A MATHEMATICAL FOUNDATION FOR WATERMARKING

Nigel Boston

Department of Mathematics and Coordinated Science
Laboratory, University of Illinois, Urbana, IL 61801

0. Introduction.

Watermarks are embedded data that tell us who is the owner of the object watermarked (and maybe provide a time stamp). This provides copyright protection. Physical watermarks have been used for many years, for example in letterheaded note paper. With the current enormous increase in digital communications the need for reliable digital watermarks is paramount. Ideally, the audio or video file should have both fragile and robust marking. Fragile marks are ones that are destroyed as soon as the file is modified too much - these can detect tampering such as if the file has been compressed. Robust marks are ones that it is infeasible to remove without destroying the object at the same time. We focus on robust marks. A historical survey of watermarking (and information-hiding in general) can be found in [1].

Because of the importance of this subject, there has been plenty of activity. This is a major issue in audio, image, and video processing, with standardization efforts for JPEG, MPEG, and Digital Video Disks underway. There are commercial products, international workshops, and special issues of major technical journals. The problem is that so far only ad hoc methods of digital watermarking have been discovered and all of these have flaws. In particular, there is such a huge variety of possible attacks, that produce imperceptible distortions but ones potentially fatal to the mark. This includes JPEG compression, additive Gaussian noise, low pass filtering, rescaling, cropping, rotation, audio restoration, color quantization, shearing, removal of a few lines or columns, horizontal flip, random geometric distortions, GIF conversion, ...

In fact, the paper [2], which attempts to provide a benchmark for judging the success of various schemes, finds that the survival rate for marks, even provided by the important commercial product Digimarc 1.51, is 0% against the freely downloadable StirMark 2.2. Faced with such problems, a group of 180 companies and organizations formed the Secure Digital Music Initiative (see http://www.sdmi.org/). On September 6, 2000, they issued the HackSDMI challenge to hackers to attempt to remove watermarks from some audio files (without significantly changing the audio). Many hackers boycotted this challenge, but some took it on and succeeded. For instance, two students in Leuven successfully removed the marks (see http://www.julienstern.org/sdmi/).

The other problem with the state of the theory is that there is a lack of a solid foundation for it. Contrast this with coding theory and cryptography, which

are well-established. The paper [2], providing a benchmark, is one step in the right direction. Another step was taken by Moulin and O'Sullivan [3], when they introduced an analysis of the information-theoretic limits of watermarking. In particular, they decribed the fundamental notion of a hiding capacity. Just as in coding theory, where one seeks codes whose rates approach Shannon's capacity with probability of error tending to zero (and successfully so, of late, thanks to turbo codes and low density parity check codes), one can seek to approach hiding capacity. This was exploited in [4].

My idea is to develop a theory of watermarking that in some ways parallels coding theory. The two main steps in laying a foundation for coding theory were Shannon's 1948 paper [5] defining capacity and hence a measure of goodness for codes and showing that good codes exist, and Hamming's 1950 paper [6] giving a mathematical description of block codes. Hamming's codes fell well short of Shannon's capacity but they created a new mathematical field, provided a source of practical examples (e.g. BCH codes, Reed-Solomon codes), and allowed us to prove theorems about whole families of codes. Fifty years later, we have a parallel to Shannon's work in [5] - my aim is to parallel Hamming's work by developing a new mathematical field that can form a basis for watermarking theory.

1. Some Definitions.

The idea below is to abstract the essence of watermarking. This naturally is over-simplified and some delicate issues are hidden. For instance, the choice of metric $d(,)$ has to capture what it means for two images to be similar or dissimilar. It seems, however, natural to look first for schemes that defeat simple choices of $d(,)$ (such as Hamming distance) and then develop that knowledge to handle the more sophisticated attacks listed in the introduction.

Definition. A watermarking scheme consists of a metric space V and a finite set W together with maps $f : V \times W \to V$ and $g : V \to W$ such that $gf(v, w) = w$ for all $v \in V, w \in W$.

V is the space of objects to be watermarked - its elements will be called images. W is the set of watermarks - its elements will be called marks. $f(v, w)$ denotes the result of watermarking image v with mark w and g denotes the recovery of the watermark from the marked image. $gf(v, w) = w$ means that in the absence of any attack the correct mark is recovered.

The space V is made into a metric space so that we might define related quantities that give a first measure of how good the scheme is. This is analogous to the competing notions of rate and (relative) minimum distance in algebraic coding theory. Just as those notions are not enough later on (weight distribution, probability of error, etc. being finer invariants), there are refinements of our invariants. For example, if one can integrate over V, then one can put probability functions on V and obtain probabilities of correct recovery of mark.

Definition. The distortion of the scheme is the infimum of the real numbers D

for which $d(v, f(v,w)) \leq D$ for all $v \in V, w \in W$ holds true.

The correctability of the scheme is the supremum of the real numbers C for which $d(v', f(v,w)) \leq C$ imples $g(v') = w$ holds true.

Note that so long as $\#W > 1$, $C \leq D$. This follows since if $D < C$, then combining the above inequalities we get $g(v) = w$ for all $v \in V, w \in W$.

For the most common examples of metric space V, e.g. subspaces of some Euclidean space or finite vector spaces with the Hamming metric, the stronger inequality $C \leq D/2$ holds. This is not universally true, however, since there exist metric spaces in which the ultrametric inequality $d(x,z) \leq \max(d(x,y), d(y,z))$ holds and for these any point in an open disk is its center. One consequence is that if $x, y \in V$ and $d(x,y) = r > 0$, then any open disks around x, y of radius r do not intersect. In practice, however, how close C/D is to $1/2$ will be a good first measure of the scheme's viability. (One wonders then if these ultrametric spaces might somehow be of use in watermarking.)

Some examples of schemes that are already used follow. We translate them into our framework.

(1) Chen-Wornell [7]. Let $V = \mathbf{R}^2$ with its usual metric and $X = \mathbf{Z}^2$. The lattice X can be partitioned into X_0 and X_1 according as $(a,b) \in X_i$ if and only if $a + b \equiv i (\mathrm{mod}\ 2)$. Let $W = \{0, 1\}$.

Define $f(v,w)$ to be the element of X_w nearest to v (ties resolved arbitrarily). Define $g(v)$ by first mapping v to the element x of X nearest to it and then to the unique w such that $x \in X_w$.

(2) Atallah-Wagstaff [8]. Let $Y = X = \mathbf{Z}$ with its usual metric. Pick a large prime p (which will be kept secret for cryptographic purposes). Partition X into X_0 and X_1, where $x \in X_0$ if and only if x is a square modulo p. Let $W = \{0, 1\}$.

As in example 1, define $f(v,w)$ to be the element of X_w nearest to v (ties resolved arbitrarily). Define $g(v)$ by first mapping v to the element x of X nearest to it and then to the unique w such that $x \in X_w$.

The method of picking X first, partitioning it into subsets X_w indexed by $w \in W$, and then defining f and g as in the above examples will be termed the standard method. Watermarking schemes can always be introduced in this way, since starting with our first definition we take X_w to be $\{f(v,w) : v \in V\}$. (Choosing $f(v,w)$ to be an element of X_w other than the nearest to v clearly yields an inferior scheme.) Following this method, there is another very natural kind of scheme which surprisingly does not appear to have found much application in the theory so far, but presumably will in time.

(3) Let V be an N-dimensional vector space over the field with 2 elements with the Hamming metric. Let $Y \subseteq X$ be linear subspaces, which we will think of as linear codes inside V. Let $W = X/Y$ and X_w be the coset w.

One way of constructing X given Y is to pick an element $x \in V$ at a deep hole with respect to Y and set X to be the union of Y and the coset $Y + x$.

As regards computing C and D for the above schemes, C is the radius of the largest circle inscribed in every Voronoi cell [9] of X (considered as a subspace of V), whereas D is the radius of the smallest circle circumscribing every Voronoi cell of every X_w (considered as a subspace of V). So, for instance, in example 1, $C = 1/2$ and $D = 1$. In example 2, computing invariants involves some analytic number theory (what e.g. is the longest chain of successive squares modulo p?). In example 3, D is the covering radius of the code Y. Perfect or close to perfect codes give good schemes.

Pursuing this further, investigating the case of $Y \subseteq V$ being the binary Golay code, we get that if there is a 1% probability of a $0, 1$ switch, then there is a 99.8% probability of a correct watermarking bit being recovered. This will be studied further (and is very suitable for computation, e.g. using the Computer Algebra System MAGMA [10]). We will, however, head in another, also practical direction. First, however, consider some mathematical issues in the theory.

2. Some Mathematics.

Examples 1 and 2 above hide only one bit at a time (in other words $\#W = 2$). They are adapted to handle multiple bits as follows. Suppose $\#W = 2^N$, considered as length-N vectors over $\mathbf{F}_2$. The image to be watermarked is broken up into a sequence $v_1, ..., v_N$ (where $\#W = 2^N$). If the mark is $(w_1, ..., w_N)$, then the watermarked image is simply the sequence $f(v_1, w_1), ..., f(v_n, w_n)$. Mathematically, this is captured by the notion of a product scheme.

Definition. Let $(f_1 : V_1 \times W_1 \to V_1, g_1 : V_1 \to W_1)$ and $(f_2 : V_2 \times W_2 \to V_2, g_2 : V_2 \to W_2)$ be watermarking schemes. The product (watermarking) scheme consists of the metric space $V_1 \times V_2$ and the set $W_1 \times W_2$ together with maps $f : (V_1 \times V_2) \times (W_1 \times W_2) \to V_1 \times V_2$ and $g : V_1 \times V_2 \to W_1 \times W_2$, given by $f(v_1, v_2, w_1, w_2) = (f(v_1, w_1), f(v_2, w_2))$ and $g(v_1, v_2) = (g(v_1), g(v_2))$.

This definition is extended to products of finitely many schemes in an obvious way. If one wishes to get fancy, one can in a Bourbaki sense think of a watermarking scheme as follows. Suppose a W-set is a set X together with a map $X \to W$. A morphism of W-sets is then a map that makes a commutative triangle. Consider $V \times W$ as a W-set via its projection onto the second component and V as a W-set via g. Then a watermarking scheme is simply a morphism f between these W-sets.

Indeed, one can define a category whose objects are watermarking schemes and whose morphisms are as follows. Let $(f_1 : V_1 \times W_1 \to V_1, g_1 : V_1 \to W_1)$ and $(f_2 : V_2 \times W_2 \to V_2, g_2 : V_2 \to W_2)$ be watermarking schemes. A morphism from one to the other is a morphism $\phi : V_1 \to V_2$ of metric spaces (so a map that preserves order between relative distances - these occur most usually in the theory of normed spaces) together with a set map $\psi : W_1 \to W_2$ such that the obvious diagrams commute.

We then have all kinds of tools from mathematics to apply (for instance, we can talk of two schemes being isomorphic, define the automorphism group of a scheme, have group actions on schemes, etc.). Some of these ideas, such as the product scheme above, have physical meaning.

3. A Worked Example.

We wish to inject some mathematical ideas into the current theory. One thing we can do is to try out the Chen-Wornell scheme [7] with other lattices. For instance, if we take $V = \mathbf{R}^N$ and $X = \mathbf{Z}^N$ and partition X into X_0 and X_1 according as $(a_1, ..., a_N) \in X_i$ if and only if $\sum a_j \equiv i (\text{mod } 2)$, then we get a higher dimensional version of Chen-Wornell. If $N \geq 4$, then the point $(1/2, ..., 1/2)$ is further from any point of X_0 than any other point of V (it's a deep hole [9]), and so $D = \sqrt{N}/2$. However, $C = 1/2$, and so $C/D = 1/\sqrt{N}$, which can be much smaller than $1/2$.

Using some of the lattices in SPLAG [9], we can obtain better examples. In particular, the E_8 and Leech lattices have excellent sphere-packing properties and should be investigated. Here we begin investigation of applying the E_8 lattice.

The E_8 lattice consists of all the points $\{(x_1, ..., x_8) : \text{all } x_i \in \mathbf{Z} \text{ or all } x_i \in \mathbf{Z} + 1/2, \sum x_j \equiv 0 (\text{mod } 2)\}$. We let $V = \mathbf{R}^8$ and X_0 be the above lattice. The covering radius of the lattice $D = 1$.

To construct X_1 and hence $X = X_0 \cup X_1$, we throw in a point such as $(1, 0, 0, ..., 0)$. Thus $X_1 = X_0 + (1, 0, 0, ..., 0)$. Then $C = 1/2$ and so $C/D = 1/2$, as desired.

For comparison, if we take the product scheme coming from 4 copies of Chen-Wornell, we obtain $V = \mathbf{R}^8, X = \mathbf{Z}^8$, and X_0 a sublattice of index 16 in X. Then, just as in the first computation in this section, $D = \sqrt{2}$ and $C = 1/2$, so that $C/D = 1/(2\sqrt{2})$, worse than the $1/2$ from the E_8 example above.

One might object,however, and point out that the Chen-Wornell approach hides 4 bits, whereas our E_8 approach only hides 1 bit, but this is easily remedied. In addition to $(1, 0, 0, ..., 0)$, we can throw in other independent points at distance 1 from the origin, e.g. $(1/2, 1/2, ..., 1/2), (1/2, 0, 1/2, 0, ..., 1/2, 0), (3/4, 1/4, 1/4, ..., 1/4)$. The lattice generated by these cosets of E_8 contains E_8 with index 16 but C has not been lowered, since the new lattice has no nonzero vectors of norm < 1. C is still $1/2$, and $C/D = 1/2$.

BIBLIOGRAPHY

[1] Fabien A.P. Petitcolas, Ross J. Anderson, and Markus G. Kuhn, "Information hiding - a survey." Proc. IEEE, special issue on protection of multimedia content, 87 (7): 1062-1078, July 1999.

[2] Fabien A.P. Petitcolas and Ross J. Anderson, "Evaluation of copyright marking systems." Proc. IEEE Multimedia Systems '99, vol. 1, 574-579, 7-11 June 1999, Florence, Italy.

[3] P.Moulin and J.A.O'Sullivan, "Information-theoretic analysis of watermarking." Proc. ICASSP '00, Istanbul, Turkey, June 2000.

[4] M.Kesal, M.K.Mihcak, R.Koetter, and P.Moulin, "Iteratively decodable codes for watermarking applications." Proc. 2nd Int. Symp. on Turbo Codes and Related Topics, Brest, France, Sep. 2000.

[5] C.E.Shannon, "A mathematical theory of communication." Bell System Technical Journal, vol. 27, 379-423, 1948 (Part I), 623-656 (Part II) reprinted in book form with postscript by W.Weaver, University of Illinois Press, Urbana, IL, 1949; Anniversary edition, 1998.

[6] R.W.Hamming, "Error detecting and error correcting codes." Bell System Technical Journal, vol. 29, 147-160, 1950.

[7] B.Chen and G.W.Wornell, "An information-theoretic approach to the design of robust digital watermarking systems." Proc. ICASSP '99, Phoenix, AZ, March 1999.

[8] S.S.Wagstaff and M.J.Atallah, "Watermarking with quadratic residues." Proceedings of the IS&T/SPIE Conference on Security and Watermarking of Multimedia Contents, SPIE–The International Society for Optical Engineering, San Jose, California, January 1999, vol. 3657, 283–288.

[9] J.H.Conway and N.J.A.Sloane, "Sphere Packings, Lattices, and Groups." Third Edition, Springer, 1999.

[10] W.Bosma and J.Cannon, "Handbook of MAGMA functions." Sydney: School of Mathematics and Statistics, University of Sydney, 1993.