Considering Security and Robustness Constraints for Watermark-based Tardos Fingerprinting

Benjamin Mathon, Patrick Bas, François Cayre, Benoît Macq

Considering Security and Robustness Constraints for Watermark-based Tardos Fingerprinting

Benjamin Mathon [#*1], Patrick Bas [+2], François Cayre [#3], Benoît Macq [*4]

[#] *GIPSA-Lab, Grenoble-INP*
961 rue de la Houille Blanche - BP 46, 38402 GRENOBLE CEDEX , FRANCE
[1] benjamin.mathon@grenoble-inp.fr
[3] francois.cayre@grenoble-inp.fr

[+] *LAGIS, Ecole Centrale Lille*
Avenue Paul Langevin - BP 48, 59651 VILLENEUVE D'ASCQ CEDEX , FRANCE
[2] patrick.bas@ec-lille.fr

[*] *TELE, Université Catholique de Louvain*
Place du Levant 2, Bâtiment Stévin, 1348 LOUVAIN-LA-NEUVE, BELGIUM
[4] benoit.macq@uclouvain.be

***Abstract*—This article is a theoretical study on binary Tardos' fingerprinting codes embedded using watermarking schemes. Our approach is derived from [1] and encompasses both security and robustness constraints. We assume here that the coalition has estimated the symbols of the fingerprinting code by the way of a security attack, the quality of the estimation relying on the security of the watermarking scheme. Taking into account the fact that the coalition can perform estimation errors, we update the Worst Case Attack, which minimises the mutual information between the sequence of one colluder and the pirated sequence forged by the coalition. After comparing the achievable rates of the previous and proposed Worst Case Attack according to the estimation error, we conclude this analysis by comparing the robustness of no-secure embedding schemes versus secure ones. We show that, for low probabilities of error during the decoding stage (e.g. highly robust watermarking schemes), security enables to increase the achievable rate of the fingerprinting scheme.**

I. Introduction

Active fingerprinting or traitor tracing consists of marking copies of a digital content such as a Video On Demand movie that a distributor wants to provide to its users. Each copy is marked by the distributor with a sequence which identify one specific user and the fingerprinting code is used to trace any illegal copy of the document on file sharing networks. However, a coalition of malicious users can try to mix their copies in order to forge a new content so the distributor could not trace any users of the coalition. This attack is called a collusion attack. Probabilistic codes have been introduced by Boneh and Shaw [2] and Gabor Tardos' probabilistic codes [3] have been proposed in order to offer an optimal solution for collusion-secure fingerprinting because length m of the codes meets the Peikert's theoretical lower bound [4]: $m = O\left(c^2 \log\left(\frac{n}{p_{fa}}\right)\right)$, where c is the size of the coalition, n the number of users, and p_{fa} the probability of accusing an innocent (the probability of false alarm).

Recent works on Tardos' probabilistic codes [1] propose an attack which minimises the mutual information between a pirated sequence forged by the colluders and the initial sequence of one of these colluders. This attack is called the "Worst Case Attack" (WCA) and it also both maximises the p_{fa} and minimises the probability of not accusing an adversary.

It is important to point out that practically any fingerprinting code used to trace a multimedia content such as a movie has to be embedded using a watermarking technique. The transparency of the embedding guaranties that the quality of the content is not degraded by the fingerprinting code. The inherent robustness of watermarking techniques enables to extract the fingerprinting code whenever the media suffers recoding, noise addition, or transcoding. Examples of such practical implementations have already been proposed in the literature [5][6][7].

One aspect that have not been studied in the framework mixing watermarking and fingerprinting technologies is the possible impact of the security of the watermarking scheme [8] on the strategies available to the coalition of colluders. Tardos' fingerprinting codes have been designed assuming that the colluders are able to know exactly the positions where the symbols of their codes differ. Nevertheless, the use of secure embedding schemes such as proposed in [9][10][11] enables to lure the colluders by implying prediction errors on the symbols of their codes. Note that the security of a watermarking scheme can be assessed by performing a estimation of secret key using unsupervised learning techniques such as ICA[12][13][10][14], PCA[15] or clustering [16].

We propose in this article Worst Case Attacks on Tardos' codes with a constraint given by the security of the watermarking scheme used for data-hiding. We consider an estimation error ϵ of the sequences decoded by the colluders

MMSP'10, October 4-6, 2010, Saint-Malo, France.
US June 30, 2010 20:42:00 CEST

and we propose attacks in this framework. We compare it with classical WCA by computing false alarm probabilities and mutual information. Moreover we simulate robustness attacks, given a channel bit error rate η, we quantify the robustness of watermarking methods offering different degree of security compared to totally insecure ones.

II. Notations

We first list some notational conventions used in this article. Functions are noted in roman fonts, sets in calligraphy fonts and variables in italic fonts. Vectors and matrices are set in bold fonts, vectors are written in small letters and matrices in capital ones. $\mathbf{x}(i)$ is the i-th component of a vector $\mathbf{x}$. As for the C programming language, all indexes start from 0. We write $(\mathbf{x}(0)\dots\mathbf{x}(m-1))$ the content of a vector $\mathbf{x}$ of length m. If n is an integer, $[n]$ denotes the set $[\![0; n-1]\!]$. $\#\mathcal{A}$ is the cardinality of the set $\mathcal{A}$. $\binom{n}{k}$ denotes the number of k-combinations from a set of n elements. $\mathcal{M}_{n,m}(\mathbb{K})$ denotes the set of n-by-m matrices whose components are in the division ring $\mathbb{K}$.

III. Basics on Tardos' traitor tracing codes

A. Construction

The binary fingerprinting codes constructed by G. Tardos [3] for n users are constructed as follows: we consider a matrix $\mathbf{X} \in \mathcal{M}_{n,m}(\mathbb{F}_2)$. Each row $\mathbf{x}_j$ of the matrix $\mathbf{X}$ is a sequence of m bits which identify the user $j \in [n]$. The columns of $\mathbf{X}$ (the i-th bits of users) are generated according to a Bernoulli distribution $\mathcal{B}(p_i)$. $\{p_i\}_{i\in[m]}$ are distributed in the set $[0,1]$, according to the random variable P with p.d.f. $\mathrm{f}_P(p)$:

$$\mathrm{f}_P(p) = \frac{1}{\pi\sqrt{p(1-p)}}. \tag{1}$$

A group of c colluders $\mathcal{C} = \{j_0 \dots j_{c-1}\}$ work together in order to forge a pirated sequence $\mathbf{y}$ of m bits by combining the bits of their sequences:

$$\mathbf{y} = \left(\mathbf{x}_{j'_0}(0)\dots\mathbf{x}_{j'_{m-1}}(m-1)\right), \tag{2}$$

with $\left(j'_0\dots j'_{m-1}\right) \in \mathcal{C}^m$ and are chosen according to a strategy beforehand defined by the colluders. This property enables the "marking assumption": if the colluders have the same symbol at $i \in [m]$, so will the pirated sequence. The construction of the codes enables the identification of at least one of the colluders.

B. Accusation process

For identifying at least one of the colluders, we use the accusation function proposed by G. Tardos and improved by Skoric *et al.* [17]. The accusation is based on the construction of a matrix $\mathbf{U} \in \mathcal{M}_{n,m}(\mathbb{R})$:

$$\mathbf{U}(j,i) = \begin{cases} \mathrm{g}_1(p_i), & \text{if } \mathbf{y}(i)=1, \mathbf{x}_j(i)=1, \\ \mathrm{g}_0(p_i), & \text{if } \mathbf{y}(i)=1, \mathbf{x}_j(i)=0, \\ \mathrm{g}_0(1-p_i), & \text{if } \mathbf{y}(i)=0, \mathbf{x}_j(i)=1, \\ \mathrm{g}_1(1-p_i), & \text{if } \mathbf{y}(i)=0, \mathbf{x}_j(i)=0, \end{cases} \tag{3}$$

with:

$$\mathrm{g}_1(p) = \sqrt{\frac{1-p}{p}},\ \mathrm{g}_0(p) = -\sqrt{\frac{p}{1-p}}. \tag{4}$$

The score of a user $j \in [n]$ is defined by S_j:

$$\mathrm{S}_j = \sum_{i=0}^{m-1} \mathbf{U}(j,i). \tag{5}$$

An user $j \in [n]$ is accused if $\mathrm{S}_j > T$ where T is a specified threshold.

IV. Worst case attacks and security

A. Worst Case Attack for insecure watermarking schemes

In [18], the authors define embedding security classes in the WOA (Watermarked Only Attack) framework, adversaries have only access to several marked contents and try to estimate the secret key used for embedding. If the scheme is in the *insecurity* class, users could be able to estimate embedded messages. If we use the terminology of [19] for the classes of collusion, the colluders would be *sighted*. On the other hand, with a totally secure watermarking scheme, the colluders would not be able to say if their symbols are "0" or "1" and the only strategy they will have would be to randomly chose one symbol among the whole coalition for each position $i \in [m]$.

A strategy defines the process used by the colluders to generate a pirated sequence $\mathbf{y}$. For each $i \in [m]$, the value of $\mathbf{y}(i)$ depends on the number of "1" symbols that the coalition have at this position. A strategy is completely defined by the vector $\theta = (\theta(0)\dots\theta(c))$ where $\theta(k) = \Pr\left(\mathbf{y}(i)=1 | \sum_{j\in\mathcal{C}} \mathbf{x}_j(i) = k\right)$. We assume that the coalition always uses the same strategy for each bit of the pirated sequence.

In [1] the authors compute the "Worst Case Attack" (WCA), they propose the strategy θ which minimises the achievable rate of a fingerprinting scheme $\mathrm{R}_s(\theta) = \mathbb{E}_P[I(Y; X_{j_0})|P = p]$, where I denotes the mutual information, Y and X_j are two random variables which respectively denote the binary symbol at one position in the pirated sequence and in the sequence of the colluder j_0.

B. Worst Cases Attack for secure watermarking schemes

The security of the watermarking scheme is mathematically expressed by the fact that each colluder estimates his symbols with an error ϵ. The more secure the embedding scheme is, the closer ϵ is to 0.5. The former case of classical embedding is insecure and consequently $\epsilon = 0$ in this case. Considering this new assumption, the colluders are only able to decode properly their sequences $\mathbf{x}_j$ if $\epsilon = 0$. Moreover, they cannot say if one colluder has the same symbol than another one. Note that because of the Kerckhoffs' principle, we can assume that each member of the coalition knows the security of the used watermarking embedding scheme and consequently knows the error ϵ made by their estimation process. Consequently, we denote $\mathbf{z}_j(i)$ the symbol decoded by the colluder j at position i and Z_j the associated random variable with the property:

$$\Pr\left(Z_j = 1 | X_j = 0\right) = \Pr\left(Z_j = 0 | X_j = 1\right) = \epsilon. \tag{6}$$

The coalition then forges a sequence $\mathbf{z}$ by using the strategy θ which is now defined for each bit $i \in [m]$ by:

$$\theta(k) = \Pr\left(\mathbf{z}(i) = 1 \middle| \sum_{j\in\mathcal{C}} \mathbf{z}_j(i) = k\right). \tag{7}$$

The pirated sequence $\mathbf{y}$ is constructed as follows:

$$\forall i \in [m],\ \mathbf{y}(i) = \mathbf{x}_{j'}(i), \tag{8}$$

where j' is uniformly chosen in $\{j \in \mathcal{C} : \mathbf{z}_j(i) = \mathbf{z}(i)\}$. Fig. 1 shows an example of this process for $c = 5$ colluders.

Note that secure watermarking schemes imply also the marking assumption: whenever the coalition receives identical symbols, the results of the strategy will still output the very same symbol because of Eq. (8).

The achievable rate R_s (in bits/sample) of the fingerprinting scheme is defined by [1][20]:

$$\begin{aligned}\mathrm{R}_s(\theta,\epsilon) &= \mathbb{E}_P[I(Y;X_{j_0})|P=p]\\ &= \mathbb{E}_P[H(Y) - H(Y|X_{j_0})|P=p]\\ &= \mathbb{E}_P[H(Y) - (pH(Y|X_{j_0}=1)\\ &\quad + (1-p)H(Y|X_{j_0}=0))|P=p],\\ &= \mathbb{E}_P[H_b(p_1) - (pH_b(p_2) + (1-p)H_b(p_3)|P=p].\end{aligned} \tag{9}$$

where $H(.)$ and $H_b(.)$ denote respectively the entropy and the binary entropy, the probability p_1, p_2 and p_3 are given by:

$$p_1 = \Pr(Y=1), \tag{10}$$

$$p_2 = \Pr(Y=1|X_{j_0}=1), \tag{11}$$

$$p_3 = \Pr(Y=1|X_{j_0}=0). \tag{12}$$

Given:

- X_{j_0} the bit of the colluder j_0,
- Z_{j_0} the estimated bit of the colluder j_0,
- Z the estimated bit chosen by the strategy,
- Y the bit of the pirated sequence,
- $\sum_{j\in\mathcal{C}} X_j = \Sigma_X$,
- $\sum_{j\in\mathcal{C}} Z_j = \Sigma_Z$,
- $\theta(k) = \Pr(Z=1|\Sigma_Z=k)$ the strategy.

We now compute the analytic expressions of p_1, p_2 and p_3.

1) Derivation of p_1: we have:

$$\begin{aligned}p_1 = \sum_{l=0}^{c}\sum_{k=0}^{c} &\Pr(\Sigma_X=l,\Sigma_Z=k)\\ &\times \Pr(Y=1|\Sigma_X=l,\Sigma_Z=k).\end{aligned} \tag{13}$$

We introduce the random variable V, which corresponds to the number of $X_j = 1$ which have been decoded to $Z_j = 0$, $V = \#\{j \in \mathcal{C} : X_j = 1, Z_j = 0\}$. For $l, k \in [c+1]$; V gets its values in the set $\Omega = \{i \in \mathbb{N} : i \le l; i \le c-k; i \ge l-k\}$. We obtain:

$$\begin{aligned}p_1 = \sum_{l=0}^{c}\Bigg(&\Pr(\Sigma_X=l)\sum_{k=0}^{c}\Big(\Pr(\Sigma_Z=k|\Sigma_X=l)\\ &\times\sum_{i\in\Omega}\big(\Pr(Y=1|V=i,\Sigma_X=l,\Sigma_Z=k)\\ &\times\Pr(V=i|\Sigma_X=l,\Sigma_Z=k)\big)\Big)\Bigg),\end{aligned} \tag{14}$$

where the four probabilities involved in the equation can be computed using classical combinatorial analysis and definition of conditional probabilities:

$$\Pr(\Sigma_X=l) = \binom{c}{l}p^l(1-p)^{c-l},$$

$$\begin{aligned}&\Pr(\Sigma_Z=k|\Sigma_X=l)\\ &=\sum_{i\in\Omega}\binom{l}{i}\binom{c-l}{k-l+i}\epsilon^i(1-\epsilon)^{l-i}\epsilon^{k-l+i}(1-\epsilon)^{c-k-i},\end{aligned}$$

$$\begin{aligned}&\Pr(Y=1|V=i,\Sigma_X=l,\Sigma_Z=k)\\ &=\theta(k)\frac{l-i}{k} + (1-\theta(k))\frac{i}{c-k},\end{aligned}$$

$$\begin{aligned}&\Pr(V=i|\Sigma_X=l,\Sigma_Z=k)\\ &=\frac{\binom{l}{i}\binom{c-l}{k-l+i}\epsilon^i(1-\epsilon)^{l-i}\epsilon^{k-l+i}(1-\epsilon)^{c-k-i}}{\sum_{t\in\Omega}\binom{l}{t}\binom{c-l}{k-l+t}\epsilon^t(1-\epsilon)^{l-t}\epsilon^{k-l+t}(1-\epsilon)^{c-k-t}}.\end{aligned}$$

2) Derivation of p_2 and p_3: we look now for:

$$p_2 = \Pr(Y=1|X_{j_0}=1) = \Pr_1(Y=1),$$

and:

$$p_3 = \Pr(Y=1|X_{j_0}=0) = \Pr_0(Y=1),$$

with $\Pr_1(.) \equiv \Pr(.|X_{j_0}=1)$ and $\Pr_0(.) \equiv \Pr(.|X_{j_0}=0)$. Again, using combinatorial analysis, we obtain:

$$\begin{aligned}p_2 = \sum_{l=1}^{c}\Bigg(&\Pr_1(\Sigma_X=l)\sum_{k=0}^{c}\Big(\Pr(\Sigma_Z=k|\Sigma_X=l)\\ &\times\sum_{i\in\Omega}\big(\Pr(Y=1|V=i,\Sigma_X=l,\Sigma_Z=k)\\ &\times\Pr(V=i|\Sigma_X=l,\Sigma_Z=k)\big)\Big)\Bigg),\end{aligned} \tag{15}$$

and:

$$\begin{aligned}p_3 = \sum_{l=0}^{c-1}\Bigg(&\Pr_0(\Sigma_X=l)\sum_{k=0}^{c}\Big(\Pr(\Sigma_Z=k|\Sigma_X=l)\\ &\times\sum_{i\in\Omega}\big(\Pr(Y=1|V=i,\Sigma_X=l,\Sigma_Z=k)\\ &\times\Pr(V=i|\Sigma_X=l,\Sigma_Z=k)\big)\Big)\Bigg),\end{aligned} \tag{16}$$

with:

$$\Pr_1(\Sigma_X=l) = \binom{c-1}{l-1}p^{l-1}(1-p)^{c-l},$$

and:

$$\Pr_0(\Sigma_X=l) = \binom{c-1}{l}p^l(1-p)^{c-l-1}.$$

Eq. (14), (15), (16) are consequently used to compute the binary entropy functions of Eq. (9) which have to be averaged using numerical integration.

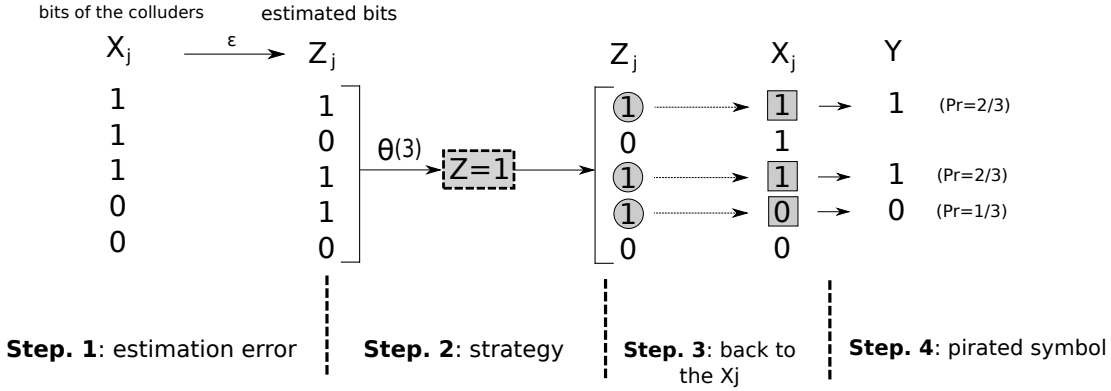

Figure 1. Collusion process for a secure watermarking scheme with $c = 5$ colluders and $\theta(3) = 1$. **Step. 1**: the colluders decode three "1" symbols Z_j. **Step. 2**: because $\theta(3) = 1$, the strategy gives $Z = 1$. **Step. 3**: the coalition looks for the X_j which correspond to the $Z_j = Z =$"1". **Step. 4**: the pirated symbol Y is randomly chosen among the selected X_j.

C. Comparison between the WCA and the $\epsilon - WCA$

We compute the ϵ-Worst Case Attack, e.g. the strategy $\theta_{\epsilon\text{-WCA}}$ which minimises the achievable rate given by Eq. (9). The minimisation step was performed using the Simplex algorithm [21]. For $c = 2$, $\forall\epsilon$, $\theta_{\epsilon\text{-WCA}}$ = (0. 0.5 1.). Tab. I shows $\theta_{\epsilon\text{-WCA}}$ for $c = 3, 4$ and several values of ϵ. Interestingly, we notice that the two different strategies converge toward an alternating *deterministic* strategy whenever the estimation error grows[1].

	$c = 3$	$c = 4$
$\epsilon = 0.$	(0. 0.651 0.349 1.)	(0. 0.487 0.5 0.513 1.)
$\epsilon = 0.05$	(0. 0.726 0.274 1.)	(0. 0.543 0.5 0.457 1.)
$\epsilon = 0.1$	(0. 0.830 0.170 1.)	(0. 0.620 0.5 0.379 1.)
$\epsilon = 0.15$	(0. 0.982 0.018 1.)	(0. 0.734 0.5 0.266 1.)
$\epsilon = 0.2$	(0. 1. 0. 1.)	(0. 0.908 0.5 0.091 1.)
$\epsilon > 0.2$	(0. 1. 0. 1.)	(0. 1. 0.5 0. 1.)

Table I
VALUES OF $\theta_{\epsilon\text{-WCA}}$ FUNCTIONS OF ϵ FOR $c = 3, 4$. FOR $c = 2$, FOR ALL ϵ, $\theta_{\epsilon\text{-WCA}}$ = (0. 0.5 1.).

Fig. 2 shows an estimation of the probability of false alarm p_{fa} (probability of accusing an innocent) functions of ϵ for three strategies: blind, WCA and ϵ-WCA. For the blind strategy, each component of $\mathbf{y}$ is chosen uniformly among the colluders, $\theta_{blind}(k) = k/c$. We estimate the p_{fa} by an expectation of 1000 observations using rare event analysis as in [1]. As expected, the performance of the WCA and ϵ-WCA attacks are better than the blind attack when $\epsilon = 0$. Moreover, performances of ϵ-WCA are better than WCA when ϵ is close to 0.25 because the difference between the p_{fa} is higher and consequently the probability for a colluder to be accused decreases.

We can see these performances on Fig. 3 which shows the values for $\mathrm{R}_s(\theta)$ functions of ϵ for the three strategies. As expected, the mutual information between the pirated sequence and the sequence of one colluder is weaker for ϵ-WCA than for WCA.

[1]The analysis has to be carried on for more important c but for this we need to deal with optimisation problems in high dimensional spaces.

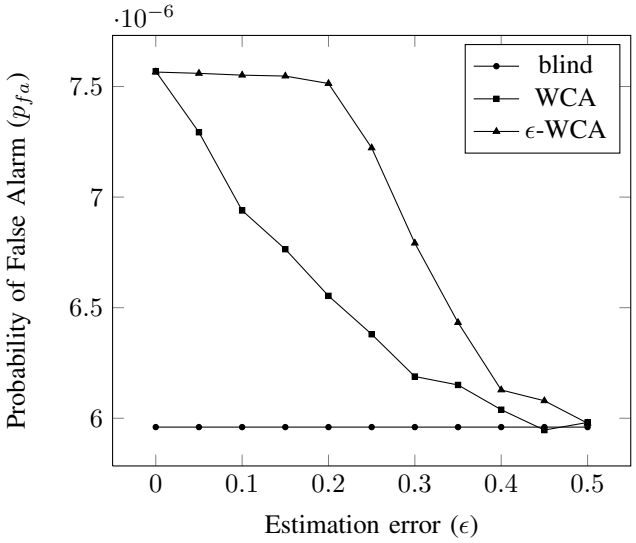

Figure 2. Estimation of the false alarm probability w.r.t the estimation error ϵ for three attacks: WCA, ϵ-WCA and blind. Parameters: $m = 400$, $T = 80, c = 4$.

V. SECURITY VS ROBUSTNESS

We now consider the effects of attacks as compression or Gaussian noise addition on digital documents where the codes are hidden. Instead of decoding each symbol $\mathbf{y}(i)$, the distributor decodes $\mathbf{y}'(i)$ with a BER (channel bit error rate) η modeling a Binary Symmetric Channel (BSC). The corresponding random variable Y' is defined by:

$$\Pr(Y' = 1 | Y = 0) = \Pr(Y' = 0 | Y = 1) = \eta. \qquad (17)$$

The goal of this section is to compare the achievable rates of embedding schemes that are insecure ($\epsilon = 0$) and schemes that are secure ($\epsilon \neq 0$) but including a BSC channel of characteristic η which takes into account the robustness of the scheme. We compute the achievable rate $\mathrm{R}'_s(\theta, \epsilon, \eta)$ (in bits/sample) defined by:

$$\mathrm{R}'_s(\theta, \epsilon, \eta) = \mathbb{E}_P[I(Y'; X_{j_0}) | P = p]. \qquad (18)$$

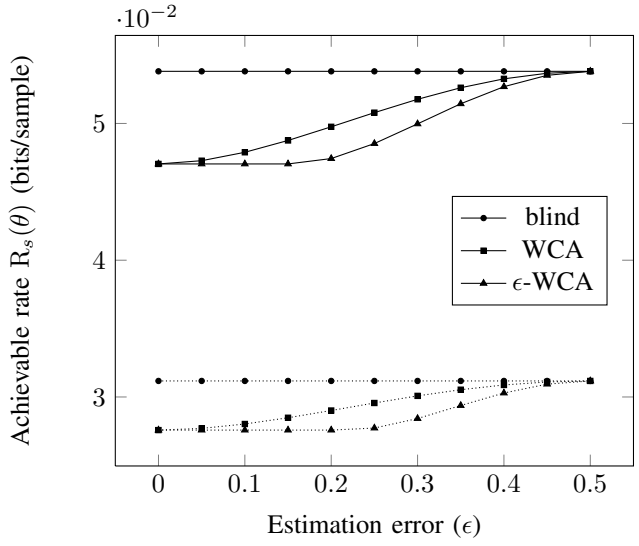

Figure 3. Values of $\mathrm{R}_s(\theta) = \mathbb{E}_P(I(X,Y))$ w.r.t the estimation error ϵ for $c = 3$ (solid) and $c = 4$ (dotted).

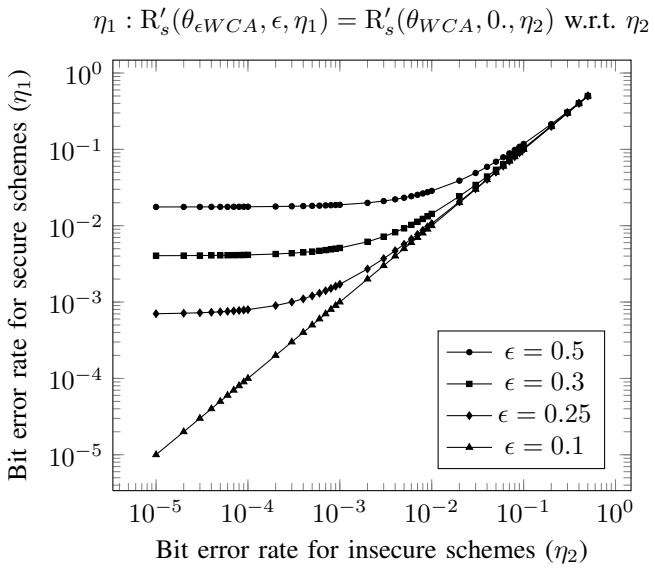

Figure 4. Bit error rate η_1 (secure schemes) w.r.t. bit error rate η_2 (insecure schemes) for $\epsilon = 0.5, 0.3, 0.25, 0.1$, $c = 4$.

We compute $\mathrm{R}'_s(\theta, \epsilon, \eta)$ with the same method as in Eq. (9) with:

$$p'_1 = \Pr(Y' = 1) = (1-\eta)p_1 + \eta(1-p_1), \quad (19)$$

$$p'_2 = \Pr(Y' = 1 | X_{j_0} = 1) = (1-\eta)p_2 + \eta(1-p_2), \quad (20)$$

$$p'_3 = \Pr(Y' = 1 | X_{j_0} = 0) = (1-\eta)p_3 + \eta(1-p_3). \quad (21)$$

In Fig. 4, for $c = 4$ colluders, given ϵ, we find the BER η_1 such as the achievable rate R'_s after ϵ-WCA (the strategy of the coalition is the one devised in IV-B) is the same for insecure schemes ($\epsilon = 0$) given a BER η_2. η_1 is tantamount to the maximum probability of error that has to handle the insecure schemes in order to offer the same transmission rate.

Formally, we look for the root η_1 which satisfies:

$$\mathrm{R}'_s(\theta_{\epsilon WCA}, \epsilon, \eta_1) = \mathrm{R}'_s(\theta_{WCA}, 0., \eta_2). \quad (22)$$

η_1 is computed using the Brent-Dekker algorithm [22][23].

This figure enables to quantify the compromise between security and robustness. When ϵ grows up, a secure watermarking scheme will be more prone to handle errors than an insecure watermarking schemes. For the same mutual information between the decoded pirated sequence and the initial sequence of a colluder, a BER η_2 of $1.e-05$ for an insecure watermarking scheme corresponds to a totally secure embedding scheme ($\epsilon = 0.5$) with a BER $\eta_1 = 1.761e-02$. Note however that the difference between secure and insecure scheme becomes negligible whenever the security of the scheme is not important ($\epsilon < 0.1$) or the BER grows.

Fig. 5 illustrates the difference Δ_{rate} (in bits/sample) between the achievable rates of secure and insecure embedding schemes that undergo the same BSC of parameter η for $c = 4$ colluders. Here again, we can see that the difference is only significant for highly secure schemes (ϵ close to 0.5) and highly robust schemes (η close to 0). Based on these final

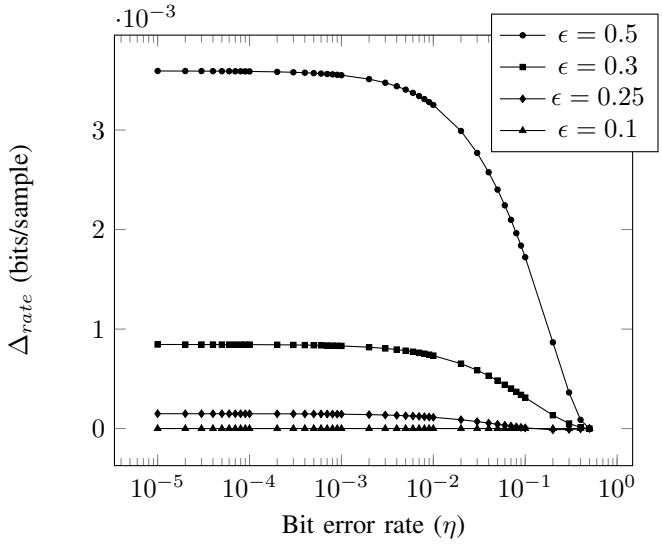

Figure 5. $\Delta_{rate} = \mathrm{R}'_s(\theta_{\epsilon WCA}, \epsilon, \eta) - \mathrm{R}'_s(\theta_{WCA}, 0., \eta)$ w.r.t. bit error rate η for $\epsilon = 0.5, 0.3, 0.25, 0.1$, $c = 4$.

results, we are able to highlight the importance of using highly secure and robust watermarking schemes in comparison with only highly robust ones. However, this advantage becomes negligible whenever the robustness of the scheme, or its security, decreases.

VI. Conclusion

In this article, we find a new collusion strategy for the colluders based on the estimation of the embedded symbols due to a security attack performed on the watermarking scheme. This new attack enables to minimise the mutual information between the pirated sequence forged by a coalition and sequences of the members of this coalition and increases

the probability of accusing an innocent user. Moreover we quantify the compromise to be done between security and robustness for data-hiding and show the advantage of using highly secure and robust watermarking schemes to reduce the coalition power. Our future works include confronting the different colluder strategies, the security attacks of specific watermarking schemes and the robustness constraints in real case scenario, e.g. on digital sequences.

ACKNOWLEDGMENT

Benjamin Mathon, Francois Cayre and Patrick Bas are partly supported by the National French projects ANR-06-SETIN-009 Nebbiano. We acknowledge fruitful discussions with Teddy Furon of IRISA on collusion attacks on Gabor Tardos' fingerprinting codes and rare event analysis.

REFERENCES

[1] T. Furon, L. Pérez-Freire, A. Guyader, and F. Cérou, "Estimating the minimal length of tardos code," in *Proc. of the 11th Information Hiding Workshop*, vol. LNCS, Darmstadt, Germany, jun 2009.

[2] D. Boneh and J. Shaw, "Collusion-secure fingerprinting for digital data," *Information Theory, IEEE Transactions on*, vol. 44, no. 5, pp. 1897 – 1905, sep 1998.

[3] G. Tardos, "Optimal probabilistic fingerprint codes," *J. ACM*, vol. 55, no. 2, pp. 1–24, 2008.

[4] C. Peikert, A. Shelat, and A. Smith, "Lower bounds for collusion-secure fingerprinting," in *SODA '03: Proceedings of the fourteenth annual ACM-SIAM symposium on Discrete algorithms*. Philadelphia, PA, USA: Society for Industrial and Applied Mathematics, 2003, pp. 472–479.

[5] M. Wu and Z. Wang, "Collusion resistance of multimedia fingerprinting using orthogonal modulation," in *IEEE Trans. on Image Proc*. Citeseer, 2005, pp. 804–821.

[6] W. Trappe, M. Wu, Z. Wang, K. Liu *et al.*, "Anti-collusion fingerprinting for multimedia," *IEEE Transactions on Signal Processing*, vol. 51, no. 4, pp. 1069–1087, 2003.

[7] F. Xie, T. Furon, and C. Fontaine, "On-off keying modulation and tardos fingerprinting," in *Proceedings of the 10th ACM workshop on Multimedia and security*. ACM, 2008, pp. 101–106.

[8] T. Kalker, "Considerations on watermarking security," *Proc. MMSP*, pp. 201–206, Oct. 2001.

[9] B. Mathon, P. Bas, F. Cayre, and B. Macq, "Comparison of secure spread-spectrum modulations applied to still image watermarking," *Annals of Telecommunications*, vol. 64, no. 11-12, pp. 801–813, Dec. 2009.

[10] F. Cayre, T. Furon, and C. Fontaine, "Watermarking security: Theory and practice," *IEEE Trans. Sig. Proc.*, vol. 53, no. 10, pp. 3976–3987, Oct. 2005.

[11] L. Perez-Freire and F. Perez-Gonzalez, "Spread-spectrum watermarking security," *Information Forensics and Security, IEEE Transactions on*, vol. 4, no. 1, pp. 2 –24, Mar. 2009.

[12] A. Hyvärinen, J. Karhunen, and E. Oja, *Independent Component Analysis*. John Wiley & Sons, 2001.

[13] A. Hyvarinen, "Fast and robust fixed-point algorithm for independent component analysis," *IEEE Trans. Neur. Net.*, vol. 10, no. 3, pp. 626–634, 1999.

[14] P. Bas and J. Hurr, "Vulnerability of dm watermarking of non-iid host signals to attacks utilising the statistics of independent components," *Information Security, IEE Proceedings*, vol. 153, no. 3, pp. 127 –139, Sep. 2006.

[15] P. Bas and A. Westfeld, "Two key estimation techniques for the broken arrows watermarking scheme," in *MM&Sec '09: Proceedings of the 11th ACM workshop on Multimedia and security*. New York, NY, USA: ACM, 2009, pp. 1–8.

[16] P. Bas and G. J. Doërr, "Practical security analysis of dirty paper trellis watermarking," in *Proc. Information Hiding*, 2007, pp. 174–188.

[17] B. Skoric, T. Vladimirova, M. Celik, and J. Talstra, "Tardos fingerprinting is better than we thought," *Information Theory, IEEE Transactions on*, vol. 54, no. 8, pp. 3663 –3676, Aug. 2008.

[18] F. Cayre and P. Bas, "Kerckhoffs-based embedding security classes for woa data hiding," *IEEE Trans. Inf. For. Sec.*, vol. 3, no. 1, pp. 1–15, Mar. 2008.

[19] T. Furon, A. Guyader, and F. Cérou, "On the design and optimisation of tardos probabilistic fingerprinting codes," in *Proc. of the 10th Information Hiding Workshop*, Springer-Verlag, Ed., vol. LNCS, Santa Barbara, Cal, USA, may 2008.

[20] P. Moulin, "Universal fingerprinting: Capacity and random-coding exponents," in *IEEE International Symposium on Information Theory, 2008. ISIT 2008*, 2008, pp. 220–224.

[21] J. A. Nelder and R. Mead, "A Simplex Method for Function Minimization," *Computer Journal*, vol. 7, no. 4, pp. 308–313, 1965.

[22] R. P. Brent, "An algorithm with guaranteed convergence for finding a zero of a function," *Computer Journal*, vol. 14, pp. 422–425, 1971.

[23] J. C. P. Bus and T. J. Dekker, "Two efficient algorithms with guaranteed convergence for finding a zero of a function," *ACM Trans. Math. Softw.*, vol. 1, no. 4, pp. 330–345, 1975.