Multimedia Content Screening using a Dual Watermarking and Fingerprinting System

Darko Kirovski, Henrique Malvar, and Yacov Yacobi
Microsoft Research
One Microsoft Way
Redmond, WA 98052
{darkok,malvar,yacov}@microsoft.com

ABSTRACT

We present a new dual watermarking and fingerprinting system, where initially all copies of a protected object are identically watermarked using a secret key, but individual detection keys are distinct. By knowing a detection key, an adversary cannot recreate the original content from the watermarked content. However, knowledge of any one detection key is sufficient for modifying the object so that a detector using that key would fail to detect the marks. Detectors using other detection keys would not be fooled, and such a modified object necessarily contains enough information about the broken detector key – the fingerprint. Our dual system limits the scope of possible attacks, when compared to classic fingerprinting systems. Under optimal attacks, the size of the collusion necessary to remove the marks without leaving a detectable fingerprint is superlinear in object size, whereas classic fingerprinting has a lower bound on collusion resistance that is approximately fourth root in object size. By using our scheme one can achieve collusion resistance of up to 900,000 users for a two hour high-definition video.

1. INTRODUCTION

With the growth of the Internet, unauthorized copying and distribution of digital media has never been easier. The music industry, for example, claims a $5B annual revenue loss due to piracy [21], which is likely to increase due to file-sharing Web communities such as Gnutella. As the Internet bandwidth increases, the movie industry is expected to encounter the same piracy problem. Legal attempts to alleviate the problem have shown limited success so far, in view of the complexity of the issues involved.

One source of hope for copyrighted content distribution on the Internet lies in technological advances that would provide ways of enforcing copyright in server-client scenarios. Traditional data protection methods such as scrambling or encryption cannot be used, since the content must be played back in the original form, at which point it can always be re-recorded and then freely distributed. One approach for this problem is marking the media signal with a secret, robust, and imperceptible watermark. The media player at the client side can detect this mark and consequently enforce the corresponding e-commerce policy. Although the effectiveness of such a system requires global adoption of many standards, the industry is determined to carry out that task [22].

1.1 General

Watermarks (WM) and fingerprints (FP) are marks hidden in an object for two distinct purposes. The former are used to designate an object as protected, and signal to the client machine that some license is needed in order to use the object. The latter are used to trace piracy to its origins. The detection process of a WM is done "blindly" (without the presence of the original recording) and in real-time, even on small devices. FPs are detected by powerful machines that can devote significant resources to the forensic process. Watermarks are identical in all the copies, while FPs are individualized. If necessary, the FP detector can have access to a copy of the original unmarked object, using it to improve its likelihood of success in detecting the FPs, even from content modified by malicious attacks.

In classic WM systems, the detection key is identical to the embedding key, and hence the whole system collapses after breaking just one client machine. In this paper, we propose a WM system where, as usual, all the copies of a protected object are identically watermarked, but where each user has a distinct secret detection key. All such detection keys are different from the secret embedding key. Even by gaining the knowledge of a relatively large number of detection keys, an adversary cannot remove the secret marks from the protected content. We assume that the WM system is robust against signal-processing attacks on the protected object and focus on collusion attacks against the detection keys. We show that an attacker who has access to one detection key can always fool the corresponding WM detector, but not other WM detectors. More importantly, we also show that in that process the attacker necessarily inserts a fingerprint in the modified content. The main entities of the WM/FP system and their interactions are illustrated in Figure 1.

In the subsequent sections, we quantify the security properties of the proposed scheme through the following aspects:

- Construction of distinct detection keys from a secret watermark key,

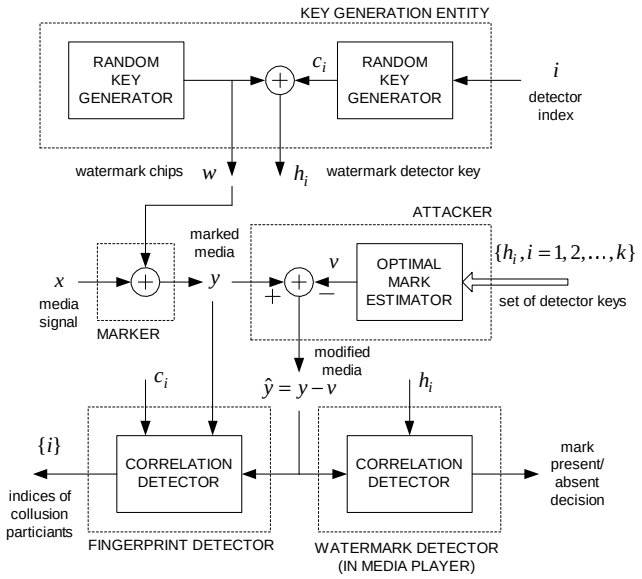

Figure 1: General system block diagram for the proposed WM/FP system. Note that each watermark detector i uses a different key h_i. In the attack model, a set of detection keys is colluded to form an estimate v of the watermark w.

- The probabilities of false positive and false negative decisions for detectors using a fixed-length fixed key,
- The size of a collusion clique that would fool: (*i*) a single WM detector, (*ii*) all WM detectors, and (*iii*) the FP detector, respectively, and
- The probabilities of false positive and negative decisions for the three (*i-iii*) respective types of collusion.

A main contribution of this paper is to show that our proposed WM/FP system can achieve a minimum collusion size K that grows linearly with the size N of the marked object. A second contribution is that we can augment our WM/FP system with a segmentation layer. The media content is partitioned into S segments, where a watermark or fingerprint can be reliably detected within each segment. Only detection keys that belong to the same segment, can participate in the collusion clique. With segmentation, the minimum collusion size K grows as $\mathcal{O}(N \log N)$. Therefore, with or without segmentation, our WM/FP system significantly improves on the best known asymptotic resistance to (fingerprint) collusion attacks of $\mathcal{O}(N^{1/4})$ [4]. Since we use a new protection protocol, comparison to classic FP systems may be characterized as unfair. However, comparison is important from the perspective of designing a content protection application based upon the two schemes.

1.2 Previous work

A survey of watermarking techniques is presented in [14]. We point our reader to review a watermarking technology that succeeds to imperceptibly hide data in audio while being robust with respect to numerous attacks (including sequence desynchronization) specifically designed to prevent detection of spread-spectrum watermarks [16]. Asymmetric watermarking schemes that try to alleviate the problem of storing the embedding sequence at the client, have been proposed in [13], [8], and [12].

In the fingerprinting domain, Ergun et. al. [9] have considered embedding distinct spread-spectrum sequences per copy and have modeled collusion attacks as averaging of copies with additive noise. Boneh and Shaw [4] have defined a lower bound on the collusion size with a proposal for collusion-secure encoding and an improved "majority attack" model. The previous two works put an upper bound at $\mathcal{O}((N/\log(N))^{1/2})$ and a lower bound at $\mathcal{O}(N^{1/4})$ respectively on collusion-secure fingerprinting with respect to object size N.

Fiat and Tassa have introduced a dynamic traitor-tracing mechanism where the set of users is randomly grouped into r subsets, each receiving a distinct symbol [11]. After a subset is identified as the one that includes the pirate(s), the search continues within that subset only. The assumption is that, per round of the tracing process, the pirates simply choose one of the multi-bit symbols available to them. The assumption of [4] is that for the bits where a collusion disagrees, the colluders may choose any value.

2. SYSTEM DESCRIPTION

In this section, we review the basics of spread-spectrum watermarking, and introduce our WM/FP system.

2.1 Spread-Spectrum Watermarking

The media signal to be watermarked $x \in \mathcal{R}^N$ can be modeled as a random vector, where each element of x is a normal (Gaussian) random variable with standard deviation A, i.e. $x_j = \mathcal{N}(0, A^2)$. For example, for audio signals marked in a log magnitude domain, the normal assumption is reasonably accurate, typically with $A \in [5, 15]$, after necessary media preprocessing steps [16]. A *watermark key* w is defined as a spread-spectrum sequence vector $w \in \{\pm 1\}^N$, where each element w_j is usually called a "chip." The marked signal y is created by vector addition $y = x + w$.

Let $w \cdot v$ denote the normalized inner product of vectors w and v, i.e. $w \cdot v \equiv N^{-1} \sum w_j v_j$, with $w^2 \equiv w \cdot w$. For example, for w as defined above we have $w^2 = 1$. We assume that the media player contains a watermark (WM) detector that receives a modified version $\hat{y}$ of the watermarked signal y. The WM detector performs a correlation (or matched filter) test $d_T = \hat{y} \cdot w$, and using a classical Neyman-Pearson hypothesis test, it decides that the watermark is present if $d_T > \delta_T$, where δ_T is the detection threshold that controls the tradeoff between the probabilities of false positive and false negative decisions. We recall from modulation and detection theory that such a detector is optimal [24].

Under no malicious attacks or other signal modifications, i.e. $\hat{y} = y$, if the signal y has been marked, then $d_T = 1 + g_T$, where the *detection noise* g_T is a normal zero-mean random variable with variance $\sigma^2_{g_T} = A^2/N$. Otherwise, the correlation test yields $d_T = 0 + g_T$. For equal probabilities of false positives and false negatives, we should set $\delta_T = 1/2$. For robustness against attacks, the signal domain x must be appropriately chosen, and some small modifications on the watermark pattern may be necessary [16]. In this paper, we assume that such precautions have been taken care of in the design of the WM detector [16], so we can disregard media attacks. See [14] for an overview of techniques that use this paradigm for hiding data in audio and video.

2.2 The Dual WM/FP System

Traditional spread-spectrum WM systems detect watermarks using a key w that is in essence a *secret watermarking key* (SWK). In copyright enforcement schemes, the watermark detection is done at the client (the media player), which must then have access to the SWK. An adversary can recreate the original content if they succeed in obtaining the SWK. This can be achieved in several ways: by breaking into a detector (i.e. reverse engineering the detection software or hardware), or using the sensitivity attack [17].

In our dual WM/FP system, depicted in Figure 1, the *watermark detection key* (WDK) is different from the SWK, so breaking into a single detector does not provide enough information to remove the watermark w. The media signal x is watermarked in the same way as in traditional spread-spectrum watermarking. However, for each media player i an individualized WM/FP detection key WDK h_i is created from a SWK w in the following way. Let $C = \{c_{ij}\}$ denote an $m \times N$ matrix, where $c_{ij} \in \mathcal{R}$, $c_{ij} = \mathcal{N}(0, B^2)$, i.e. each entry is a zero-mean normal random variable with standard deviation $\sigma_c = B$. Each row i contains a *watermark carrier*, denoted by c_i. The ith WDK is defined as $h_i = w + c_i$. The goal of the watermark carrier c_i is to hide the SWK w in h_i so that knowledge of h_i does not deterministically imply knowledge of w, as long as B is large enough. In other words, no player contains the SWK w, but rather a modified version of it. Because the players use a correlation-based WM detection, they should still be capable of detecting the watermark in a marked content y, as long as the number of chips N is large enough to attenuate the noise introduced by the watermark carriers c_i.

The detection process is carried out by correlating the received media file $\hat{y}$ with h_i, generating a detector output $d_W = \hat{y} \cdot h_i$. Similarly to traditional spread-spectrum watermarking, if $\hat{y}$ was marked, then $d_W = 1 + g_W$; otherwise $d_W = 0 + g_W$. The difference is that now g_W is a function of both the media x and the watermark carrier c_i. If there are no attacks, i.e. $\hat{y} = y = x + w$, then

$$\begin{aligned} d_W &= y \cdot h_i = (x + w) \cdot (w + c_i) = 1 + g_W, \quad \text{where} \\ g_W &= x \cdot (w + c_i) + w \cdot c_i \end{aligned}$$

is a zero-mean noise component of d_W. For this case, we derive the detection noise variance as $\sigma^2_{g_W} = (A^2 + B^2 + A^2B^2)/N$. The detection noise variance is significantly increased due to the watermark carrier c_i, i.e. if $\hat{y}$ is watermarked then $g_W = g_T + x \cdot c_i + w \cdot c_i$ or else $g_W = g_T + x \cdot c_i$, where $\mathrm{Var}\{x \cdot c_i\} \gg \mathrm{Var}\{g_T + w \cdot c_i\}$. Thus, our WM/FP system requires larger N than traditional spread-spectrum, for the same WM detector performance. The performance of the detector can be slightly improved without any tangible effect on system security by generating watermark carriers such that $(\forall c_i \in C) w \cdot c_i = 0$, which reduces the detection noise.

2.3 Copyright Enforcement using WM/FP

In this subsection, we identify the main entities in the dual WM/FP system and describe their roles.

2.3.1 Watermark Detector (WMD)

The WMD correlates a potentially marked signal $\hat{y}$ with client's WDK h_i, i.e. $d_W = \hat{y} \cdot h_i$. It decides that the content is marked if $d_W > \delta_W$. The probability of false positives (identifying an unmarked content as marked) is denoted as ε_1, which must be relatiively small, e.g. $\varepsilon_1 = 10^{-9}$.

2.3.2 Attacker

As part of an optimal attack to the system, the adversary breaks K clients and extracts their WDKs $\{h_i, i = 1 \dots K\}$. Next, the adversary creates an attack vector v as an optimal estimate of the SWK w given the collusion key set $\{h_i, i = 1, \dots K\}$, and generates an attacked signal as $\hat{y} = y - v$. The closer v estimates w, the more the attacker will clean the watermark. We use ε_2 to denote the probability that a watermark chip is incorrectly estimated by the attacker, i.e. $\varepsilon_2 = \Pr[v_j \neq w_j]$. The attacker aims at having ε_2 as small as possible, whereas we design the system parameters to force ε_2 to be close to 1/2.

2.3.3 Fingerprint Detector (FPD)

The FPD recovers the attack vector v from an attacked content $\hat{y}$ and the originally marked content y simply by $v = \hat{y} - y$. Unlike the WMDs, the FPD has access to the watermark carrier matrix C. Thus, the FPD correlates v with a suspect watermark carrier c_i, i.e. it computes $d_F = v \cdot c_i$, and decides that the ith client is part of the collusion if $d_F > \delta_F$, i.e. δ_F is the FPD threshold. Compared to the WMD, the FPD has less noise in the correlated vectors, and thus the collusion resistance of the FPD is much higher than that of the WMD. We use ε_3 to denote the probability of false positives in the FPD, i.e. incriminating a player that was not in the collusion set. Therefore, ε_3 must be very small, just like ε_1. We use η to denote the probability of false negatives at the FPD. We would like it to be small, but do not have to insist that it is as small as ε_1 and ε_3.

Ultimately, the goal of the adversary is to create an optimal attack vector $v \approx w$ based on a collection of K WDKs such that:

- v reduces the expected $E[d_W]$ to a level where the probability of detecting a true positive at the WMD is relatively low ($\approx \varepsilon_1$) and
- v reduces the expected $E[d_F]$ to a level where the likelihood of false negative at the FPD is relatively high (e.g. $\eta \geqslant 0.9$).

3. ATTACKS WITHOUT COLLUSION

In this section, we discuss the attacks that can be performed on an object with knowledge of at most one WDK.

3.1 Attacks on a Protected Object

Here we elaborate on a basic assumption for our WM/FP mechanism mentioned in the previous section: that there exists a spread-spectrum watermarking mechanism that can be broken only by modifying the marked content beyond the threshold for low fidelity of the attacked copy with respect to the original recording [16]. Typical attacks in this domain range from compression, filtering, resampling, equalization, and various other editing procedures [14], to desynchronization (or data shifting) techniques that aim at misaligning the embedded spread-spectrum sequence in the content (e.g. the Stirmark attack [2]).

Having a robust watermarking technology is not the only requirement for secure e-commerce of content. Traditional watermarking assumes that the watermarking key (SWK) is hidden at the client side. By breaking a single client, the adversary can create the original content and thus enable all

clients to play that content as unmarked. We refer to that as BORE – break once run everywhere. In our WM/FP system, we assume that the attacker will eventually break at least one client and capture that machine's WDK. This can be accomplished by physically breaking the machine (code debugging, reverse engineering) or by using the sensitivity attack [17], in which the sign of each chip of the detection key is iteratively estimated. Note that in the WM/FP system, the sensitivity attack does not reveal w; it reveals just the sign of each chip of of the WDK: $\text{sign}(h_i) = \text{sign}(c_i + w)$.

Our scheme is generally BORE-resistant at the protocol level. By breaking a single client, the adversary can play content as unmarked on that broken client, but needs to collude the extracted client WDKs with other clients to finally create content that can play on all players. With our dual WM/FP system, we significantly improve collusion resistance through a fingerprinting mechanism that can identify the members of the clique if its cardinality is smaller than a relatively large lower bound, which is determined in the next section.

3.2 The Subtraction Attack

Suppose that an adversary breaks client i and extracts its WDK $h_i = c_i + w$. Then, the adversary can create an attack vector $v = \alpha h_i$ such that the modified media $\hat{y} = y - v$ produces $E[d_W] = E[\hat{y} \cdot h_i] \ll \delta_W$, and thus defeating that client's WM detector. To determine α, we note that $d_w = \hat{y} \cdot h_i = [x + w - \alpha(c_i + w)] \cdot (c_i + w) = 1 - \alpha(1 + c_i^2) + x \cdot c_i + x \cdot w + (1 - 2\alpha) c_i \cdot w$, from which we have $\text{E}[d_W] = 1 - \alpha(1 + B^2)$. Thus, by setting $\alpha = (1 + B^2)^{-1}$ we get $\text{E}[d_W] = 0$, that is $d_W = 0 + g_W$. Also, we see that $\sigma^2_{g_W} \simeq (3 + A^2 + B^2 + A^2 B^2)/N$, and that $\sigma_v^2 = \alpha^2(1 + B^2) = \alpha \ll 1$.

Therefore, we see that given knowledge of the client's detection key, the subtraction attack can drive the detector correlation all the way to zero, with just a slight increase in the detector noise $\sigma^2_{g_W}$ and a negligible increase in distortion in the content (since $\sigma_v^2 \ll w^2 = 1$). If the attacker tries to use a key h_l to break a detector $i \neq l$, it is easy to see that to drive $\text{E}[d_W] = 0$ the attacker would then need to set $\alpha = 1$. However, that would drive $\sigma_v^2 = (1 + B^2) \gg 1$, causing too much distortion to the content. Also, it would make $\sigma^2_{g_W}$ increase by an amount equal to $3B^4/N$, which would make the decisions in the ith WM detector erratic. In other words, even by driving $\text{E}[d_W] = 0$ the ith detector would not be broken with probability much better than $1/2$.

3.3 Resemblance to Public-Key Systems

We have concluded that if the attacker knows the WDK key h_i of a single detector, that information is not sufficient to break any other detector via the key subtraction attack. Knowing h_i is not enough to infer w, either. In that respect, our dual WM/FP system resembles a public-key cryptosystem, since knowledge of the verification key (in our case h_i) does not imply knowledge of the signing key (in our case w). However, as opposed to a public-key cryptosystem, WDK in the dual WM/FP system is not exposed outside an individual player. The efficacy of the entire system is based on the requirement that obtaining the WDK requires non-trivial effort by the adversary, e.g., reverse engineering the player or a lengthy, sensitivity attack [17]. This can be obtained by means of hardware support that enables software integrity controlled by the operating system. As a result, low-cost breaking alternatives such as software patching or debugging of a media player in this system are not possible.

4. COLLUSION ATTACKS ON WM/FP

Consider a collusion clique of size K that has broken their players and extracted K different WDKs h_i. We now devise the optimal attack based on that set of keys $\{h_i, i = 1, 2, \ldots, K\}$. Without loss of generality, we assume that those extracted WDKs (with indices 1 to K) are the ones in the collusion.

4.1 The Optimal Attack

The attacker's job is to estimate the SWK key w by an attack vector v, so that the modified media $\hat{y} = y - v$ will not show significant correlation in any WM detector j, i.e. even for $j > K$. The best job the attacker can possibly perform is given by the $\text{sign}(\text{mean}\{\cdot\})$ attack.

LEMMA 1. *The optimal attack vector is given by*

$$v = \text{sign}\left(\sum_{i=1}^{K} h_i\right)$$

PROOF. The optimal estimate for each element v_j of the attack vector is given by $v_j = +1$ if $\Pr[w_j = +1|\{h_i\}] \geq 1/2$ and $v_j = -1$ if $\Pr[w_j = +1|\{h_i\}] < 1/2$. That estimate is optimal because it minimizes $\Pr[v_j \neq w_j]$. Since $h_{ij} = w_j + c_{ij}$, where the c_{ij} are independent and normally distributed, we can write $\Pr[w_j = +1|\{h_i\}] = 1/(1 + \nu_j)$, where $\nu_j = \prod_{i=1}^{K} p_c(h_{ij} + 1)/p_c(h_{ij} - 1)$ and $p_c(\zeta) = (2\pi)^{-1/2} \exp[-\zeta^2/(2B^2)]$. We can write $\nu_j = \exp(-2\rho_j/B^2)$, where $\rho_j = \sum_{i=1}^{K} h_{ij}$. Thus, $\Pr[w_j = +1|\{h_i\}] \geq 1/2]$ when $s_j \geq 0$ and $\Pr[w_j = +1|\{h_i\}] < 1/2]$ when $s_j < 0$. □

4.2 WMD Performance

Given the optimal attack above, we can compute the average estimation error in the attack vector, $\varepsilon_2 = \Pr[v_j \neq w_j]$, as follows. Since the w_j chips are equally likely to be $+1$ or -1, it is clear that because of symmetry of w, $\varepsilon_2 = \Pr[s_j \geq 0 | w_j = -1]$. Since for $w_j = -1$ we have $s_j = -K + \bar{c}_j$, where $\bar{c}_j = \sum_{i=1}^{K} c_{ij}$. Therefore, $\varepsilon_2 = \Pr[\bar{c}_j \geq K]$, where $\bar{c}_j$ has a normal distribution with zero mean and variance $B\sqrt{K}$.

COROLLARY 1. *A collusion of size K produces*

$$\varepsilon_2 = \frac{1}{2} \text{erfc}\left(\frac{\sqrt{K}}{B\sqrt{2}}\right).$$

where $\text{erfc}(\cdot)$ is the complementary error function.

Given ε_2, we can evaluate the efficiency of the subtraction attack $\hat{y} = y - v$ for the optimal attack vector v. Since $\text{E}[v \cdot w] = \Pr[v_j = w_j] - \Pr[v_j \neq w_j] = 1 - 2\varepsilon_2$, it is easy to see that after attack the expected output of the WM correlation detector drops to $\text{E}[d_W] = 2\varepsilon_2$. Figure 2 depicts the resulting probability density functions (pdfs) for d_W when computed against an original, marked, and attacked signal. The attacker may attempt a stronger subtraction attack, of the form $\hat{y} = y - \beta v$, with $\beta > 1$, because that would bring the WMD output further down to $\text{E}[d_W] = 2\beta\varepsilon_2 - (\beta - 1)$. As long as β is not too large, the attacked content $\hat{y}$ may be acceptable to users.

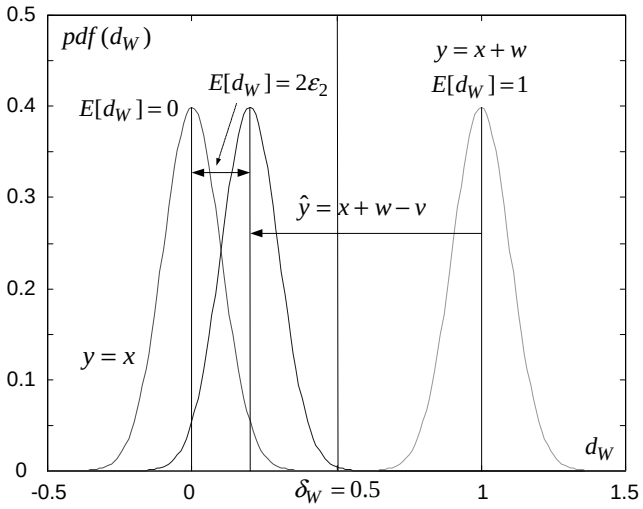

Figure 2: Illustration of the probability density functions for correlation tests against three different signals: a unmarked $\hat{y} = x$, marked $\hat{y} = x + w$, and attacked using the $v = \text{sign}(\text{mean}(\cdot))$ attack $\hat{y} = x + w - v$. The deviations of all correlation tests approximate $\sigma_{g_W} \approx AB/\sqrt{N}$. Exemplary detection threshold is set to $\delta_W = 1/2$.

4.3 Collusion Size That Defeats The WMD

In order to reduce the expected correlation to $\mathrm{E}[d_W] = \theta$, the adversary needs to achieve an attack vector error rate of $\varepsilon_2 = (\theta + \beta - 1)/(2\beta)$ through collusion. From Corollary 1, we see that for fixed θ and β the minimum collusion size grows proportional to B^2.

COROLLARY 2. *In order to reduce the correlation value to $E[d_W] = \theta$, the adversary needs to collude K WDKs, with*

$$K = 2B^2 \left[\mathrm{erf}^{-1} \left(\frac{1-\theta}{\beta} \right) \right]^2 .$$

EXAMPLE 1. *For $B = 10$, $\theta = 0.25$ and $\beta = 2$, the attacker must collude at least $K = 24$ keys. For $\beta = 1$, the attacker must collude at least $K = 133$ keys.*

We note that the attacker needs to set θ much smaller than δ_W, otherwise the probability that a WMD will still detect the watermark is not low enough to justify the attacker's effort. In other words the attack is successful only if it makes $\varepsilon_1 \simeq 1$. For that it is not necessary to set θ all the way to zero, because it would require K to be excessively large. By setting $\beta > 1$, though, it is possible to force $\theta = 0$.

To make the attacker's job more difficult, we need to increase the parameter B, the standard deviation of the watermark carrier c, since K grows with B^2. In doing so, however, we increase the detection noise variance $\sigma^2_{g_W} = (A^2 + B^2 + A^2B^2)/N$, where we recall that A is the standard deviation of the original content x and N is the object size. For a given σ_{g_W}, we determine that the probability of false positives $\varepsilon_1 = \Pr[d_w > \delta_w \,|\, \text{object is not marked}]$ is given by:

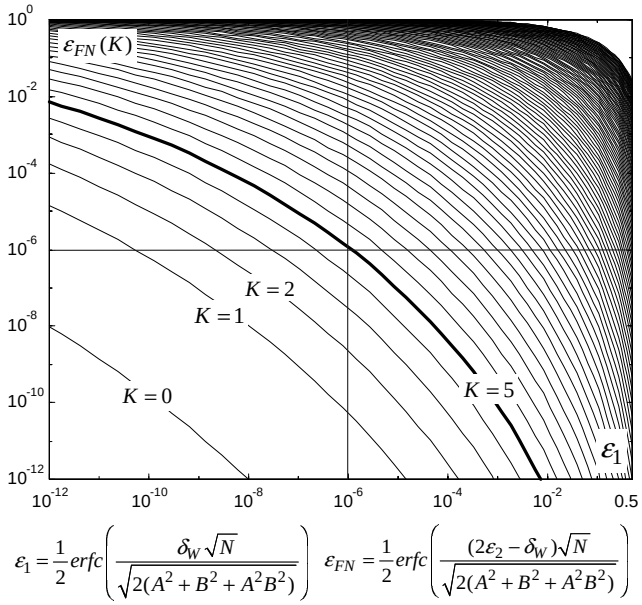

Figure 3: Receiver operating characteristic graph for the WMD. The graph considers the false positive probability ε_1 and false negative probability $\varepsilon_{FN}(K)$ after an attack with v averaged from K WDKs. In the example, $A = B = 7$ and $N = 4 \cdot 10^5$. One possible system design decision is $\varepsilon_1 = \varepsilon_{FN} \leqslant 10^{-6}$ with $K = 5$ collusion resistance and $\delta_W = \varepsilon_2(K = 5)$. Note that if we keep $\delta_W = 1/2$, the likelihood that the attacked clip will not be detected as marked drops to $\varepsilon_{FN}(K = 5, \delta_W = 1/2) \approx 10^{-3}$ at fixed $\varepsilon_1(K = 5, \delta_W = 1/2) \approx 10^{-10}$.

COROLLARY 3. *An object of size N produces*

$$\varepsilon_1 = \frac{1}{2} \,\mathrm{erfc} \left(\frac{\delta_W \sqrt{N}}{\sqrt{2(A^2 + B^2 + A^2B^2)}} \right).$$

We note that if $\delta_W = 1/2$, then ε_1 is also the probability of false negatives, i.e. the probability of a WMD not detecting a marked object that was not attacked. Figure 3 illustrates in more detail the receiver operating characteristic graph of the WMD under the assumption that the marked signal has been attacked by averaging K WDKs. From Corollary 3 we can compute N by:

COROLLARY 4. *The object size N required to achieve a given ε_1 is*

$$N = \frac{2[A^2 + B^2(1 + A^2)]}{\delta_W^2} \left[\mathrm{erfc}^{-1}(2\varepsilon_1) \right]^2 .$$

By combining the result above with that in Corollary 2, we arrive at one of the main results in this paper:

THEOREM 1. *Minimal collusion size K_W that achieves a fixed $E[d_W] = \theta$ at a WMD with fixed δ_W, β, and ε_1 grows linearly with object size N, i.e. $K_W = \mathcal{O}(N)$.*

PROOF. As N grows, for a given ε_1, B also grows, and thus $\sigma^2_{g_W} \to B^2(1 + A^2)/N$. Combining this asymptotic expression for σ_{g_W} with the results in Corollary 4 and Corollary 2,

we get

$$K_W = N\frac{\delta_W^2}{1+A^2}\left[\frac{\text{erf}^{-1}\left(\frac{1-\theta}{\beta}\right)}{\text{erf}^{-1}(1-2\varepsilon_1)}\right]^2.$$

The equation above allows us to compute the object size N necessary to achieve any desired collusion resistance K_W for given WMD performance. □

It is important to note that the result above is so far determined only by the WMD performance. In the next section, we confirm the linear relationship between K and N when considering the FPD performance.

5. FINGERPRINT DETECTION

As we mentioned in Section 2, the FPD has less noise in its correlation output. Therefore, it should be able to identify the indices i corresponding to all the WDKs h_i used in the collusion by the attacker, even if the collusion size K is large enough to fool all clients, as computed above. In this section, we evaluate the error probabilities for the FPD.

We recall that the FPD knows the marked content y, the attacked version $\hat{y}$, and the watermark carriers c_i. It computes the correlation $d_F = (\hat{y} - y) \cdot c_i$, and decides that the the ith client participated in the collusion if $d_F > \delta_F$. Assuming the attack model of the previous section, $\hat{y} = y - \beta v$, the FPD output can be written as

$$d_F = (\hat{y} - y) \cdot c_i = \beta(v \cdot c_i) = \text{E}[d_F] + g_F$$

where g_F is the zero-mean FPD correlation noise. The most critical error for the FPD is a false positive, i.e. incriminating a WDK i that did not participate in the collusion. The probability ε_3 of that error is given by the following:

LEMMA 2. *An object of size N produces*

$$\varepsilon_3 = \frac{1}{2}\,\text{erfc}\left(\frac{\delta_F\sqrt{N}}{\sqrt{2}\beta B}\right).$$

PROOF. If c_i is not in the collusion, it is independent of the attack vector βv. Thus, $\sigma_{g_F}^2 = \text{E}[\beta^2 v_{ij}^2 c_{ij}^2]/N = \text{E}[\beta^2 c_{ij}^2]/N = \beta^2 B^2/N$, which follows from $\varepsilon_3 = \Pr[g_F > \delta_F]$ and the fact that g_F has normal distribution. □

It is clear that, as expected $\varepsilon_3 \ll \varepsilon_1$ (usually by several orders of magnitude), since the argument in erfc($\cdot$) for ε_3 is approximately $(A\delta_F)/(\beta\delta_W)$ times larger than the argument in erfc($\cdot$) for ε_1. Thus, by choosing B and N for a sufficiently low ε_1, we achieve a negligibly low probability ε_3 of false positives in the FPD.

In order to compute the detection performance of the FPD we need to determine its expected output when we correlate with a carrier c_i such that h_i was part of the collusion. We see that $\text{E}[d_F] = \beta\text{E}[z_j]$, where $z_j = v_j c_{ij} = \text{sign}\,[s_j]\,c_{ij}$, with $s_j = w_j + b_j$, and $b_j = \frac{1}{K}\sum_{m=1}^{K} c_{mj}$.

LEMMA 3. *A collusion of size K produces*

$$E[d_F] = \beta\frac{B}{\sqrt{K}}\sqrt{\frac{2}{\pi}}\exp\left(-\frac{K}{2B^2}\right).$$

PROOF. It is clear that $\text{E}[z_j] = (\text{E}[z_j|w=+1] + \text{E}[z_j|w=-1])/2$, since the w_j chips are equally likely. Also, because of the symmetry of the problem we see that $\text{E}[z_j|w=+1] = \text{E}[z_j|w=-1]$, and so $\text{E}[z_j] = \text{E}[z_j|w=+1]$.

Assuming $w_j = +1$, $\text{E}[z_j] = \text{E}[z_j|s_j \geq 0]\Pr[s_j \geq 0] + \text{E}[z_j|s_j < 0]\Pr[s_j < 0] = \text{E}[c_{ij}|s_j \geq 0]\Pr[s_j \geq 0] - \text{E}[c_{ij}|s_j < 0]\Pr[s_j < 0]$. Under each of the conditions $s_j \geq 0$ or $s_j < 0$, we see that $s_j = 1 + b_j$ and c_{ij} are all jointly-normal variables, with variances $\sigma_s^2 = \sigma_b^2 = B^2/K$ and $\sigma_c^2 = B^2$. Furthermore, the correlation coefficient between b_j and c_{ij} is equal to one, since c_{ij} is part of the average that defines b_j. Thus, computing the conditional expectations above is just an exercise of computing expectations of a normal random variable, conditioned on minimum or maximum values for that variable. □

5.1 Collusion Size That Defeats The FPD

Given the expected FPD output, in order to concurrently minimize the likelihood of false negatives η (i.e. the probability that a key index i in the collusion is not detected) and false positives ε_3, it would be desired to set $\delta_F = \text{E}[d_F]/2$. However, the FPD does not know K at detection time. Since the probability of a false positive does not depend on K, we set ε_3 to a constant value $\varepsilon_3 = \tau$ (typically $\tau \leqslant 10^{-12}$) which determines the detection bound δ_F.

COROLLARY 5. *To achieve $\varepsilon_3 \leqslant \tau$, FPD must set:*

$$\delta_F \geqslant \beta B\sqrt{\frac{2}{N}}\text{erfc}^{-1}(2\tau).$$

The detection threshold uniquely determines the probability of a false negative η. Since the FPD output, d_F, is normal with expected value $\text{E}[d_F]$ and variance $\sigma_{d_F}^2 = \sigma_{g_F}^2 = \beta^2 B^2/N$, we conclude that:

COROLLARY 6. *An object of size N produces*

$$\eta = \frac{1}{2}\text{erfc}\left(\frac{(E[d_F] - \delta_F)\sqrt{N}}{\sqrt{2}\beta B}\right), \text{or}$$

$$\eta = \frac{1}{2}\text{erfc}\left(\sqrt{\frac{N}{\pi K}}\exp\left(-\frac{K}{2B^2}\right) - \text{erfc}^{-1}(2\tau)\right).$$

The imminent goal of the collusion is to avoid detection at the FPD. From the latter equation, we can compute the minimal size of a collusion clique K_F that would have the probability of individual clique member detection $\eta(K_F)$ above a desired threshold $\eta \geqslant \tau_C$, where typically $\tau_C \geqslant 0.9$. Figure 4 illustrates an example of how $\eta(K)$ changes with the increase of K. For the example system depicted in Figures 3 and 4, we can identify that cliques of $K \geqslant K_F = 318$ WDKs would make every colluder virtually unidentifiable by our FPD as $\eta(K_F) \approx 0.9$.

We can compare this result to the K_W computed earlier in Corollary 2 and Theorem 1. Since $K_W = 2B^2\{\text{erf}^{-1}[(1-\theta)\beta^{-1}]\}^2$ WDKs are required to reduce the expected correlation of the WMD to $E[d_F] = \theta$, for the example illustrated in Figure 3 ($A = B = 7$ and $N = 4{\cdot}10^5$), we see that the collusion size that drops down $E[d_W] = 0$ is $K_W \geqslant 47$ WDKs (at $\beta = 2$).

Although the adversary can create a signal that can play as unmarked on almost all players in the world with colluded K_W WDKs, it would be foolish to expect such a collusion as

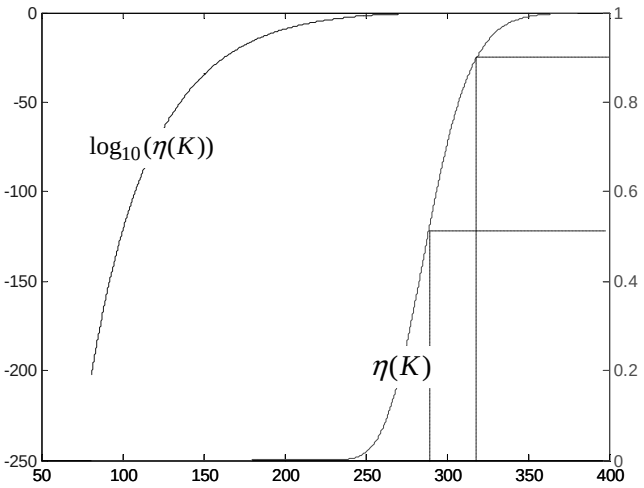

Figure 4: Diagram of $\eta(K)$ for $B = 7$, $\tau = 10^{-12}$, $N = 4 \cdot 10^5$. Collection of $K \geqslant 288$ WDKs imposes likelihood of detection $\eta \geqslant \tau_C = 0.5$.

each colluder would be easily identified.[1] Thus, we assume that the ultimate goal of the collusion clique is to average K_F WDKs such that each WDK is virtually undetectable at FPD time.

Using Corollary 6, we can compute the object size N necessary to achieve a desired probability η of false negatives in the FPD. We recall from the previous Section that the minimal collusion size to drive down $E[d_W] = \theta$ can be computed as $K_W = 2B^2\mu^2$, where $\mu = \text{erf}^{-1}[(1-\theta)\beta^{-1}]$ is fixed for a fixed attack efficiency (i.e. a fixed θ and fixed β). Therefore, as we increase B the attacker has to increase K_F proportionally to B^2, which imposes:

COROLLARY 7. *The object size N required to achieve a given η for fixed τ and μ is:*

$$N = K_F \pi \left[(\text{erfc}^{-1}(2\eta) + \text{erfc}^{-1}(2\tau))\exp(\mu^2)\right]^2.$$

This result thus confirms Theorem 1, i.e. that collusion size and object size are linearly related. We note that in fixing the WMD performance we obtained one constant of proportionality, whereas in fixing the FPD performance we obtained another. Therefore, in designing a practical system we determine the desired error probabilities, and select N as the largest of the values computed from the WMD and FPD equations.

Finally, during the forensic search, the FPD must test individually for each user's WDK h_i whether it is "contained" in the pirated copy $\hat{y}$. This search can be lengthy, considering the length of a single watermark $|w| > 0.5 \cdot 10^6$ and the cardinality of the user space $|C| > 10^9$. However, the search can be parallelized and effectively distributed over a network of computers, which can significantly reduce the search time. We estimate that for a large user space, the FPD search can be performed within a week using the idle cycles of a typical enterprize network of 1000+ computers.

[1]For fixed η, ε_1, and τ, and realistic β and θ parameters - $\beta \sim 2$ and $\theta \sim 0.1$ - it can be derived from Corollary 6 and Theorem 1 that $K_F > K_W$.

6. SEGMENTATION

In the WM/FP system, watermarks protect the content and fingerprints enable the copyright owner to identify a clique of users that launched an attack to remove the watermark. This unique property of the protection system, enables us to add multiple watermarks in the object and enforce the adversary to create cliques independently for each watermark. More formally, we divide the protected object into S segments and watermark each of them with a distinct spread spectrum sequence. For each segment i, we publish m distinct WDKs $h_{i,j}, j = 1..m$, created in accordance to the described dual WM/FP system. Each client gets a single WDK $h_{i,j}$ to exactly one segment.

6.1 Collusion Resistance

Object and collusion of any realistic size result in a probability of false positives (ε_3) close to zero such that it can be neglected. Because of this, we conveniently conclude that "a segment can resist K colluders" without mentioning error probabilities. A protected object is defeated if watermarks are removed from all segments, while no fingerprints are introduced in the process. In order to break the system, the adversary has to collect at least K WDKs for each segment. When published, WDKs are uniformly assigned to random segments. We assume that the total number of published WDKs mS surpasses significantly $mS \gg KS$. Thus, the effort of the adversary to collect a set of WDKs that would break the WM/FP system can be modeled as "coupon collecting" [10], [18].

DEFINITION 1. *The collusion resistance M of a segmented dual WM/FP system with S segments equals the* expected *number of WDKs needed to be selected from an infinite pool of WDKs, such that each segment has a collusion clique of at least K WDKs.*

Let X be a r.v. denoting the collusion size in a given segment, when we have S segments and overall M broken clients (i.e. extracted WDKs). X has a Poisson distribution with mean $\mu = M/S$. Let $p = \Pr[X \leq K]$. From ([1] Th.A.15 pp.239) $\Pr[X \leq \mu(1-\gamma)] \leq \exp(-\gamma^2\mu/2)$. In our case $K = \mu(1-\gamma)$, so $\gamma = 1 - K/\mu$ and $p \leq \exp[-(1 - K/\mu)^2\mu/2]$. Let $q = \Pr[\text{all segments contain more than K keys}]$. Then, assuming independence among segments, $q = (1-p)^S$, which for a small p becomes $q = 1 - pS$. If we shoot for $q = 1 - \varepsilon_s$, then $\varepsilon_s = pS$. So, $\varepsilon_s/S \leq \exp[-(1 - K/\mu)^2\mu/2]$. Plugging in $\mu = M/S$, we get $\ln(S/\varepsilon_s) \geq \frac{M}{2S} - K + \frac{SK^2}{2M}$, and solving for M, we get:

LEMMA 4. *If $M = S\left(\ln\frac{S}{\varepsilon_s} + K + \sqrt{\ln^2\frac{S}{\varepsilon_s} + 2K\ln\frac{S}{\varepsilon_s}}\right)$, then $q > 1 - \varepsilon_s$.*

THEOREM 2. *A dual WM/FP system with segmentation has super-linear collusion resistance.*

PROOF. For a fixed collusion resistance per segment K, the number of segments S is linear in object size $S = \mathcal{O}(N)$, and so, using Lemma 4, we get overall super-linear collusion resistance with respect to object size $M = \mathcal{O}(N \log N)$.

An alternate direction: when the asymptotic case of the "coupon collecting" problem is analyzed for $S \to \infty$ and minimal K collected WDKs per segment, collusion resistance M has a well known [18] sharp threshold at:

$$\lim_{S\to\infty} \Pr\left[M_K > S(\ln S + (K-1)\ln\ln S + c)\right] = e^{-e^{-c}}$$

for any real number $c \in \mathbb{R}$. This points to the fact that the number of WDKs, M, that the adversary needs to collect to cover at least K keys in S segments, is expected to be centered at $M = S \ln S + (K-1) S \ln \ln S$ with exceptionally small variance at both tails. $\square$

Note that the variance of the solution to the "coupon collecting" problem is exceptionally small [18]; thus, it is expected that for large number of segments within an object, two distinct attacks to the system require similar number of collected WDKs (within S keys) with exceptionally high probability. Thus, although the collection of WDKs during the attack is probabilistic, with great certainty we can assume that the resulting super-linearity of the collusion resistance is almost deterministic because of this sharp transition.

Segmentation is impossible in classic fingerprinting systems as they all require some form of a "marking assumption" [4], [9].

7. KEY COMPRESSION

The major drawback of the basic dual system is the requirement for a relatively large storage space for the detection keys. A brief problem overview: it is hard to compress the sum of two independent pseudo-random sequences, such that it is hard to infer the individual sequences. Let $g(s,n)$ denote the output of length n of generator g, given seed s. We need a way to create two generators g_1, g_2 with two seeds s_1, s_2 such that $\exists (g,s) \,|\, g_1(s_1,n) + g(s,n) = g_2(s_2,n)$ and the sequences $g_1(s_1,n)$ and $g(s,n)$ are mutually independent. This remains as an open problem. The current situation is that we need to create $g_1(s_1,n)$ and $g(s_2,n)$ independently in a secure machine and store their sum on a client. For realistic loads to the system, the length of the key is in the order of 10^5 bytes, which may be too much data for certain embedded devices.

7.1 Proposed Solution

Recall that the WDK of user i is created as $h_i = c_i + w$, where c_i and w are mutually independent. Instead, we can generate the key from a short seed using any standard cryptographically secure pseudo-random key generator, and per chosen w do sieving and select only those seeds for which the resulting long sequence (we denote it as s) has the property that $s \cdot w \geq 1$, thus, inferring $h_i = s$. The deviation of $s \cdot w$ is roughly $\sigma^* = B\sqrt{N_o}$, so the probability for a randomly chosen seed to meet this criteria is $\varepsilon^* = \frac{1}{2}\mathrm{erfc}(N_o/(B\sqrt{2}))$. For example, for $\varepsilon^* < 10^{-6}$ we get $N_o = 2B^2[\mathrm{erfc}^{-1}(2\varepsilon^*)]^2 = 2000$. Since $N = 10^5$, we partition the generation of h_i into N/N_o segments, where for each segment we perform sieving expected $1/\varepsilon^*$ times. For a seed size of $\xi = 100$ bits, we obtain a compression ratio of $N_o/\xi \sim 20$.

8. WM/FP SUMMARIZING DISCUSSION

The dual WM/FP technology aims at building practical secure content protection mechanisms. Although the main underlying theoretical concepts have been presented so far, in this section we focus on their interaction and practical implications. Solid overview of the mutual impact of scheme parameters can be obtained from Table 1. When designing a realistic protection system, several parameters are given as constants: total object size N_o and media variance A^2. All other parameters can be chosen such that the overall detection mechanisms are of desired quality.

The primary decision is to determine the number of segments S per object. Since collusion resistance within a single segment is $K = \mathcal{O}(N)$, where $N = N_o/S$ is the length of the segment, and collusion resistance achieved over S segments is $M = \mathcal{O}(S \ln(S))$, then the objective is to have as short as possible segments in order to: (i) maximize overall collusion resistance M and (ii) reduce the storage space for a single WDK. On the other hand, due to security measures for hiding w within a watermark carrier c_i, there exists a lower bound on the watermark carrier amplitude B, commonly set to $B \geqslant A$. Selection of B uniquely identifies the segment length N with respect to a desired probability of a false alarm ε_1 under the optimal sign(mean$(\cdot)$) attack. Such a setup directly impacts the maximal collusion size per segment K and maximal efficacy of the adversary in guessing SWK bits $1 - \varepsilon_2$. It also traces the guidelines for FPD detection performance ε_3 and η. Finally, η and N imply the collusion size K (computed from Corollary 6) required to make all colluders invisible at FPD time.

For realistic loads to the system, such as high-definition television, the number of bits per object ranges in the order of 10^{11} bytes. Assuming, one chip is embedded per 100 pixels, we derive an object size of $N \approx 10^9$ chips. On the other hand, from $B = A \approx 7$ and $\varepsilon_1 = 10^{-10}$, we derive $N \approx 4 \cdot 10^5$ chips. This boosts the number of segments to $S \approx 2.5 \cdot 10^3$. The resulting error probabilities are: (i) desired likelihood of an incorrectly guessed w_j bit during the sign(mean$(\cdot)$) attack of $\varepsilon_2 < 0.40$ can be achieved for $K = 3$ and (ii) for fixed false negative rate of $\varepsilon_3 = 10^{-12}$, the false positive rate follows the diagram in Figure 4, thus yielding a per-segment collusion resistance of $K \geqslant 318$ for $\eta \geqslant 0.9$. Most importantly, the achieved overall collusion resistance is lower-bounded by $M > 9.1 \cdot 10^5$ users. One can hardly expect that, under realistic piracy scenarios, such a clique could be established to oppose the protection of the proposed dual WM/FP system.

One disadvantage of the dual WM/FP system is content collusion, where L media clips marked with an identical watermark w are used to estimate w using the optimal collusion attack (see Subsection 4.1): $v = \sum_{i=1}^{L} x_i + w$. In order to reduce the sensitivity to this type of an attack, the set $\{w, C\}$ needs to be renewed after 100 or so movies. The WDKs are distributed to user players using standard cryptographic tools for authenticated communication.

9. COMPARISON WITH OTHER COPYRIGHT PROTECTION SYSTEMS

In this section, we outline the main characteristics of the existing technologies for copyright enforcement and compare them to the proposed dual WM/FP system (see Table 2 for brief comparison overview).

9.1 Public-Key Watermarking

Public-key WM systems have been mainly focusing on providing a solution to the "prisoners' problem" introduced by Simmons [23]. This problem requires two trusted parties (i.e. prisoners) to establish a covert communication channel in the presence of a "warden". Anderson suggested to encrypt the embedded message with the public key of the

WM/FP Parameter	Related Parameter Dependencies
$\varepsilon_1 = \Pr[d_w > \delta_w \mid \text{object is not marked}]$	$\sim \text{erfc}\left(\frac{\sqrt{N}}{AB}\right)$
Segment length: N	$\sim B^2A^2\left[\text{erf}^{-1}(1-2\varepsilon_1)\right]^2$
$\varepsilon_2 = \Pr[v_j \neq w_j]$	$\sim \text{erfc}\left(\frac{\sqrt{K}}{B}\right)$
Collusion resistance per segment: K	$\mathcal{O}(N)$
$\varepsilon_3 = \Pr[d_F(c_i) > \delta_F \mid c_i \notin \mathcal{K}]$	$\sim \text{erfc}\left(\frac{\sqrt{N}}{B}\right)$
System collusion resistance: M	$\mathcal{O}(N(\log(N)))$

Table 1: Dependencies among main parameters of the WM/FP system.

	Traitor Tracing	Fingerprinting	Dual WM/FP
Primary target application	Detection of pirated players. Content considered free after delivery.	Copyright enforcement.	
How to create a clean copy of the content?	Decrypt and capture content.	Users collude their content copies.	Users collude their keys to remove the protection mark.
Content distribution	Real-time broadcast. Single encrypted copy of content distributed.	Each user receives a unique copy.	A single watermarked copy. Each user has a distinct detection key.
Collusion resistance	Low (hundreds)	Low (tens)	High (millions)
Trace-back mechanism	Player confiscation. Player response to a probe with invalid ciphertext reveals colluders.	Analysis of pirated content. Fingerprint detector can compare the pirated content to the original copy and the individual marks.	
Advantages	Protocols can be based on provably hard problems.	No action required at client side. Players remain unchanged.	High collusion resistance. Copyright is enforced through prevention.
Disadvantages	Difficult to enforce player confiscation. Low collusion resistance.	Exceptionally low collusion resistance. Fraud is not prevented but only detected.	Marking key needs to be replaced after marking 100+ media clips.

Table 2: Comparison of main characteristics of content screening technologies: traitor tracing, fingerprinting, and the dual WM/FP system.

recipient prior to WM embedding, so that a warden would not be able to understand it [3]. Craver has extended this protocol to the case of an active warden [7]. Both protocols do not fit the requirements of a content screening system which aims at achieving a much harder task: a server sending one bit to a set of clients such that if the adversary fully controls any client the adversary cannot interfere with the communication of the server with other clients.

9.2 Fingerprinting

Ergun et al. consider embedding distinct spread sequences per copy and are among the first to formalize the metrics of attacks (the limits beyond which a copy would be considered too corrupt to be useful) [9]. Their paper considers one attack: averaging of fingerprinted copies with additional noise. This attack is weaker than the attacks considered by Boneh and Shaw [4], and accordingly the upper bound on collusion size that can be overcome is much higher in [9]. Boneh and Shaw construct fingerprint codes which in the worst case produce collusion resistance $K = \mathcal{O}(N^{1/4})$. Pfitzmann and Waidner have introduced a FP scheme where buyers can buy digital content anonymously, but they can be identified if they redistribute the fingerprinted content [19].

The copyright protection approach of Fiat and Tassa [11] is similar to that of [4]. Users are randomly sub-grouped into r subsets, each getting a distinct symbol out of r symbols. After some subgroup is identified as including pirates the search continues with that subset only. It is repartitioned into r smaller subsets, and so on. This is called dynamic traitor tracing. It is slightly more efficient than static tracing, where the whole universe of users is repartitioned, but due to the fast convergence of the search process the difference is not dramatic. The approach of [11] is less realistic than that of [4] in the following aspect. The former assumes that per round of the above tracing process the pirates simply choose one of the symbols available to them. The assumption of [4] is that on bits where a collusion disagrees they may choose any value. Symbols are composed of many bits. Thus, the collusion may create new symbols not in the original alphabet.

9.3 Traitor Tracing

It is important to distinguish the differences between a traitor-tracing (TT) and an FP system for copyright protection.[2] The goal of a TT system (as stated in [6]) is the same as in FP systems, but the scenario and the means are different. The content is usually broadcast in real-time and has very little value afterwards. Pirates are assumed to be unable to manipulate and re-broadcast content in (near) real-time. The content is encrypted, and legitimate clients have distinct sets of keys that when combined enable decryption.

[2]In particular, one cannot blindly export error correction ideas from TT to classic digital FP, where some form of the Boneh-Shaw "marking assumption" is necessary.

Each legitimate set of keys is uniquely associated with a single client. If a pirate resells his keys to others, and the box of a suspect client is confiscated, then law-enforcement can use the set of keys in the confiscated box to trace back the leak. However, a large enough collusion can create a good set of keys that will not incriminate any of the culprits.

Boneh and Franklin constructed a public key encryption scheme in which there is one public encryption key, and many private decryption keys [5]. If a broadcaster encrypts once with the public key, then each legitimate receiver can decrypt with a different private key. If a coalition of receivers collude to create a new decryption key, then there is an efficient algorithm to trace the new key to its creators.

Kiayias and Yung have established a black-box traitor tracing model in which the pirate-decoder employs a self-protection technique [15]. They proved that any system that does not meet certain well defined combinatorial conditions cannot overcome collusion of size superlogarithmic in object size. To the best of our knowledge the Chor, Fiat, Naor system [6] is currently the only TT scheme for which the Kiayias-Yung conditions hold.

The main drawback of all TT systems (including black-box TT) is that they require physical confiscation of a suspect client machine in order to examine it, and the assumption is that the protected content itself cannot be traded by pirates. This limits significantly the scenarios in which TT can be applied.

10. CONCLUSION

We have introduced a new dual WM/FP system, where all copies of a protected object are identically watermarked using an SWK, but where individual WDKs are distinct. By knowing a WDK, an adversary cannot recreate the original from the marked content. However, knowledge of any one WDK is sufficient for modifying the object so that a detector using that key does not detect marks. Such a modified object necessarily contains a fingerprint: sufficient information to point at the WDK used to break the detector.

Our dual WM/FP system limits the scope of possible attacks, when compared to classic fingerprinting systems. Under optimal attacks, the size of the collusion necessary to remove the marks without leaving a detectable fingerprint is asymptotically $K = \mathcal{O}(N)$ without segmentation, and $M = \mathcal{O}(N \log(N))$ with segmentation, where N denotes object size. Classic fingerprinting has a lower bound on collusion resistance that is roughly $\mathcal{O}(N^{1/4})$. Thus, by using the dual WM/FP system one can achieve content protection with collusion resistance of up to 900,000 users for a two-hour high-definition video, for example.

REFERENCES

[1] N. Alon, J. H. Spencer, and P. Erdős, "The Probabilistic Method," Wiley-Interscience series in Discrete Mathematics and Optimization. New York: Wiley, 1992.

[2] R. J. Anderson and F. A. P. Petitcolas, "On the limits of steganography," IEEE Journal in Selected Areas in Communications, vol.16, pp.474–481, 1998.

[3] R. J. Anderson, "Stretching the limits of steganography," Workshop on Information Hiding, pp.39–48, 1996.

[4] D. Boneh and J. Shaw. "Collusion secure fingerprinting for digital data," IEEE Transactions on Information Theory, vol.44, pp.1897–1905, 1998.

[5] D. Boneh and M. Franklin, "An efficient public key traitor tracing scheme," Crypto, LNCS, vol.1666, pp.338–353, New York: Springer-Verlag, 1999.

[6] B. Chor, A. Fiat, and M. Naor, "Tracing traitors," Crypto, LNCS vol.839, pp.257–270, New York: Springer-Verlag, 1994.

[7] S. Craver, "On Public-key Steganography in the Presence of an ActiveWarden," Workshop on Information Hiding, pp.355–368, 1998.

[8] J.J. Eggers, J.K. Su, and B. Girod, "Public Key Watermarking By Eigenvectors of Linear Transforms," European Signal Processing Conference, vol.3, 2000.

[9] F. Ergun, J. Kilian, and R. Kumar, "A Note on the Limits of Collusion-Resistant Watermarks," Eurocrypt, 1999.

[10] W. Feller, "An introduction to probability theory and its applications," New York: Wiley - Series in Probability and Mathematical Statistics, 1968.

[11] A. Fiat and T. Tassa, "Dynamic traitor tracing," Crypto, LNCS, vol.1666, pp.354–371. New York: Springer-Verlag, 1999.

[12] T. Furon and F. P. Duhamel, "Robustness of An Asymmetric Watermarking Method," IEEE International Conference on Image Processing, vol.III, pp.21–24, 2000.

[13] F. Hartung and B. Girod, "Fast public-key watermarking of compressed video," IEEE International Conference on Image Processing, 1997.

[14] S. Katzenbeisser and F. A. P. Petitcolas, Eds. "Information Hiding Techniques for Steganography and Digital Watermarking," Boston, MA: Artech House, 2000.

[15] A. Kiayias and M. Yung, "Self Protecting Pirates and Black-Box Traitor Tracing," Crypto, LNCS, vol.2139, pp.63–79, New York: Springer-Verlag, 2001.

[16] D. Kirovski and H. S. Malvar, "Robust spread-spectrum audio watermarking," IEEE International Conference on Acoustics, Speech, and Signal Processing, 1999.

[17] J.-P. M. G. Linnartz and M. Van Dijk, "Analysis of the sensitivity attack against electronic watermarks in images," Workshop on Information Hiding, pp. 258–72, 1998.

[18] R. Motwani and P. Raghavan, "Randomized Algorithms," Cambridge: University Press, 1995.

[19] B. Pfitzmann and M. Waidner, "Anonymous Fingerprinting," Eurocrypt, vol.1233, pp.88–102, 1997.

[20] W. H. Press, B. P. Flannery, S. A. Teukolsky, and W. T. Vetterling, "Numerical Recipes in C: The Art of Scientific Computing," Cambridge University Press, Cambridge, 1988.

[21] Recording Industry Association of America. See `http://www.riaa.org`.

[22] Secure Digital Music Initiative. See `http://www.sdmi.org`.

[23] G. J. Simmons, "Prisoners' Problem and the Subliminal Channel," Crypto, pp.51–67, New York: Springer-Verlag, 1984.

[24] H. L Van Trees. "Detection, Estimation, and Modulation Theory," Part I, New York: John Wiley and Sons, 1968.

[25] Y. Yacobi, "Improved Boneh-Shaw fingerprinting," RSA Conference, 2001.