An Asymmetric Image Watermarking Scheme Resistant against Geometrical Distortions

Dariusz Bogumił
Institute of Computer Science, Warsaw University of Technology
ul. Nowowiejska 15/19, 00-665 Warszawa
dbogumil@ii.pw.edu.pl

Abstract

In this paper, a novel asymmetric public-key watermarking scheme is proposed. The watermark is embedded with the use of a private key, while the decoding uses only simple public key. The asymmetric watermark carries certain number of bits of information. However, neither the knowledge of a public key, nor the knowledge of the hidden information, does not allow the attacker to remove the watermark. That information can be read from an attacked image, in particular from a geometrically transformed image. Moreover, the proposed watermark could be used as a synchronization template for another private-key watermark.

1. Introduction

The growth of network environments and the ease of copying digital media without the loss of quality caused the need to find techniques to prevent (or at least deter) the unauthorized copying, forgery and distribution of digital data. Watermarking is one of the tools that can help to enforce copyright protection and authentication of digital audio, images and video. Robust digital watermarks used for copyright protection can be described as imperceptible information hidden in a digital media. That information should be detectable as long as the quality of the content is considered acceptable. Although many algorithms, methods, techniques and fully functional systems hiding information have been elaborated, the main problem is that the majority of these methods use the symmetric key.

The symmetric watermarks are similar to symmetric cryptography: the same key is used to encode the content and to verify the watermark. As long as the key is secret it does not make any problem. However, to decode (or detect) the hidden data one has to know the secret key. Once the secret key is known, the watermark can be not only decoded, but also easily estimated and removed from the content. In such a situation, a decoder used for copy protection has to be either implemented as a tamper-proof device, or located in a trusted third party. Both solutions are expensive and hardly feasible. In the case of symmetric watermarks the management of the keys is another challenge.

On the contrary, asymmetric watermarks use the secret key in the encoding step only. It is not necessary to know the secret key to verify the watermark. However, the watermarking has some requirements other than the asymmetric cryptography, which make it difficult to use well-known, secure methods from cryptography in watermarking applications. In

cryptography a change of only one bit in the clear text should cause a random change of many bits in the cipher. In watermarking, a small change of even all samples in the signal should not affect the decoder; the same watermark should be read [1].

Hachez and Quisquater outlined the main features of a perfect asymmetric watermark [2]:

- the watermark used for copyright protection should be robust against any attack,
- it should be embedded with a secret key,
- the verification of the watermark should be possible without the knowledge of the secret key (however a public key can be used),
- the watermark can be verified without contacting any authority (an off-line verification),
- similarly to asymmetric cryptology, the knowledge of the verification algorithm and the public key cannot allow to remove or alter the watermark.

It is not clear if such a scheme is feasible, having in mind the limits related to perceptible quality requirements. O'Ruanaidh says that a watermark on its own does not provide any legal proof of ownership [3]. The use of a given digital watermark to protect intellectual property must be registered with a trusted third party to be of any value.

This work presents a method for asymmetric watermark embedding and decoding for static images. It can be considered both as a standalone watermarking application as well as a part of a hybrid watermarking protocol using the trusted third party. The scheme is resistant against geometrical deformations and can be used as a synchronization template for another watermark.

2. Asymmetric watermarking

Eggers *et al.* showed that symmetric methods in combination with public detectors are insecure [4]. Either the watermark can be completely removed by simultaneously improving the quality of the attacked image or the watermark can be made unreadable at a very little expense. So, ensuring the secrecy of the private keys is crucial for the security of those schemes. To allow public verification of watermarks another methods have to be found.

One of the first attempts to create an asymmetric watermark was proposed by Hartung and Girod [5] as an extension of their private spread-spectrum watermarking scheme [6]. The watermark is embedded with the use of a private key p, but its presence is verified with another key - p_k. The public key p_k is a subset of the samples of p, while other samples are randomly replaced. The watermark can be successfully verified with p_k thanks to the redundancy of embedded key p. However, knowing p_k it is very easy to remove the part of the watermark connected with p_k. The benefits of public key are lost as soon as the key is made public.

Smith and Dodge inserted the same watermark into two halves of an image [7]. The decoder correlated these two parts to decide whether a watermark was embedded or not. The watermark did not carry any information beside its existence.

A method based on N-length Legendre sequences was proposed by Shyndel et al. [8]. The Legendre sequence correlates with its conjugate Fourier transform, it is Fourier invariant. Therefore it is used as a private key - watermark embedded into the image. The detector correlates the signal with its conjugate Fourier transform to verify the presence of a Legendre sequence of length N. So, the private key is a Legendre sequence, and the public key – the sequence length. Unfortunately, the method is exposed to exhaustive search attack, as only $N-2$ Legendre sequences exist for length N. There are also other malicious attacks [9].

A modification of transformation invariance idea was introduced by Eggers et al. [10]. The authors used the peculiarity that an eigenvector of a matrix correlates with its

transformation by the matrix. Verification of the watermark is similar to previous technique. The eigenvector of the matrix is the private key and the matrix itself serves as the private key. Also for that scheme effective attacks are known [11].

Choi at al. [12] proposed a transformed-key asymmetric watermarking system. A linear transformation A is applied to the primitive key u to form the private encoding key Au and the public decoding key $A^{-t}u$, where $^{-t}$ stands for inverse transpose. Then a secret key Au is embedded into the image. The watermark can be detected using either the public key $A^{-t}u$ or the private one - Au. The knowledge of the public key does not compromise the security of the private key. To some degree similar approach was presented by Fu et al. [13].

Furon and Duhamel introduced a watermark in the power density spectrum of a signal [14][15]. The power density spectrum (PDS), although it describes the signal to some extent, in general does not allow perfect reconstruction due to the loss of the signal phase. First the signal is permuted to make its power density spectrum flat. Then a private watermark with specified PDS is embedded into the host signal. This PDS is a public key. The public detection process is based on the specific shape of the PDS of the public signal. After the PDS is computed, a hypothesis test is used to verify if the PDS is the same as the public key. The original watermark sequence is not needed to verify the power density spectrum.

All approaches presented above can be exposed to a sensitivity attack and other malicious attacks [16][17]. In such a situation we can use a zero-knowledge protocol [18][19][20][21]. However, the integration of cryptographic protocols with the watermarking systems is difficult and should be done very carefully [22]. The interaction of all components of the system can be an aim of new attacks.

3. Proposed scheme

The proposed method is based on the self-reference concept. A specially designed watermark is inserted using a private secret key. The watermark carries a few tens of bits of information. The hidden information can be decoded without the use of any key. Furthermore the watermark autocorrelation features enable the possibility to identify and recover from geometrical transformations of the watermarked image.

The watermark encoder constructs two 2-dimensional templates: T_1 and T_2, initialized with zero-mean pseudo-random values. The watermark component T_1 is a matrix with the size m x n and T_2 is n x m, where m is much greater than n. The templates can be depicted as narrow rectangles. The length of m should be comparable or bigger then the length of the longer side of a watermarked image, and it is constrained only with the performance of the embedding process. The length of n should be relatively short, to allow detection of the watermark in a small fragment of an image.

The templates are copied in a specific way to cover the whole surface of the watermarked image:

$$w_1(x,y) = t_1(x \bmod m, y \bmod n)(-1)^{\left\lfloor \frac{y}{n} \right\rfloor}$$
$$w_2(x,y) = t_2(x \bmod n, y \bmod m)(-1)^{\left\lfloor \frac{x}{n} \right\rfloor} \quad \textbf{(1)}$$

where $t_1(x, y)$ and $t_2(x, y)$ correspond to the matrixes T_1 and T_2.

The watermark is defined as:

$$w(x,y) = w_1(x,y) + w_1(x-x_1, y-y_1) + w_2(x,y) + w_2(x-x_2, y-y_2) \quad \textbf{(2)}$$

where the shifts x_1, y_1, x_2 and y_2 are the encoded watermark information.

In order to achieve better auto-correlation response and minimize the visible artifacts, the w_1 and w_2 were embedded in separate sets of pixels, using the formula:

$$w(x,y)=\begin{cases} w_1(x,y)+w_1(x-x_1,y-y_1), & \left\lfloor\frac{x}{p}\right\rfloor+\left\lfloor\frac{y}{p}\right\rfloor=2k, \quad k\in N \\ w_2(x,y)+w_2(x-x_2,y-y_2), & \left\lfloor\frac{x}{p}\right\rfloor+\left\lfloor\frac{y}{p}\right\rfloor=2k+1, \quad k\in N \end{cases} \tag{3}$$

where p is a pattern size and it can vary in a range from 1 to n. That expression can be depicted as a "chessboard" in which the fields have the side's length equal p. Such a separation, however, does not interfere with the following analysis.

The watermark W is embedded then into the image I using the additive scheme:

$I' = I + W$

and I' is the watermarked image.

The detection operation is performed as an auto-correlation of the potentially watermarked image I*"*:

$$c = I''I'' = (I+W)(I+W) = (I+W_1+W_{1S}+W_2+W_{2S})(I+W_1+W_{1S}+W_2+W_{2S}) \tag{4}$$

where W_1 and W_2 are the components of the watermark corresponding to T_1 and T_2, and W_{1S}, W_{2S} are the shifted copies $w_1(x-x_1,y-y_1)$ and $w_2(x-x_2,y-y_2)$ respectively. Assuming that I, W_1 and W_2 are orthogonal and do not correlate with each other, we obtain:

$$c = {W_1}^2+{W_{1S}}^2+{W_2}^2+{W_{2S}}^2+W_1W_{1S}+W_{1S}W_1+W_2W_{2S}+W_{2S}W_2 \tag{5}$$

Thanks to watermark templates design we know that the correlation values for the products are:

$$\begin{aligned} {W_1}^2 &= {W_{1S}}^2 = (-1)^{\left\lfloor\frac{y}{n}\right\rfloor}, \quad (x,y)=(0,kn), \quad k\in I \\ {W_2}^2 &= {W_{2S}}^2 = (-1)^{\left\lfloor\frac{x}{n}\right\rfloor}, \quad (x,y)=(kn,0), \quad k\in I \\ W_1W_{1S} &= (-1)^{\left\lfloor\frac{y}{n}\right\rfloor}, \quad (x,y)=(x_1,y_1+kn), \quad k\in I \\ W_{1S}W_1 &= (-1)^{\left\lfloor\frac{y}{n}\right\rfloor}, \quad (x,y)=(x_1,y_1-kn), \quad k\in I \\ W_2W_{2S} &= (-1)^{\left\lfloor\frac{x}{n}\right\rfloor}, \quad (x,y)=(x_2+kn,y_2), \quad k\in I \\ W_{2S}W_2 &= (-1)^{\left\lfloor\frac{x}{n}\right\rfloor}, \quad (x,y)=(x_2-kn,y_2), \quad k\in I \end{aligned} \tag{6}$$

so after substitution and normalization:

$$c=\begin{cases} 1, & (x,y)=(0,0) \\ (-1)^k/2, & (x,y)\in\{(0,kn),(kn,0)\}, \quad k\neq 0 \\ (-1)^k/4 & (x,y)\in\{(x_1+kn,y_1),(x_2,y_2+kn),(-x_1+kn,-y_1),(-x_2,-y_2+kn)\} \\ 0 & other \end{cases} \tag{7}$$

Such a shape of the autocorrelation function allows the decoder to find the hidden watermark information: x_1, y_1, x_2 and y_2. Regularly repeated peaks increase the probability of successful decoding the hidden information and minimize false positive errors.

Summing up, the private key in the proposed scheme is composed of the templates T_1 and T_2 or it is a key used to construct T_1 and T_2. A watermark can be detected without the use of any key. It carries some information encoded in the shifts x_1, y_1, x_2 and y_2. That information can be decoded with a public key (n, p). However, the knowledge of the hidden watermark information does not allow the attacker to remove that watermark from the image, because autocorrelation function does not describe the source image perfectly due to the loss of the signal phase.

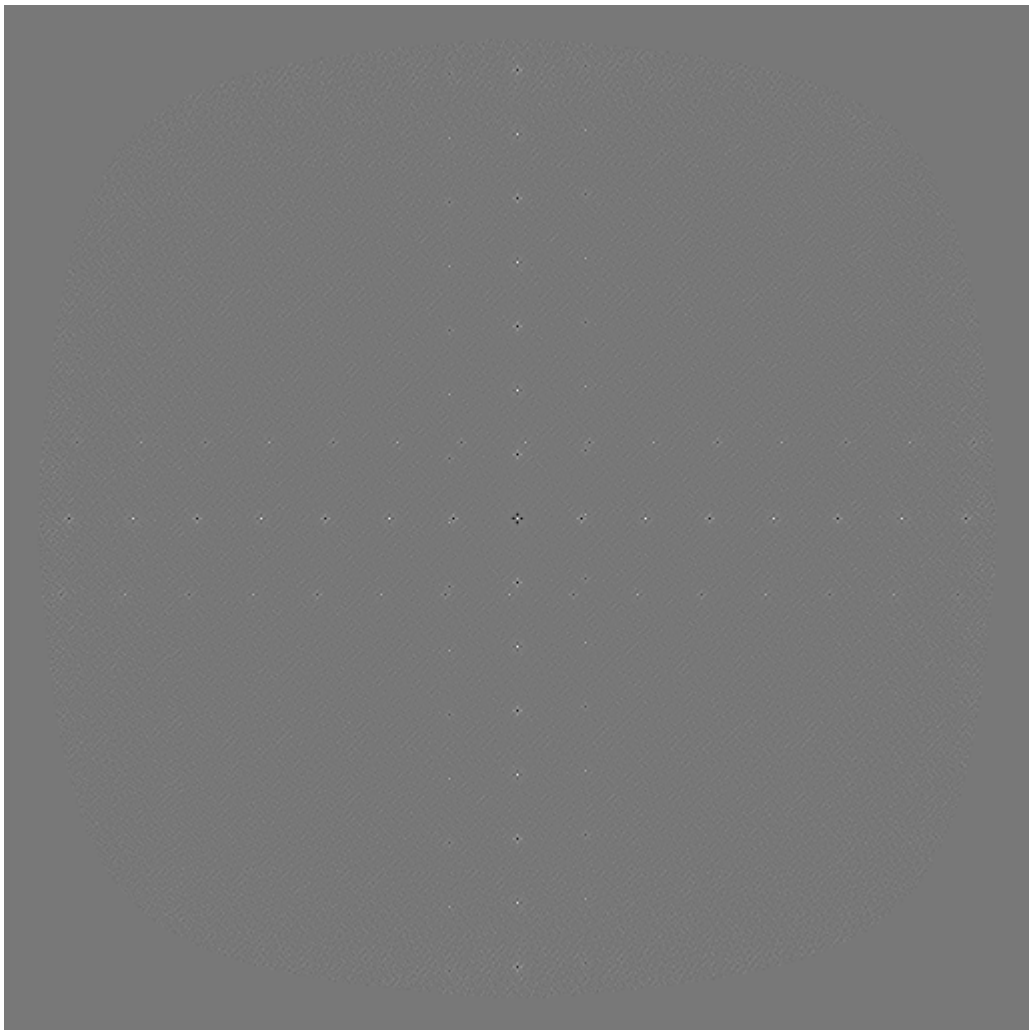

Fig. 1 – The autocorrelaction function graph for a watermarked image.

The encoding/decoding process of the hidden information in x_1, y_1, x_2 and y_2 needs some further explanations. According to the equation **(3)** the templates T_1 and T_2 are embedded in separate positions with the use of the parameter p. Taking into consideration that parameter, the autocorrelation of the watermark depends on remainders of x_1, y_1, x_2 and y_2 divided by $2p$. That autocorrelation reaches the maximum value when the remainders are equal to 0, because all pixels from $w_1(x, y)$ and $w_1(x-x_1, y-y_1)$ are correlating with each other. For other remainders, the autocorrelation is usually smaller and equal to 0 in the worst case. That function for one axis can be described as a graph Fig. 2.

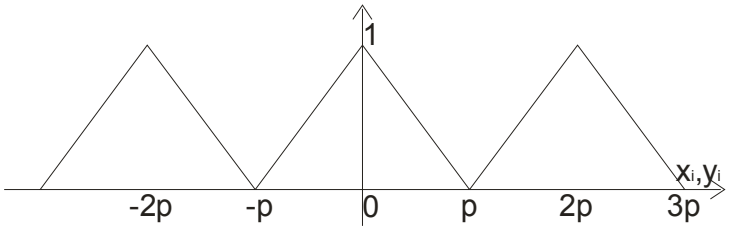

Fig. 2. The function of the autocorrelation of the templates dependent on x_i or y_i and p.

Such a shape of autocorrelation function induces that the shifts x_1, y_1, x_2 and y_2 should be divisible by $2p$. There are also other constraints for x_1, y_1, x_2 and y_2. The y_1 and x_2 are restricted to the range $(0, 2n)$ due to equations **(1)** and **(6)**. The x_1 and y_2 also can not be too big, because we need a good autocorrelation response also for small image fragments. We propose to use the same range as for y_1 and x_2. Each of the shifts has to be greater than 0, for the reason that an autocorrelation function is centrally symmetric so the response would be ambiguous for the shift equal to 0. Last but not least, the shifted autocorrelation peaks can not be too close to the peaks lying on the axes and caused by repeated copies of T_1 and T_2 – in an adverse situation they could be masked with these stronger peaks **(7)**.

Taking into account these all conditions the shifts x_1, y_1, x_2 and y_2 are positive integer numbers, divisible by $2p$ and less than $2n$. That allows encoding almost $4 \cdot \log_2((n-1)/p)$ bits of information in a public watermark. A public key (n, p) is needed to decode such information.

The presented scheme allows the decoder to read the watermark even in case of geometrical distortions [23]. Geometrical transformations, such as scaling, rotation, shearing or any combination of them, can be represented as an affine transform:

$$\begin{bmatrix} x' \\ y' \end{bmatrix} = A \cdot \begin{bmatrix} x \\ y \end{bmatrix} + T, \quad A = \begin{bmatrix} a & b \\ c & d \end{bmatrix}, \quad T = \begin{bmatrix} t_x \\ t_y \end{bmatrix} \tag{8}$$

The expression **(8)** maps each point of the original image from Cartesian coordinates (x, y) to new coordinates (x', y') in the transformed image, where a, b, c, d are the components of the transformation matrix A and t_x, t_y are the components of the translation vector T. An autocorrelation function spectrum undergoes the same transformations as the image, so when we find the autocorrelation peaks, it is possible to estimate the transform matrix:

$$A^{-1} = \begin{bmatrix} 0 & n \\ n & 0 \end{bmatrix} \cdot \begin{bmatrix} x_{n1} & x_{n2} \\ y_{n1} & y_{n2} \end{bmatrix}^{-1} \tag{9}$$

where (x_{n1}, y_{n1}), (x_{n2}, y_{n2}) are the coordinates of periodical peaks caused multiple copies of the templates T_1 and T_2 (second equation of **(7)**). Then the geometrical transformation can be reversed. In other words, the stronger periodical peaks are the reference and synchronization pattern for information part – the peaks caused by shifting of the templates. These parts are inseparably linked and together form a public-key asymmetric watermark.

4. Watermark security

The watermarking system is as insecure as its weakest part. Here, although the template is created with a secret key, its autocorrelation could be read without the knowledge of the key. It creates a possibility for an attack, aiming to remove synchronization template from the watermarked image [24]. Especially, an autocorrelation attack is threatening [25]. The

proposed scheme to some degree lowers the risk of a successful autocorrelation attack, because the template is constructed from two distinct parts, which correlate independently. The naive implementation of the autocorrelation attack would introduce too big artifacts to accept the results. To provide against more sophisticated attacks, taking into account the specific template design, it is possible to introduce a different, key-dependant method of merging the parts of the template.

Another threat is an attack that introduces some strong peaks into the autocorrelation function of the watermarked image. This could mislead the decoder. The remedy for such an attack is trying all the combinations of the autocorrelation function peaks that have the properties compliant with equation **(7)**.

Every watermarking system where the detection can be public is insecure, as an attacker can modify the image until the decoder shows no watermark or a different watermark. However, because of the loss of the signal phase, automatic watermark removing usually is combined with significant quality degradation.

5. Experimental results

The performance of the proposed scheme was tested with Stirmark 3.1. The standard test images "Lena", "Bear", "Skyline arch", "Watch", "Fishing boat" and "Barb" were watermarked with 10 bits of information, without any error correction. Throughout the evaluation, the template parameters were set to:

$n = 32$ – the shorter side of the template rectangle,
$m = 1024$ – the longer side of the template rectangle,
$p = 1$ – the pattern size.

The watermark information was represented as a number "654321" from a range of about 1'000'000 numbers. The PSNR of the watermarked image was kept not less than 38 dB. The detection was done without the presence of the original image. The test case was marked as successful, if all of the bits of hidden information were decoded correctly.

Table 1. Stirmark 3.1 benchmark results

Image modifications class	**Average response**
Signal enhancement	**0.92**
Gaussian	1.00
Median	1.00
Sharpening	1.00
FMLR	0.67
Compression	**0.80**
JPEG	0.60
GIF	1.00
Scaling	**0.49**
Without JPEG 90	0.50
With JPEG 90	0.47
Cropping	**0.82**
Without JPEG 90	0.87
With JPEG 90	0.78
Shearing	**0.71**
X axis w/wo JPEG	1.00
Y axis w/wo JPEG	0.42
Rotation	**0.83**
Auto-crop	0.83
Without JPEG 90	0.83
With JPEG 90	0.82
Auto-scale	0.84
Without JPEG 90	0.85
With JPEG 90	0.82
Other geometric trans.	**0.86**
Col & line removal	0.86
Random Geometric Dist.	**0.17**
Overall Performance	**0.70**

Total 984 images attacked with Stirmark 3.1 were tested and the all 20 bits of the watermark was successfully decoded from 728 images, what gives 74% of successful answers. The score calculated according to Stirmark methodology is 0.70. This is a very good result for an asymmetric scheme. Additionally, we can expect further improvement of an overall system performance for bigger value of p. The amount of watermark information would be less in such a situation, however.

6. Conclusions

In this paper, a novel blind asymmetric public-key watermarking scheme is proposed. The watermark is embedded with the use of a private key. The detection and decoding can be performed with a public key (n, p). The search space for the public key is very small, so the scheme can be also considered as an asymmetric system with private key and no public key. The knowledge of a public key, however, does not allow the attacker to remove the watermark. The asymmetric watermark carries $4 \cdot \log_2\left((n-1)/p\right)$ bits of information. That information can be read from an attacked image, in particular from a geometrically transformed image. Moreover, the proposed watermark could be used as a synchronization template for another private-key watermark. Such properties of the proposed watermark enable new applications for it. An example application is using the public information from the asymmetric watermark to query the Trusted Third Party what other private-key secure watermark is hidden in the analyzed image. In such a situation the decoded information helps to choose the symmetric key or choose a party that will perform further actions.

7. References

[1] Ingemar J. Cox, Matthew L. Miller, Jeffrey A. Bloom. *Digital Watermarking*, pages 299-302. Morgan Kaufmann, 2002.

[2] Gael Hachez, Jean-Jacques Quisquater. *Which directions for asymmetric watermarking?* Proceedings of the XI European Signal Processing Conference (EUSIPCO 2002), Toulouse, France, September 2002. http://www.dice.ucl.ac.be/crypto/index.php?page=pdf92.pdf

[3] Joseph J. K. Ó Ruanaidh, Gabriella Csurka, Watermarking methods, In 26th International Conference on Computer Graphics and Interactive Techniques (SIGGRAPH'99), Los Angeles, CA, USA, 8-13 August 1999. Presented in the Panel ``Digital watermarking: what will it do for me? And what it won't!"

[4] J.J. Eggers, J.K. Su and B. Girod, "Asymmetric Watermarking Schemes," Tagungsband des GI Workshops "Sicherheit in Mediendaten,", Berlin, Germany, 19. September 2000, Springer Reihe: Informatik Aktuell. Invited paper

[5] F. Hartung, B. Girod. Fast Public-Key Watermarking of Compressed Video. In Proc. Of the IEEE Intl. Conf. on Image Processing 1997, Santa Barbara, CA, USA, October 1997.

[6] F. Hartung, B. Girod. Watermarking of uncompressed and compressed video. Signal Processing, 66(3):283–301, May 1998.

[7] J. Smith, C. Dodge, “Developments in Steganography,” 3rd International Workshop on Information Hiding, pp.77-87, 1999.

[8] R. G. van Schyndel, A. Z. Tirkel, I. D. Svalbe. Key Independent Watermark detection. In IEEE International Conference on Multimedia Computing and Systems, volume 1, 1999.

[9] J. J. Eggers and B. Girod. Robustness of Public Key Watermarking Schemes. In V^3D^2 Watermarking Workshop, Erlangen, Germany, October 1999.
[10] J.J. Eggers, J.K. Su and B. Girod, "Public Key Watermarking By Eigenvectors of Linear Transforms," European Signal Processing Conference (EUSIPCO2000), Tampere, Finland, September 2000
[11] Yongdong Wu, Feng Bao, Changsheng Xu. On the Security of Two Public Key Watermarking Schemes. TODO WuYongdong_2003_pcm03b.pdf
[12] Hyuk Choi, Kiryung Lee, Taejeong Kim, "Transformed-key Asymmetric Watermarking System", Proc. SPIE Vol. 4314, Security and Watermarking of Multimedia Contents III, pp.280-289, 2001.
[13] Yonggang Fu, Ruimin Shen, Liping Shen. A Novel Asymmetric Watermarking Scheme. TODO
[14] T. Furon, P. Duhamel. An Asymmetric Public Detection Watermarking Technique. In Workshop on Information Hiding, Dresden, Germany, October 1999.
[15] T. Furon, P. Duhamel. Robustness of an Asymmetric Watermarking Technique. In IEEE International Conference on Image Processing, volume 3, pages 21-24, 2000.
[16] I. J. Cox, J.-P.M. G. Linnartz, "Some general methods for tampering withwatermarks", IEEE J. Select. Areas Commun., vol. 16, pp. 587–593, May 1998.
[17] J.-P. M. G. Linnartz, M. van Dijk, "Analysis of the sensitivity attack against electronic watermarks in images", in Proc. 2nd Int. Information Hiding Workshop, Apr. 1998, pp. 258–272.
[18] Scott Craver. Zero knowledge watermarking detection. In Workshop on Information Hiding, pages 101-116, Dresden, Germany, October 1999.
[19] Scott Craver and Stefan Katzenbeisser. Copyright Protection Protocols Based on Asymmetric Watermarking. In M. Steinebach R. Steinmetz, J. Dittmann, editor, Proceedings of the Fifth Conference on Communication and Multimedia Security (CMS'01), pages 159-170. Kluwer Academic Publishers, 2001.
[20] Scott Craver and Stefan Katzenbeisser. Security Analysis of Public-Key Watermarking Schemes. In Proceedings of the SPIE, Mathematics of Data/Image Coding, Compression, and Encryption IV, volume 4475, pages 172-182, July 2001.
[21] A. Adelsbach and A.-R. Sadeghi, "Zero-knowledge watermark detection and proof of ownership," in Information Hiding—4th International Workshop, IHW 2001, I. S. Moskowitz, ed., Lecture Notes in Computer Science 2137, pp. 273–288, Springer-Verlag, Berlin Germany, (Pittsburgh, PA, USA), 2001.
[22] Stefan Katzenbeisser. On the Integration of Watermarks and Cryptography. TODO
[23] Dariusz Bogumił. Reversing Global and Local Geometrical Distortions in Image Watermarking. In Proceedings of the 6th Information Hiding Workshop. Toronto, Canada, June 2004.
[24] S. Voloshynovskiy, A. Herrigel, and Y. B. Rytsar: Watermark template attack. In Ping Wah Wong and Edward J. Delp, editors, EI'2001: Security and Watermarking of Multimedia Content III, SPIE Proceedings, San Jose, California USA, 22–25 January 2001.
http://vision.unige.ch/publications/postscript/2001/HerrigelVoloshynovskiyRytsar_spie2001.pdf
[25] Dariusz Bogumił: Removing digital watermarks based on image autocorrelation features (in Polish). TPO 2002, Serock, Poland, November 2002.
http://www.ii.pw.edu.pl/~dbogumil