Digital Watermarking for DVD Video Copy Protection

What Issues Play a Role in Designing an Effective System?

Maurice Maes, Ton Kalker, Jean-Paul Linnartz, Joop Talstra, Geert Depovere, and Jaap Haitsma

One of the most quoted applications for digital watermarking is in the context of copy-protection of digital (multi-)media. In its most basic form a watermark is used as a single bit that indicates whether or not content is copyrighted. By the nature of a watermark, this copy-protection bit is (theoretically) strongly tied to the content: it can only be removed when the content suffers serious degradation. As simple as this idea may seem, the actual design of an appropriate watermarking scheme and its application in a copy protection system is quite complicated.

In this article we illustrate the various issues that play a role in designing a copy-protection system for digital versatile disk (DVD) video as perceived by Millennium, one of the two contenders in the DVD-video copy protection standardization activity. We present the Millennium watermark system, the systems proposed for DVD video copy protection by Philips, Macrovision, and Digimarc. We also address some specific system aspects, such as watermark detector location and copy generation control.

AU: Please add subhead here

Digital multimedia technology paves the way for new applications, features, and services. The transition from analog to digital, however, has been seriously affected by a slow release of content. Film and music content owners are afraid to lose revenues as digital content, if unprotected, can be copied rapidly, perfectly, at large scale, and without limitations on the number of copies. Copy control issues have come on the "critical time path" of the market introduction of several digital products, including DVD video [1], [2], the IEEE 1394 (firewire) digital interface [3], digital broadcasting, and improved digital audio carriers such as super audio-CD, DVD-audio, and secure solid-state audio carriers [4]. The standardization of DVD video has unleashed an unprecedented debate over copy protection, which has influenced the entire digital multimedia landscape. Recent security breaches of the DVD-video encryption proved once again that encryption with essentially fewer than 40 key bits, relying on the secrecy of the algorithms, does not satisfy content owners requirements. This is particularly true in a situation with dozens of manufacturers, each employing hundreds of designers, to say nothing of the thousand-plus hackers intent on breaking the system just for kicks. Meanwhile, improved cryptographic protection and additional techniques including watermarking are being considered.

Given the current status, this article cannot describe a fully defined and mature system. It must be regarded as a status report from an ongoing discussion. We present our findings from active participation in several forums. Portions of this article have previously appeared as white papers or responses to Calls for Proposals, e.g., [5]-[9]. We strongly believe in open, publicly evaluated systems and solutions which have been discussed not only in industrial standardization meetings but also at academic symposia. Some technologies, such as encryption on DVD video discs have been standardized in the Copy Protection Technical Working Group for DVD. Other technologies described in this paper are currently under discussion or are most likely to become topic of discussion any time

soon. Although the underlying technologies are mostly well understood, it appeared less trivial to standardize a complete copy protection system.

Encryption Alone Does Not Suffice

Copy management can not easily be formalized into "Alice and Bob" protocols, as commonly studied for other fields of security and cryptography [13]. In fact, Alice, in our case the content owner, intends to sell information to an unreliable customer Bob, without allowing Bob to further disseminate this information. Evidently there is no cryptographic or information theoretical solution to this problem. Nonetheless, international standardization efforts have recognized that a workable way to redefine the problem is as follows: Alice sells digital data to an unreliable Bob, who can only process this data on a trusted device. The protection relies on Bob's inability to access the data directly.

Protection by encryption leads to the notion of a compliant world of consumer devices which communicate over authenticated and encrypted digital links, using frequently updated session keys. A device is compliant when its manufacturer has agreed to follow specific copy protection rules described in a licensing agreement, in return for knowledge of cryptographic keys to get access to protected digital content. Hence, noncompliant devices never get access to the digital content in the clear. Without claiming to be exhaustive in our summary, important consequences of this approach are:

▲ Protected digital content must be encrypted on any "open interface." This includes digital interconnects (e.g., IEEE 1394, USB), over the air broadcast, PCMCIA connectors, and internal PC busses. The licensing agreement prohibits the use of "insecure" interfaces.

▲ Encryption as such is not sufficient. An attacker can copy data, which compliant devices inherently would understand during playback. Thus

▲ An authentication and session-key generation mechanism is needed for all interfaces.

▲ Content on a digital storage medium, such as a recordable CD or DVD disc, needs protection against the bit-by-bit copying of encrypted data. One way of dealing with this is through binding the content to the storage medium using unique media (disc) identifiers, which, by definition, may not be changeable by hackers.

▲ Internally, playback and copy devices often need to interpret and process data. Examples are navigation through content (fast forward), reformatting for display, and conversion for storage and transmission. Therefore, end-to-end encryption, though favored from a security point of view, is less workable in a consumer environment. This is a partial explanation of why for the moment link-by-link encryption seems to be preferred for the IEEE 1394 firewire.

▲ Content eventually needs to be presented in decrypted form (or "in-the-clear") to the human consumer (at least, until humans have decryption electronics implanted at birth). While digital protection can be extended all the way to digital monitors and speakers, eventually an analog signal, vulnerable to (noncompliant) copying must be created to allow consumption. Additional protection is needed to prevent that this analog signal can return to the compliant world, i.e., successfully offered to a (compliant) player or recorder as if it were the user's personal creation. This is where watermarking enters the world of copy protection. Note that in general the noncompliant world has to be considered as a lost case for copy protection.

Digital Watermarking

Digital video watermarking is a technique that is used to prevent copy-protected video content from re-entering the compliant world after having been copied or transmitted by noncompliant devices. By imperceptibly hiding information into the video content it will be possible to prevent copying or playback of such content.

The basic requirements on the watermarking method include that the watermark is *invisible* and *difficult to remove*. Detection of the watermark should be *fast* (within the declared DVD detection interval of 10 seconds) and *cheap*, i.e., requiring only limited additional hardware in players/recorders. Detection should also be *robust* with respect to common image processing, or transformations applied to video such as compression, noise addition, logo insertion, shifts, format conversions, etc. Furthermore, there is a requirement on the *payload* of the watermark to be equal to or larger than 8 bits per detection interval. Another important requirement is that the probability of a *false alarm* (the situation where a watermark is detected while there was no watermark embedded) is extremely small (less than 10^{-12} per detection).

The above requirements are mutually conflicting, and in the design of a watermarking system, compromises and tradeoffs have to be made. In the remainder of this section, we will sketch the Millennium watermark system and illustrate in what way the basic requirements have been met.

Basic Video Watermarking Philosophy

The basic premise at the start of the development was to design a watermarking system, which was at the same time simple and satisfied all the requirements with respect to perceptual quality and robustness.

Several issues had to be addressed. First we had to decide upon the basic format in which the watermark was to be detected. It seemed inevitable that the watermarking scheme should at least be able to detect in base-band domain. A consequence of this decision is that, without special tricks, watermark detection on digital MPEG video needs at least a partial MPEG decoder.

Second, we had to decide in which representation to detect the watermark. Browsing through the literature,

one finds basically three kinds of approaches. In the simplest approach no transformation is performed, and the watermark is directly detected in the base-band video using some correlation-like method. At the other end of the spectrum, watermarks are embedded and detected in some type of frequency domain. Embedding and detection is therefore preceded by a frequency domain transform. Well-known transforms are: the Fourier transform (FT), the discrete cosine transform (DCT) and the wavelet transform (WT). Again by using some correlation-like method, the watermark is then detected in the transform domain. Although these latter approaches tend to yield very reliable watermark detection, we decided not to pursue this direction due to the complexity of the global transform, very likely prohibiting real-time detection. The third approach addresses this complexity issue by performing frequency transforms on a block-by-block basis. The problem with such an approach is its vulnerability to spatial image shifts, a very common and cheap processing step. Spatial shifts cause a misalignment of block-boundaries and therefore a failure to detect the watermark. Based upon this analysis we decided to rely on the first approach, i.e., simple spatial correlation. Representing a pixel (luminance) value at position i (both spatial and temporal) by the symbol y_i and the correlation pattern by w_i, watermark detection can succinctly be described by the formula

$$d = \frac{1}{N}\sum_i y_i w_i,$$

where N is the number of pixels involved in the correlation. The system is designed such that a large value of d indicates the presence of the watermark $W = \{w_i\}$ and a small value indicates the absence of the watermark. In this manner it is possible to embed a one-bit payload. Note that watermark detection is not performed on chrominance values, as the system is required to be robust to gray-scale conversions. For the remainder of this we will therefore ignore any chrominance data and assume that all content is gray-scale only.

Third, we had to decide upon the exploitation of the temporal axis. For reasons of complexity we decided upon the use of a purely spatial watermark pattern W and to embed W repeatedly in every frame of the video. This choice amounts to treating video as a sequence of still images. Watermark detection can now succinctly be described by

$$d = \frac{1}{NT}\sum_i \left(\sum_t y_{t,i} \right) w_i$$

where t and i denote the temporal and spatial position of a pixel, respectively. The symbols N and T denote the number of pixels in a single video image (note the difference with the previous interpretation of N) and the number of video frames, respectively. By first accumulating in time, the complexity of watermark detection is reduced by decreasing the number of multiplications. The following section goes into some more details of watermark embedding.

Basic Watermark Embedding

Previously we concluded that, for watermarking purposes, video is best considered as a sequence of stills. Embedding the same watermark in a number of consecutive frames then constitutes a mark for a video sequence. By changing the watermark pattern at a low rate, we can also realize payload along the temporal axis, but for the current discussion this is of no relevance. We therefore focus on watermark embedding in a single video frame.

Given our preferred watermark detection scheme, viz. correlation with a watermark pattern $W = \{w_i\}$, a well-known and efficient embedding scheme consists of adding a scaled version of W to an original image $X = \{x_i\}$. That is, a watermarked image $Y = \{y_i\}$ is obtained

$$y_i = x_i + s w_i$$

where s is a global scaling parameter. In other words, a watermark is simply additive noise. The samples of the watermark pattern W are independently drawn from a normal distribution $N(0,1)$ with mean and standard deviation equal to 0 and 1, respectively. In particular, the sample values of W are multibit values.

At this point we need to remark that the embedding formula needs to be modified slightly to include rounding and clipping to reflect the fact that luminance can only be integer valued (typically between 16 and 235)

$$y_i = \text{RoundAndClip}(x_i + s w_i).$$

As the sample values of the watermark pattern are independently drawn, it follows that the watermark pattern W is spectrally white. It is *a priori* not clear that "white" is the optimal choice for the spectral color of watermark patterns. On the one hand, as (natural) images tend to be highly correlated, one might argue for using a correlated watermark pattern. Such a pattern can be obtained, for example, by low-pass filtering a spectrally white pattern. Experimentally we have found that low-pass watermarks are indeed more robust than white watermarks, but also that it is difficult to avoid visual artifacts. On the other hand, as the human visual system (HVS) is less sensitive to high frequency patterns than it is to low frequency patterns, one could also argue for using a high-pass watermark. A drawback of such an approach is that watermark detection is less robust. After weighing the pros and cons of the low- and high-pass approaches, we decided to compromise upon a spectrally white watermark pattern.

If watermark embedding is performed directly as described, one easily finds that artifacts appear in image regions where there is little activity, e.g., in regions with

little texture. A solution to this problem is the incorporation of a local scaling factor $\Lambda = \{\lambda_i\}$,

$$y_i = \text{RoundAndClip}(x_i + s\lambda_i w_i).$$

The value of λ_i should be small in image regions where there is little activity (e.g., flat regions in cartoons) and large in regions where there is much activity (e.g., in textured regions or at edges). A satisfactory local scaling factor is obtained by filtering the image with a Laplacian high-pass filter L and taking absolute values, i.e.,

$$\Lambda = |L \otimes X|,$$

where "⊗" denotes convolution, and where is L is defined by

$$L = \begin{bmatrix} -1 & -1 & -1 \\ -1 & 8 & -1 \\ -1 & -1 & -1 \end{bmatrix}.$$

Basic Watermark Detection

We recall above that watermark detection is performed by spatial correlation. If watermark embedding is performed as in the previous section, we can write

$$\Upsilon = X + s \times \Lambda \times W,$$

where X is the original image, s is the global scaling parameter, Λ the local activity measure, and W the watermark pattern. Performing watermark detection by correlation, the resulting decision d consists of two terms

$$d = d_{\text{org}} + d_{\text{wmk}} = \frac{1}{N}\sum_i x_i w_i + \frac{1}{N}\sum_i \lambda_i w_i^2 .$$

It is not difficult to show that the expected value $E[d_{\text{org}}]$ contributed by the original unmarked image is equal to zero, and that under very general conditions (we assume normalization to zero mean before correlation) the standard deviation of d_{org} is given by

$$d_{\text{org}} = \frac{\sigma_X}{\sqrt{N}},$$

where σ_X is the standard deviation of the original image. The contribution of the watermark is given by

$$d_{\text{wmk}} = s\mu_1(\Lambda),$$

where $\mu_1(\Lambda)$ denotes the first moment (or mean) of the local activity. It follows that for a given false positive rate (the probability that a watermark is being detected without one having been embedded) ρ and associated threshold $T_\rho = \text{erfc}^{-1}(\rho)$ the value of s should be larger than

$$s \geq \frac{\sigma_X T_\rho}{\mu_1(\Lambda)\sqrt{N}}.$$

In practice s has to be chosen considerably larger in order for the watermark to survive common video processing.

Shift Invariance

In the foregoing we have assumed that during detection the watermark and the image are perfectly aligned. In practice we cannot rely on this. During normal processing the position of the image may easily vary a little. Moreover, to circumvent watermark detection, a malevolent hacker can easily and cheaply induce spatial shifts, even on a frame-by-frame basis. It is, therefore, strictly required that the watermark system be resistant to spatial shifts. The simplest approach to achieve this invariance is exhaustive search for the correct alignment of the watermark. That is, for each allowed spatial shift k the decision variable d_k,

$$d_k = \frac{1}{N}\sum_i y_i w_{i-k}$$

has to be computed (note: i and k are *vectors*). For ease of presentation, we have neglected boundary problems in this formula.

This search over all possible spatial shifts is computationally prohibitive if we aim for real-time watermark detection. The solution adopted in the Millennium system is to introduce translational symmetry in the watermark pattern W. The particular choice made here requires that

$$w_{i+k} = w_i$$

for every vector k whose components are multiples of M, where M is referred to as the tile size. A practical choice for the value of M is $M = 128$, which is what we have settled on for the Millennium system. In other words, the watermark pattern $\{w_i\}$ is completely determined by an $M \times M$ matrix $\{\underline{w}_i\}$ of (pseudo) random values. The full watermark pattern $\{w_i\}$ is obtained by *tiling* (possibly with truncation) the matrix $\{\underline{w}_i\}$ over the extent of the image.

With these assumptions the exhaustive search over all possible shifts is greatly simplified. As the watermark is repeated over vectors which are multiples of M, one can first fold the suspect image data Υ to a matrix $B = \{b_i\}$ of size $M \times M$

$$b_i = \text{fold}(\Upsilon)_i = \sum_{j=(j_1 M, j_2 M)} y_{i+j},$$
$$i \in \{(i_1, i_2) \,|\, i_1 = 0,\ldots,M-1, i_2 = 0,\ldots,M-1\}.$$

Due to the folding (and neglecting boundary problems), we now only need to search over *cyclic shifts* k. More mathematically this is expressed as

$$d_k = \frac{1}{M^2}\sum_i b_i \underline{w}_{i-k},$$

where the subtraction in the index $i-k$ of $\underline{w}$ is computed modulo M.

In fact, it is easy to see that we have to compute a two-dimensional cyclic convolution. Letting $\underline{w}_i^*$ denote the spatial inversion of $\underline{w}_i$, i.e., $\underline{w}_i^* = \underline{w}_{-i}$, we can write

$$D = B \otimes \underline{W}^*,$$

where, by abuse of notation, "⊗" now denotes cyclic convolution. It is well known that a cyclic convolution is most efficiently computed in the frequency domain. The computation of the matrix $\boldsymbol{D}$ then proceeds as follows.

▲ 1. Pre-compute, using a fast Fourier transform (FFT), the Fourier transform $\underline{W}$ of the matrix $\underline{W}$.

▲ 2. Compute the Fourier transform $\boldsymbol{B} = \mathrm{FFT}(B)$ of the fold buffer B.

▲ 3. Perform a point-wise multiplication of $\boldsymbol{B}$ and $\underline{\boldsymbol{W}}^*$ to obtain the matrix $\boldsymbol{D}$. Note that in this context the superscript "*-operator" denotes complex conjugation.

▲ 4. Compute D by applying the inverse FFT (IFFT) to $\boldsymbol{D}$.

More concisely,

$$D = \mathrm{IFFT}\left(\mathrm{FFT}(B) \times \mathrm{FFT}(\underline{W})^*\right).$$

The performance can be improved by preceding correlation by *matched filtering*. The goal of matched filtering is to decorrelate the suspect image Υ to obtain an approximately spectrally white version of Υ. Matched filtering is usually performed in the spatial domain (using some simple and cheap decorrelation filter), but can in our current setup also be computed in the Fourier domain. Moreover, we need not be satisfied with an approximately white signal. By only retaining the phases of $\boldsymbol{B}$ we obtain a purely white signal. Experimentally we have found that the best detection is obtained by also ignoring the magnitude information in $\underline{\boldsymbol{W}}$, resulting in the detection formula

$$D = \mathrm{IFFT}\left(\mathrm{phase}\left(\mathrm{FFT}(B)\right) \times \mathrm{phase}\left(\mathrm{FFT}(\underline{W})\right)^*\right),$$

where $\mathrm{phase}(\alpha) = \alpha / |\alpha|$, for complex α. This method of detection is actually well known in the field of pattern recognition and is referred to as symmetrical phase only filtering (SPOMF) [14].

Increasing the Payload

Previously we mentioned that SPOMF detection is an excellent method to detect the presence/absence or sign of a watermark, whether the watermark is shifted or not. This allows us to embed a one-bit payload. For the DVD application at hand, this one-bit payload is not sufficient. The payload can be increased along essentially two-non-exclusive axes: the spatial axis and the temporal axis, respectively. More precisely we can increase the payload spatially by using more basic patterns and temporally by varying the absence/presence or sign of these patterns over time. Disregarding the temporal domain, a payload of n bits can be obtained by using n basic patterns W_l of size $M \times M, l = 0,\ldots,n-1$. Each pattern will then correspond to one bit.

There are two disadvantages to the above approach. First, the energy of the total embedded watermark is linear in the number of bits of the payload. This may be a cause for visible artifacts. Second, n SPOMF detections are required to retrieve an n-bit payload. For complexity reasons this is not a favored solution. Experimentally three, or maybe four, is found to be the maximum number of basic patterns. An outcome to this dilemma was the insight that the inherent shift invariance of the Millennium system could be used to increase to payload to n bits with less than n basic patterns!

Because of the shift invariance, an embedded pattern W will be found whatever its position in the image is. The same is true if this pattern is embedded several, say m, times, but at different positions. Performing detection by SPOMF, all m copies of the watermark will be found. If the whole image is shifted before detection, the absolute positions of the correlation values will change cyclically. The relative positions *will remain unchanged*, however, at least if computed with modular arithmetic to the base M. We can therefore embed information in the relative position of the correlation peaks. This basic idea needs some refinement to really make it work.

Let us consider the example of one pattern which is embedded twice and that the tile size M is equal to 128. Moreover, let's assume that the pattern is embedded at the origin (0,0) and at position (8,8). Upon detection two correlation peaks will be detected (see also Fig. 1). We don't know, however, which of the two peaks corresponds to the pattern embedded at the origin. Therefore we can only determine the relative position of the two peaks up to a sign, i.e., we cannot distinguish between (−8,−8) and (8,8). A simple calculation shows that therefore we can only distinguish between $M^2/2+1 = 8193$ different relative positions. This amounts to a little more than 13 bits for a single SPOMF detection. In practice this payload cannot be achieved because we have to exclude a few interference-sensitive constellations (when the peaks are located too close to each other).

To boost the false-positive reliability we can exploit two degrees of freedom very cheaply:

▲ 1. We only allow peak constellations where the relative position of the peaks is a multiple of the so-called *grid size* G. To be able to embed at least 8 bits, we settled on $G = 4$.

▲ 2. We note that we are at liberty to embed a watermark with either a positive or negative sign. This sign is correctly retrieved by SPOMF detection. Clearly the sign is shift invariant and can therefore also be used as part of the information carrier. Continuing our example, we now

embed the pattern at the origin with a positive sign and the shifted pattern with a sign, depending on its position on the grid. In this setup, the SPOMF detector is able to distinguish between the peak corresponding to the pattern at the origin and the peak corresponding to the shifted pattern.

False Positive Analysis

We recall that the retrieval of watermark payload is in essence achieved by looking for large positive or negative peaks in the correlation buffer D. We are interested in the rate of false positives. There are actually two types of false positives. A true false positive occurs when a watermark is detected when no watermark has been embedded. An invalid positive occurs in case a watermark has been embedded but the wrong payload is retrieved. Both types of false positives are highly undesirable because they may lead to "unhappy" customers and will therefore have serious implications both for the manufacturing industry as well as the content providers.

An intermediate result in a watermark detection event is an $M \times M$ buffer D of correlation values. The payload of the watermark is determined by the (relative) positions of a number of extreme values in that buffer. The key insight is now that the nonextreme values can be considered as watermark detections for nonwatermarked images: by correlating the watermark with the image at nonembedding positions, the image appears to the watermark as an original, nonmarked image. Experiments have confirmed that these nonextreme values are normally distributed. In fact one can prove that for SPOMF correlation, under very general conditions, the mean and the standard deviation are 0 and $1/M$, respectively. By setting the threshold for peak detection at $6/M$ we achieve a probability for invalid peak detection of $P = 2.0 \times 10^{-9}$.

Using the DVD setting with one pattern of multiplicity two, a grid size G equal to four, the false positive rate for unmarked images can be computed. A false positive occurs if there are precisely two extreme values in the correlation buffer at positions, which differ by a multiple of four. The relative sign of the peaks also has to be compatible with their relative position. It is not difficult to derive that the false positive rate Q_0 is (approximately) given by

$$Q_0 \approx \frac{1}{4G^2} M^4 P^2 \approx 1.6 \times 10^{-11}.$$

The factor 1/4 comes from the fact that a) the signs of the peaks have to agree, and b) we reject half of the relative peak-positions as illegal (peaks too close together). As we are dealing with video we accumulate several (say T_1) of these *micro-decisions*. The probability that more than T_2, $0 < T_2 \le T_1$, of these micro-decisions yield the same result (i.e., not just a valid payload, but all payloads the same) is given by the formula

$$Q_1 \approx \sum_{T_2 \le T_3 \le T_1} \frac{T_1!}{(T_1 - T_3)! \cdot T_3!} \times Q_0^{T_3} \times 2^{-8(T_3 - 1)}.$$

The factor $2^{-8(T_3-1)}$ comes from enforcing the 8-bit payload of all valid micro-decisions to be the same. For DVD copy protection, the choices $T_1 = 10$ and $T_2 = 2$ have been made with false positive probability 4.7×10^{-23}. For all practical purposes this false probability rate is more than sufficient. But if necessary, it can be reduced even more by choosing appropriate values for T_1 and T_2.

A similar reasoning can be applied to estimate the probability of invalid positives. In fact, it is not difficult to see that the similar reasoning as above can be made.

Note that the previous reasoning assumes that micro-decisions are independent events. Experiments have confirmed that for most video scenes this is true. For certain scenes, such as extremely long stills, this assumption might not hold true.

Complexity Analysis

Figure 2 gives an overview of the watermark embedding procedure. Given a payload K, a pattern W is computed from a fixed and universal basic pattern W_0. The pattern W is then tiled over the extent of a video frame and locally scaled by means of the local activity measure. After globally scaling with the parameter s, the result is added to the video frame. Finally rounding and clipping obtain a watermarked video frame. The payload K needs to be kept constant for a sufficient number of video frames to allow reliable detection. By changing the payload at a sufficiently low rate (as not to violate the constraint of the previous sentence), payload can be embedded along the temporal axis.

The most complex operation for watermark embedding is the computation of the local activity measure. The computational complexity per pixel is quite low, but the computations have to be performed at video rate. To show feasibility, a real-time watermark embedder has been implemented both on a TriMedia and FPGA platform. This shows that with "modest" means the complexity of embedding can be surmounted.

Figure 3 gives an overview of the watermark detection procedure. Detection starts with the accumulation of sufficiently many video frames. The frames are folded, summed and stored in an $M \times M$ buffer B. When a sufficient amount of data has been accumulated, SPOMF correlation with the basic pattern W_0 is applied. Note that SPOMF computations can be done in place. The resulting correlation buffer is examined for extreme values, and, if present, a payload K is returned. A nice feature of Millennium detection is that the computations at video rate are simple (mainly additions) and that the complex computations (SPOMF) operate at a much lower rate.

In DVD the memory resources for watermark retrieval are restricted, then several ways exist to reduce the amount of memory.

▲ 1. One can compute the phase-only representations of the basic pattern W_0 on the fly. This will increase the computational logic, but remove the need for a large and insecure ROM buffer.
▲ 2. It is possible to stop taking in video after accumulation and do all subsequent processing in place. After payload retrieval, another round of accumulation and folding can start. If memory resources are not scarce, two buffers in ping-pong mode can be used: at all times one buffer is being used for accumulation and folding and the other is being processed for payload retrieval. After each payload retrieval, the buffers change function. In the one buffer case, the payload retrieval has to be as fast as possible to reduce the amount of missed data. In the two-buffer case, the payload retrieval process may take as long as the accumulation process. It is without saying that in the last case the computing requirements are less severe than the requirements in the first case.

The computational power, which is needed for the FFT, is quite modest. For example, the TriMedia signal processor can calculate the FFT in 15 ms. For phase extraction, 128×128 divisions by magnitude are required. The division and the square-root function (which is needed for calculating the magnitude) are costly in terms of processing power. In software these calculations take up many machine cycles, and in hardware they will occupy quite a large area of silicon. We have implemented adequate approximations of the square-root and division functions that do not suffer from this drawback. For the convolution 128×128 complex multiplications have to be performed.

To show feasibility, a real-time watermark detector has been built on three different platforms, viz. on a high-end Silicon Graphics workstation, on a TriMedia processor board, and on an FPGA based board.

The FPGA platform is most relevant for DVD. The FPGA implementation is characterized by the numbers in the first row of Table 1. The numbers in the second row characterizes an IC implementation of the watermark detector, where the functionality of the ROM (the secret watermark noise pattern) has been replaced by a random number generator. Not only is a random number generator a more secure solution, it is also less costly in terms of silicon area. (The gates associated with a random generator can be hidden extremely well. ROM content, on the other hand, is relatively easy to reverse engineer.)

Robustness

Many experiments have been performed to test the robustness of this system. It has been shown that Millennium survives MPEG-2 compression down to at least 2.5 Mb/s, MJPEG compression, DA/AD conversion, PAL conversion, noise addition, quantization, subtitling and logo insertion, cropping, frame erasure, speedups, and transmission errors. The robustness results have formally been verified within two contexts. First, robustness tests in the DVD context have confirmed that the Millennium watermarking system meets all the requirements of the proposal. Second, a variant of this watermarking algorithm is being used for broadcast monitoring. Also in this application, it has been shown to preserve content quality and to be robust to all common processing in the broadcast environment. No public information is available on the Millennium robustness performance, but for an overview of the broadcast monitoring performance see [15].

Issues

This section discusses a few aspects that are still on the agenda for copy protection standardization.

Location of Watermark Detector and Copy Control

Security requirements for copy protection sometimes conflict with the architecture of PCs and consumer electronic devices. From a security perspective, the effectuation of play control can best be located in the drive, i.e., as early as possible in the chain of circuits that handle digital video coming from a storage medium. This suggests that one would also like to include a watermark detector in the playback drive, where sectors of data are read from the disc surface. PC DVD drives are designed to obediently deliver sector data to the PC bus, however, without having any natural ability to interpret the (video audio or other) data. Watermark detection in the drive involves recognition of the type of data in the sectors, concatenation of data from multiple sectors, decryption, demultiplexing, and (partial) MPEG decompression. None of these tasks occur naturally within the drive. It has been proposed to skip watermark checks whenever data is encrypted, but evidently this opens many circumvention methods. Another solution [7] is to outsource the watermark detection to a device that can perform this task more naturally, such as an MPEG decoder, and to rely on a secure authenticated link between the drive and decoder. Such a link is already available in the DVD-ROM concept, but would require some additional features. In particular an integrity mechanism is needed to ensure that the drive and decoder negotiate about the same video data. It would allow the drive to effectuate play control, based on watermarks checked by the decoder. This also prevents the "local scrambling" or "bit inversion" attack [9].

Copy Generation Control

Having covered the case of content that may never be copied, we must also deal with the much less straightforward implementation of "(only) one (generation of) copy allowed." Because of the nature of this "Copy Once" requirement, information has to be passed along with the content to allow a recorder to distinguish between original and copy. Two basic principles are known:

▲ Embedding of a secondary watermark by the recorder (the remarking concept)
▲ Removal of a "volatile" piece of information from the content during recording (the ticket concept).

Both solutions have their own pros and cons. Remarking requires that a consumer recorder must be able to embed a watermark. This implies that content must made accessible in a form that allows embedding (e.g., partial MPEG decoding). Reliable and invisible embedding may require evaluation of the content using a perceptual model. Another disadvantage is that pirates can compare the input and output of such storage device, and find the embedded secondary watermark. Almost inevitably that provides information on how to remove the watermark. The ticket approach [6], [10] avoids the above issues. The volatile piece of information, i.e., the "ticket" acts as an authorization identifier. It can either be embedded in the content or passed on as a separate signal. Failure of a device to handle the ticket leads to a loss of rights to copy. The remarking and ticket concept have fundamentally different failure modes. In particular, remarking tends to allow recorder to make copies in cases when a legacy or modified recorder has failed to add the secondary watermark, whereas the ticket concept may deny the user rights to copy when a legacy device has accidentally mishandled the ticket.

To ensure that the ticket is specific for a particular title or for any specific transfer (e.g., copy) of the content, the ticket is cryptographically tied to the watermark payload in the content. As the watermark is (required to be) preserved under processing, the ticket can remain the same. The ticket is used as a proof that the source of the content has prior knowledge of the watermark [6], [10]. A random number is generated by the copyright owner, which then becomes the versatile ticket. The ticket acts as a cryptographic counter that can be decremented, but not incremented. Depending on how many generations of recording and playback the content owner desires to grant to the user, she sets the system by passing the ticket through a cryptographic function $F(.)$, n times. Here $F(.)$ is a publicly known cryptographic one-way function. Neither the player nor the recorder passes T transparently. Instead, the ticket is clipped, i.e., the counter is decremented by passing the ticket data through a one-way function, on every passage through a recorder or player (see Fig. 2). Verification of the ticket occurs in players and in recorders. It is done checking for a watermark. If that is present, the ticket data is passed through the one-way function m times and compared with the watermark data. Players check for $m = 1$ or 3. Recorders check for $m = 2$. Mastering equipment checks for $m = 2$ or $m = 4$ before creating stampers for "copy never" or "copy once" discs, respectively. A real-life analogy would be a movie theater where the entrance ticket is stripped by the attendant at the entrance (record control), but where viewers have to hang on to the stub to allow wardens to check whether nobody sneaked into the theater through the emergency exit (playback control). The ticket concept also allows play control of copy-once material. In the remarking system, the first and any further generation of copies would all carry the both the primary and secondary watermark. Thus play control can not distinguish between these.

The cryptography behind the ticket system does not rely on a global secret. From a cryptographic point of view it is not necessary that $F(.)$ is kept secret to potential attackers. Compliant consumer devices check for the watermark. If it is present and has payload W, it also interprets the ticket data T to verify whether $F^m(T)$, with $m = 1,2,...$ equals W. If $m = 1$ the device is entitled to playback the content. If $m = 2$ the device is entitled to record the content, and to store $T' = F(T)$ along with the content.

Media Type Recognition

Several reasons exist why recordable storage media should be distinguished from pressed media and need a unique identifier that may not be modifiable in a consumer device.
▲ The "play control" system needs information about whether the disc is an original premastered (stamped) disc or a recordable.
▲ To prevent that both the encrypted content and the associated keys can be bit copied from pressed discs to recordables, some uncopyable data should be stored on pressed media.
▲ Copy-Once content stored on a recordable disc must be encrypted in a way such that cloning to another recordable disc is not possible. A solution is to use a unique disc identifier to generate the encryption key. If the encrypted content, but not the ID, is transferred to another disc with a different ID, a player will not be able to generate the appropriate decryption key.

Many proposals have been brought up to distinguish between pressed (ROM) and recordable discs. To some extent, the DVD standard relies on data stored in ROM sectors which should not be write-accessible by recorders. This is now recognized as being both too weak to stop hackers and inadequate from a licensing point of view.

The measurement of physical parameters such as the disc reflectivity initially was one of the solutions discussed extensively, but this idea is viewed with suspicion as it is not very reliable (fingerprints!) and because it conflicts with the current development of better materials for recordable discs.

Also, the pregroove wobble, a positioning technique used by all known *recordable* disc formats appears less suitable. Different wobble frequencies are used by different standards. Thus a pregroove wobble detector does not necessarily recognize recordable disc using a new format. None of these two concepts are future proof, in that they inherently deal with new formats.

The most secure solution proposed thus far is the *pit wobble* of *pressed* (i.e., DVD-ROM) media. As illustrated

in Fig. 3, the wobble is a rapid radial deviation from the track spiral on the disc. The deviations are at tens of kilohertz or faster and can be detected electronically in the servo control circuitry of the player. However, the mechanics of the optical pick do not allow the laser head to precisely follow the deviations. The optical head thus follows an (unwobbled) spiral, and the wobble is experienced as a minor detracking which does not affect the detector of the video data that resides in the pits, as was evidenced in clock-jitter measurements on wobbled evaluation disks. The security resides in the fact that although consumer readers can detect it, consumer recorders fundamentally cannot write a wobble.

Data embedded in the wobble carries a payload of cryptographic data that is specific for every title produced on ROM. This is tied to the watermark in the same manner described in the previous section for the ticket.

Conclusions

The DVD copy protection problem can not be solved by encryption alone. Digital watermarking is needed to prevent copy protection being circumvented by noncompliant devices. In this paper we have described the Millennium watermark system as proposed for DVD copy protection purposes, and we have illustrated how the basic requirements for that application are met.

Even though copy protection has received ample attention in the standardization of digital video in the past five years, several issues have not yet been fully resolved. It may be unlikely that a bullet-proof solution will ever be found, but the discussions are converging on what technical mechanisms should be involved and against what these can protect. We identified several issues that will be on the agenda in the coming year(s). We also discussed solutions to some of these problems.

Jaap Haitsma was born in 1974 in Easterein, The Netherlands. He received his B.Sc. in electronic engineering from the Noordelijke Hogeschool Leeuwarden in 1997. In 1995 he had a internship at Holland Signaal, Hengelo, the Netherlands. In 1996 he interned with the Universitat Polytèchnica de Catalunya in Barcelona, Spain. Since 1997, he had been with Philips Research Laboratories, Eindhoven, The Netherlands, where he has been doing research into digital video and audio watermarking . He is working toward his M.Sc. in electronic engineering at the Technical University of Eindhoven. His areas of interest include digital image, video and audio signal processing, digital signal processors, and copy protection.

Ton Kalker was born in The Netherlands in 1956. He received his M.S. in mathematics in 1979 from the University of Leiden, The Netherlands and his Ph.D. in 1986. From 1979 until 1983, while he was a Ph.D. candidate, he was a Research Assistant at the University of Leiden. From 1983 until December 1985 he was a Lecturer at the Computer Science Department of the Technical University of Delft. In December 1985 he joined the Philips Research Laboratories Eindhoven, where he is currently a member of the Digital Signal Processing group. He is a Senior Member of IEEE. He has a part-time position as a Professor at the Eindhoven University of Technology, teaching signal processing methods for data protection. His research interests include wavelets, multirate signal processing, motion estimation, psycho physics, digital video compression, digital watermarking and multimedia security.

Jean-Paul M.G. Linnartz is Principal Scientist with Philips Natuurkundig Laboratorium (Nat. Lab.), Eindhoven, The Netherlands. In 1994, he was with Delft University of Technology in The Netherlands, as an Associate Professor. From 1992 to 1998, he was with the Department of Electrical Engineering and Computer Science, University of California at Berkeley, and from 1994-1998. From 1988-1991, he was Assistant Professor at Delft University of Technology. During 1987-1988, he worked with the Physics and Electronics Laboratory (F.E.L-T.N.O., The Hague) of the Netherlands Organization for Applied Scientific Research. He received his Ir. (M.Sc. E.E.) degree in electrical engineering (cum laude) from Eindhoven University of Technology, The Netherlands, in 1986 and his Ph.D. (cum laude)from Delft University of Technology in 1991. His main research interests are in conditional access and information security, electronic watermarks, (wireless) multi-media communications. He received the Dutch Veder Prize 1991 for his research on tele-traffic aspects in mobile radio networks. In 1993, he published the book *Narrowband Land-Mobile Radio Networks*. He is Editor-in-Chief of *Wireless Communications, The Interactive Multi-Media CD-ROM*.

Maurice Maes was born in Maastricht, the Netherlands, on September 1, 1963. In 1987 he received his Master's degree in mathematics with honors and in 1997 he received his Ph.D. from the University of Amsterdam. In October 1987, he joined Philips Research Laboratories (PRL) as a member of the Mathematics Group. Presently, he is a member of the Digital Signal Processing group at PRL, and his main interests include content protection for digital video and audio by means of digital watermarking and other copy protection techniques.

Geert F.G. Depovere was born in Roeselare, Belgium, on March 26, 1965. He received the M.Sc. and Ph.D. degrees in electrical engineering from the University of Gent, Belgium, in 1988 and 1992, respectively. He joined Philips Research Laboratories, Eindhoven, The Netherlands, in 1992 where he worked in the field of high bit rate coherent multi-channel systems and broadband optical network architectures. He co-ordinated the Philips Research activities in a number of European Projects: CMC R1010, COBRA R2065 and PHOTON A22. He was involved in the CPTWG (copy protection), EBU and MPEG standardisation activities. In April 2000

he was appointed department head of the "Video Processing and Visual Perception" group in Philips Research. He has authored and co-authored over 30 regular and invited publications in scientific journals and at international conferences. He has served as chairman at conferences and he has filed a number of patent applications.

Joop Talstra received his B.A. in physics from the University of Nijmegen in 1990 and his M.A. and Ph.D., both in physics, from Princeton University in 1991 and 1995, respectively. He has been with Philips Research Labs, Eindhoven, Netherlands, since 1997. His primary field of interest is copy-protection methods for optical recording systems, working on such content-management techniques as watermarking, encryption, and bit-detection of optical subchannels.

Table 1. Implementation costs for the Millennium watermarking system.

	LOGIC	ROM	RAM
FPGA	17 kG	34 kB	36 kB
Silicon	14 kG	1 kG	36 kB/3.6 mm^2

Note that figures in this table reflect the situation of a stand-alone detector. In general the detector will be integrated e.g., into an MPEG-decoder or the host-interface chip of a DVD-drive. In either case, the main functionality of these ICs (MPEG-decoding and PCI-bus streaming, respectively) require huge amounts of buffer memory, of which 36 kB can easily be borrowed. This increases logic cost by 5kG-mostly for protecting the memory/watermark-detector interface against malicious attackers trying to upset the watermark detection process or glean information about the (secret) watermark pattern.

References

1. J.A. Bloom, I.J. Cox, T. Kalker, J.P. Linnartz, M. Miller, and C. Traw, "Copy protection for DVD video," *Proc. IEEE.*, vol. 87, pp. 1267-1276, July 1999.
2. A.E. Bell, "The dynamic digital disk," *IEEE Spectrum*, vol. 36, pp. 28-35, Oct. 1999.
3. 5C digital transmission content protection white paper. [Online] Available WWW: http://www.dtcp.com.
4. B. Ponce, "The impact of MP3 and the future of digital entertainment products," *IEEE Commun. Mag.*, vol. 37, pp. 68-70, Sept. 1999.
5. G. Wirtz, *Philips Electronics Response to Call For Proposals Digital Transmission,* Copy Protection Technical Working Group, Burbank, CA.
6. J.P. Linnartz, A. Kalker, and G. Depovere, *Philips Electronics Response to Call For Proposals Data Hiding (Watermarking),* Copy Protection Technical Working Group, Burbank, CA.
7. J.P. Linnartz, J. Talstra A. Kalker G. Depovere, and M. Maes, *Philips Electronics Response to Call For Proposals WG 9*, Tokyo, Japan, Sept. 1998.
8. J.P. Linnartz, J. Talstra A. Kalker G. Depovere, and M. Maes, *Philips Electronics Response to Call For Proposals WG 6,* Tokyo, Japan, Oct. 1998.
9. I.J. Cox and J.P.M.G. Linnartz, "Some general methods for tampering with watermarks," *IEEE J. Select. Areas Commun.*, vol. 16, pp. 587-593, May 1998.
10. J.P.M.G. Linnartz, "The ticket concept for copy control based on embedded signalling," in *Proc. ESORICS '98, 5th. European Symposium on research in Computer Security,* (Lecture Notes in Computer Science, vol. 1485). Louvain-La-Neuve, Sept. 1998, pp. 257-274.
11. T. Kalker, "Digital video watermarking for DVD copy protection," in *Proc. SPIE Multimedia Systems and Applications*, 1999. **AU: please provide page numbers**
12. T. Kalker, "Digital video watermarking for DVD copy protection," in *Pro. Erlangen Watermark Workshop '99*, 1999. [Online]. Available: http://www.lnt.de/~watermarking
13. B. Schneier, *Applied Cryptography.* New York: Wiley, 1997.
14. L. Brown, "An overview of image registration techniques," *ACM Computing Surveys*, vol. 24, pp. 325-376, Dec. 1992.
15. P. Termont, L. De Strycker, J. Vandewege, J. Haitsma, A. Kalker, M. Maes, and G. Depovere, "An implementation of a real-time digital watermarking process for broadcast monitoring on a trimedia VLIW Processor," *Proc. IPA*, pp. 775-779, July 1999. **AU: Is this a journal or a conference proceedings? Please provide a volume number if it is a journal.**

▲ *Fig. 1. SPOMF detection for one pattern with multiplicity two.*
▲ *Fig. 2. Overview of watermark embedding.*
▲ *Fig. 3. Overview of watermark detection. For DVD N_1=720, N_2=480; T=27.*
▲ *Fig. 4. The ticket is clipped (cryptographically modified) during each playback or recorder passage.* ***AU: please cite this figure in the text.***
▲ *Fig. 5. Artist impression of wobbled pits on DVD disc.* ***AU: please cite this figure in the text.***
▲

Callouts:

The standardization of DVD video has unleashed an unprecedented debate over copy protection.

The basic requirements on the watermarking method include that the watermark is invisible and difficult to remove.

In other words, a watermark is simply additive noise.

The DVD copy protection problem can not be solved by encryption alone.