KEY LENGTH ESTIMATION OF ZERO-BIT WATERMARKING SCHEMES

Patrick Bas [*]

CNRS-LAGIS,
Lille, France

Teddy Furon

INRIA Research Centre
Rennes Bretagne Atlantique
Rennes, France

ABSTRACT

This paper proposes a new definition of Watermarking Security which is more in line with the cryptographic viewpoint. To this end, we derive the *effective key length* of a watermarking system from the probability of guessing a key equivalent to the original key. The effective key length is then computed for two zero-bit watermarking schemes based on normalized correlation by estimating the region of equivalent keys. We show that the security of these schemes is related to the distribution of the watermarked contents inside the detection region and is not antagonist with robustness. We conclude the paper by showing that the key length of the system used for the BOWS-2 international contest was indeed equal to 128 bits.

***Index Terms*—** Watermarking, security, key length

1. MOTIVATIONS

Security in watermarking is usually enabled by a secret key $\mathbf{k}$ shared by the embedding and the detection algorithms: the secret key can grant *security*, defined by T. Kalker as *"the inability by unauthorized users to have access to the raw watermarking channel"* [13], since the adversary cannot have access to the watermarking channel without knowing $\mathbf{k}$. This definition is very close to the definition of a secure encryption scheme which grants the inability by unauthorized users to have access to the clear message. However, contrary to symmetric cryptography, a secret key may not be unique in watermarking, i.e. the access to the watermarking channel is sometimes possible when the adversary uses an approximation of $\mathbf{k}$. As an illustration, if we take the example of binary Quantization Index Modulation [7], the dither component d represents the secret key: to be able to decode the watermark, the adversary doesn't need to know exactly d, but only an approximation $\hat{d} \in [d - \Delta/4, d + \Delta/4]$. By drawing a random key the adversary has one out of two chances to pick a close enough key enabling the decoding of the watermark (see Fig. 1).

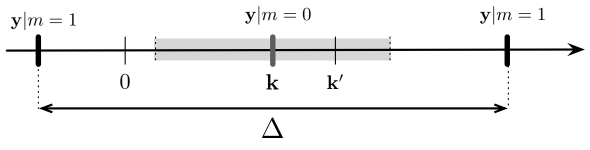

Fig. 1. Non-unicity of decoding keys for binary Quantization Index Modulation [7]: key $\mathbf{k}'$ decodes content $\mathbf{y}$ watermarked with the key $\mathbf{k}$.

If this distinction between cryptography and watermarking is important, the evaluation of the security is also different so far between the two domains. In symmetric cryptography, the key-length defines the security of the algorithm when facing a brute force attack, but in watermarking the security was typically measured using the entropy $h(\mathbf{k})$ or the equivocation $h(\mathbf{k}|O^{N_o})$ w.r.t. a set of N_o observations O^{N_o} [9, 6]. As firstly outlined in [10], the assessment of security in watermarking is not straightforward and not related to the length of the seed of the random generator generating the watermark. This is mainly due the fact that the watermarking scheme has to deal with robustness beside security. Additionally, the measures given by entropy and equivocation take only into account parameters coming from the embedding scheme ($\mathbf{k}$ and O^{N_o}) and completely ignores the decoding part which is yet fundamental in order to define the *access to the raw watermarking channel.*

The typical setup of a brute force attack in watermarking is depicted in Fig. (2) and will be used in thereafter in the paper: the adversary challenges a watermarking detector using a key derived from a set of possible observations O^{N_o}, in order to have access to the true detector output at least $1 - \epsilon$ of the times. We denote by P the probability of success. The effective key length $\ell = -\log_2 P$ defines the average maximum number of keys 2^ℓ that needs to be tested during the brute force attack.

This paper proposes the *effective key length* as a new measure of security in watermarking. The definition of this measure is given in section II and section III gives estimation of the key length for two popular zero-bit watermarking schemes.

[*] This work was partly founded by the French National Research Agency program referenced ANR-10-CORD-019 under the Estampille project.

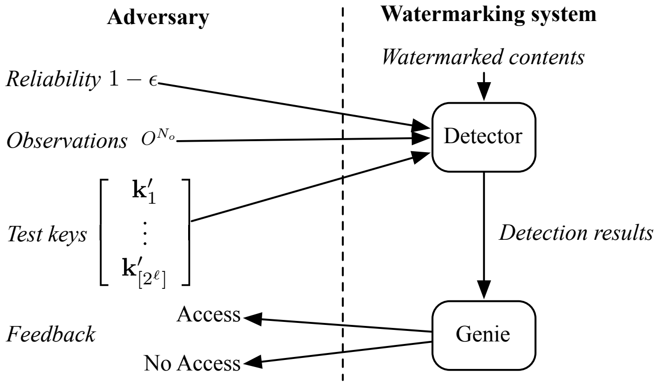

Fig. 2. An example of a brute force attack in Watermarking

2. DEFINITION OF THE EFFECTIVE KEY LENGTH

In [2], the effective key length is defined for multi-bit watermarking schemes whereas this paper proposes a translation for zero-bit watermarking. Given a key $\mathbf{k}\in\mathcal{K}$, a host vector $\mathbf{x}\in\mathcal{X}$ (here $\mathcal{X}=\mathbb{R}^{N_v}$) and a watermarked vector $\mathbf{y}$, we first define by $\mathcal{D}(\mathbf{k})\subset\mathcal{X}$ the detection region for the key $\mathbf{k}$ by:

$$\mathcal{D}(\mathbf{k})\triangleq\{\mathbf{y}\in\mathcal{X}:d(\mathbf{y},\mathbf{k})=1\},\tag{1}$$

where $d(.)$ is the detection function: $d(\mathbf{y},\mathbf{k})=1$ if the watermark is detected and $d(\mathbf{y},\mathbf{k})=0$ if not. We consider the embedding function $e(.)$ such that $\mathbf{y}=e(\mathbf{x},\mathbf{k})$, and the embedding region, i.e. the set of all the watermarked contents as:

$$\mathcal{E}(\mathbf{k})\triangleq\{\mathbf{y}\in\mathcal{X}:\mathbf{x}\in\mathcal{X}\text{ s.t. }\mathbf{y}=e(\mathbf{x},\mathbf{k})\}.\tag{2}$$

Because of the robustness constraint, the embedding region (or a large proportion of it) is included in the detection region. This implies that there might be several detection regions (and associated keys) that can the watermark. We can consequently define in the set $\mathcal{K}$, the subset $\mathcal{K}_{eq}$ of equivalent keys associated with a reliability $1-\epsilon$, as the subset of keys that enables to detect the watermark with a probability $1-\epsilon$:

$$\mathcal{K}_{eq}(\mathbf{k},\epsilon)=\{\mathbf{k}'\in\mathcal{K}:\mathbb{P}\left[d(e(\mathbf{X},\mathbf{k}),\mathbf{k}')=0\right]\le\epsilon\},\tag{3}$$

where $\mathbf{X}$ denotes the random variable representing a host content[1]. Fig. 3 proposes an illustration of embedding and detection regions together with the set of equivalent keys.

In order to derive the expression of the effective key length, we first compute the average probability that a random key $\mathbf{K}'$ belongs to the equivalent region $\mathcal{K}_{eq}(\mathbf{K},\epsilon)$ given the set of observations $\mathbf{O}^{N_o}$:

$$P(\epsilon,N_o)=\mathbb{E}_{\mathbf{K}}[\mathbb{E}_{\mathbf{O}^{N_o}}[\mathbb{E}_{\mathbf{K}'}[\mathbf{K}'\in\mathcal{K}_{eq}^{(d)}(\mathbf{K},\epsilon)|\mathbf{O}^{N_o}]]],\tag{4}$$

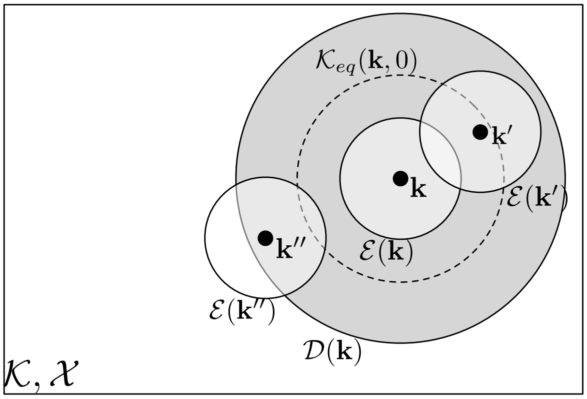
$\mathcal{K}_{eq}(\mathbf{k},0)$
$\mathbf{k}'$
$\mathbf{k}$
$\mathcal{E}(\mathbf{k}')$
$\mathbf{k}''$
$\mathcal{E}(\mathbf{k})$
$\mathcal{E}(\mathbf{k}'')$
$\mathcal{D}(\mathbf{k})$
$\mathcal{K},\mathcal{X}$

Fig. 3. Illustrations of the detection region $\mathcal{D}(\mathbf{k})$ and 3 embedding regions $\mathcal{E}(\mathbf{k})$, $\mathcal{E}(\mathbf{k}')$ and $\mathcal{E}(\mathbf{k}'')$ respectively associated to three keys $\mathbf{k}$, $\mathbf{k}'$ and $\mathbf{k}''$. In this example, $\mathbf{k}'\in\mathcal{K}_{eq}(\mathbf{k},\epsilon=0)$ but $\mathbf{k}''\notin\mathcal{K}_{eq}(\mathbf{k},0)$. The equivalent region $\mathcal{K}_{eq}(\mathbf{k},0)$ and the key space $\mathcal{K}$ are also represented in this toy example but usually these 2 regions live in a different space than $\mathcal{E}(\mathbf{k})$ and $\mathcal{D}(\mathbf{k})$.

and the effective key length is:

$$\ell(\epsilon,N_o)\triangleq-\log_2 P(\epsilon,N_o)\text{ bits}.\tag{5}$$

This definition is very close to the definition of the min entropy $H_\infty(\mathbf{K})$ or the Shannon entropy $H(\mathbf{K})$ for equiprobable keys in cryptography [5], with the particularity that for watermarking we consider the possibility of having a plurality of equivalent keys.

3. APPLICATION ON ZERO-BIT WATERMARKING

To illustrate how it is practically possible to compute an estimation of the key length, we analyze two similar zero-bit embedding methods proposed by Comesaña et al. [8] and Furon and Bas [11]. Both schemes use the normalized correlation as a detection function and the detection function is given by:

$$\begin{array}{ll}d(\mathbf{y},\mathbf{k})=1, & \text{if }\ \frac{|<\mathbf{y},\mathbf{k}>|}{|\mathbf{y}|.|\mathbf{k}|}\ge\cos\alpha,\\ d(\mathbf{y},\mathbf{k})=0 & \text{else.}\end{array}\tag{6}$$

The angle α is computed according to the probability of false-alarm $p_{fa}=\mathbb{P}[d(\mathbf{X},\mathbf{k})=1)]$. The decoding region for these two schemes is a double hyper-cone of axis $\mathbf{k}$ and angle α. Without lost of generality, we set $|\mathbf{k}|=1$ and the set of all possible keys $\mathcal{K}$ is consequently represented by an unitary hypersphere of dimension N_v.

The two embeddings consist in moving the host vector $\mathbf{x}$ into the closest cone by pushing it by a distance D. Using information theoretic arguments [8] propose the OBPA embedding (for Orthogonal to the Boundary and Parallel to the Axis) which first pushes $\mathbf{x}$ in a direction orthogonal to the

[1] Note that we focus here our attention on equivalent keys $\mathbf{k}'$ that can detect a watermark embedded using $\mathbf{k}$, but similar definitions can be derived for equivalent keys $\mathbf{k}'$ that can embed a watermark detected using $\mathbf{k}$.

cone boundary, and then moves afterwards the content parallel to the cone axis. This is proven to maximize the robustness regarding the AWGN channel. The embedding proposed in [11] called BA embedding (for Broken Arrows, the name of the watermarking system), uses worst case attack arguments to first push $\mathbf{x}$ in a direction orthogonal to the cone boundary and continue in a direction orthogonal to the boundary. If the cone axis is reached however, it goes parallel to the axis. Represented in a plan $\mathcal{P} = (O, \mathbf{k}, \mathbf{e}_2)$ with $\mathbf{e}_2$ a vector such that $\mathbf{x} = x_1\mathbf{k} + x_2\mathbf{e}_2$ and $\mathbf{y} = y_1\mathbf{k} + y_2\mathbf{e}_2$, we can illustrate using Fig. 4 the geometrical representations of these two embedding strategies in $\mathcal{P}$. Note that the main difference between these two embeddings is the fact that for a given distortion D, the watermarked contents will be closer to the cone axis using the BA embedding than using the OBPA embedding.

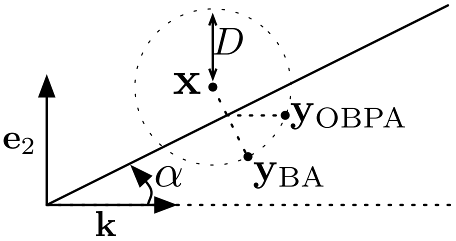

Fig. 4. Embeddings proposed by Comesaña et al. (OBPA) and Furon and Bas (BA).

3.1. No observation ($N_o = 0$)

Contrary to watermarking schemes proposed in [2], it is not possible to derive a literal expression of (8) and we want here to infer an approximation of the equivalent region $\mathcal{K}_{eq}(\mathbf{k}, \epsilon)$ from a set of watermarked contents. Our goal is consequently to compute the maximum possible deviation $\mathbf{k}'$ of the secret key $\mathbf{k}$, such that at least a ratio $(1 - \epsilon)$ of the watermarked contents is included into the hyper-cone of axis $\mathbf{k}'$ and angle α. Let us denote by θ the angle between $\mathbf{k}'$ and $\mathbf{k}$. The equivalent region $\mathcal{K}_{eq}(\mathbf{k}, \epsilon)$ is consequently the union of two spherical caps which is the intersection of the double hyper-cone of axis $\mathbf{k}$ and solid angle θ and the unitary N_v-D hypersphere. $P(\epsilon, 0)$ corresponds to the ratio between the surface of one spherical cap of solid angle θ and the surface of half the sphere (see eq. (8) of [12]).

$$P_{NC}(\epsilon, 0) = 1 - I_{\cos^2\theta}(1/2, (N_v - 1)/2), \qquad (7)$$

Where NC stands for a detection using Normalized Correlation. Applying (8), the key length is given by

$$\ell_{NC}(\epsilon, 0) = -\log_2\left(1 - I_{\cos^2\theta}(1/2, (N_v - 1)/2)\right), \qquad (8)$$

where $I_x(a, b)$ is the regularized incomplete beta function.

Our problem now consists in finding $\hat{\theta}$ such that:

$$\hat{\theta} = \max\{\theta : \mathbb{P}(d(\mathbf{y}, \mathbf{K}') = 1) = \epsilon, \mathbf{k}^t\mathbf{k}' = \cos\theta\}. \qquad (9)$$

This can be estimated in practice using a set of N_c watermarked contents included in $\mathcal{D}(\mathbf{k})$:

$$\begin{aligned}\hat{\theta} \quad = \quad & \max\left\{\theta : |\{\mathbf{y}_i : \mathbf{y}_i \in \mathcal{D}(\mathbf{k}')\}| = [(1-\epsilon)N_c],\right.\\ & \left.\text{and } \mathbf{k}^t\mathbf{k}' = \cos\theta\right\}, \qquad (10)\end{aligned}$$

where $1 \leq i \leq N_c$ and [.] denotes the nearest integer function.

It is possible to perform this estimation in a 3D space instead of a N_v-D space by picking a random unitary basis vector $\mathbf{e}_r$, orthogonal to $\mathbf{k}$, and computing the rotation of $\mathbf{k}$ in the plane $(\mathbf{k}, \mathbf{e}_r)$. The test $\mathbf{y} \in \mathcal{D}(\mathbf{k}')$ is then performed in two steps:

1. each content $\mathbf{y}$ is projected onto the orthonormal basis $(\mathbf{k}, \mathbf{e}_r, \mathbf{e}_3)$ where $\mathbf{e}_3$ is such that $\mathbf{y} = y_1\mathbf{k} + y_2 b_i \mathbf{e}_r + y_3\mathbf{e}_3$. Note that it is still possible to perform the test $\mathbf{y}_i \in \mathcal{D}(\mathbf{k}')$ using this particular projection.

2. the coordinates of $\mathbf{y}$ in the basis $(\mathbf{k}', \mathbf{e}'_r, \mathbf{e}_3)$, i.e. the basis related to the cone of axis $\mathbf{k}'$, are computed by $\mathbf{k}' = (\cos\theta, \sin\theta, 0)$ and $\mathbf{e}'_r = (-\sin\theta, \cos\theta, 0)$.

A geometric illustration of the two cones in the 3D space is illustrated on Fig. (5).

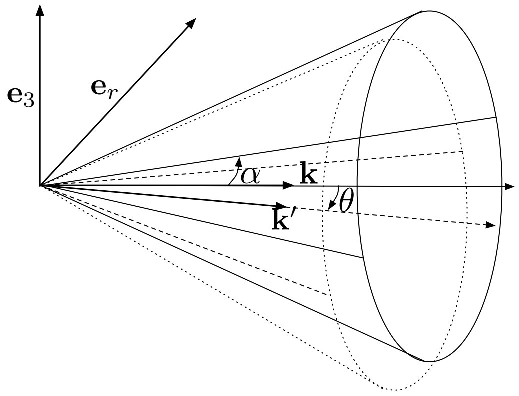

Fig. 5. Geometry when the detection region is a hyper-cone.

The test $\mathbf{y} \in \mathcal{D}(\mathbf{k}')$ is then equivalent to:

$$\frac{y'_1}{\sqrt{y_1^2 + y_2^2 + y_3^2}} \geq \cos\alpha \qquad (11)$$

with $y'_1 = y_1\cos\theta + y_1\sin\theta$. The search of $\hat{\theta}$ satisfying (10) can be done iteratively using a dichotomic search because the number of contents satisfying (11) is a decreasing function w.r.t. θ. Note that in order to increase the accuracy of $\hat{\theta}$, we can sequentially draw several vectors $\mathbf{e}_r$ and average the results of each estimation.

Fig. 6 shows the evolution of the key length w.r.t. the DWR (Document to Watermark power Ratio) for the two embeddings with $N_v = 128$, $\epsilon = 0.05$, and $p_{fa} = 10^{-4}$. The key lengths are computed using (8) and Monte-Carlo simulations with rare event analysis [12] on 1000 watermarked vectors. As expected the key length grows according to the DWR and can reach sizes over 100 bits for $DWR > 6dB$. Notice that the key length of the BA embedding is smaller than the one of OBPA. This is due to the fact that, with BA, the watermarked contents are closer to the hyper-cone axis and consequently the size of the equivalent region is bigger than the one of OBPA. For BA, when the embedding distortion is very important ($DWR \to -\infty$) all the contents tend to be located on the cone axis which means that $\theta \to \alpha$ and $\ell_{NC}(0,0) \to -\log_2 p_{fa}$. The gap between the two key lengths decreases w.r.t. the embedding distortion because both embedding tends to behave the same way for small distortion since the first step, moving toward the boundary, is identical.

Note also that robustness and security are not antagonist here: OBPA, the most robust scheme w.r.t. the AWGN channel, provides also the longest effective key length.

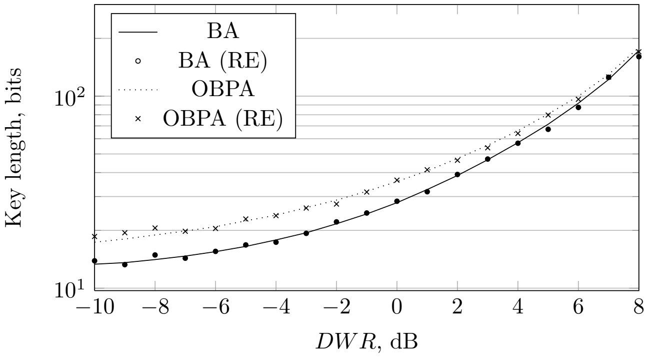

Fig. 6. Key-length evolution according to the embedding distortion using the proposed approximation and Rare Event estimation (RE) for $N_v = 128$, $\epsilon = 0.05$ and $p_{fa} = 10^{-4}$.

3.2. $N_o \neq 0$

In this setup $O^{N_o} = Y^{N_o} = (Y_1, Y_2, \ldots, Y_{N_o})$ and we propose to use the principal component of the observations Y^{N_o}, i.e. the eigenvector associated to the most important eigenvalue of the covariance matrix $\mathbf{C_Y} = N_o^{-1} Y^{N_o} (Y^{N_o})^t$ as a guessing key $\mathbf{k}'$. A similar strategy was previously used to evaluate the security of Broken-Arrows during the BOWS-2 challenge [3]. If $N_o < N_v$, we can compute the Eigen decomposition of the Gram matrix $\mathbf{G_Y} = (Y^{N_o})^t Y^{N_o}$ instead (see [4], sec. 12.1.4).

Fig. 7 presents the evolution of the key size in the same setup than in the previous subsection ($N_v = 128$, $\epsilon = 0.05$, $p_{fa} = 10^{-4}$) for two embedding distortions ($DWR = 5dB$ and $DWR = 7dB$). Monte-Carlo simulation using 10^8 sets of N_o contents where used in this experiment. We can observe the tremendous reduction of the key size for these two schemes when watermarked contents are available to the attacker. The key length of BA decreases faster than the key length of OBPA. This is due to the fact that variance of the contents along directions orthogonal to $\mathbf{k}$ is smaller for BA than OBPA and this favors an accurate estimation of the most principal component.

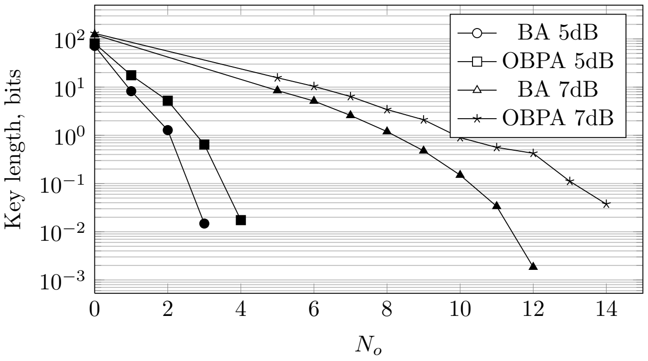

Fig. 7. Key length evolution according to N_o ($N_v = 128$, $\epsilon = 0.05$, $p_{fa} = 10^{-4}$) for $DWR = 5dB$ and $DWR = 7dB$.

3.3. The practical example of the BOWS-2 contest

This subsection applies our methodology to approximate the effective key length of the BA scheme of the BOWS-2 international contest [1]. This helps understanding if it was possible to find a key by random guess. We recall that the host vector was extracted from a 512×512 image and $N_v = 258,048$. In a subspace of dimension 256, the detection region was built as the union of 30 double-hyper-cones with orthogonal axis. The secret key was consequently defined by the basis vectors of the 256 dimensional subspace. By assuming that the subspace is public (but not its basis vectors), we can compute an equivalent region which is larger than the real one and consequently find an upper bound of $P_{BOWS}(\epsilon, 0)$ and a lower bound of $\ell_{BOWS}(\epsilon, 0)$. $P_{BOWS}(\epsilon, 0)$ is upper bounded by the probability of drawing 30 orthogonal vectors falling each in a different equivalent region associated to a true axis:

$$\begin{aligned} P_{BOWS}(\epsilon, 0) &< \Pi_{i=1}^{30} P_{NC}(\epsilon, 0, N_v = 256 - i + 1) \\ &< P_{NC}(\epsilon, 0, N_v = 226)^{30}, \end{aligned} \tag{12}$$

with the parameter $N_v = 256 - i + 1$ coming from the fact that the vector axes are orthogonal. Hence:

$$\ell_{BOWS}(\epsilon, 0) > 30.\ell_{NC}(\epsilon, 0, N_v = 226). \tag{13}$$

In practice, we compute $\ell_{NC}(\epsilon, 0, N_v = 226)$ using (8) on $10,000$ images watermarked using the same embedding setup than during the contest ($PSNR = 43dB$, $p_{fa} = 3.10^{-6}$). We obtain $\ell_{NC}(0.05, 0, N_v = 226) \approx 35$ bits and $\ell_{BOWS}(0.05, 0) > 1050$ bits. On the other hand, since the pseudo random-generator used a 128 bits long seed within the C implementation of the algorithm, this length is an implicit upper bound of the true key length and we can finally conclude that $\ell_{BOWS}(0.05, 0) = 128$ bits. This confirms the idea that the random exhaustive search was impossible during the contest.

4. CONCLUSION

The paper proposes a new methodology to evaluate the security of watermarking techniques based on the computation of the effective key length. Contrary to previous security measures in watermarking, this parameter takes into account the difficulty of accessing the watermarking channel. Moreover the key length brings a close connection with the specification of cryptographic algorithms even if for watermarking it strongly relies on the embedding and the robustness of the scheme. The nature and number of the observations available to the adversary has also an important impact because it dramatically reduces the effective key length.

This paper also proposes the computation of the effective key length for two zero-bit watermarking schemes based on normalized correlation. Whereas we knew for a long time that OBPA is more robust than BA, we show that it is also more secure.

5. REFERENCES

[1] P. Bas and T. Furon. Bows-2. `http://bows2.ec-lille.fr`, July 2007.

[2] P. Bas, T. Furon, and F. Cayre. Practical key length of watermarking systems. In *Proceedings of ICASSP*, Kyoto, Japan, March 2012.

[3] 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*, pages 1–8, New York, NY, USA, 2009. ACM.

[4] C.M. Bishop. Neural networks for pattern recognition. 1995.

[5] C. Cachin. *Entropy measures and unconditional security in cryptography*. Zürich, 1997.

[6] F. Cayre, C. Fontaine, and T. Furon. Watermarking security: theory and practice. *IEEE Trans. Signal Processing*, 53(10), oct 2005.

[7] B. Chen and G. W. Wornell. Quantization index modulation : a class of provably good methods for digital watermarking and information embedding. In *IEEE Transaction on information theory, Vol. 47, N. 4*, pages 1423–1443, may 2001.

[8] P. Comesaa, N. Merhav, and M. Barni. Asymptotically optimum universal watermark embedding and detection in the high-snr regime. *Information Theory, IEEE Transactions on*, 56(6):2804–2815, 2010.

[9] P. Comesaña, L. Pérez-Freire, and F. Pérez-González. Fundamentals of data hiding security and their application to spread-spectrum analysis. In *7th Information Hiding Workshop, IH05*, Lecture Notes in Computer Science, Barcelona, Spain, June 2005. Springer Verlag.

[10] I. Cox, G. Doerr, and T. Furon. Watermarking is not cryptography. In *Proc. Int. Work. on Digital Watermarking*, volume 4283 of *LNCS*, Jeju island, Korea, Nov. 2006. Springer-Verlag.

[11] T. Furon and P. Bas. Broken Arrows. *EURASIP Journal on Information Security*, 2008:1–13, 2008.

[12] T. Furon, C. Jégourel, A. Guyader, and F. Cérou. Estimating the probability fo false alarm for a zero-bit watermarking technique. In *Digital Signal Processing, 2009 16th International Conference on*, pages 1–8. IEEE, 2009.

[13] T. Kalker. Considerations on watermarking security. In *Proc. of MMSP*, pages 201–206, Cannes, France, October 2001.