To appear in the Proc. of Electronic Imaging'97, February 1997

A review of watermarking and the importance of perceptual modeling

Ingemar J. Cox and Matt L. Miller
NEC Research Institute
4 Independence Way
Princeton, NJ 08540

ABSTRACT

A watermark embeds an imperceptible signal into data such as audio, video and images, for a variety of purposes, including captioning and copyright control. In this paper, we first outline the desirable characteristics of digital watermarks. Previous work in digital watermarking is then reviewed. Early work identified redundant properties of an image (or its encoding) that can be modified to encode watermarking information. The early emphasis was on hiding data, since the envisioned applications were not concerned with signal distortions or intentional tampering that might remove a watermark. However, as watermarks are increasingly used for purposes of copyright control, robustness to common signal transformations and resistance to tampering have become important considerations. Researchers have recently recognized the importance of perceptual modeling and the need to embed a signal in perceptually significant regions of an image, especially if the watermark is to survive lossy compression. However, this requirement conflicts with the need for the watermark to be imperceptible. Several recent approaches that address these issues are discussed.

1. INTRODUCTION

There has been significant recent interest in watermarking. This is primarily motivated by a need to provide copyright protection to digital content, such as audio, images and video. Digital representations of copyrighted material such as movies offer many advantages. However, the fact that an unlimited number of perfect copies can be illegally produced is a serious threat to the rights of content owners. Watermarking can be used for owner identification, to identify the content owner, fingerprinting, to identify the buyer of the content, for broadcast monitoring to determine royalty payments, and authentication, to determine whether the data has been altered in any manner from its original form. The latter purpose is somewhat different from those of copyright control and the characteristics thereof may be different and are therefore not discussed further here.

A number of technologies are being developed to provide protection from illegal copying. Two complimentary techniques are encryption and watermarking. Encryption protects content during the transmission of the data from the sender to receiver. However, after receipt and subsequent decryption, the data is no longer protected and is in the clear. Watermarking compliments encryption by embedding a signal directly into the data. Thus, the goal of a watermark is to always remain present in the data. It should be noted that embedded signaling or watermarking can be used for a variety of other purposes other than copyright control, but we restrict our discussion here to issues related to copyright control.

In the next section, we outline desirable properties of a watermark for copyright control, which can be quite different from watermarks for authentication purposes, for example, and explain why perceptually modeling is important to watermarking. Section 3introduces a framework in which to discuss the many different proposed watermarks that are described in Section 4.

2. PROPERTIES OF WATERMARKS

There are a number of desirable characteristics that a watermark should exhibit. These include that it be difficult to notice, robust to common distortions of the signal, resistant to malicious attempts to remove the watermark, support a sufficient data rate commensurate with the application, allow multiple watermarks to be added and that the decoder be scalable. These characteristics are discussed in more detail next.

Difficult to notice The watermark should not be noticeable to the viewer nor should the watermark degrade the quality of the content. In earlier work,[1,2] we had used the term "imperceptible", and this is certainly the ideal. However, if a signal is truely imperceptible, then perceptually-based lossy compression algorithms should, in principle, remove such a signal. Current state-of-the-art compression algorithms probably still leave room for an imperceptible signal to be inserted. This may not be true of next generation compression algorithms. Thus, to survive the next generation of lossy compression algorithms, it will probably be necessary for a watermark to be noticeable to a trained observer.

Of course, a just noticeable difference is usually observed by comparing two signals, e.g. compressed and uncompressed or watermarked and original. However, a typical observer will not be comparing two signals, so while a song may sound different from the original, the observer may have no way of knowing this and will probably be satisfied provided the difference is not displeasing.

Early work on watermarking focused almost exclusively on designing watermarks that were imperceptible and therefore often placed watermark signals in perceptually insignificant regions on the content. However, other properties of a watermark conflict with this choice.

Robustness Music, images and video signals may undergo many types of distortions. Lossy compression has already been mentioned, but many other signal transformations are also common. For example, an image might be contrast enhanced and colors might be altered somewhat, or an audio signal might have its bass frequencies amplified. In general, a watermark must be robust to transformations that include common signal distortions as well as digital-to-analog and analog-to-digital conversion and lossy compression. Moreover, for images and video, it is important that the watermark survive geometric distortions such as translation, scaling and cropping.

It has been argued[1,2] that robustness can only be attained if the watermark is placed in perceptually significant regions of an image. This is because the image fidelity is only preserved if the perceptually significant regions of the image remain intact. Conversely, perceptually insignificant regions can be removed without affecting the image quality. Consequently, watermarks that are placed in perceptually insignificant regions will not be robust and can be easily removed. Note that robustness actually comprises two seperate issues: (1) whether or not the watermark is still present in the data after distortion and (2) whether the watermark detector can detect it. For example, watermarks inserted by many algorithms remain in the data after geometric distortions such as scaling, but the corresponding detection algorithms can only detect the watermark if the distortion is first removed. In this case, if the distortion cannot be determined and/or inverted, the detector cannot detect the watermark.

Tamper-resistance As well as requiring the watermark to be robust to legitimate signal distortions, a watermark may also be subject to signal processing that is solely intended to remove the watermark. In addition, when many copies of the same content exist with different watermarks, as would be the case when a watermark is used for buyer identification, further attacks are possible based on collusion amongst several buyers.

It is important that a watermark be resistant to tampering. There are a number of possible ways this may be achieved:

1. **Private watermark**: We believe that a private watermark, i.e. where either the decoder requires knowledge of the unwatermarked content or the pseudo-random noise sequence that constitutes the watermark is only known to the sender and receiver, are inherently more tamper resistant that public watermarks in which anybody is free to decode the watermark.

   For the case in which only a single watermarked copy of the content is available, the only attack appears to be to add noise to the image in the hope of destroying the watermark. However, it can be shown that the magnitude of noise that needs to be added to be confident that the watermark is destroyed is so large that the image fidelity will be severely degraded. For the case of multiple watermarked copies of the same content, more powerful collusion attacks are possible, the most obvious being to average together all n copies.

   In the case where all knowledge to decode the watermark is public, the most obvious attack is to simply invert the encoding process.

2. **Asymmetric encoder/decoder**: If removal of a public watermark requires inverting the encoding process, then it is highly desirable to make the encoder as complex as possible, especially if the watermark is only to be applied once. However, if decoders must run in real-time, then it is necessary for the decoding process to be significantly simpler than the encoding.

Bit rate The bit rate of a watermark refers to the amount of information a watermark can encode in a signal. This is especially important for public watermarks.

Modification and multiple watermarks In some circumstances, it is desirable to alter the watermark after insertion. For example, in the case of digital video discs, a disc may be watermarked to allow only a single copy. Once this copy has been made, it is then necessary to alter the watermark on the original disc to prohibit further copies. Changing a watermark can be accomplished by either (i) removing the first watermark and then adding a new one or (ii) inserting a second watermark such that both are readable, but one overrides the other. The first alternative does not allow a watermark to be tamper resistant since it implies that a watermark is easily removable. Allowing multiple watermarks to co-exist is preferable and also facilitates the tracking of content from manufacturing to distribution to eventual sales, since each point in the distribution chain can insert their own unique watermark.

Scalability In commercial applications, the computational costs of the encoder and decoder are important. In some applications, the insertion is only done once and can be performed off-line. Consequently, the cost of encoding may be less important than the cost of decoding, which may have to occur at real-time video rates, for example. Computational requirements constrain a watermark to be simple, but this simplicity may significantly reduce the resistance to tampering. Further, it is well known that computer speeds are approximately doubling every eighteen months, so that what looks computationally unreasonable today may very quickly become a reality. It is therefore very desirable to design a watermark whose decoder is scalable with each generation of computers. Thus, for example, the first generation of decoder might be computationally inexpensive but might not be as reliable as next generation decoders that can afford to expend more computation to deal with issues such as geometric distortions.

In the next section, we summarize early work on watermarking and then describe more recent work which attempts to insert a watermark into the perceptually significant regions of an image.

3. A FRAMEWORK FOR WATERMARKING

The process of watermarking an image can be represented by the addition of a noise term that is a function of the watermark signal, w, and possibly of the original image, I. The watermarked image, I' is then given by:

$$I' = I + f(I, w) \tag{1}$$

The watermarked image may then be subject to any number of distortions due to tampering or common use which can also be represented as a noise process, n. In many cases, the noise may be approximated by a linear additive process. However, distortions such as geometric transforms of an image may be highly non-linear and image dependent, i.e. $n = n(I)$. The image presented at the decoder, I'', is then given by:

$$I'' = I' + n = I + f(I, w) + n(I) \tag{2}$$

At the decoder, we wish to extract the watermark signal, w, i.e. the unwanted signal (or noise) is the image, I. It should be noted that the magnitude of I is very much larger than the inserted watermark, $f(I, w)$, and the distortions, n, otherwise the image fidelity would not be preserved. Consequently, the signal-to-noise ratio at the input to the decoder, where the signal is now the watermark, w, is much less than one. It is immediately apparent that methods that use the original image as part of the decoding process can greatly improve the SNR by simply subtracting the original image, I from (2).

There are many ways to characterize the numerous proposed watermarking methods. Two properties which we think are important are (1) whether the watermark is inserted into perceptually significant regions of the image and (2) whether the inserted signal, $f(I, w)$, is independent of the image, I.

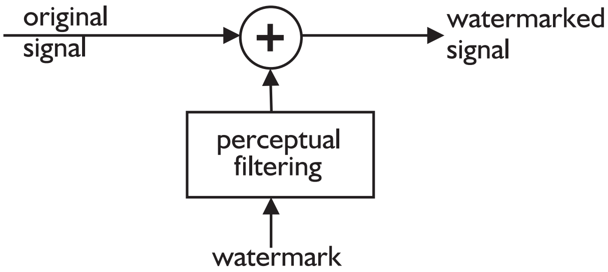

Figure 1. Block diagram of linear insertion method when the watermark signal is independent of the image

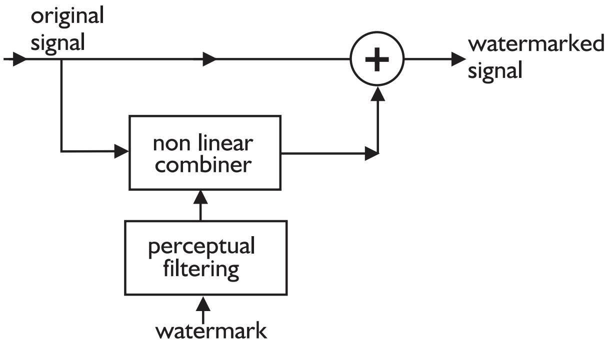

Figure 2. Block diagram of non-linear insertion method when the watermark signal is a function of the image.

Consider first the case in which the inserted signal is independent of the image, i.e.$f(I, w) = w$. In this case, Equation 2 reduces to:

$$I'' = w + I + n \tag{3}$$

where the signal is w and the noise is $I + n$, and the signal can be extracted using traditional matched filtering. In this case, if the watermark is to be placed in perceptually significant regions of the image, then it must be bandpass filtered based on existing knowledge of the human auditory or visual systems. This is illustrated in Figure 1.

However, there is a (possible) disadvantage to shaping the watermark spectrum independently from the image to match currently known human auditory or visual systems. The power present in these frequency bands varies greatly from image to image. Consequently, if simple linear addition of the shaped watermark and image occurs then the magnitude of the watermark must be very low to avoid worst case scenarios in which the image energy in a particular band is very low and artifacts are created because the watermark energy was too strong relative to the image. Conversely, if the image energy is very strong in a particular channel, there is an opportunity to add relative more watermark energy without affecting the image fidelity.

Inserting a signal that is a function of the image leads to a non-linear insertion procedure, as illustrated in Figure 2. For example, Cox *et al*[1,2] proposed scaling the watermark to a fixed fraction of the energy present in a particular frequency coefficient, such that:

$$I' = I(1 + \alpha w) \tag{4}$$

Such a procedure has the advantage that when the image energy in a particular frequency channel is small, the watermark energy is also reduced, thereby avoiding artifacts, and when the image energy is large, the watermark energy is increased, thereby improving the robustness of the procedure. In general, if the watermark is chosen so that its spectrum is white, then multiplication or scaling by the corresponding image coefficient can be thought of as shaping the watermark to the spectrum of the image. The α term then scales the shaped watermark to an acceptable level that is a compromise between robustness and perceptibility. Of course, shaping the watermark across the entire image spectrum, including perceptually insignificant regions, is unnecessary, and the two procedures should be combined such that the watermark is first filtered based on knowledge of the human perceptual system, and subsequently shaped in those regions by an non-linear process prior to insertion in the image.

4. PROPOSED METHODS

In this section, we provide a review of watermarking methods that have been proposed. This is unlikely to be a complete list and omissions should not be interpreted as being inferior to those described here. Recent collections of papers can be found in.[3,4]

Early work on watermarking focused on hiding information within a signal but without considering the issues discussed earlier. In an application in which a covert channel between two parties is desired, tamper resistance may not be an issue if only the communicating parties are aware of the channel. Thus, early work can be thought of as steganography.[5]

Turner[6] proposed inserting an identification code into the least significant bits of randomly selected words on compact discs. Decoding is accomplished by comparison with the original unwatermarked content. Although the method is straightforward, it is unlikely to be robust or tamper resistant. For example, randomizing the least significant bits of all words would remove the watermark. Later van Schyndel *et al*[7] proposed a similar method as well as a spread spectrum method that linearly adds a watermark to an image.

Brassil *et al*[8] describe several methods for watermarking text, based on slightly altering the character or line spacings on a page or by adding/deleting serifs from characters. This approach is further refined in.[9] Unfortunately, as the authors note, these approaches are not resistant to tampering. For example, a malicious attacker could randomize the line or character spacing, thereby destroying the watermark. In general, text is particularly difficult to watermark based on adding noise, since optical character technology is, in principle, capable of eliminating it. An alternative approach is to insert the watermark at the symbolic level, by, for example, inserting spelling errors or by replacing words or phrases with alternatives in a predetermined manner, e.g. substituting "which" for "that". However, these approachs also appear susceptible to tampering.

Caronni[10] describes a procedure in which faint geometric patterns are added to an image. The watermark is therefore independent of the image, but because the watermark is graphical in nature, it has a spatial frequency distribution that contains perceptually significant components. However, it is unclear whether such a method is preferable to adding a pre filtered PN noise sequence.

Tanaka *et al*[11] proposed a method to embed a signal in an image when the image is represented by dithering. Later, Matsui and Tanaka[12] suggested several different methods to encode a watermark, based on whether the image was represented by predictive coding, dithering (monotone printing) or run-lengths (fax). A DCT-based method is also proposed for video sequences. These methods make explicit use of the representation and it is unclear whether such approaches are robust or tamper resistant.

Koch *et al*[13,14] describe several procedures for watermarking an image based on modifying pairs or triplets of frequency coeffecents computed as part of the JPEG compression procedure. The rank ordering of these frequency coefficients is used to represent the binary digits. The authors select mid-range frequencies which typically survive JPEG compression. To avoid creating artifacts, the DC coefficient is not altered. Several similar methods has recently been proposed. Bors and Pitas[15] suggest an alternative linear constraint among selected DCT coefficients, but it is unclear whether this new constraint is superior to that of.[13,14] Hsu and Wu[16] describe a method in which the watermark is a sequence of binary digits that are inserted into the mid-band frequencies of the 8×8 DCT coefficients. Swanson *et al*[17] describe linearly adding a PN sequence that is first shaped to approximate the characteristics of the human visual system to the DCT coefficients of 8×8 blocks. In the latter two cases, the decoder requires access to the original image. It is interesting to note that a recently issued patent[18] describes the general principle of extracting a watermark based on comparison of the watermarked and unwatermarked image.

Rhoads[19] describes a method in which N pseudo random (PN) patterns, each pattern having the same dimensions as the image, are added to an image in order to encode an N-bit word. The watermark is extracted by first subtracting a copy of the unwatermarked image and correlating with each of the N know PN sequences. The need for the original image at the decoder was later relaxed. While Rhoads does not explicitly recognize the important of perceptual modeling, experiments with image compression led him to propose that the PN sequences be spectrally filtered, prior to insertion, such that the filtered noise sequence was within the passband of common image compression algorithms such as JPEG.

Bender *et al*[20] describe several possible watermarking methods. In particular, "Patchwork" encodes a watermark by modifying a statistical property of the image. The authors note that the difference between any pair of randomly chosen pixels is Gaussian distributed with a mean of zero. This mean can be shifted by selecting pairs of points and incrementing the intensity of one of the points while decrementing the intensity of the other. The resulting watermark spectrum is predominantly high frequency. However, the authors recognize the importance of placing the watermark in perceptually significant regions and consequently modify the approach so that pixel patches rather than individual pixels are modified, thereby shaping the watermark noise to significant regions of the human visual system. While the exposition is quite different from Rhoads,[19] the two techniques are very similar and it can be shown that the Patchwork decoder is effectively computing the correlation between the image and a binary noise pattern.

Paatelma and Borland[21] propose a procedure in which commonly occurring patterns in images are located and target pixels in the vicinity of these patterns are modified. Specifically, a pixel is identified as a target if it is preceded by a preset number of pixels along a row that are all different from their immediate neighbors. The target pixel is then set to the value of the pixel a fixed offet away, provided the intensity difference between the two pixels does not exceed a threshold. Although the procedure appears somewhat convoluted, the condition on target pixels assures that the watermark is placed in regions that have high frequency information. Although the procedure does not explicitly discuss perceptual issues, a commercial implementation of this process is claimed to have survived through the printing process.

Holt *et al*[22] describe a watermarking procedure in which the watermark is first nonlinearly combined with an audio signal to spectrally shape it and the resulting signal is then high pass filtered prior to insertion into the original audio signal. Because of the high pass filtering, the method is unlikely to be robust to common signal distortions. However, Preuss *et al*[23] describe an improved procedure that inserts the shaped watermark into the perceptually significant regions of the audio spectrum. The embedded signaling procedure maps an alphabet of signals to a set of binary PN sequences whose temporal frequency response is approximately white. The audio signal is analyzed through a window and the audio spectrum in this window is calculated. The watermark and audio signals are then combined nonlinearly by multiplying the two spectra together. This combined signal will have a shape that is very similar to the original audio spectrum. The resulting signal is then inverse transformed and linearly weighted and added to the original audio signal. This is referred to as spectral shaping. To decode the watermark, the decoder first applies a spectral equalizer that whitens the received audio signal prior to filtering through a bank of matched filters, each one tuned to a particular symbol in the alphabet. While the patent does not describe experimental results, we believe that this is a very sophisticated watermarking procedure that should be capable of surviving many signal distortions.

Cox *et al*[1,2] describe a somewhat similar system for images in which the perceptually most significant DCT coefficients are modified in a non-linear fashion that effectively shapes the watermark spectrum to that of the underlying image. The decoder requires knowledge of the original unwatermarked image in order to invert the process and extract the watermark. This constraint has been subsequently relaxed. The authors also note that binary watermarks are less resistant to tampering by collusion than watermarks that are based on real valued, continuous pseudo random noise sequences.

O Ruanaidh *et al*[24] describe an approach similar to[1,2] in which the phase of the DFT is modified. The authors note that phase information is perceptually more significant than the magnitude of Fourier coefficients and therefore argue that such an approach should be more robust to tampering as well as to changes in image contrast. The inserted watermark is independent of the image and is recovered using traditional correlation without the use of the original image.

In several cases,[23,13,14,1,2] the authors draw upon work in spread spectrum communications. Smith and Comiskey[25] analyze watermarking from a communications perspective. They propose a spread spectrum based technique that "predistorts" the watermark prior to inserting. However, the embedded signal is not a function of the image, but rather is pre-filtered based on expected compression algorithms such as JPEG.

5. SUMMARY

In this paper we have outlined desirable characteristics of a watermark whose purpose is to encode copyright information. An important property of watermarks is that they be robust to common signal distortions and resistant to tampering. To achieve this, the watermark must be placed in perceptually significant regions of the audio, video or image data. However, because the watermark should also be unnoticeable, the signal (watermark) to noise (image) ratio must be small, i.e. much less than one.

We then outlined a simple mathematical framework in which many proposed watermarking methods can be analyzed. In particular, we focused on two important properties of the insertion process, namely (1) whether the watermark was prefiltered to remove perceptually insignificant components and (2) whether the watermark was inserted linearly or nonlinearly. We believe that prefiltering and nonlinear insertion are to be preferred.

Finally, we surveyed many of the numerous recent proposal for watermarking and attempted to identify their strengths and weaknesses with respect to prefiltering and image spectral shaping.

REFERENCES

1. I. Cox, J. Kilian, F. T. Leighton, and T. Shamoon, "Secure spread spectrum watermarking for images, audio and video," in *IEEE Int. Conference on Image Processing*, vol. 3, pp. 243–246, 1996.
2. I. Cox, J. Kilian, F. T. Leighton, and T. Shamoon, "A secure, robust watermark for multimedia," in *Information Hiding: First Int. Workshop Proc.*, R. Anderson, ed., vol. 1174 of *Lecture Notes in Computer Science*, pp. 185–206, Springer-Verlag, 1996.
3. R. Anderson, ed., *Information Hiding*, vol. 1174 of *Lecture Notes in Computer Science*, Springer-Verlag, 1996.
4. *IEEE Int. Conf. on Image Procesing*, 1996.
5. D. Kahn, "The history of steganography," in *Information Hiding*, R. Anderson, ed., vol. 1174 of *Lecture Notes in Computer Science*, pp. 1–5, Springer-Verlag, 1996.
6. L. F. Turner, "Digital data security system." Patent IPN WO 89/08915, 1989.
7. R. G. van Schyndel, A. Z. Tirkel, and C. F. Osborne, "A digital watermark," in *Int. Conf. on Image Processing*, vol. 2, pp. 86–90, IEEE, 1994.
8. J. Brassil, S. Low, N. Maxemchuk, and L. O'Gorman, "Electronic marking and identification techniques to discourage document copying," in *Proc. of Infocom'94*, pp. 1278–1287, 1994.
9. J. Brassil and L. O'Gorman, "Watermarking document images with bounding box expansion," in *Information Hiding*, R. Anderson, ed., vol. 1174 of *Lecture Notes in Computer Science*, pp. 227–235, Springer-Verlag, 1996.
10. G. Caronni, "Assuring ownership rights for digital images," in *Proc. Reliable IT Systems, VIS'95*, Vieweg Publishing Company, 1995.
11. K. Tanaka, Y. Nakamura, and K. Matsui, "Embedding secret information into a dithered multi-level image," in *Proc, 1990 IEEE Military Communications Conference*, pp. 216–220, 1990.
12. K. Matsui and K. Tanaka, "Video-steganography," in *IMA Intellectual Property Project Proceedings*, vol. 1, pp. 187–206, 1994.
13. E. Koch, J. Rindfrey, and J. Zhao, "Copyright protection for multimedia data," in *Proc. of the Int. Conf. on Digital Media and Electronic Publishing*, 1994.
14. E. Koch and Z. Zhao, "Towards robust and hidden image copyright labeling," in *Proceedings of 1995 IEEE Workshop on Nonlinear Signal and Image Processing*, June 1995.
15. A. G. Bors and I. Pitas, "Image watermarking using dct domain constraints," in *IEEE Int. Conf. on Image Processing*, 1996.
16. C.-T. Hsu and J.-L. Wu, "Hidden signatures in images," in *IEEE Int. Conf. on Image Processing*, 1996.
17. M. D. Swanson, B. Zhu, and A. H. Tewfik, "Transparent robust image watermarking," in *IEEE Int. Conf. on Image Processing*, 1996.
18. D. C. Morris, "Embedding hidden indentification codes in digital objects." US Patent 5,530,751, 1996.
19. G. B. Rhoads, "Indentification/authentication coding method and apparatus." World Intellectual Property Organization, WIPO WO 95/14289, 1995.
20. W. Bender, D. Gruhl, N. Morimoto, and A. Lu, "Techniques for data hiding," *IBM Systems Journal* **35**(3/4), pp. 313–336, 1996.

21. O. Paatelma and R. H. Borland, "Method and apparatus for manipulating digital data works." WIPO Patent WO 95/20291, 1995.
22. L. Holt, B. G. Maufe, and A. Wiener, "Encoded marking of a recording signal." UK Patent GB 2196167A, 1988.
23. R. D. Preuss, S. E. Roukos, A. W. F. Huggins, H. Gish, M. A. Bergamo, P. M. Peterson, and D. A. G, "Embedded signalling." US Patent 5,319,735, 1994.
24. J. J. K. O. Ruanaidh, W. J. Dowling, and F. Boland, "Phase watermarking of digital images," in *IEEE Int. Conf. on Image Processing*, 1996.
25. J. R. Smith and B. O. Comiskey, "Modulation and information hiding in images," in *Information Hiding: First Int. Workshop Proc.*, R. Anderson, ed., vol. 1174 of *Lecture Notes in Computer Science*, pp. 207–226, Springer-Verlag, 1996.