate. Therefore, conflicts arise when many watermarks are embedded in a common host for respectively different purposes. Qualitative rules to avoid the conflicts are first proposed in [2]. It is argued that watermarks ought to be embedded in the order of decreasing robustness. Otherwise, detection/decoding of more fragile watermarks are likely to fail due to the interference from more robust watermarks. Unlike [2], in [3], two watermarks are embedded in a *time-sharing* manner, which means that information-carrying coefficients are randomly divided into two groups, and each group is modulated by either one watermark or the other. Interestingly, due to the mutual exclusiveness of [1]time sharing, it does not matter which of the two watermarks is embedded first.

Inspired by the two complementary approaches mentioned above, we intend to provide a unified mathematical framework for *2-watermarking*, the [2]simultaneous embedding of two watermarks. We shall model it as broadcast communication with side information known at the encoder, and derive performance bounds in terms of data hiding rates.

This paper is organized as the following. Sec. 2 reviews an existing information-theoretic model of watermarking as communication with side information. In Sec. 3, we examine the underlying assumptions of the model, and describe ways to bridge the gap between the theory and the practice. In Sec. 4, a broadcast communication model for 2-watermarking is presented. Regions of achievable rates are plotted and implications are discussed. A power-sharing 2-watermarking system is presented and briefly evaluated in Sec. 5. Conclusions are stated in Sec. 6.

2. REVIEW OF A THEORETIC MODEL

Watermarking has been treated as communication using a watermark of limited power subject to a dominating interference from the host signal. As shown in Fig. 1, a binary message $m \in \mathcal{M}, |\mathcal{M}| = 2^{NR}$ modulates a watermark Z_0. Here, N is the encoding block length, and R is an attempted data hiding rate in bits/sample. Although R is not necessarily an integer, we will also use the more sloppy notation $m \in \{0,1\}^{NR}$. The watermark Z_0 is added to a host signal S. The host is assumed to be independent identically distributed (IID) Gaussian $S \sim \mathcal{N}(0, \sigma^2)$. Let $X = S + Z_0$ be a watermarked version of S, which is subject to a blind attack Z_1, modeled as an additive noise to X. Let $Y = X + Z_1$ be the blindly attacked version of X. A message $\hat{m} \in \{0,1\}^{NR}$ is decoded from Y, and the decoding is successful if $\hat{m} = m$.

Because of the transparency requirement, we must assume $\mathbf{E}[Z_0^2] \ll \sigma^2$. From Z_0's perspectives, the interference from S is very strong, which may seem to severely limit the data hiding rate R. However, it is now clear that the interference is a side information known to the watermark encoder [4], which can be utilized to achieve higher data hiding rates than otherwise ignoring it. Particularly, if human perception imposes power constraints $\mathbf{E}[Z_0^2] \leq D_0$ and $\mathbf{E}[Z_1^2] \leq D_1$ on the watermark and the attack respectively, the data hiding capacity has a close-form expression

[1]Here, we hesitated to use the term "time"-sharing, because coeffi cients can be actually in non-temporal domains such as frequency or space. However, we will keep using this term for its conciseness.

[2]Again, simultaneity is not necessarily in time.

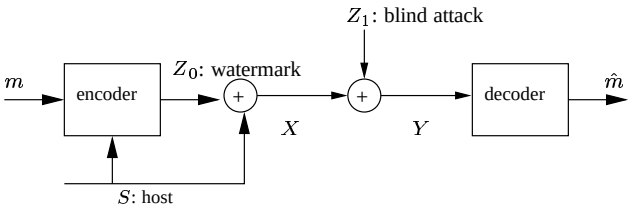

Fig. 1. Watermarking as writing on dirty paper

given in [5]. Based on this important article called *writing on dirty paper* (WDP), watermarking is modeled as a capacity game between an embedder and an attacker [6, 7].

To briefly summarize, the game is formulated as a $\max\min$ game over the data hiding capacity C for general informed coding [8],

$$C(D_0, D_1) = \max_{\substack{\mathbf{E}Z_0^2 \le D_0 \\ p(u,s)p(z_0|u,s)}} \min_{\mathbf{E}Z_1^2 \le D_1} I(U;Y) - I(U;S) \quad (1)$$

where $I(\cdot;\cdot)$ denotes the mutual information [9] of two random variables, and U is an auxiliary random variable such that $(U, S) \to X \to Y$ form a Markov chain. To find the capacity, it is in general difficult to optimize over all possible choices of probability distributions $p(u, s)$ and $p(z_0|u, s)$. However, Costa proved that [5], when both Z_0 and Z_1 are IID Gaussian with variances D_0 and D_1, respectively, an optimal U has the following form,

$$U = \alpha S + Z_0 \quad (2)$$

The proof can be shown by simply substituting $\alpha = D_0/(D_0 + D_1)$ and calculating that

$$C(D_0, D_1) = I(U;Y) - I(U;S) = \frac{1}{2}\log_2(1 + \frac{D_0}{D_1}) \quad (3)$$

We can look at this result in many ways. First, eq. (3) is Shannon's capacity with a signal to noise ratio (SNR) of D_0/D_1. It does not depend on S at all. In other words, the watermarking game has *public-private equivalence* (PPE)[7], which means that data hiding rates can be as high as if the the interference from the host signal were not present. Secondly, similar to Shannon's channel coding game (see Chap. 10 of [9]), Gaussianity is the best strategy for both the embedder and the attacker regardless of each other's choice of probability distribution. Finally and strangely, the game has the property of $\max\min = \min\max$. One may interpret that it does not matter if the embedder or the attacker goes first in the game. However, we will not emphasize this interpretation because it is not realistic to swap positions of the embedder and the attacker.

More recently, it is shown that the Gaussian assumption on the host can be loosened. PPE holds as long as the host signal is ergodic [10]. Also, watermarking schemes inspired by WDP have been proposed to achieve high data hiding rates (e.g. [11, 12, 13]).

3. WDP MODELING UNDER SCRUTINY

Having reviewed the literature, however, we might question if WDP is a realistic model for watermarking.

3.1. IID of host

For obvious reasons, multimedia objects in the original media space are far from being IID. Transform domain watermarking (See e.g. [14][12]) has been a common practice to make the IID assumption valid. A host signal is partitioned into quasi-stationary blocks. A human perception motivated transform, such as FFT for audio, is applied to the blocks. With a good choice of the transform, coefficients in the transform domain can be treated as mutually independent channels. If the block size is also chosen properly so that inter-block correlation is negligible, IID of host is a good assumption for each of the channels.

3.2. Watermark power constraints: time-invariant or not?

After the human perception motivated transform, if cross-channel masking is negligible, we can model the perceptual masking phenomena as channel-wise power constraints D_{0j} for channel j,

$$D_{0j} = f_j(S_j) \quad (4)$$

where $f_j(S_j)$ is a masking function. Hereafter, the channel index j is omitted for the sake of simplicity.

Eq.(4) complicates WDP in two subtly different ways. First, D_0 depends on S. This can be resolved using a nonlinear mapping that removes the dependency of D_0 on S. For instance, in audio watermarking, since human hearing masking is approximately log-scaled, we can write

$$D_0 = 10^{-\Delta/10} \cdot |S|^2$$

where Δ is the masking scaling factor in dB. In this case, a valid nonlinear mapping is $\log_{10}(\cdot)$. Let Z_0' be a new watermark variable defined in the log domain,

$$\log_{10} X = \log_{10} |S| + Z_0'$$

Trivially, it can be shown that the new constraint for Z_0' becomes

$$Z_0'^2 \le D_0' = \Delta/10$$

independent to S and invariant.

The second complication turns out to be more problematic and needs to be addressed next.

3.3. Watermark power constraints: L_2 v.s. L_∞?

We now assume that the constraint D_0 can be made independent to S. Nevertheless, to model the human perceptual masking more faithfully, we usually have to assume that a watermark Z_0 is not perceptible only if the following is true sample-wise,

$$(Z_0[k])^2 \le D_0, \forall k \quad (5)$$

This is an L_∞ (peak-power) constraint to the watermark, which is more strict than the corresponding L_2 (average-power) constraint,

$$\frac{1}{N}\sum_{k=1}^{N}(Z_0[k])^2 \le D_0 \quad (6)$$

Unfortunately, after such modification, we don't know much about the capacity beyond Gel'fand-Pinsker's [8] general expression $C = \max\min I(U;Y) - I(U,S)$, which is just a rewrite of

eq. (1) without specifying the constraints. Because of the difficulties to optimize U, the capacity generally does not have a simple close-form expression for arbitrary input constraints, except for L_2.

Realizing that it is not the best way to model human perception, nevertheless, we will keep using the L_2 approach for the good insights that come with the concise expression of data hiding capacities.

4. A MATHEMATICAL MODEL FOR 2-WATERMARKING

This section applies recent findings in network information theory to derive the data hiding capacity region for 2-watermarking with the watermarks jointly satisfying a power constraint. As shown in Fig. 2, the system diagram is a straightforward extension from WDP, and is called *broadcast on dirty paper* (BDP) here. The watermark encoder has to embed a fragile watermark and a robust watermark carrying messages $m_1 \in \{0,1\}^{NR_1}$ and $m_2 \in \{0,1\}^{NR_2}$, respectively. Let Z_0 be the sum of two watermarks with an L_2 power constraint $\mathbf{E}[Z_0^2] \le D_0$ subject to two different strengths of blind additive Gaussian attacks $Z_1 \sim \mathcal{N}(0, D_1)$ and $Z_2 \sim \mathcal{N}(0, D_2)$, $D_1 \le D_2$. From Y_1 and Y_2, messages $\hat{m}_1, \hat{m}_2$ are decoded at rates R_1 and R_2, separately. The decoding is successful if $\hat{m}_1 = m_1$ and $\hat{m}_2 = m_2$.

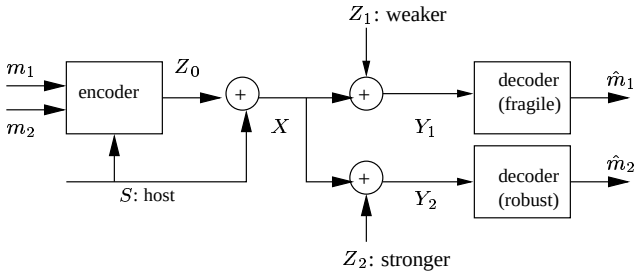

Fig. 2. Simultaneous robust and fragile watermarking

It has been shown that the following rate region is achievable for BDP,

$$\begin{cases} R_1 < c\left(\frac{\beta D_0}{D_1}\right) \\ R_2 < c\left(\frac{(1-\beta)D_0}{\beta D_0 + D_2}\right) \end{cases} \tag{7}$$

where

$$c(x) = \frac{1}{2}\log_2(1+x) \tag{8}$$

is Shannon's capacity function at an SNR of x. The proof of achievability [15, 16] of (7) is sketched here. Let $(U_1, U_2) = (\alpha_1 S + V_1, \alpha_2 S + V_2)$ be auxiliary random variables, where $V_1 \sim \mathcal{N}(0, \beta D_0)$ and $V_2 \sim \mathcal{N}(0, (1-\beta)D_0)$ are independent watermarks with $V_1 + V_2 = Z_0$. The optimal choice of coefficients is $\alpha_1 = (1-\alpha_2)\beta D_0/(\beta D_0 + D_1)$ and $\alpha_2 = (1-\beta)D_0/(D_0 + D_2)$. Then, it can be similarly derived, as was done in [5], that any rate pair within the region defined by (7) is achievable.

Conversely, any rate pair outside the region is actually not achievable even without the interference from S [17]. Therefore, we can conclude that PPE also holds for BDP.

The achievability of (7) has a power-sharing interpretation. βD_0 of power is assigned to the fragile watermark V_1, and $(1-\beta)D_0$ is assigned to V_2. More interestingly, by inspecting the SNR in (7), it can be inferred that the robust watermark V_2 is embedded first. This coincides with the conjecture by Mintzer et al [2]. Then, in terms of data hiding rates, how much is power sharing better than time sharing?

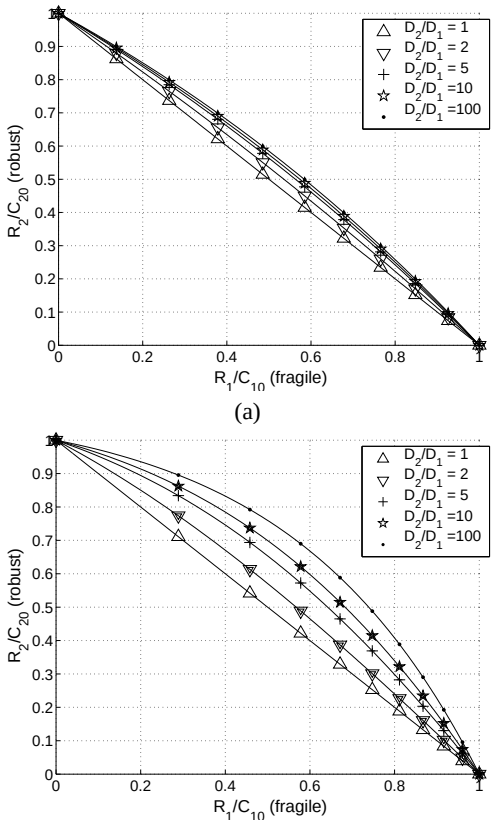

Fig. 3. Capacity regions of BDP: (a) $D_0/D_1 = 1$ (b)$D_0/D_1 = 10$

In Fig. 3, capacity regions given by (7) are plotted at various energy ratios $(D_0/D_1, D_2/D_1)$. The axes are normalized with respect to $C_{10} = c(D_0/D_1)$ and $C_{20} = c(D_0/D_2)$, the individual single watermarking capacities. Each convex curve connecting $(1,0)$ and $(0,1)$ shows the boundary of a capacity region achievable by power sharing. Particularly, if $D_2/D_1 = 1$, the region is triangular and the same as the region achievable by time sharing. Nominally, a power-sharing region is significantly larger than the time-sharing triangle when both D_0/D_1 and D_2/D_1 are large. However, when D_0/D_1 is small, the improvement is marginal.

These results suggest that it may worth considering power sharing, instead of time sharing, if the total power of the watermarks can be as much as the masking threshold allows. In this case, one can possibly assume a large D_0/D_1 because the fragile watermark attacker has very little room to play without ruining the perceptual quality. The improvement is especially effective if D_2/D_1 is also large and most of the power is to be assigned to one watermark. For instance, as shown in Fig. 3(b), at $D_0/D_1 = 10$ and $D_2/D_1 = 100$, the rate pair $(0.28, 0.90)$ is achievable, nearly three times better than the time-sharing rate pair $(0.10, 0.90)$.

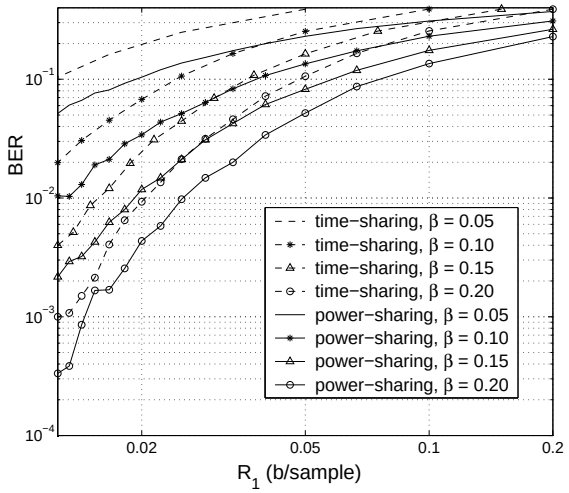

Fig. 4. Data hiding rates: power-sharing vs. time-sharing

5. JOINT SS-QIM WATERMARKING: PRELIMINARY RESULTS

Spread spectrum (SS) [18] and quantization index modulation (QIM) [11] have become two of most popular watermarking techniques. Because SS and QIM each has its own advantages [19], we propose to simultaneously use both by power sharing. Let V_1, V_2 be two watermarks. We intend that V_1 carries a high payload with low bit error rate (BER), and that V_2 is detected with low probability of false alarm. V_2 is embedded first and is based on SS. Afterwards, V_1 is embedded based on QIM. In the preliminary experiments being conducted, we have chosen the quantization technique to be *scalar Costa scheme* [13] with *spread transform* (ST-SCS) [11]. The detailed design of the quantizer follows the guidelines given by Eggers et al [13].

Fig. 4 shows BER versus data hiding rates R_1 performed by V_1, at various sharing factor β. As shown, at four different β, power sharing consistently has lower BER than time sharing at all data hiding rates. The attack power level is $D_1/D_0 = 1$ in this simulation. Each BER data point is obtained by averaging over 10^6 samples. Here, performances of V_2 are not presented. A more thorough comparison between power sharing and time sharing of joint SS-QIM watermarking, evaluated by performances of both watermarks, is left as a future research direction.

6. CONCLUSIONS

We presented BDP as a mathematical model for multiple watermarking. For simultaneous embedding of robust and fragile watermarks, we have shown that optimal data hiding requires embedding the robust watermark first. Also, we have demonstrated that, both in theory and in practice, power sharing between watermarks achieves higher data hiding rates than time sharing.

7. REFERENCES

[1] Ingemar J. Cox and Matthew L. Miller, "Electronic watermarking: the first 50 years," in *Proc. IEEE 2001 Int. Workshop on Multimedia Signal Processing*, pp. 225–230.

[2] Fred Mintzer and Gordon W. Braudaway, "If one watermark is good, are more better?," in *Proc. Int. Conf. on Acoustics, Speech, and Signal Processing, Phoenix*, May 1999, pp. 2067–2069.

[3] Chun-Shien Lu and Hong-Yuan M. Liao, "Multipurpose watermarking for image authentication and protection," *IEEE Trans. Image Processing*, vol. 10, no. 10, pp. 1579–1592, Oct. 2001.

[4] Ingemar J. Cox, Matthew L. Miller, and Andrew L. McKellips, "Watermarking as communications with side information," *Proc. IEEE*, vol. 87, no. 7, pp. 1127–1141, Jul. 1999.

[5] Max H. M. Costa, "Writing on dirty paper," *IEEE Trans. Information Theory*, vol. IT-29, no. 3, pp. 439–441, May 1983.

[6] Pierre Moulin, "A mathematical approach to watermarking and data hiding," May 2002, tutorial notes for IEEE Int. Conf. on Acoustics, Speech, and Signal Processing.

[7] Aaron S. Cohen and Amos Lapidoth, "the Gaussian watermarking game," *IEEE Trans. Information Theory*, vol. 48, no. 6, pp. 1639–1667, Jun. 2002.

[8] S. I. Gel'fand and M.S. Pinsker, "Coding for channel with random parameters," *Problems of Control and Information Theory*, vol. 9, no. 1, pp. 19–31, Jan 1980.

[9] T. M. Cover and J. A. Thomas, *Elements of Information Theory*, Wiley-Interscience, New York, 1991.

[10] Aaron S. Cohen and Amos Lapidoth, "Generalized writing on dirty paper," in *Proc. 2002 Int. Symp. on Information Theory*, Lausanne, Switzerland, June 2002, p. 227.

[11] Brian Chen and Gregory W. Wornell, "Quantization index modulation: a class of provably good methods for digital watermarking and information embedding," *IEEE Trans. Information Theory*, vol. 47, no. 4, pp. 1423–1443, May 2001.

[12] Jim Chou, K. Ramchandran, and A. Ortega, "Next generation techniques for robust and imperceptible audio data hiding," in *Proc. Int. Conf. on Acoustics, Speech, and Signal Processing, Salt Lake City*, May 2001, pp. 1349–1352.

[13] Joachim J. Eggers, Robert Bauml, Roman Tzschoppe, and Bernd Girod, "Scalar Costa scheme for information embedding," *IEEE Trans. Signal Processing*, vol. 51, no. 4, Apr 2003.

[14] M. Swanson, M. Kobayashi, and A.H. Tewfik, "Multimedia data-embedding and watermarking technologies," *Proc. IEEE*, vol. 86, no. 6, pp. 1064–1087, June 1998.

[15] Yossef Steinberg, "On the broadcast channel with random parameters," in *Proc. 2002 Int. Symp. on Information Theory*, Lausanne, Switzerland, June 2002, p. 225.

[16] Young-Han Kim and Styrmyr Sigurjonsson, "Capacity theorems for channels with state information," private communication, Dec. 2002.

[17] P.P. Bergmans, "A simple converse for broadcast channels with additive white gaussian noise (corresp.)," *IEEE Trans. Information Theory*, vol. IT-20, pp. 279–280, Mar. 1974.

[18] Ingemar J. Cox, Joe Killan, F. Thomson Leighton, and Talal Shamoon, "Secure spread spectrum watermarking for multimedia," *IEEE Trans. Image Processing*, vol. 6, no. 12, pp. 1673–1687, Dec. 1999.

[19] Mauro Barni, "What is the future for watermarking? (part ii)," *IEEE Signal Processing Magazine*, vol. 20, no. 6, Nov. 2003.