

A Framework for the Design of Good Watermark Identification Codes

Pierre Moulin and Ralf Koetter
University of Illinois
Beckman Inst., Coord. Sci. Lab & ECE Dept.
405 N. Mathews Ave., Urbana, IL 61801

ABSTRACT

The fundamental difference between the data-hiding and watermark signature verification problems was highlighted in a 2001 paper by Steinberg and Merhav. In data hiding, the maximum number of messages that can be reliably decoded is essentially 2^{nC} , where n is the host sequence length and C is the data-hiding capacity. A dramatically different result is obtained for signature verification: in principle one can discriminate between a doubly exponential number of signatures: $2^{2^{nC'}}$, where C' is the *identification capacity*. This paper proposes a practical design of codes for the latter application and compares the results with current designs in the literature.

1. INTRODUCTION

The last five years have seen much research on the design of data-hiding systems that can reliably communicate a message to a decoder. The basic setup is the following. Given a length- n host signal $\mathbf{s} \in \mathcal{S}^n$, a message $m \in \mathcal{M}$ and a key $v \in \mathcal{V}$, the transmitter produces a marked sequence $\mathbf{x} = f(\mathbf{s}, m, v)$, using an embedding function f . Throughout this paper, we assume for simplicity that the attacker passes \mathbf{x} through a discrete memoryless channel $p(y|x)$, producing a degraded output \mathbf{y} . The receiver knows v , observes \mathbf{y} , and produces an estimate $\hat{m} = g(\mathbf{y}, v)$ of the transmitted message, where g is the decoding function. When both the embedder and the attacker are subject to distortion constraints, a lot is known about the fundamental performance limits for this transmission problem (in terms of capacity and error exponents for transmission rates below capacity).¹ A lot is known about the structure of optimal codes as well. For instance, if $\mathcal{S} = \mathbb{R}$ and distortion is measured in the squared-error sense, lattice quantization index modulation (QIM) schemes are nearly optimal in the sense of approaching the aforementioned fundamental limits as the lattice dimension tends to infinity. These schemes may be viewed as codes for the so-called Modulo Lattice Additive Noise (MLAN) channel.^{1,2}

Now the problem is quite different when the receiver must perform the simpler binary decision: Is the received signal \mathbf{y} marked using a *given* signature $m \in \mathcal{M}$ or not? This is a signature verification problem, sometimes referred to as watermark detection. This problem has a variety of possible applications. Here are two examples.

1. Each m corresponds to a user of the system, and \mathbf{s} is an image. By marking his image according to $\mathbf{x} = f(\mathbf{s}, m, v)$, the user claims ownership of the image.
2. Let \mathbf{s} be the picture of a speaker presenting slides at a professional meeting. Here m is an index to the slides file, perhaps as large as 2 Megabytes. By marking his image according to $\mathbf{x} = f(\mathbf{s}, m, v)$, the speaker claims that he was presenting the said file m and not another one, no matter how minutely different the other file may be.

In the first case, the size of the message set is at most 33 bits (the whole human population is currently smaller than 9 billion $\approx 2^{33}$); in the second case, the size of the message set is very large. It is impossible to hide and reliably communicate 2 Megabytes of data in a host image, but is it possible to reliably solve the verification problem?

E-mails: moulin@ifp.uiuc.edu, koetter@comm.csl.uiuc.edu.

2. MATHEMATICAL MODEL

In the signature verification problem, the receiver has access not just to the degraded data \mathbf{y} and the key v , but also to a test signature $m^* \in \mathcal{M}$. The decoding function $\hat{m} = g(\mathbf{y}, v)$ is replaced with a *binary decision rule* $g(\mathbf{y}, v, m^*)$ taking values in $\{0, 1\}$ and indicating the absence or presence of the tested signature, respectively. If m^* is embedded in the signal, \mathbf{y} follows a conditional probability distribution $p_{m^*}(\mathbf{y}|v)$ that is induced by the probability distribution of S , the choice of f , and the attack channel $p(y|x)$.

Given (v, m^*) , the binary hypothesis test takes the form

$$\begin{cases} H_0 : & \mathbf{Y} \sim p_m(\cdot|v), \quad \text{for some } m \neq m^* \\ H_1 : & \mathbf{Y} \sim p_{m^*}(\cdot|v) \end{cases} \quad (1)$$

where H_0 and H_1 are respectively the “signature absent” (negative ID) and “signature present” (positive ID) hypotheses. The challenge is to design a *good* embedding code. How should we measure goodness of the code f ?

The two possible error events at the detector are *false positives* (deciding H_1 when H_0 is true) and *false negatives* (deciding H_0 when H_1 is true). Denote by λ_1 and λ_2 respectively, the probabilities of false negatives and false positives. A natural measure of goodness of the code is the largest size of the message set \mathcal{M} for which reliable detection is still possible (that is, λ_1 and λ_2 are arbitrarily small.)

Notation: in this paper, all logarithms are in base 2. Boldface letters denote sequences, and calligraphic letters denote sets.

3. INFORMATION-THEORETIC BACKGROUND

The problem defined in the previous section has been studied in the information theory literature, starting from Ahlswede and Dueck’s award-winning paper on identification (ID) codes.³

For the standard decoding problem (in which the receiver must estimate the transmitted m), the largest size of \mathcal{M} is essentially 2^{n^C} , where C is the data-hiding capacity, expressed in bits per host signal sample. For the signature verification problem (1), the receiver’s task is dramatically easier, and the answer is dramatically different. As shown by Ahlswede and Dueck,³ the largest size of \mathcal{M} is *doubly exponential* in n , i.e., it is of the form $2^{2^{n^{C'}}$. This remarkable result was later extended in several ways.⁴⁻⁶ In particular, Steinberg and Merhav’s paper⁶ applies to the problem of private watermarking, in which the encoder and decoder have access to a common source of randomness (the host signal).

The papers^{3,6} shed some light on what might be the structure of optimal identification codes via a random coding design, but are not concerned about practical constructions. In 1993, Verdú and Wei⁷ proposed the first (and only one known by us to date) explicit construction of identification codes using a three-layer concatenated constant-weight code designed for a *noiseless channel*, in conjunction with a conventional channel code. This design is fundamentally different from the designs used in conventional coding problems. Verdú and Wei emphasized achievability of capacity bounds, but their paper was not concerned about practical algorithms for implementing their three-layer codes.

The goal of this paper is therefore to present a practical design of good identification codes for watermarking and assess whether current signature verification schemes (e.g., those based on lattice QIM⁸⁻¹⁰) possess the desired structure.

4. DEFINITIONS

The following definition may be found in.^{3,7} Let W be a discrete memoryless channel with input alphabet \mathcal{A} and output alphabet \mathcal{B} . We shall frequently use the *binary noiseless channel* in our examples:

$$\mathcal{A} = \mathcal{B} = \{0, 1\}, \quad W(b|a) = 1_{\{b=a\}}. \quad (2)$$

We also use the standard notation W^n to denote the n -th order extension of the channel W , i.e., $W^n(\mathbf{b}|\mathbf{a}) = \prod_{i=1}^n W(b_i|a_i)$. The ID theory can be extended to the case of continuous alphabets.⁵ In the context of QIM watermarking, which will be developed in Section 11, \mathcal{A} and \mathcal{B} are intervals on the real line.

Definition. An $(n, N, \lambda_1, \lambda_2)$ ID code for the channel $W : \mathcal{A} \rightarrow \mathcal{B}$ is a collection $\{\mathbf{q}_m, \mathcal{D}_m, 1 \leq m \leq N\}$ where λ_1 and λ_2 are respectively the type-I and type-II error probabilities, \mathbf{q}_m is a pmf on \mathcal{A}^n , and $\mathcal{D}_m \subset \mathcal{B}^n$, such that

$$P_c(m) \triangleq \sum_{\mathbf{a} \in \mathcal{A}^n} W^n(\mathcal{D}_m | \mathbf{a}) \mathbf{q}_m(\mathbf{a}) \geq 1 - \lambda_1 \quad \forall m \quad (3)$$

$$P_e(m \rightarrow m') \triangleq \sum_{\mathbf{a} \in \mathcal{A}^n} W^n(\mathcal{D}_{m'} | \mathbf{a}) \mathbf{q}_m(\mathbf{a}) \leq \lambda_2 \quad \forall m' \neq m \quad (4)$$

An ID code works as follows. Given a message m , the encoder *randomly* generates a codeword $\mathbf{a} \in \mathcal{A}^n$ according to a certain probability mass function (pmf) \mathbf{q}_m . The decoder, given the test message m^* and the observed channel output \mathbf{b} , declares H_1 (positive ID) if \mathbf{b} belongs to the decoding region \mathcal{D}_{m^*} . A key feature of such codes is that they are stochastic, i.e., given the message m , the codeword \mathbf{a} is generated randomly. For a deterministic code, \mathbf{a} would be a deterministic function of m . As discussed by Ahlswede and Dueck,³ deterministic ID codes generally have *very poor* performance; see Section 5 for an illustration.

In (3), $P_c(m)$ represents the probability of *correct identification* of m , and $1 - \lambda_1$ is a requirement on the worst-case $P_c(m)$. In (4), $P_e(m \rightarrow m')$ represents the probability that m is confused with another *fixed* message m' , and λ_2 is a requirement on the worst-case $P_e(m \rightarrow m')$ (over all m and m'). The probability $P_e(m \rightarrow m')$ should not be confused with the *average* probability of misidentification, where the average is taken over all $m' \neq m$:

$$\bar{P}_e(m) = \frac{1}{N-1} \sum_{m' \neq m} P_e(m \rightarrow m') \quad (5)$$

Example: for the binary noiseless channel (2), the number of possible subsets of $\mathcal{B}^n = \{0, 1\}^n$ is 2^{2^n} . So potentially, N could be as large as 2^{2^n} . If the pmf's \mathbf{q}_m are uniform (within their support set \mathcal{D}_m), then according to (4), the decoding subsets should have small relative overlap:

$$\frac{|\mathcal{D}_m \cap \mathcal{D}_{m'}|}{|\mathcal{D}_m|} \leq \lambda_2 \quad \forall m \neq m'.$$

Computational issues. We are interested in ID codes for which N is extraordinarily large: $\log N$ could be of the order of 10^6 , or perhaps even larger. The theory of ID codes will not be useful to the practitioner unless we find a convenient way to specify all N pmf's \mathbf{q}_m and decoding regions \mathcal{D}_m ; generate \mathbf{a} according to the selected pmf; and verify whether \mathbf{b} belongs to the selected decoding region. This is a substantial challenge, and some of our design ideas (not presented in this paper) failed this test.

5. A NAIVE DESIGN: DETERMINISTIC ID CODE

Consider the binary noiseless channel (2). Assume $\mathcal{M} = \{1, 2, \dots, N\}$. Our naive design is a *deterministic* mapping f from \mathcal{M} to \mathcal{A}^n . That is, \mathbf{q}_m is a zero/one pmf over \mathcal{A}^n :

$$\mathbf{q}_m(\mathbf{a}) = 1_{\{\mathbf{a}=f(m)\}}, \quad \forall \mathbf{a} \in \mathcal{A}^n, m \in \mathcal{M}.$$

The corresponding decoding regions are singletons:

$$\mathcal{D}_m = \{f(m)\}, \quad \forall m \in \mathcal{M}.$$

If $N = 2^n$, then obviously perfect identification is possible: each m can be mapped to a different bitstring, i.e., $f(m)$ is invertible. Assume now that $N > 2^n$, and that the set \mathcal{M} is partitioned into $2^{-n}N$ equivalence classes, all of which have the same size, 2^n . That is, all m in any given equivalence class are mapped to the same bitstring. For any m , the *average* probability of misidentification takes the form

$$\begin{aligned} \bar{P}_e(m) &= |\mathcal{M}|^{-1} \sum_{m' \neq m} 1_{\{f(m')=f(m)\}} \\ &= \frac{1}{N} (2^n - 1) \end{aligned}$$

which is independent of m and can be made arbitrarily small by choosing $N \gg 2^n$.

This result looks attractive, but a clear drawback of this scheme is that the receiver systematically confuses m with *all* m' in the same equivalence class as m . For such m, m' , we obtain $P_e(m \rightarrow m') = 1$ in (4). Therefore this scheme does not satisfy the requirements for an identification code, except in the trivial case $\lambda_2 = 1$.

6. AN IMPROVED DESIGN: STOCHASTIC ID CODE

Let the message set \mathcal{M} have cardinality 2^{rK} , where $r = \epsilon n$, for some $\epsilon \in (0, 1)$. Partition the message $m \in \mathcal{M}$ into k submessages m_1, m_2, \dots, m_K , each viewed as a binary string of length r . Equivalently, we view the binary representation of m as a $r \times K$ matrix, and m_u as the u -th column of that matrix.

$$m \sim \underbrace{\left(\begin{array}{cccccc} 1 & 0 & 0 & 0 & 0 & \cdots \\ 0 & 0 & 0 & 0 & 0 & \\ 0 & 0 & 0 & 0 & 0 & \\ \vdots & & & & & \ddots \end{array} \right)}_K \Bigg\} r$$

The encoding goes as follows.

1. Generate a random variable u uniformly distributed over $\{1, 2, \dots, K\}$.
2. Transmit the pair (u, m_u) .

The number of bits used to represent $\mathbf{a} = (u, m_u)$ is $n = \log K + r$. Note that $\log K = \frac{1-\epsilon}{\epsilon}r$, i.e., the submessages are very short and there a huge number of them (for large n); as well, encoding u is more expensive than encoding m_u .

The message set \mathcal{M} is potentially very large, and its size is conveniently measured using the *second-order rate* of the code, which is defined as

$$\begin{aligned} \frac{1}{n} \log \log |\mathcal{M}| &= \frac{1}{n} \log(rK) \\ &= \frac{1}{n} [\log r + n - r] \\ &= 1 - \epsilon + \frac{\log(\epsilon n)}{n}. \end{aligned} \tag{6}$$

The decoder observes the binary noiseless channel output $\mathbf{b} = \mathbf{a} = (u, m_u)$ and tests for the presence of message m^* . To this end, the decoder just compares submessage m_u with m_u^* . If $m_u = m_u^*$, the decoder declares H_1 (positive ID); otherwise it declares H_0 .

What is the statistical performance of this scheme? If $m = m^*$, the receiver always decides H_1 , so the probability of correct ID is exactly one. For $m \neq m^*$, an error occurs if and only if $m_u = m_u^*$, i.e., we have a collision for the selected submessage. For a *randomly selected* m , the probability of this event is upper bounded by $2^{-r} = 2^{-\epsilon n}$. * This looks very good.

Unfortunately, there exists a message m whose matrix representation differs from that of m^* by a single bit, so that the unwanted collision $m_u = m_u^*$ occurs with probability $1 - \frac{1}{K}$! This is catastrophic when K is large. This case is easily seen to be the worst possible one, so we have an ID code with type-II error probability

$$\lambda_2 = 1 - \frac{1}{K} = 1 - 2^{-(1-\epsilon)n}. \tag{7}$$

Therefore we only obtain a trivial improvement over the result $\lambda_2 = 1$ obtained for deterministic ID codes.

*The bound is tight as $K \rightarrow \infty$.

7. THE NEED FOR REDUNDANCY

The key problem with the scheme presented above is that for any given message m^* , there exists another message m that has the same matrix representation as m^* , except for one column. Hence one would like a representation of messages that increases the “distance” between messages – measured here in the number of distinct columns. This requires introducing redundancy in the matrix representation of the messages. Several ideas are possible in this regard, let’s begin with the simplest one.

7.1. Reed-Solomon Codes

A $r \times K$ matrix of 0’s and 1’s may be viewed as a sequence of K symbols (matrix columns), each symbol taking on 2^r possible values. Now view m as such a sequence of source symbols, and consider an (L, K) Reed-Solomon (RS) code over an alphabet of size $q = 2^r$. Such a code has the property that if any K of the L transmitted symbols are received, then the original K source symbols can be recovered; Reed-Solomon codes exist for $L \leq q$.

Reed-Solomon codes are based on the arithmetic of finite fields. Given a length- K source sequence $\mathbf{z} = (z_0, z_1, \dots, z_{K-1})$, where each z_k is a r -bit symbol, a Reed-Solomon code outputs the length- L codeword $\mathbf{c} = (c_0, c_1, \dots, c_{L-1})$ where $c_k = c_k(\mathbf{z}) = P(g^k | \mathbf{z})$ is a r -bit symbol; g is a generator of the (cyclic) group of nonzero elements in $\{0, 1, \dots, q-1\}$; and $P(x | \mathbf{z}) = z_0 + z_1x + \dots + z_{K-1}x^{K-1}$ for any field element x .

In the context of our ID problem, the RS code maps each possible message m to a unique sequence of source symbols $\mathbf{z}[m]$:

$$\mathbf{z}[m] \sim \underbrace{\left(\begin{array}{cccccccc} 1 & 0 & 0 & 0 & 1 & 0 & 1 & 0 & \dots \\ 0 & 0 & 1 & 1 & 0 & 0 & 0 & 0 & \\ 0 & 1 & 0 & 1 & 0 & 1 & 0 & 1 & \\ \vdots & & & & & & & & \ddots \end{array} \right)}_{L > K} \Bigg\}^r$$

When asked to produce an identification string for m , the encoder does the following:

1. Choose a field element u according to the uniform distribution on the field $\{0, 1, \dots, L-1\}$.
2. Transmit the string $\mathbf{a} = (u, c_u(\mathbf{z}[m]))$.

The number of bits used to represent $(u, c_u(\mathbf{z}[m]))$ is $n = \log L + r$.

For the noiseless binary channel, the receiver observes the string $\mathbf{b} = \mathbf{a} = (u, c_u)$ and must declare whether the message was m^* or equivalently, whether the source sequence was $\mathbf{z}^* = \mathbf{z}[m^*]$. To this end, the receiver evaluates $c_u(\mathbf{z}^*)$. If $c_u(\mathbf{z}^*)$ equals the received c_u , we declare H_1 (positive ID), otherwise we declare H_0 .

Let us explicitly identify the pmf’s \mathbf{q}_m and the decoding sets \mathcal{D}_m for this ID code. Write the codeword \mathbf{a} as $(\mathbf{a}^{(1)}, \mathbf{a}^{(2)})$, where $\mathbf{a}^{(1)}$ is the binary representation of u ($\log L$ bits), and $\mathbf{a}^{(2)}$ is the binary representation of m_u ($r = n - \log L$ bits). The set of all sequences \mathbf{a} has log cardinality equal to n . For any m , the pmf $\mathbf{q}_m(\mathbf{a})$ puts uniform probability over a set of log cardinality equal to $\log L$. All pmf’s \mathbf{q}_m are distinct. The decoding set \mathcal{D}_m coincides with the support set of \mathbf{q}_m .

7.2. Performance Analysis

What is the statistical performance of the above scheme? If $m = m^*$, the receiver always decides H_1 , so the probability of correct ID is exactly one. For $m \neq m^*$, the analysis proceeds as follows. The sequences $\mathbf{c}(\mathbf{z})$ and $\mathbf{c}(\mathbf{z}^*)$ coincide in at most K positions. We will thus make an error (decide H_1) if u happens to be a position where $c_u(\mathbf{z})$ and $c_u(\mathbf{z}^*)$ coincide. The probability of this happening is K/L . Therefore, we have constructed an ID code with

$$n = \log L + r, \quad N = 2^{rK}, \quad \lambda_1 = 0, \quad \lambda_2 = \frac{K}{L}.$$

In other words we can accommodate 2^{rK} users using bit strings of length $n = \log L + r$ and achieving an error probability of K/L .

Let us evaluate the second-order rate of this code. Defining

$$E_2 = -\frac{1}{n} \log \frac{K}{L} > 0, \quad (8)$$

we may write

$$\lambda_2 = 2^{-nE_2}$$

and

$$\begin{aligned} R &= \frac{1}{n} \log \log |\mathcal{M}| \\ &= \frac{1}{n} \log(rK) \\ &= -E_2 + \frac{1}{n} \log(rL) \\ &= -E_2 + \frac{1}{n} [\log r + n - r] \\ &= -E_2 + 1 - \frac{r}{n} + \frac{\log r}{n}. \end{aligned}$$

For a given error probability, we want to minimize r in order to maximize R . For any RS code, we need $L \leq 2^r$ and therefore

$$r \geq \log L = n - r$$

which implies $r \geq \frac{n}{2}$. Therefore we choose $r = \frac{n}{2}$ and $L = 2^r$, achieving the desired bound. The second-order rate of this code is

$$R = -E_2 + \frac{1}{2} + \frac{\log(n/2)}{n}. \quad (9)$$

For a sequence of RS codes in which $n \rightarrow \infty$, the last term in the right side vanishes, and the sum $R + E_2$ converges to $\frac{1}{2}$. This provides the fundamental tradeoff between second-order rate R and error exponent E_2 for this construction.

Numerical Examples. Choose codelength $n = 32$ and error probability $\lambda_2 = 2^{-14} \approx 0.000061$. Then $E_2 = \frac{14}{32}$. The second-order rate of the code is obtained from (9) as $R = -\frac{14}{32} + \frac{1}{2} + \frac{\log 16}{32} = \frac{3}{16}$. Therefore the log cardinality of the code is $\log N = 2^{nR} = 64$, which is good but not impressive.

If we increase n to 64, performance becomes rather spectacular: now $E_2 = \frac{14}{64}$, $R = -\frac{14}{64} + \frac{1}{2} + \frac{\log 32}{64} = \frac{23}{64}$, and $\log N = 2^{nR} = 8,388,608$. The effect of a mere doubling of n is an improvement of $\log N$ by five orders of magnitude! And the parameters of the RS code, $r = 32$ and $K = 262,144$, are computationally manageable. Finally, while $L = 2^{32} \approx 4 \times 10^9$ is much larger than K , only *one* of the L output symbols needs to be computed.

8. COMMON RANDOMNESS

If the encoder and receiver share common randomness, the performance of the authentication scheme can be improved. Consider for instance the design of Section 7. If the random variable u was known to the receiver prior to the transmission, then the authentication code could consist of c_u alone. In fact we can do even better.

In the following scheme, encoder and receiver share randomness in the form of a random variable v , uniformly distributed over $\{1, 2, \dots, L'\}$. Denote by

$$H(V) = \frac{1}{n} \log L' \quad (10)$$

the entropy rate of V . Represent the message $m \in \mathcal{M}$ by a length- K sequence $\mathbf{z}[m]$ of symbols defined over the alphabet $\{0, 1, \dots, 2^r - 1\}$. Then apply a (K, L) RS code to $\mathbf{z}[m]$ and denote by $\mathbf{c}(\mathbf{z}[m])$ the output, as described in Section 7. If L is chosen as a multiple of L' , we may view the position of each symbol of \mathbf{c} as indexed by the pair (u, v) , where $u \in \{1, 2, \dots, L/L'\}$ and $v \in \{1, 2, \dots, L'\}$.

The encoder observes v , generates a uniform random variable u , applies the RS code to $\mathbf{z}[m]$, and then transmits the pair $(u, c_{u,v}(\mathbf{z}[m]))$. The binary representation of this pair has length $n = \log(L/L') + r$.

The decoder knows v , observes u and $c_{u,v}$, evaluates $c_{u,v}^* = c_{u,v}(\mathbf{z}[m^*])$, and tests whether $c_{u,v} = c_{u,v}^*$. The type-I error probability is zero, and the type-II error probability is $\lambda_2 = \frac{K-1}{L}$.

Choose the following parameters for the RS code:

$$r = \frac{1 + H(V)}{2}n, \quad L = 2^r.$$

Recalling (8) and (10), the second-order rate of the code is

$$\begin{aligned} R &= \frac{1}{n} \log \log |\mathcal{M}| \\ &= \frac{1}{n} \log(rK) \\ &= -E_2 + \frac{1}{n} \log(rL) \\ &= -E_2 + \frac{1}{n} \left[\log \left(\frac{1 + H(V)}{2}n \right) + r \right] \\ &= -E_2 + \frac{1 + H(V)}{2} + \frac{\log \left(\frac{1 + H(V)}{2}n \right)}{n}. \end{aligned} \tag{11}$$

Compared with (6), the rate is increased by a factor equal to $\frac{H(V)}{2}$, half the entropy rate of the common randomness. Remarkably, just two bits of common randomness ($nH(V) = 2$) suffice to *double* $\log |\mathcal{M}|$, for the same values of n and λ_2 .

Hashing. The availability of common randomness is reminiscent of hash codes.¹¹ A hash function h maps a length- n' bitstring m and a length- k bitstring v (secret key) to a length- n output bitstring $h(m, v)$. Hash functions are frequently used in applications where $n' \gg n$, e.g., $n \approx 100$ and n' is extremely large, possibly unbounded. We may view $h(m, v)$ as the output of a deterministic ID code (because v is known to the receiver). Upon observing $h(m, v)$, the receiver tests the hypothesis $m = m^*$ by computing $h(m^*, v)$ and comparing with $h(m, v)$. If these two bitstrings are identical, the receiver decides H_1 ; otherwise H_0 .

Hashing codes are deterministic, and very much like the code of Section 5, they cannot be expected to be good ID codes. [†] Indeed, while the *average error probability* (over all m) is 2^{-k} and can be made very low by appropriately choosing k , collisions occur when $n + k < n'$ (the normal mode of operation of hashing codes). Therefore the worst-case error probability is close to 1.

9. ID CAPACITY

Ahlswede and Dueck³ have studied the fundamental limits of ID codes. In this section, we first state those limits; in the next one, we discuss how to approach them.

Definition. An ID rate R is achievable if for any $\lambda_1 > 0$, $\lambda_2 > 0$, $\epsilon > 0$, and sufficiently large n , there exist $(n, N, \lambda_1, \lambda_2)$ ID codes with rate $\frac{1}{n} \log \log N \geq R - \epsilon$.

Definition. The ID capacity of the channel is the maximum achievable ID rate.

Theorem 1.³ The ID capacity of the channel $W(b|a)$ without common randomness is equal to its Shannon capacity, $C(W) = \max_{p(a)} I(A; B)$.

Theorem 2.⁶ Assume an i.i.d. source of randomness with entropy $H(V)$ is available to the encoder and decoder. In this case, the ID capacity of the channel $W(b|a)$ is equal to $C(W) + H(V)$.

As an immediate application of Theorem 1, we consider the noiseless binary channel W , which has capacity $C(W) = 1$. The RS construction of Section 7 achieves a second-order rate $\frac{1}{2}$ (obtained by making the error exponent E_2 arbitrarily small). This sequence of ID codes is therefore quite good (due to its ability of encoding a double exponential number of messages) but not optimal.

[†]The practical justification for hashing codes is that it is *believed to be computationally difficult* to find two messages m and m^* that can be confused. In contrast, ID codes offer an *absolute guarantee* on the probability of confusion.

10. GOOD ID CODES FOR NOISY CHANNELS

A natural idea to design an ID code for a general noisy channel is to concatenate a standard transmission code and an ID code for a noiseless channel. We show that not only is this construction easy to implement, but also it is optimal.

Definition. An (n, N_{TR}, λ) transmission code is a collection of codewords $\phi(i) \in \mathcal{A}^n$ and nonoverlapping decoding regions $\mathcal{E}(i) \subset \mathcal{B}^n$, for $1 \leq i \leq N_{TR}$, such that the worst-case (over all i) probability of correct detection is at least $1 - \lambda$:

$$W^n(\mathcal{E}(i)|\phi(i)) \geq 1 - \lambda, \quad 1 \leq i \leq N_{TR}.$$

The rate of an (n, N_{TR}, λ) transmission code is $\frac{1}{n} \log N_{TR}$.

If the Shannon capacity of the noisy channel $W(b|a)$ is C , then for any arbitrarily small $\lambda_{TR} > 0$, $\eta_{TR} > 0$, and large enough n , there exists an $(n, N_{TR}, \lambda_{TR})$ transmission code for the channel W^n , with code rate

$$R_{TR} = \frac{1}{n} \log N_{TR} = C - \eta_{TR}.$$

Now consider the (nR_{TR}) -th order extension of the binary noiseless channel, which has capacity $C = 1$. According to Theorem 1, for any arbitrarily small $\lambda_2^{\text{noiseless}}$ and $\eta_{\text{noiseless}}$, there exists an $(nR_{TR}, N, 0, \lambda_2^{\text{noiseless}})$ ID code for the binary noiseless channel, with second-order rate

$$R_{\text{noiseless}} = \frac{1}{nR_{TR}} \log \log N = 1 - \eta_{\text{noiseless}}.$$

The output of the above ID code belongs to an alphabet of size $2^{nR_{TR}}$ and is fed into the input of the transmission code.

Proposition 1 below states that reliable ID is guaranteed upon observing the output of the transmission code. The statement makes good sense: the codewords of the transmission code can be reliably decoded, and therefore, with high probability, the decoder will make the ID decision based on the correct codeword. The proof of this statement is analogous to the proof of the direct part of Theorem 1a in.³

Proposition 1. The above construction results in an $(n, N, \lambda_{TR}, \lambda_{TR} + \lambda_2^{\text{noiseless}})$ ID code for the noisy channel. The second-order rate of the code is given by

$$R = \frac{1}{n} \log \log N = R_{\text{noiseless}} R_{TR}.$$

Proposition 2. The above construction is optimal, in the sense that for R arbitrarily close to $C(W)$ and λ arbitrarily small, there exists such an (n, N, λ, λ) ID code with n large enough and $N = 2^{2^{nR}}$.

Motivated by this theory, we propose the following practical scheme. Given target type-I and type-II error probabilities λ_1 and λ_2 , choose n “large enough”. Then select a length- n transmission code for channel $W(b|a)$, with rate R_{TR} and error probability $\lambda_{TR} < \min(\lambda_1, \lambda_2)$. This code has $N_{TR} = 2^{nR_{TR}}$ codewords. Also choose $\lambda_{\text{noiseless}} \leq \lambda_2 - \lambda_{TR}$, and evaluate the following parameters:

$$r = \frac{1}{2} n R_{TR}, \quad L = 2^r, \quad K = L \lambda_{\text{noiseless}}, \quad N = 2^{rK}.$$

Define a one-to-one mapping ψ from $\{1, \dots, L\} \times \{1, \dots, 2^r\}$ to $\{1, \dots, N_{TR}\}$. To encode $m \in \{1, 2, \dots, N\}$, perform the following operations.

1. Generate a random variable u uniformly distributed over $\{1, 2, \dots, L\}$;
2. Evaluate $c_u = c_u(\mathbf{z}[m])$, the u -th symbol of the RS codeword corresponding to input m ;
3. Let $i = \psi(u, c_u)$;

4. Transmit $\mathbf{a} = \phi(i)$.

The decoder is given the test message $m^* \in \{1, 2, \dots, N\}$. Upon observing the channel output \mathbf{b} , the decoder performs the following operations:

1. Find i such that $\mathbf{b} \in \mathcal{E}(i)$; declare an error if no such i can be found.
2. Let $(u, c_u) = \psi^{-1}(i)$;
3. Evaluate $c_u^* = c_u(\mathbf{z}[m^*])$;
4. If $c_u = c_u^*$, declare H_1 (positive ID); otherwise declare H_0 .

11. WATERMARK IDENTIFICATION CODES

Here we restrict our attention to public watermarking schemes (host signal is not available to the receiver) based on scalar quantization index modulation (QIM). The quantizer step size $\Delta = \sqrt{12D_1}$ is determined by the embedding per-sample mean-squared distortion constraint, D_1 . The marked samples are subjected to additive i.i.d. Gaussian noise with mean zero and variance $D_2 = D_1$. For the Costa parameter, we select $\alpha = \frac{1}{2}$, which approximately minimizes the error probability of the scheme (see below). For scalar QIM, the channel input and output alphabets are the intervals $\mathcal{A} = \mathcal{B} = [-\Delta/2, \Delta/2]$. The combined effect of self-noise due to quantization and the Gaussian noise (scaled by α) results in a Modulo Lattice Additive Noise (MLAN) channel $W(b|a)$. The capacity of this channel is approximately 0.2.

Select target n and error probability λ of transmission code. The Bhattacharyya coefficient for the MLAN channel with $D_2 = D_1$ and $\alpha = \frac{1}{2}$ is 0.232.¹ For instance, to embed a single bit in a block of length 32, the Bhattacharyya upper bound on error probability is $\frac{1}{2}e^{-0.232 \times 32} \approx 0.0003$. So consider the following design:

- A transmission code for the MLAN channel, consisting of 64 blocks of size 32; one bit is embedded in each block. Hence the parameters of the transmission code are $n = 2048$, $N_{TR} = 2^{64}$ (rate $R_{TR} = \frac{1}{32}$), and $\lambda \approx 0.0003$.
- The ID code of Section 7 for the binary noiseless channel, with the following parameters: $n_{ID} = nR_{TR} = 64$, $N = 2^{8,388,608}$ (hence second-order rate $R_{\text{noiseless}} = \frac{23}{64} \approx 0.36$), $\lambda_1 = 0$, and $\lambda_2^{\text{noiseless}} \approx 0.000061$.

The concatenation of these two codes gives an ID code for the MLAN channel with the following parameters: $n = 2048$, $N \approx 2^{8,388,608}$, $\lambda_1 \approx \lambda_2 \approx 0.0003$. The second-order rate of the code is $R = \frac{23}{2048} \approx 0.011$.

It is worth noticing that the error probability of this scheme is dominated by the error probability for the transmission code ($\lambda_2^{\text{noiseless}}$ is almost an order of magnitude lower than λ_2). Performance could be improved almost “for free” by increasing N in exchange for an increase in $\lambda_2^{\text{noiseless}}$. We have not bothered with such tinkering of performance because N is rather large already!

Comparison with conventional designs. Previous papers on QIM watermark ID used simply an (n, N, λ) transmission code, with rate $R_{TR} = \frac{1}{n} \log N$. They mapped each message $m \in \{1, 2, \dots, N\}$ to a different codeword. These schemes are (n, N, λ, λ) ID codes. Their second-order rate is equal to $\frac{1}{n} \log \log N = \frac{1}{n} \log(nR_{TR})$ and vanishes as $n \rightarrow \infty$, falling far short of the potential of ID codes.

12. CONCLUSION

We have developed practical ID codes that can achieve roughly one half of the ID capacity. For instance, we have proposed a computationally simple scheme based on a Reed-Solomon code and a scalar QIM code with the following parameters: codelength $n = 2048$, watermark-to-noise ratio $D_1/D_2 = 1$, number of messages $N = 2^{8,388,608}$, and worst-case error probability guaranteed to be below 0.0003. We are currently working at several extensions of this work, including private watermarking (exploiting the presence of a source of common randomness in the form of the host signal), and some algebraic-geometric codes that may be useful alternatives to the Reed-Solomon construction proposed here.

ACKNOWLEDGMENTS

This work was supported by NSF grants CCR 02-08809 and CCR 03-25924.

REFERENCES

1. P. Moulin and R. Koetter, “Data-Hiding Codes,” *Proc. IEEE*, Vol. 93, No. 12, pp. 2083–2127, Dec. 2005.
2. R. Zamir, S. Shamai (Shitz), and U. Erez, “Nested Linear/Lattice Codes for Structured Multiterminal Binning,” *IEEE Trans. Information Theory*, Vol. 48, No. 6, pp. 1250–1276, June 2002.
3. R. Ahlswede and G. Dueck, “Identification via Channels,” *IEEE Trans. Information Theory*, Vol. 35, No. 1, pp. 15–29, Jan. 1989.
4. T. S. Han and S. Verdú, “New Results in the Theory of Identification via Channels,” *IEEE Trans. Information Theory*, Vol. 38, No. 1, pp. 14–25, Jan. 1992.
5. M. Burnashev, “On Identification Capacity of Infinite Alphabets or Continuous-Time Channels,” *IEEE Trans. Information Theory*, Vol. 46, No. 7, pp. 2407–2414, Nov. 2000.
6. Y. Steinberg and N. Merhav, “Identification in the Presence of Side Information with Application to Watermarking,” *IEEE Trans. Information Theory*, Vol. 47, No. 4, pp. 1410–1422, May 2001.
7. S. Verdú and V. K. Wei, “Explicit Construction of Optimal Constant-Weight Codes for Identification via Channels,” *IEEE Trans. Information Theory*, Vol. 39, No. 1, pp. 30–35, Jan. 1993.
8. J. J. Eggers and B. Girod, “Blind Watermarking Applied to Image Authentication,” *Proc. ICASSP*, pp. III. 1977–1980, Salt Lake City, May 2001.
9. E. Martinian and G. W. Wornell, “Authentication with Distortion Constraints,” *Proc. IEEE Int. Conf. on Image Processing*, pp. II.17–20, Rochester, NY, 2002.
10. T. Liu and P. Moulin, “Error exponents for watermarking game with squared-error constraints,” *Proc. Int. Symp. on Info Theory*, Yokohama, Japan, July 2003.
11. D. R. Stinson, *Cryptography: Theory and Practice*, CRC Press, Boca Raton, FL, 1995.