SECURE DETECTION OF PUBLIC WATERMARKS WITH FRACTAL DECISION BOUNDARIES

Mohamed F. Mansour and Ahmed H. Tewfik
Department of Electrical and Computer Engineering, University of Minnesota,
Minneapolis, MN 55455
{*mmansour, tewfik*} *@ ece.umn.edu*

ABSTRACT

In this paper we will address the problem of constructing a nonparameteric decision boundary for watermark detection. Most current watermarking algorithms have a parametric decision boundary that can be estimated if the pirate has unlimited access to the detector. In this work we propose a fractal decision boundary which can not be estimated. That boundary is obtained by processing the decision boundary corresponding to the underlying watermarking algorithm. The performance of the new technique is essentially similar to any watermarking algorithm from which it is derived.

1. INTRODUCTION

Watermarking has become an essential tool for multimedia copyright protection. Numerous techniques have been proposed for watermarking audio and image contents [1]. For copyright protection purposes, the watermark should remain in the host media regardless of any reasonable processing that it may undergo. It may be removed only at the expense of excessive image or audio distortion.

However, most algorithms have security pitfalls especially if the detector becomes public and unlimited access to it is possible. For example, Secure Digital Music Initiative (SDMI) organization motivated the wide scale deployment of detectors. Many attacks have been proposed to remove the watermark from the host media or at least render it undetectable. These attacks take advantage of the nonempty overlap between the permissible region for image or audio modification without noticeable distortion and the decision boundary of the detector. The larger the overlap region, the easier the successful attack on the watermarking scheme. Another serious pitfall that is important for our particular purpose is the use of the same secret key for watermarking an unlimited number of copies. This enables the pirate, If she has unlimited access to the detector as a black box, to take advantage of the mutual information between watermarked pieces to estimate the watermark or the decision boundary. This unlimited access is typical in some applications, e.g. the Digital Versatile Disk (DVD) player which in the future will have the capability of identifying the existence of the watermark in the video content to decide whether to play the video or not. This unlimited access is a real challenge to the next generation design of the watermarking algorithms.

Most watermarking techniques for copyright protection solve a binary hypothesis test. They answer a yes/no question about the existence of the watermark in the host media, and hence the watermark itself is immaterial and what is important is to be well structured to optimize the performance of the detector and increase its robustness. All watermarking algorithms take advantage of the inherent redundancy in representing the image or audio to modify the signal values without introducing noticeable distortion. This modification can take place either in the time domain (spatial domain for image) or in the transform domain, e.g. the DFT or the DCT coefficients. In many cases, especially for copyright protection the detection is based on thresholding the correlation coefficient between the test signal and the watermark. This test statistic is optimal if a Gaussian distribution is assumed for the host signal or the transform coefficients, which is not always the case. In [2], the authors proposed a detector scheme based on a Laplacian probability density functions (pdf) of the DCT coefficients, and they show its superiority over the Gaussian representation. However, both approaches result in a hyper plane decision boundary which can be estimated if sufficient number of watermarked samples are available.

In [3], the authors described the asymmetric watermarking technique that is used a test statistic in a quadratic form. Although the decision boundary is more complicated in this case, it is still parametric and can be estimated, for example using the least square techniques with a *finite* number of samples.

In this work we propose an alternative detection technique that has a fractal decision boundary to avoid the mentioned problems. This boundary is nonparametric and cannot be estimated with unlimited access to the detector. The robustness degradation after this detector is minor, and can be tolerated. To our knowledge the only work that addresses this problem is [9]. In that work, a new structure for the detector that employs a nonempty transition band between the two hypothesis. For example for a correlation test statistic, this band is [A, B]. Denote the correlation coefficient by "*y*". If y < A then H_0 (no watermark) is decided, if y > B, then H_1 (watermark exists) is decided, and if y∈[A,B], then the detector choose H_1 with probability $p(y)$, where $p(y)$ is smoothly increasing in y. The estimation of the boundary is increased by orders of magnitude but it is still linear in the signal size.

The paper is organized as follows. In section 2, we give a brief analysis of the pitfalls of the current detectors when the pirate has unlimited access to the detector. In section 3, we propose our new idea of employing a nonparametric decision boundary in the form of fractal curves. We study the possible implementations and modifications to the test statistic. In addition, we discuss briefly the fractal generation for our particular purpose. Finally, in section 4 we give an overview of the results that show that the distortion is essentially similar to the original detector.

2. PITFALLS IN CURRENT ALGORITHMS

2.1 Introduction

In this section we give brief analysis of the current detector schemes. These schemes share a common feature that the decision boundary is *parametric*, i.e. it can be fully specified by a finite set of parameters. The parameters estimation is possible in theory, for example using least square techniques, if sufficient samples (points on the decision boundary) are available. In this section we will review the common detector schemes and describe their security gaps.

We will use the following notations in the remainder of the paper. $\underline{U}$ is the original (non-watermarked) signal, $\underline{W}$ is the watermark signal, $\underline{X}$ is the watermarked signal, and $\underline{R}$ is the test signal. The individual items will be referenced by lower case letters and referenced by the discrete index n, where n is a two-element vector in case of image watermarking, for example samples of the watermark will be denoted by w[n].

The detector of our problem can be formulated as a binary hypothesis test,

$$H_1 : \underline{X} = \underline{U} + \underline{W}$$
$$H_0 : \underline{X} = \underline{U} \qquad (1)$$

The optimal detector depends on the assumed underlying probability density function (pdf). For the exponential family the detector becomes a minimum distance detector, and for the Gaussian distribution it becomes a correlation detector. In the following subsections, we will analyze these detectors.

2.2 Correlation Based Detectors

This detector is the most common and it has been proposed as the optimum detector for the class of additive watermark. The log-likelihood test statistic is reduced after removing the common terms to

$$l(\underline{R}) = \underline{R}^T.\underline{W} = (1/N) \sum_n r[n].w[n] \qquad (2)$$

And in this case the decision boundary is a hyper-plane in R^N and it requires N distinct points to be completely specified. If more points are available a least square can be applied to get the best estimate of the decision boundary.

2.3 Minimum Distance Detector

In [2], the authors proposed a more accurate pdf for the DCT coefficients that has the form of a generalized exponential family:

$$f_x(x) = A \, . \, exp(- \, | \, bx|^c \,) \qquad (3)$$

where A and b can be expressed in terms of c and the standard deviation σ. For example for the Gaussian pdf, c=2, $b = 1/2\sigma^2$, and $A = (2\pi\sigma)^{-1/2}$.The resultant test statistic has the general form:

$$l(\underline{R}) = \sum_n b\,[n]^{c[n]} \, (\, |r[n]|^{c[n]} - | \, r[n]- w[n] \, |^{c[n]} \,) \qquad (4)$$

This is more complicated than the direct correlator in (2). However, the boundary can still be completely specified in terms of {w[n]}. Again least square minimization can be applied to get the best estimate if sufficient samples are available.

2.4 Asymmetric Detector

Asymmetric detectors were suggested to make the problem of estimating the secret key for unauthorized parties more difficult. In [3], four asymmetric techniques were reviewed and an unified form for the detector is introduced. The general form of the test statistic is:

$$l(\underline{R}) = R^T.A.R = \sum_{n,m} a_{n,m} \, r[n].r[m] \qquad (5)$$

The decision boundary in this case will be a two dimensional elliptic surface. This may be more complicated than the correlator case but again it can be estimated using a finite number of samples. However, in this case we will need at least N^2 samples to estimate the boundary rather than N samples. The estimation problem of this boundary is ill-posed because the data matrix will have a structure similar to the *Vandermonde* matrix. The round-off errors in the computations of this matrix can produce excessive numerical errors. However, this problem can be solved either by finding sufficiently large number of points on the curve so that it can be approximated numerically or to use orthogonal discrete polynomial such as Chebeshev polynomials for boundary approximation [6, ch. 22].

2.5 Quantization Based Detector

The same situation occurs when multiple codebooks are used for zero/one embedding. In this case the decision boundary will have the shape of spheres with centers at the centroid of each codebook. Again this shape can be well specified by the centroid and its diameter, and hence the whole partition of the watermark space can be identified. Also the schemes that incorporate even/odd quantization for zero/one embedding can be modeled as a finite set of hyper planes

2.6 Generalized Attack

In the previous subsections we described the shape of the decision boundary in R^N of the most common watermarking schemes. All these boundaries can be specified using least square techniques if sufficient number of points on the boundary are available. This is typical when the pirate has unlimited access to the detector device even as a black box. In this case, she can make *slight* changes to the watermarked signal until reaching a point at which the detector is not able to detect the watermark. At this point, she can go back and forth around the boundary until identifying a point on it with the required precision. Generating a large number of these points is sufficient to estimate the decision boundary in the above cases. However, the evaluation of the coefficients of a polynomial is in general a difficult problem from the numerical standpoint, but one cannot rely on this for the security of the detector. Specifically, the boundary shape can be equivalently evaluated using methods described in subsection 2.4.

Once the boundary is specified any watermarked signal can be projected to the nearest point on the boundary to render the watermark undetected with the smallest possible distortion. The illustration of this idea is shown in figure 1.

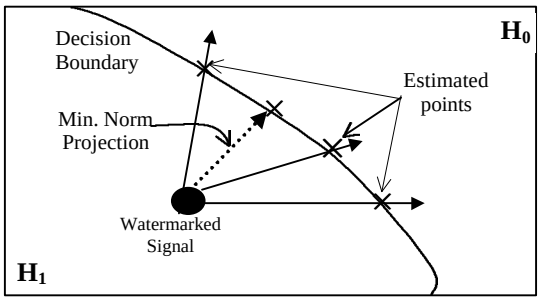


Figure 1. Illustration of the generalized attack

It should be mentioned that the watermark needs not to be extracted prior to removal. It can be completely be undetected

without estimating it. However, for the correlator in (2), estimating the decision boundary is equivalent to estimating the embedded watermark.

This problem motivated this work to search for decision boundaries that cannot be parameterized. In this case even if the pirate can change individual watermarked signals to make the watermark undetected, the modification will be random and the minimum distortion modification cannot be found as before. The only choice for the attacker is to try to approximate it numerically and this will require extra cost by several orders of magnitude.

3. PROPOSED ALGORITHM

3.1 Algorithm

Rather than formulating the test statistic in a functional form, it will be described by a nonparameterized function. We select *Peano* curves to represent this decision boundary. So instead of the test statistic of the forms (2), (4), and (5) we employ a fractal test statistic whose argument is $\underline{R}$ and has the general form

$$f(\underline{R}) > threshold \qquad (6)$$

where *f(.)* is a random walk or a fractal function.

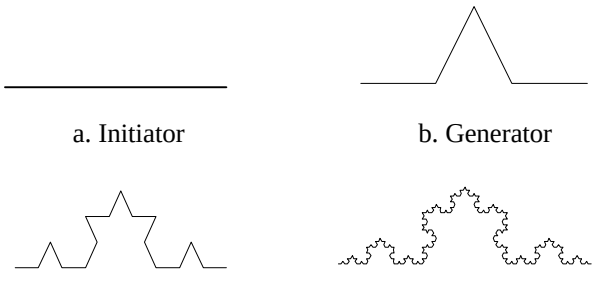
The basic steps of the proposed algorithms are:

1. Start with a given watermarking algorithm and a given test statistic *f(x)*, e.g. a correlation sum (2), which has a general form f(x) > c
2. Fractalize the boundary using the fractal generation technique that will be discussed in subsection 3.2.
3. Use the same decision inequality but with a new test statistic using the new fractal boundary.
4. Modify the watermarked signal if necessary to assure the same distance from the boundary after modification.

It should be noted that other nonparametric curves can be used as well. The fractal curves are used because of its relatively straightforward generation. This is important because the boundary should by stored at the decoder with high precision. So instead the procedure for generating the curve can be used.

3.2 Fractal Generation

This class of curves, called *Peano* curves [5], has been well studied for representing self-similar structures. Very complicated curves can be constructed using simple repeated structures. The generation of these curves is done by repetitive replacing of each straight line by the generator shape. So each shape can be completely constructed by the initiator (figure 2a), and the generator (figure 2b). Many combinations of the generator and the initiator can be employed to generate different curves.



a. Initiator b. Generator

c. After 2 iterations d. After 4 iterations

Figure 2. Peano Curve Example

This family of curves can be used to represent statistically self-similar random processes, e.g. the brownian random walk [7, ch. 11], and some fractal curves can approximate a random walk process.

3.3 Modifying the decision boundary

The most important step in the algorithm is modifying the decision boundary to have the desired fractal shape. In figure (3), we give an example with a decision boundary of the correlator in R^2. The original boundary is a line, and the modified decision boundary is as shown in the figure.

There is a tradeoff in designing the decision boundary. The maximum difference between the old decision and the modified one should be large enough so that the new boundary cannot be approximated by the old one. On the other hand it should not be very large to avoid excessive distortion.

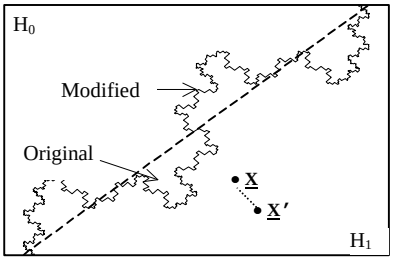


Figure 3. Example of modifying the decision boundary

After modifying the decision boundary, the watermarked signal $\underline{X}$ *may* need some modification to sustain the same shortest distance from $\underline{X}$ to the decision boundary. This is done by measuring the shortest distance between $\underline{X}$ and the new boundary, and moving $\underline{X}$ along the direction of the shortest distance in the opposite direction (from $\underline{X}$ to $\underline{X}'$). However this modification is not critical for the performance especially if the variance of the test statistic is small or if the distance between $\underline{X}$ and the original boundary is large compared to the maximum oscillation of the fractal curve.

Instead of evaluating the log likelihood function of the forms (2), (4), and (5), the chaotic curve is evaluated for the received signal and is compared with the same threshold to detect the existence of the watermark.

3.4 Practical Implementation

Instead of applying multidimensional fractalization, a simplified practical implementation that achieves the same purpose is discussed in this section.

If the Gaussian assumption is adopted, then the test statistic in (2) is optimal according to the Neyman-Pearson theorem [8,ch. 3]. Instead of this test statistic, two test statistics are used for the even and odd indexed random subsequences $r[2k]$ and $r[2k+1]$:

$$T_1(\underline{R}) = (2/N).\ \Sigma_k\, r[2k].w[2k]\ ,$$
$$T_2(\underline{R}) = (2/N).\ \Sigma_k\, r[2k+1].w[2k+1]$$
$$\underline{T} = (T_1\ \ T_2) \qquad (7)$$

Under H_0, $E(\underline{T}) = (0,0)$, and under H_1, $E(\underline{T}) = (1,1)$ and in both cases $cov(\underline{T}) = s^2 I$. The Gaussian assumption of both T_1 and T_2 is reasonable if N is large by invoking the central limit theorem. Also if the original samples are mutually independent, then T_1 and T_2 are also independent. The decision boundary in this case is a line (with slope –1 for the given means). If this line is fractalized as discussed in the previous subsection, then the

corresponding decision boundary in the multidimensional space will be also nonparametric.

The detection process is straightforward in principle but nontrivial. The vector $\underline{T}$ is classified to either hypothesis if it falls in the corresponding partition. However, due to the nonparameteric characteristic of the boundary this classification is not trivial. First the unambiguous region is defined as shown in figure 4, that is outside the oscillation of the fractal curve, which are the regions outside the doted lines. For points in the ambiguous area, we extend two lines between the point and both centroids. If one of them does not intersect with the boundary curve, then it is classified to the corresponding hypothesis. Here it should be emphasized that the boundary curve is stored at the detector and it should be kept secret.

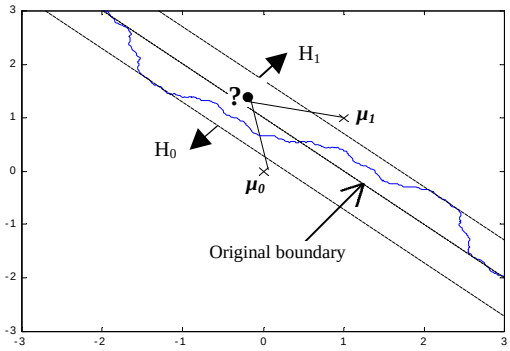


Figure 4. Detector operation

This detector may look similar to the one proposed in [9] using a different approach. However, the fundamental difference is that in our case there is no uncertainty in the decision, i.e. it is a pure deterministic operation. Also the attacker in [9] needs a linear number of iterations (with the signal size) to estimate the decision criterion. For our detector the boundary is non-differentiable anywhere (at least theoretically), and hence estimating it using tangents as described in [9] will not work, and the estimation will not be linear with the signal size.

4 RESULTS

The technique proposed in this paper is quite general, and can be applied to any watermarking scheme without changing the embedding algorithm. The algorithm performance is in general very similar to the underlying watermarking algorithm with its optimal detector. However, for some watermarked samples, we may need to increase the watermark strength as illustrated in figure 3.

The Receiver Operating Characteristic (ROC) of the proposed algorithm is close to the ROC of the optimal detector, and it depends on the maximum *oscillation* of the fractal boundary around the original one. In figure 5, we illustrate the performance for the system discussed in section 3.4, when the mean under H_0 is (0,0) and under H_1 is (1,1), and the variance in both cases is 1. As noticed from the figure the performance of the system is essentially the same as the optimal performance especially for small curve oscillation.

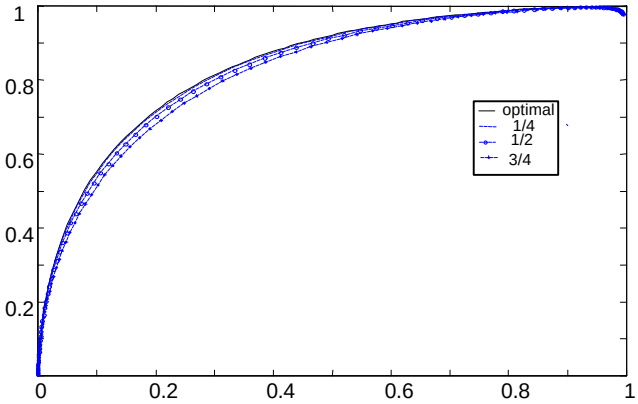


Figure 5. ROC of the proposed algorithm

5 CONCLUSION

In this work, we proposed a new class of watermark detectors based on a nonparametric decision boundary. This detector is more secure than traditional detectors especially when it is publicly available. The performance of the new algorithm is similar to the traditional ones, but it has the advantage of filling the security gap of the detector.

The proposed detector can work with any watermarking schemes without changing the embedder structure. However, sometimes the strength of the watermark should be increased slightly to compensate for the new boundary.

Future work includes applying the proposed algorithm to common audio and image watermarking and studying the optimal structure of the modified decision boundary to optimize the detector performance.

6 REFRENCES

1. Langelaar G., Setyawan I., and Lagendijk R., "*Watermarking digital image and video Data"* IEEE Signal processing magazine, Vol. 17, No. 5, pp. 20-46, September 2000.
2. Hernandez, J.R.; Amado, M.; Perez-Gonzalez, F, "*DCT-domain watermarking techniques for still images: detector performance analysis and a new structure*", IEEE Transaction on Image processing, pp. 55-68, January 2000.
3. Furon T., Venturini I., and Duhamel P., "*An unified approach of asymmetric watermarking schemes*", Proc. SPIE conference on security and watermarking multimedia contents, pp. 269-279, January 2001.
4. Press. W. et. al., "*Numerical Recipes in C* ", Cambridge University Press, 1992.
5. Mandelbrot B., "*The Fractal Geometry of Nature*", W.H. Freeman and company, 1983
6. Abramowitz M., and Stegun A., "*Handbook of mathematical functions*", Dover Books on Advanced mathematics, 1972.
7. Papoulis A., "*Probability, Random Variables, and Stochastic Processes*", 3rd edition, McGraw Hill, 1991.
8. Kay S., "Fundamentals of statistical signal processing: Detection Theory", Prentice Hall, 1998.
9. Linnartz J., and Dijk M., "Analysis of the sensitivity attack against electronic watermarks in images", Proc. 2nd International Workshop on Information Hiding, pp.258-272, 1998.