of Multimedia Contents,
January 1999.

Spread Spectrum Watermarking: Malicious Attacks and Counterattacks

Frank Hartung, Jonathan K. Su, and Bernd Girod

Telecommunications Laboratory,
University of Erlangen-Nuremberg,
Cauerstrasse 7, 91058 Erlangen, Germany
{hartung,su,girod}@nt.e-technik.uni-erlangen.de

ABSTRACT

Most watermarking methods for images and video have been proposed are based on ideas from spread spectrum radio communications, namely additive embedding of a (signal adaptive or non-adaptive) pseudo-noise watermark pattern, and watermark recovery by correlation. Even methods that are not presented as spread spectrum methods often build on these principles. Recently, some scepticism about the robustness of spread spectrum watermarks has arisen, specifically with the general availability of watermark attack software which claim to render most watermarks undetectable. In fact, spread spectrum watermarks and watermark detectors in their simplest form are vulnerable to a variety of attacks. However, with appropriate modifications to the embedding and extraction methods, spread spectrum methods can be made much more resistant against such attacks. In this paper, we systematically review proposed attacks on spread spectrum watermarks. Further, modifications for watermark embedding and extraction are presented to avoid and counterattack these attacks. Important ingredients are, for example, to adapt the power spectrum of the watermark to the host signal power spectrum, and to employ an intelligent watermark detector with a block-wise multi-dimensional sliding correlator, which can recover the watermark even in the presence of geometric attacks.

Keywords: digital watermarking, spread spectrum watermarking, robust watermarking, attacks, counterattacks

1. INTRODUCTION

With digital multimedia distribution, intellectual property rights (IPR) are more threatened than ever, due to the possibility of unlimited copying without fidelity loss. Encryption and copy protection mechanisms do not fully solve the issue. Encryption usually protects the data only on the transport channel, and as soon as the data are decrypted for display or playback, they can be copied. Copy protection mechanisms are difficult to realize in open systems. Even in proprietary systems they are often circumvented sooner or later. Thus, both encryption and copy protection offer only limited security. Watermarking has been proposed as a last line of defense in the protection of IPR. Watermarking methods embed information unremovably and invisibly into the host data. The watermark typically contains information about source and recipient of the distributed data. Thus, if pirated copies of the data are distributed, it can still be determined who owns the copyright and who was the original, authorized recipient. Thus, it allows tracing back illegally produced copies of the data.[1]

2. SPREAD SPECTRUM WATERMARKING

Many different watermarking methods for images and video have been proposed (for an overview see e.g.[2,3]). Most of them are based on ideas known from spread spectrum radio communications,[4,5] namely additive embedding of a (signal adaptive or non-adaptive) pseudo-noise watermark pattern, and watermark recovery by correlation.[6–8] Even techniques that are not presented as spread spectrum methods often build on these principles.[9–11]

Fig. 1 illustrates a simple, straightforward example of spread spectrum watermarking. The watermark bits to be embedded* are each repeated $N/4$ times, where N is the number of pixels in the image to be watermarked. The spread information bits are then modulated with a cryptologically secure pseudo noise signal, scaled according to visibility criteria, and added to the image or video pixels.

*in this example for simplicity only 4 bits are embedded. Watermark data rates are higher for many proposed schemes.

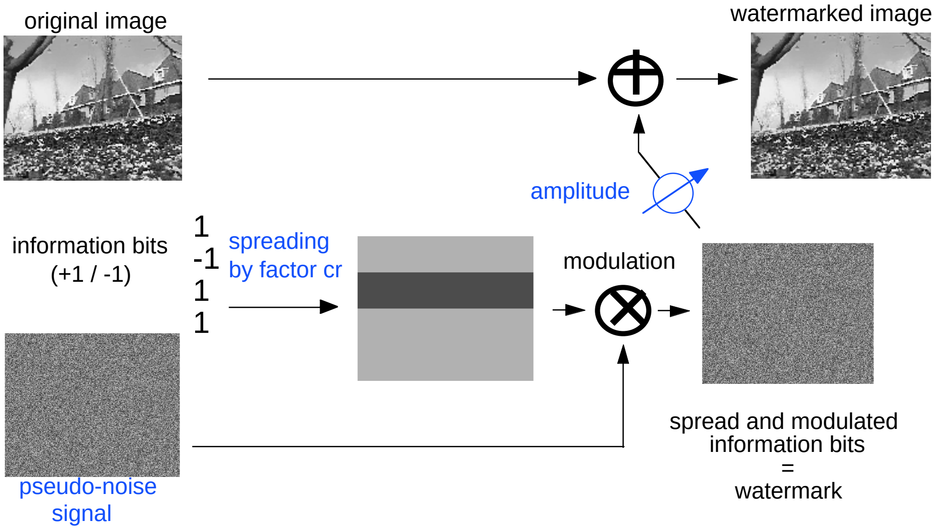


Figure 1. Spread spectrum watermark embedding.

Fig. 2 illustrates the corresponding watermark detector based on the principle of a correlation receiver (or matched filter). In order to reduce cross-talk between the image and the watermark, a pre-filter is applied in order to remove low frequencies from the signal, specifically in order to remove the local mean. If the original, unwatermarked data are available to the watermark detector, it is advantageous to replace the filtering by subtraction of the original (yielding exactly the embedded watermark). The filtered watermarked image (or the watermark, if the original is available) is then demodulated using exactly the same pseudo-noise signal previously used for watermark embedding. The samples of the correlation signal, shown on the right-hand side, are summed for each embedded watermark bit, and a threshold decision yields the output bits. Thus, the result of the watermark decoder are the same watermark information bits that have been embedded.

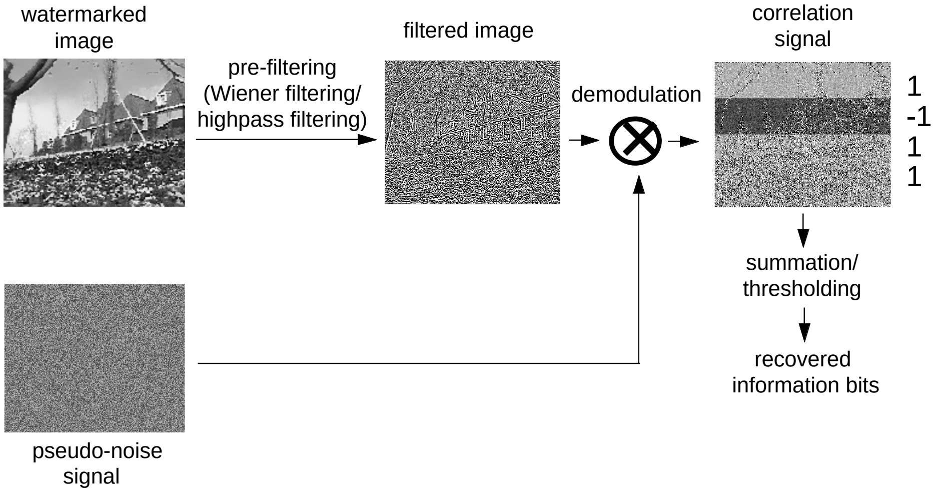


Figure 2. Spread spectrum watermark retrieval.

In practical systems, the watermarked data is sent to the receiver through a channel. We refer to the channel as the watermarking channel to distinguish it from a broadcast channel. Specific for the watermarking channel is that it includes attacks on the watermarks.

Robustness of the watermarks against attacks is a key requirement. While most proposed methods are supported with results showing the robustness against simple signal processing manipulations like linear filtering, cropping or JPEG compression (which are the least effective attacks), very few results are given proving robustness against sophisticated malicious attacks. Some of the few exceptions are results for watermarking methods specifically designed to be robust against certain attacks like spatial shift, zoom, or rotation.[12–14]

In contrast, several different watermark attacks have been proposed[15–26] that seem to prove that the known watermarking methods are in fact vulnerable to attacks and do not satisfy the robustness and security requirements. Moreover, computer programs for watermark removal have been published on the WWW.[27–29] Some of the authors claim that their software defeats all or almost all commercial watermark products.[27,28]

In the next sections, attacks on watermarks are classified and analyzed. Furthermore, remedies and counterattacks are explained that help making spread-spectrum watermarking more resistant against attacks.

3. CLASSIFICATION AND ANALYSIS OF ATTACKS ON WATERMARKS

3.1. Review of Proposed Attacks

In the following, some of the attacks that have been proposed in the literature are reviewed and discussed.

Stone[15] discusses collusion attacks in detail. He shows that advanced collusion attacks against spread spectrum watermarks like the DCT-domain method[6] can already be successful with "one to two dozen" differently watermarked versions of the same data.

Craver et al.[16–18] have shown that many previously proposed watermark methods[6] are vulnerable to so-called inversion attacks that render the embedded watermarks ambiguous. The idea of inversion attacks is that an attacker who receives watermarked data can claim that the data contains also the attacker's watermark by declaring parts of the data to be his watermark. He can easily generate an according fake original by subtracting that claimed watermark. In his fake original, the real watermark can be found. However, also in the real original the fake watermark can be found, since the attacker declared parts of the original signal to be his fake watermark. Thus, a watermark deadlock situation is produced. Craver et al. also propose remedies and methods to ensure non-invertibility. The most important ingredient is to make the watermark signal-dependent using a one-way function.[18]

Cox and Linnartz[19,20] describe and discuss a variety of attacks on watermarks and specifically on the DVD (digital versatile disk) copy protection mechanism. The discussed attacks include geometrical attacks by affine transformations, addition of noise, compression, attacks based on detector observations (see also below in the discussion of Kalker et al.[25,26]), collusion attacks, and attacks based on the availability of a black box watermark inserter. The last mentioned attack compares input and output of the watermark inserter and assumes that the difference is the embedded watermark which can be used to tamper with. Furthermore, specific attacks that circumvent the DVD copy protection mechanism are described, like tampering with the hardware or software DVD player, or scrambling of the watermarked data such that the watermark detector does not find the watermark and allows copying.

Langelaar et al.[21,29] describe an attack on spread spectrum watermarks that is based on estimation of the watermark. They propose to subtract an amplified version of the estimated watermark with an amplification factor of 2. The proposed attack does in fact work on white spread spectrum watermarks. The reason is that the high frequency components of the watermark signal can be estimated very accurately with the estimator, while the low frequency components cannot.[30] Thus, the amplified subtraction of the estimated watermark leads to a negative correlation contribution from the high frequencies, while the remaining low frequency components of the watermark give a positive contribution. If watermark estimator and amplification factor are well matched to each other, the overall correlation is zero or close to zero. However, the method does not work well for lowpass watermarks, and can, even for white watermarks, be counterattacked by lowpass filtering.

Holliman and Memon[22] propose an attack that applies only to linear block-based watermarking schemes. The idea is to exploit cryptological weaknesses of certain watermarking methods[31,32] which allow to transmit the embedded watermark to any other host data such that it also seems to contain the watermark. The Koch and Zhao[31] watermarking method works in the DCT domain and embeds the information by forcing triplets of DCT coefficients from specific blocks into a certain magnitude order. In the attack, the DCT coefficients of all blocks in the image to be fake watermarked that may carry watermark information are forced into the same magnitude order as the coefficients of the corresponding block of the watermarked image.

Barnett and Pearson[23] propose simple attack operators like MPEG quantization, addition of white noise, or linear filtering in order to attack watermarks.

Petitcolas et al.[24,27] analyze the weaknesses of watermarking schemes and propose several attacks on watermarks. The most general attack, also available as software,[27] uses a combination of non-linear geometric distortions and compression. The authors claim that their attack software makes ‘most of the watermarks embedded with commercial software unreadable’. Figure 3 illustrates the effect of the attack. Another proposed attack is tailored to combat a specific audio watermark called ’echo hiding’.[33] In echo hiding, inaudible echoes are introduced that carry watermark information. The attack removes echoes using cepstrum analysis. A third attack, called ’mosaic attack’ attempts to confuse watermark web crawlers that search the WWW for images with embedded watermarks. It simply splits images into several smaller sub-images that are displayed seamlessly in a web-browser, thus looking undistorted, but are individually too small to convey the watermark information.

Kalker et al.[25,26] analyze the vulnerability to watermark estimation through detector observations for spread spectrum watermarking schemes that provide a black box, but publicly available, watermark detector. They assume that the output of the detector is a binary decision “watermark present” or “watermark not present”. Given watermarked data and the detector, they propose a method that allows to estimate the secret spread spectrum watermark with a complexity that is quadratic in the number of samples, thus feasible, and yields a good estimate of the watermark. The basic idea is to gradually degrade the watermarked data X_0, e.g. by compression, until the detector cannot find the spread spectrum watermark W in the degraded version X_1. This version is then known to be near the decision threshold of the detector. To this version X_1 different random signals V_i are added. If the detector detects the watermark in $X_1 + V_i$, then V_i is taken as a hypothesis of the watermark W. If the detector does not detect the watermark in $X_1 + V_i$, then $-V_i$ is taken as a hypothesis of the watermark W. This is repeated N times. In the end, the stored hypotheses $V_i, i = 1 \dots N$ are averaged and give an estimate of W. According to the presented results, this estimate is typically very close to W. Kalker et al. conclude that spread spectrum watermarking techniques that provide a public detector are not secure.

The UNZIGN watermark removal software[28] available on the internet uses a similar strategy as the StirMark software and applies geometric distortions combined with filtering.

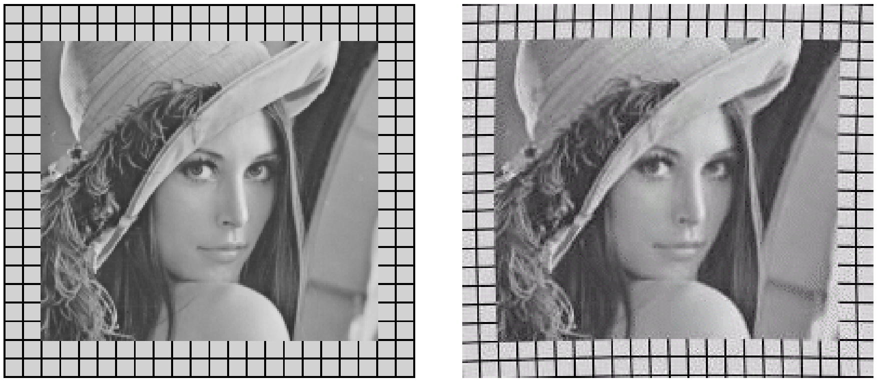


Figure 3. Demonstration of the StirMark 2.3 attack. Left: unattacked image, right: same image after the attack.

3.2. Classification of Attacks

In this section, we list conceivable attacks on spread spectrum watermarks. Most have been described in the literature, while others are new. We classify the attacks into four main groups, as explained below. The terms that we use in the classification are only partly terms that are commonly used; where a common nomenclature is missing we have ourselves coined new technical terms.

In the classification of attacks, we consider only attacks that do not significantly impair the perceived fidelity of the host data. If this assumption is relaxed, there are always successful attacks available, including total erasure of the watermarked data.

We distinguish between the following four groups of attacks:

A. 'simple attacks' (other possible names include 'waveform attacks', 'noise attacks') are conceptually simple attacks that attempt to impair the embedded watermark by manipulations of the whole watermarked data (host data plus watermark), without an attempt to identify and isolate the watermark. Examples include linear and general non-linear filtering, waveform-based compression (JPEG, MPEG), addition of noise, addition of an offset, cropping, quantization in the pixel domain, conversion to analog, and γ correction.

B. 'detection-disabling attacks' (other possible names include 'synchronization attacks') are attacks that attempt to break the correlation and to make the recovery of the watermark impossible or infeasible for a watermark detector, mostly by geometric distortion like zooming, shift in spatial or temporal (for video) direction, rotation, shear, cropping, pixel permutations, sub-sampling, removal or insertion of pixels or pixel clusters, or any other geometric transformation of the data. A typical property of this type of attacks is that the watermark remains in fact in the attacked data and can typically be recovered with increased intelligence (and thus, complexity) of the watermark decoder, as discussed below in section 4.4.1.

C. 'ambiguity attacks' (other possible names include 'confusion attacks', 'deadlock attacks', 'inversion attacks', 'fake-watermark attacks', 'fake-original attacks') are attacks that attempt to confuse by producing fake original data or fake watermarked data. An example is the inversion attack by Craver et al.[16–18] (sometimes referred to as the 'IBM attack') that attempts to discredit the authority of the watermark by embedding one or several additional watermarks such that it is unclear which was the first, authoritative watermark of the IPR owner.

D. 'removal attacks' are attacks that attempt to analyze the watermarked data, estimate the watermark or the host data, separate the watermarked data into host data and watermark, and discard only the watermark. Examples are collusion attacks,[15] denoising, certain non-linear filter operations,[21] or compression attacks using synthetic modeling of the image (e.g. using texture models or 3D models). Also included in this group are attacks that are tailored to a specific watermarking scheme and combat it by exploiting conceptual cryptographic weaknesses of the scheme that make it vulnerable to a specific attack (that however does not impair other watermarking schemes that do not have the same conceptual weakness). If for example the positions of single watermark bits in the watermarked data are known (the pixels that host one specific watermark bit), this gives rise to a certain collusion attack where different fingerprinted copies of the data are interleaved on a watermark-bit-by-watermark-bit basis.

It should be noted that the transitions between the groups are sometimes fuzzy, and that some attacks do not clearly belong to one group. Cropping for example can be regarded as either a simple attack or a detection-disabling attack. Other examples are denoising and certain non-linear filtering attacks which can be regarded as either a simple attack or a removal attack.

Collusion attacks could be argued to be a group of its own, since they require, unlike the other attacks, more than one differently watermarked copy of the data. However, since they attempt to reconstruct the unwatermarked original host data, and thus remove the watermark(s), the classification as a 'removal attack' makes sense.

The attacks proposed in the literature and discussed above in section 3.1 can be categorized according to the proposed attack classes as shown in table 1.

4. PRECAUTIONS AND COUNTERATTACKS AGAINST ATTACKS

"Simple" spread spectrum systems (characterized e.g. by the use of non-adaptive or even white watermark signals and a decoder which does not tolerate spatial shift of the watermarked image) can be defeated by most of the above mentioned attacks. In order to make spread spectrum systems more resistant against attacks, the simple approach has to be extended in several aspects, as explained in the following.

4.1. General

4.1.1. Appropriate Choice of Parameters

It is a well-known property of spread spectrum systems[34,5] that they are not significantly impaired by class A attacks ('simple attacks'), since such attacks introduce mostly impairments which are uncorrelated to the watermark. Thus, filtering, compression and similar operations are not a threat as long as the parameters of the spread spectrum watermarking systems are chosen appropriately. Specifically, the number of pixels that one bit of watermark information

author(s)	keyword	classification
Stone[15]	collusion attack	D
Craver et al.[16–18]	inversion attack	C
Cox and Linnartz[19,20]	affine transformation attack	B
	noise attack, compression attack	A
	detector observation attack	D
	inserter observation attack	D
	collusion attack	D
	DVD tampering attack	D
	DVD scrambling attack	D
Langelaar et al.[21,29]	non-linear filtering attack	D
Holliman and Memon[22]	counterfeit attack	C
Barnett and Pearson[23]	attack operators	A
Petitcolas et al.[24,27]	StirMark attack	A,B
	echo attack	D
	mosaic attack	B
Kalker et al.[25,26]	detector observation attack	D
UNZIGN[28] (authors anonymous)	unzign attack	A,B

Table 1. Classification of proposed attacks.

is distributed over should not be too low and should be at least high enough to convey the watermark information with acceptable bit error rates even if strong lowpass filtering (e.g. using a 7 × 7 box filter[†]) is applied. For video watermarking applications, this can usually be met. For image watermarking applications there may be a limit in the number of pixels that are available to embed the watermark information.

4.1.2. Avoidance of Cryptological Weaknesses and Vulnerabilities

Obviously watermark systems should be cryptologically secure. This involves specifically that the used keys are secure. For example, if pseudo-random generators are used, the structure and/or seed of the pseudo-random generator should be impossible to determine or guess even by knowledgeable parties. This holds for any other key that is involved in the embedding or retrieval. There have been examples, even of commercial watermarking systems, that did not obey this basic principle.

4.2. Watermark Signal Design

4.2.1. Collusion-Secure Watermarks

Collusion attacks are a threat wherever differently watermarked versions of the same data are distributed, i.e. for fingerprinting applications. If the watermarks are mean-free, and even with moderate numbers of colluding parties,[15] spread spectrum watermarks are vulnerable to collusion attacks like averaging.

However, as Boneh and Shaw[35,36] have shown, it is possible to construct collusion-secure watermark signals. The basic idea is to compose the watermarks out of static and dynamic components. Static components do not vanish by averaging. The codes are designed such that for every possible combination of colluding parties there are parts of the codes that do not average to zero. Moreover, from the colluded (averaged) version all colluding parties can be determined. A practical drawback is that the length of the proposed collusion-secure codes increases exponentially with the number of different codes, i.e. distributed watermarks. A remedy is the use of hierarchical codes that at least can identify groups of users. Also, for video watermarking applications, the acceptable length of the codes is reasonably high.

[†] which decreases the correlation sum down to about 2 % for a white watermark

4.2.2. Non-Invertible Watermarks and Time-Stamps

Several remedies have been proposed against ambiguity attacks, like the one postulated by Craver et al..[16–18] There are two key principles in the design of non-invertible watermarks. The first is the use of signal-adaptive watermarks[16–18,37,38] that depend on the host data in a one-way fashion, for example using a hash function. It has been shown that such watermarks are non-invertible. The second principle is the use of cryptologically secure time stamps provided by trusted third parties and encoded in the watermark.[38] Trusted time-stamps should be used for real-world applications anyway, as Wolfgang and Delp[38] pointed out, to avoid other (partly non-technical) pitfalls of watermarking and copyright protection systems.

4.2.3. Adaptation of the Watermark Power Spectrum

In early watermarking publications it was sometimes proposed to use white pseudo noise signals for spreading. Other authors have argued that the watermark should have strong components in low frequencies to better survive compression and attacks.[6] Yet other authors have argued that the watermark should be in the high frequencies or middle frequencies.[39]

For resistance against attacks, it is shown in[30] that the embedded watermark signal should have strong spectral components at frequencies where also the host data has strong spectral components. The ideal watermark power spectrum is then a scaled version of the signal power spectrum:

$$\phi_{ww}(\omega) = \frac{\sigma_w^2}{\sigma_s^2}\phi_{ss}(\omega) \tag{1}$$

where $\phi_{ww}(\omega)$ and $\phi_{ss}(\omega)$ are the power spectra of watermark signal and host signal, respectively, and σ_w^2 and σ_s^2 are the variances of watermark signal and host signal, respectively. It is advantageous to adapt the watermark power spectrum according to 1 since for this watermark power spectrum it is most difficult to estimate the watermark signal using a Wiener estimator, and the variance of the estimation error $\sigma^2_{estimation\ error}$ approaches the variance of the watermark signal:

$$\sigma^2_{estimation\ error} = \frac{\sigma_s^2\sigma_w^2}{\sigma_s^2+\sigma_w^2} \approx \sigma_w^2. \tag{2}$$

In practice, the power spectrum condition for the watermark given in (1) should be adapted to the local statistics of the host data. This can be done explicitly, e.g. by power spectrum estimation, or implicitly, e.g. by compressed-domain embedding into images or video with automatic adaptation to the local power spectrum.[8]

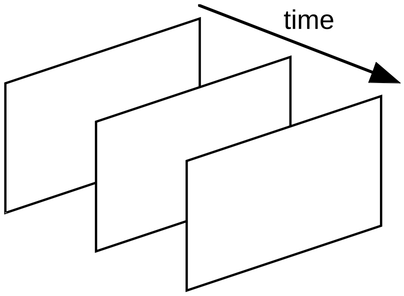
4.2.4. Registration Patterns

A precaution that anticipates detection-disabling attacks is to include registration patterns[40,41] into the watermark that can be used to detect and reverse the geometrical transformation applied by the attack. However, such registration patterns must be easy to find by definition, which again could be exploited for attacks against the registration marks.[20] Thus, the registration marks themselves must provide sufficient security against erasure and attacks.

4.3. Watermark Embedding

4.3.1. Spatial Spreading of Bits

An important precaution to avoid attacks is the randomness of the spatial position of embedded watermark bits. It is not advisable to use K adjacent pixels for embedding of one watermark bit, the next K pixels for the next watermark bit, or similar regular arrangements, as shown in the first two examples of Figure 4. This would allow an attacker to attack single bits of the watermark, thus corrupting the overall watermark, while leaving other portions of the data untouched. For example, for fingerprinting applications, an attacker could switch between different watermarked copies of the data, and thus easily scramble the embedded watermark information. Hence, it is required to distribute each bit of watermark information over pseudo-randomly chosen pixels, as shown on the right-hand side of Figure 4 for video.


time

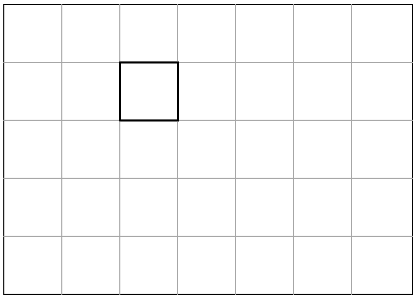
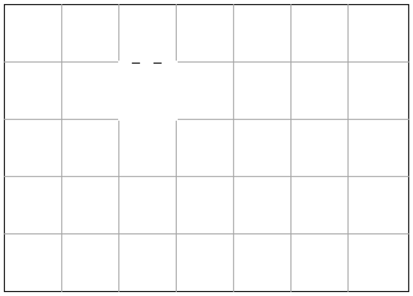


Figure 5. Subdivision of the image into blocks.



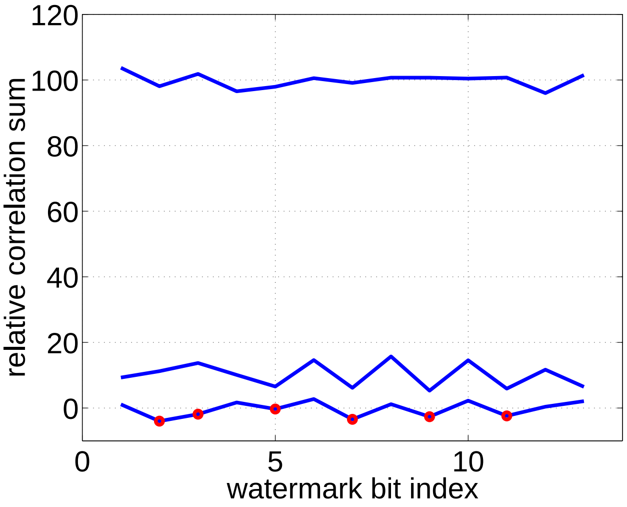


Figure 8. Correlation results for a StirMark attacked video watermark with and without block-based attack-resilient decoder.

explained above in section 4.1.1.

4.4.2. Re-Indexing Resilient Watermark Decoder

For certain detection-disabling attacks, even the proposed block-based multi-dimensional sliding correlator might fail. For example, if neighbouring pixels in an image are re-indexed (permuted), the assumption does not hold that small pixel neighbourhoods remain unchanged. However, in order to avoid visible distortion due to the attack, such re-indexing attacks[42] are typically applied in a local neighbourhood. Thus, the attack can partly be compensated by applying a prefilter to the attacked data. Even better is an optimal ML (maximum likelihood) detector that is resilient against re-indexing and has linear complexity.[42]

5. CONCLUSIONS

In this paper, we have reviewed spread spectrum watermarking principles and specifically the attacks on spread spectrum watermarks that have been discussed and proposed so far. We identified four groups that attacks can be categorized into: simple attacks, detection-disabling attacks, ambiguity attacks and removal attacks.

We have then reviewed and discussed several modifications and extensions of spread spectrum watermarking that improve resistance against the discussed attacks. As a new contribution, we have presented the block-based attack-resilient watermark decoder. Another new insight cited in this paper is that the watermark power spectrum should be a scaled version of the host signal power spectrum in order to make it most difficult for an attacker to estimate the watermark.[30]

However, though the proposed modifications improve the attack resistance of spread spectrum watermarking systems, no watermarking system can guarantee absolute security so far, and it is not clear yet whether an absolutely secure watermark exists at all. Both attacks and counter-attacks on watermarks will continue to become ever more sophisticated.

REFERENCES

1. H. Berghel and L. O'Gorman, "Protecting ownership rights through digital watermarking," *IEEE Computer* , pp. 101–103, July 1996.
2. M. Swanson, M. Kobayashi, and A. Tewfik, "Multimedia data embedding and watermarking technologies," *Proceedings of the IEEE, special issue on Multimedia Signal Processing* **86**, pp. 1064–1087, June 1998.
3. F. Hartung and M. Kutter, "Multimedia watermarking techniques," *to appear in Proceedings of the IEEE, Special Issue on Identification and Protection of Multimedia Information* **87**, June 1999.
4. R. Dixon, *Spread Spectrum Systems*, John Wiley & Sons, New York, NY, USA, 1984.

5. P. G. Flikkema, "Spread-spectrum techniques for wireless communications," *IEEE Signal Processing* **14**, pp. 26–36, May 1997.
6. I. Cox, J. Kilian, T. Leighton, and T. Shamoon, "Secure spread spectrum watermarking for multimedia," Tech. Rep. 95-10, NEC Research Institute, Princeton, NJ, USA, 1995.
7. R. B. Wolfgang and E. J. Delp, "A watermark for digital images," in *Proceedings IEEE International Conference on Image Processing 1996 (ICIP 96), Lausanne, Switzerland*, pp. 219–222, Sept. 1996.
8. F. Hartung and B. Girod, "Digital watermarking of uncompressed and compressed video," *Signal Processing (Special Issue on Watermarking)* **66**, pp. 283–302, May 1998.
9. G. Caronni, "Assuring ownership rights for digital images," in *Proceedings VIS 95, Session "Reliable IT Systems"*, Vieweg, 1995.
10. W. Bender, D. Gruhl, and N. Morimoto, "Techniques for data hiding," in *SPIE Proceedings*, vol. 2420:40, (San Jose, CA), Feb. 1995.
11. I. Pitas, "A method for signature casting on digital images," in *Proceedings IEEE International Conference on Image Processing 1996 (ICIP 96), Lausanne, Switzerland*, Sept. 1996.
12. J. Ó. Ruanaidh and T. Pun, "Rotation, scale and translation invariant digital image watermarking," in *Proceedings IEEE International Conference on Image Processing 1997 (ICIP 97), Santa Barbara, CA, USA*, vol. 1, pp. 536–539, October 1997.
13. J. Ó. Ruanaidh and T. Pun, "Rotation, scale and translation invariant spread spectrum digital image watermarking," *Signal Processing (Special Issue on Watermarking)* **66**, pp. 303–318, May 1998.
14. M. Kutter, "Watermarking resisting to translation, rotation and scaling," *Proceedings of SPIE* , November 1998.
15. H. Stone, "Analysis of attacks on image watermarks with randomized coefficients," tech. rep., NEC Research Institute, 1996.
16. S. Craver, N. Memon, B.-L. Yeo, and M. Yeung, "Can invisible watermarks resolve rightful ownerships ?," tech. rep., IBM research report RC 20509, July 1996.
17. S. Craver, N. Memon, B.-L. Yeo, and M. Yeung, "On the invertibility of invisible watermarking techniques," in *Proceedings IEEE International Conference on Image Processing 1997 (ICIP 97), Santa Barbara, CA, USA*, vol. 1, pp. 540–543, October 1997.
18. S. Craver, N. Memon, B.-L. Yeo, and M. Yeung, "Resolving rightful ownerships with invisible watermarking techniques: limitations, attacks, and implications," *IEEE Journal on Selected Areas in Communications (Special issue on Copyright and Privacy Protection)* **16**, pp. 573–586, May 1998.
19. I. Cox and J.-P. Linnartz, "Public watermarks and resistance to tampering," in *Proceedings IEEE International Conference on Image Processing 1997 (ICIP 97), Santa Barbara, CA, USA*, vol. 3, October 1997.
20. I. Cox and J.-P. Linnartz, "Some general methods for tampering with watermarks," *IEEE Journal on Selected Areas in Communications (Special issue on Copyright and Privacy Protection)* **16**, pp. 587–593, May 1998.
21. G. Langelaar, R. Lagendijk, and J. Biemond, "Removing spatial spread spectrum watermarks by non-linear filtering," in *Proceedings European Signal Processing Conference (EUSIPCO 98), Rhodes, Greece*, September 1998.
22. M. Holliman and N. Memon, "Counterfeiting attacks on linear watermarking schemes," in *Proceedings IEEE Multimedia Systems '98, Workshop on Security Issues in Multimedia Systems, Austin, TX, USA*, June 1998.
23. R. Barnett and D. Pearson, "Attack operators for digitally watermarked images," *IEE Proceedings Vision, Image and Signal Processing* **145**, pp. 271–279, August 1998.
24. F. Petitcolas, R. Anderson, and M. Kuhn, "Attacks on copyright marking systems," in *Proceedings Second International Workshop on Information Hiding, Portland, OR*, April 1998.
25. T. Kalker, J.-P. Linnartz, and M. van Dijk, "Watermark estimation through detector analysis," in *Proceedings IEEE International Conference on Image Processing 1998 (ICIP 98), Chicago, IL, USA*, October 1998.
26. T. Kalker, "Watermark estimation through detector observations," in *Proceedings IEEE Benelux Signal Processing Symposium 98, Leuven, Belgium*, March 1998.
27. F. Petitcolas and M. G. Kuhn, "StirMark 2.3 watermark robustness testing software." available at http://www.cl.cam.ac.uk/ ˜fapp2/ watermarking/ image_watermarking/ stirmark/, October 1998.
28. "UnZign watermark removal software." http:// altern.org/ watermark/, July 1997.
29. G. C. Langelaar, "Watermark Removal Software." available as a Windows95 implementation at http://www-it.et.tudelft.nl/ gerhard/watrem.zip, September 1998.

30. J. Su and B. Girod, “On the imperceptibility and robustness of digital fingerprints,” in *submitted to IEEE ICMCS 99, Florence, Italy*, June 1999.

31. E. Koch and J. Zhao, “Towards robust and hidden image copyright labeling,” in *Proceedings of 1995 IEEE Workshop on Nonlinear Signal and Image Processing*, (Neos Marmaras, Greece), June 1995.

32. M. Yeung and F. Mintzer, “An invisible watermarking technique for image verification,” in *Proceedings IEEE International Conference on Image Processing 1997 (ICIP 97), Santa Barbara, CA, USA*, vol. 2, pp. 680–682, October 1997.

33. W. Bender, D. Gruhl, and N. Morimoto, “Techniques for data hiding,” tech. rep., MIT Media Lab, 1996.

34. D. L. Nicholson, *Spread Spectrum Signal Design – Low Probability of Exploitation and Anti-Jam Systems*, Computer Science Press, 1988.

35. D. Boneh and J. Shaw, “Collusion-secure fingerprinting for digital data,” in *Advances in Cryptology – Proceedings CRYPTO '95*, D. Coppersmith, ed., vol. 963, pp. 452–465, Lecture Notes in Computer Science, Springer, August 1995. (an updated version of this paper is available at http://theory.stanford.edu/~dabo/publications.html).

36. D. Boneh and J. Shaw, “Collusion-secure fingerprinting for digital data,” *IEEE Transactions on Information Theory* **44**, pp. 1897–1905, September 1998.

37. L. Qiao and K. Nahrstedt, “Watermarking schemes and protocols for protecting rightful ownership and customer's rights,” *Journal of Visual Communication and Image Representation* **9**, pp. 194–210, September 1998.

38. R. B. Wolfgang and E. J. Delp, “A watermarking technique for digital imagery: further studies,” in *Proceedings International Conference on Imaging Science, Systems, and Applications (CISST 97), Las Vegas, NV, USA*, pp. 279–287, June 1997.

39. A. Piva, M. Barni, F. Bartolini, and V. Cappellini, “DCT-based watermark recovering without resorting to the uncorrupted original image,” in *Proceedings IEEE International Conference on Image Processing 1997 (ICIP 97), Santa Barbara, CA, USA*, vol. 1, pp. 520–523, October 1997.

40. D. Fleet and D. Heeger, “Embedding invisible information in color images,” in *Proceedings IEEE International Conference on Image Processing 1997 (ICIP 97), Santa Barbara, CA, USA*, vol. 1, pp. 532–535, October 1997.

41. A. Tirkel, C. Osborne, and T. Hall, “Image and watermark registration,” *Signal Processing (Special Issue on Watermarking)* **66**, pp. 373–384, May 1998.

42. J. Su, F. Hartung, and B. Girod, “A channel model for a watermark attack,” in *Proceedings SPIE Security and Watermarking of Multimedia Contents 99, San Jose, CA*, January 1999.