IST-2002-507932

ECRYPT

European Network of Excellence in Cryptology

Network of Excellence

Information Society Technologies

Proceedings of the 2nd WAVILA Challenge (2nd WaCha) Geneva, Switzerland, September 28th, 2006

Start date of project: 1 February 2004 Duration: 4.5 years

Lead contractor: Katholieke Universiteit Leuven, Belgium (KUL)
Document responsible: Otto-von-Guericke Universität Magdeburg, Germany (GAUSS)

Revision 1.0

ISBN: 978-3-929757-29-3
Printed at the Otto-von-Guericke University of Magdeburg, Germany
March 31st, 2007

Preface

Christian Kraetzer, Otto-von-Guericke University of Magdeburg, Germany

This document contains all papers that were presented at the "Second WAVILA Challenge" (2nd WaCha), which was held as ECRYPT workshop in Geneva (Switzerland) on September 28th, 2006. The workshop addressed two main questions, namely:

1. Is knowledge of the watermarking algorithm useful for watermark removal? Following an approach similar to that used in cryptography, the problem of watermarking security is often approached by assuming that the attacker has full knowledge of the watermarking algorithm and that he explicitly uses such a knowledge to devise a, possibly optimal, attacking strategy. The assumption underlying the above perspective is that knowing the details of the watermarking algorithm is a great help for the attacker. Whereas in general this is surely true, some recent analyses seem to point out that if the aim of the attacker is limited to watermark removal, or to make it unreadable to the detector/decoder, knowledge of the watermarking algorithm is of limited, if any, help. Some evidence of this fact is given by the effectiveness of some recently proposed blind sensitivity attacks (see for instance the blind Newton sensitivity attack described in P. Comesana, L. Perez-Freire, F. Perez-Gonzalez, The Blind Newton sensitivity attack, Proceedings of SPIE, Volume 6072, Security, Steganography, and Watermarking of Multimedia Contents VIII, Edward J. Delp III, Ping Wah Wong, Editors, 60720E (Feb. 15, 2006)), that are able to remove the watermark while keeping an extremely high PSNR (e.g. more than 50dBs) between the watermarked and the attacked version of the image. Similar results seem to stem from the BOWS contest (http://lci.det.unifi.it/BOWS, run in the period December 2005-June 2006) where very powerful attacks were devised even if the underlying algorithm was not known. A possible interpretation is that whenever the watermarking algorithm results in a very complicated detection region, no particular advantage is got by knowing the watermarking algorithm. On the contrary, such an advantage is a significant one for schemes characterized by simple detection regions. It is the aim of the second WAVILA Challenge to investigate the above problem trying to answer the following questions. Is knowledge of the watermarking algorithm of any practical help to attackers? Does the answer to the previous question depend on the complexity of the watermark detection/decoding region(s)? If knowledge of the algorithm does not help to reduce the obtrusiveness of the attack, do you think it may still be useful to reduce its complexity? Is watermark robustness more difficult to achieve than watermark security?
2. How does the output of the optimal watermarking algorithm look like? For different

application scenarios different optimal solutions and algorithms are sought for in digital watermarking. This challenge proposed here is intended to identify application scenarios with their goals and characteristics. Furthermore the metrics to measure and compare these characteristics for selected algorithms are of interest. For the identified application scenarios the question is raised: how should the benchmarking results for an optimal watermarking algorithm for this application scenario look like? Can they be described within the triangular relationship between robustness, capacity and transparency, or have other characteristics to be considered, too? How can the comparability of benchmarking results be guaranteed? Which optimisation strategies for the parameterisation of watermarking algorithms do exist and how intend to improve the output of the algorithm?

These two challenges were covered by the submitted publications, which can be found in this document and in invited talks by Scott Craver (Assistant Professor, Department of Electrical & Computer Engineering, Binghamton University, USA) and Teddy Furon (Researcher at IRISA, Rennes, France).

Scott Craver described in his talk "Noise Calipers: A Technique for Reverse-engineering Correlation Detectors" the technique of noise calipers, which employs an oracle to quickly build a pair of severe false positives from a watermarked image. By using this technique, knowledge about the detection region, detector threshold and approximate number of watermarking features can be gained by reverse-engineering a secret watermark algorithm instead of attacking the watermark itself. Being seen in the context of ECRYPT's first BOWS contest, this talk did ignite a heated discussion on the 2nd WaCha about different attack techniques including the sensitivity attacks which had strongly influenced the first BOWS.

Teddy Furon's invited talk "Is benchmarking just an academic chimera?" addressed the second challenge and the relationship between academic interest in watermark benchmarking and the benefit of such benchmarking activities to watermarking designers. This talk, together with the papers presented on the second challenge, sparked a very interesting discussion between the participants of the 2nd WaCha. A large number of high ranking watermarkers from academia and industry discussed the pros and cons of benchmarking activities in this field. The results of this discussion can be concluded as follows: benchmarking activities in digital watermarking are useful and have a future if (and only if) they produce results which can be understood by non-experts in this area. A system designer without mature knowledge in watermarking techniques should be able to choose an appropriate algorithm for a well defined application scenario from benchmarking results for a set of watermarking algorithms. If this is made possible by benchmarking techniques and appropriate presentation of the results then benchmarking is not just an "academic chimera".

Proceedings of the 2nd WAVILA Challenge (WaCha)

Editors:
Mauro Barni (National Inter-University Consortium for Telecommunications, Italy)
Jana Dittmann (Otto-von-Guericke University Magdeburg, Germany)
Christian Kraetzer (Otto-von-Guericke University Magdeburg, Germany)

Revision 1.0
Februar 28th, 2007

The work described in this report has in part been supported by the Commission of the European Communities through the IST program under contract IST-2002-507932.

Contents

Is knowledge of the watermarking algorithm useful for watermark removal ?

M. Barni[1] and A. Piva[2]

[1] CNIT (National Inter-University Consortium for Telecommunications)
Research Unit of Siena, Italy barni@dii.unisi.it
[2] CNIT (National Inter-University Consortium for Telecommunications)
Research Unit of Florence, Italy
{piva}@lci.det.unifi.it

Introduction

Since the early days of watermarking research, the Kerckhoffs principle [1] according to which the security of a cryptographic scheme should not rely on the secreteness of the cryptographic algorithm but on one or more secret keys, was taken as a model for watermarking. Since then the "security by obscurity" paradigm was rejected thus leading to the common assumption that attackers are aware of the algorithms used to embed and retrieve the watermark from the host signal. At the same time, knowledge of the watermarking algorithm has always been seen as a great advantage for the attacker that, at least in principle, could calibrate his/her attacks by considering the peculiarities of the watermarking algorithm he/she is attacking.

The recent appearance of extremely powerful blind attacks, belonging to the wide class of sensitivity attacks, that can successfully attack virtually all the watermarking algorithms proposed so far without considering their characteristics (blind attack), and the results of a recent contest among attackers, that were asked to attack an unknown watermarking scheme, raised some doubts about how the knowledge of the watermarking algorithm can help the attacker to break a watermarking system.

Specifically, the BNSA (Blind Newton Sensitivity Attack) algorithm described in [2, 3] proved to be so efficient in removing the watermark from any document, regardless of the adopted algorithm, that one may wonder how (and whether) knowledge of the watermarking algorithm may help the attacker to improve its performance. At the same time, the BOWS contest (see below for a more detailed description) explicitly aimed at evaluating whether the "security by obscurity" principle (though conceptually wrong) can help to improve the security/robustness of a watermarking system. To this aim, the contest was split in two phases: during the first one the contenders did not know the watermarking algorithm they had to break, whereas in the second phase the details of such an algorithm were made publicly available. Rather surprisingly, knowing the watermarking algorithm did not help very much the contenders (even because the results of the first phase were already rather good).

The question that naturally arises is whether and how the attacker can exploit the knowledge of the watermarking algorithm to improve the attacks.

It was the aim of the second Wavila Challenge to discuss the above question by the light of the recent research in the field. While in the final section we will briefly summarize author's answer to the challenge, the rest of the paper will be devoted to the description of the organization and the results of the BOWS contest, whose output motivated, at least in part, the formulation of the challenge. As to the BNSA, readers are referred to the original papers describing it [2, 3].

1 The first BOWS Contest (Break Our Watermarking System)

In the framework of WAVILA activities, it was proposed to launch a Contest that was named BOWS, acronym of *Break Our Watermarking System*. As suggested by the name, BOWS was designed to allow to investigate how and when an image watermarking system can be broken though preserving the highest possible quality of the modified content, in case that the watermarking system is subjected to a world-wide massive attack. BOWS contest was not intended as an attempt to prove how well-performing a watermarking system is, but it was expected by means of this action to better understand which are the disparate possible attacks, perhaps unknown at the moment of the start of the contest, the BOWS participants could carry out to perform their action and comprehend the degree of difficulty of breaking the embedded watermark. Moreover, the Contest was intended to study if and how much the knowledge of the watermarking algorithm is useful for watermark removal. The Contest was conceived in the following way: the image watermarking algorithm proposed by Miller, Doerr and Cox in [4] has been chosen to be the object of the contest. Three different generic grayscale images have been chosen. The three original images have been watermarked obtaining the watermarked versions exhibiting a Peak-Signal-to-Noise-Ratio (PSNR) included between 42 and 46 dB. These watermarked images have made available for download on the BOWS web-site at the address *http://lci.det.unifi.it/BOWS/*, whose homepage is shown in Figure 1.

Then, contenders were allowed to try to erase the watermark from all the three images by using any action they wanted while granting a minimum PSNR of 30 dB between the watermarked image and the attacked one. To verify their action, attackers had the possibility to upload each of the three images (still in raw format and size 512×512) on the BOWS web-site through an ad-hoc interface and to ask to run the detection process; finally they obtained as answer the result of the detection and the PSNR achieved. In case of successful attack, the thumbnail of the attacked image showed the word "Passed".

When a BOWS participant has succeeded to remove the watermark from all the three images, he/she has been asked to register in the Hall of Fame. In order to enter the Hall of Fame, the attacker was asked to give a brief explanation (just a few words) about the performed attack, otherwise the record was not stored

Fig. 1. The home page of the BOWS web site, available at the address *http://lci.det.unifi.it/BOWS/*.

into the database. The best performances in terms of PSNR (average PSNR on the three images when watermark deletion has been successful over all of them) were stored by the system to fill in a rank list updated from time to time.

At the beginning of the Contest, a limit of 30 upload/day was fixed. To check it and to log the Contest, all the uploads were recorded, according to the IP address of the client connecting to the BOWS server.

1.1 First phase of the Contest

The first phase of the BOWS Contest started on December 15, 2005, and ended on March 16, 2006. At the start of the Contest the participants were able to remove the watermark only on the image *Strawberry*. It was then decided to remove the limit on the maximum number of attacks a day in order to allow the attackers to carry out also a sensitivity attack (actually, the limit was not removed, but fixed to a value equal to 5000 attacks/day). Thanks to this modification and to the growing advertisement of the Contest, the number of participants and uploaded attacked images increased very much. At the end of the first phase of the Contest, from more than 300 IP addresses 72074 Attacked Images were uploaded on our server; in 10034 of them (corresponding to the 13.9% of all the received images) the watermark was erased while granting a minimum PSNR of 30 dB between the watermarked image and the attacked one. 10 participants succeeded to remove the watermark from all the 3 watermarked images, and registered their data in the Hall of Fame (some of them succeeded two or also three times). The Steering Committee responsible to rule the BOWS contest,

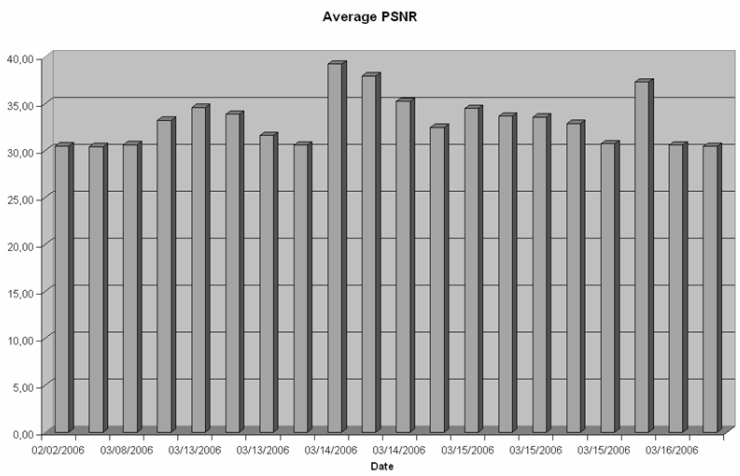

Fig. 2. The Average PSRN values obtained by the participants that entered in the Hall of Fame in the First Phase of the Contest. Note that most of the successful attack have been registered just in the last days of the Contest.

according to the recorded results, confirmed that the winner was the Team held by Scott Craver, Binghamton University, with the following results: PSNR of the image *Strawberry* = 39.69 dB, PSNR of the image *Wood Path* = 39.67 dB, and the PSNR of the image *Church* = 38.47 dB. By analyzing the Hall of Fame at the end of the first phase, it is possible to note that most of the successful attacks have been registered in the last three or four days of the Contest, as shown in Figure 2. Seven attacks obtained an average PSNR lower than 31 dB, and only three were able to exceed 36 dB. In Figure 3 the same results are shown according to the Institute the participants belong to. For each Institute the maximum value of the Average PSNR is shown, whereas the figures in top of the columns indicate the number of registrations into the Hall of Fame obtained by the same research group.

1.2 Second phase of the Contest

After the three months of the contest, it was decided to reveal that the watermarking algorithm used to embed the watermark into the three images is the one presented by Miller, Doerr and Cox in [4]. Then, the BOWS web site remained open for other three months for the second phase of the contest during which the researchers were allowed to sharpen their attacks by exploiting the knowledge about the adopted watermarking scheme. The Hall of Fame was not erased, but the participants entered in the rank in the second phase of the contest were highlighted by a different notation in the list. During these further three months the BOWS server received from more than 100 IP addresses 721734 Attacked Images; in 20666 of them (corresponding to the 2.9% of all the received images)

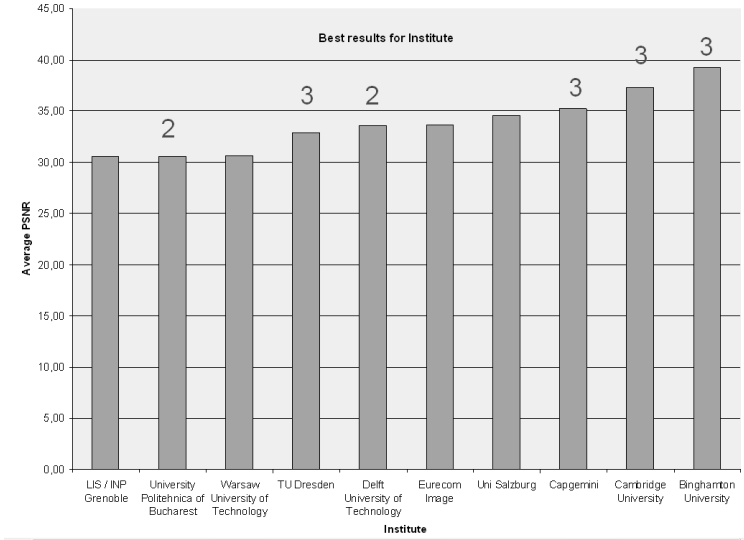

Fig. 3. The Average PSRN value obtained by the participants that entered in the Hall of Fame in the First Phase of the Contest, grouped according to the Institute they belong to. Note that some participants entered into the Hall of Fame more than once.

the watermark was removed while granting a minimum PSNR of 30 dB. In this second phase, 16 participants succeeded to remove the watermark from all the 3 watermarked images, and registered their data in the Hall of Fame (again, some of them succeeded several times). The contender reaching the highest PSNR value was Andreas Westfeld, by TU Dresden, that was also the most active in the upload of attacked images, so that we were also constrained to insert a limit, even though high (3000 attacks), to the number of images uploaded by an IP address each day. Andreas Westfeld at the end of the Contest obtained excellent values of the PSNR: for the image *Strawberry* = 60.74 dB, for the image *Wood Path* = 57.05 dB, and for the image *Church* = 57.29 dB, with an Average PSNR on the three images of 58.07 dB.

By analyzing the Hall of Fame concerning the second phase, it is possible to note that all the best results have been achieved by A. Westfeld: if we exclude him, the best result was the one by Michel Chekrallah (EPSIL Lebanon), that reached 36.76 dB, whereas all the other results are slightly higher than the minimum threshold of 30 dB, being included between 30.11 dB and 31.53 dB, as shown in Figure 4.

In Figure 5 the same results are shown according to the Institute the participants belong to. Again, for each Institute the maximum value of the Average PSNR obtained by the contenders is shown. In this phase, people coming from only six different Institutes succeeded to remove the watermarks; in particular, fifteen participants of the Eurecom institute (France) entered in the Hall of Fame, even if their results were just above the threshold.

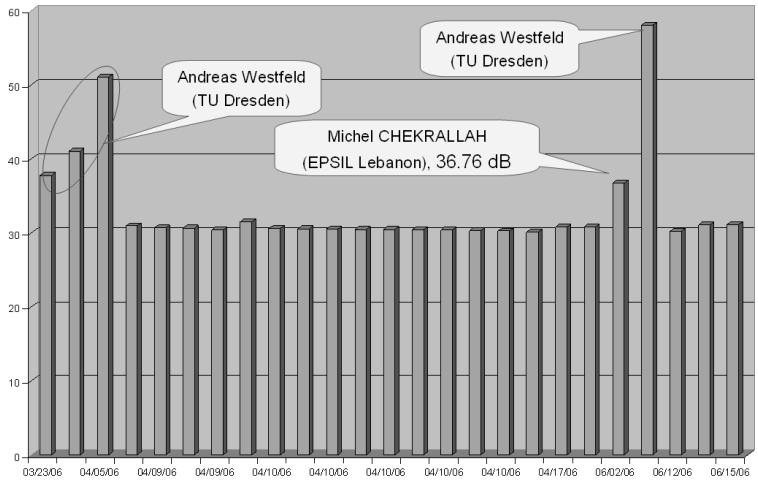

Fig. 4. The Average PSRN values obtained by the participants that entered in the Hall of Fame in the Second Phase of the Contest. Note that the best attacks have been registered by A. Westfeld.

1.3 First Phase vs. Second phase

By analyzing the results, summarized in Table 1, it is possible first of all to note that a limited number of participants succeeded to remove the watermark from all the three images, demonstrating that the adopted watermarking scheme is highly robust. In fact, at the end of the First phase of the Contest the Hall of Fame was composed by only 20 records, whereas in the second part 25 new successful attacks entered in. However, most of them registered more than once in the database, since they were able to increase the performance of their attacks, so that actually 10 participants succeeded in the first phase, and 17 in the second one. In particular, it is interesting to note that the best results were obtained by researchers expert and well known in the watermarking area. The attacks have been carried out by a high number of clients in the first phase; in many cases, only a limited number of trials were applied by the contender, without obtaining the removal of the watermark, after which the contender refrained from continuing the Contest. In the second phase a lower number of contenders participated, but with higher persuasion. The number of attacks in the second phase was ten times the attacks in the first one; however, the successful attacks were only twice as much, so that the percentage of successes decreased a lot from 13.9 % to only 2.86 %, showing that in the second part of the Contest the sensitive attack was heavily applied. This fact is confirmed if the number of attacks carried out by each IP address is analyzed. In Figure 6 the results of the First Phase and in Figure 7 the results of the Second one are given; each column represents the total number of attacked images uploaded by a given client to the BOWS server. In particular, the brighter columns indicate the IP addresses

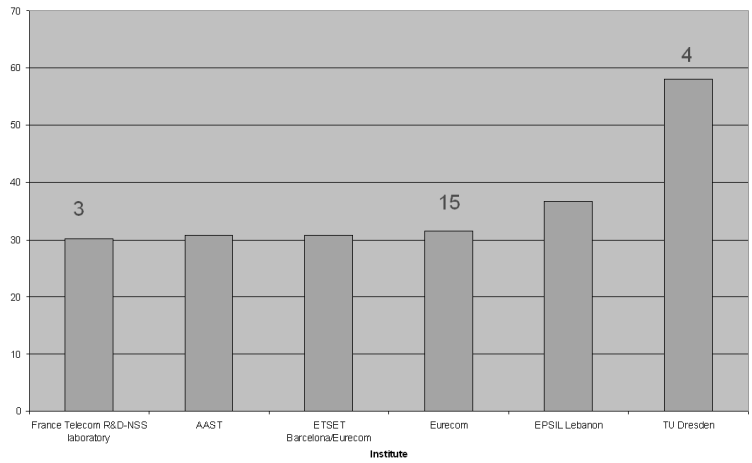

Fig. 5. The Average PSRN value obtained by the participants that entered in the Hall of Fame in the Second Phase of the Contest, grouped according to the Institute they belong to. Some research groups entered into the Hall of Fame more than once.

	First phase	**Second phase**
IP addresses	300	100
Attacks	72074	721734
Successes	10034	20666
% successes	13.90%	2.86%
Records in Hall of Fame	20	25
Participants in Hall of Fame	10	17
Best PSNR	39,22 dB	58,07 dB

Table 1. Summary of the results in the first vs. second phase of BOWS Contest.

of clients used by A. Westfeld; as it can be observed in the Figures, his attacks are prevailing in the second part of the contest, whereas in the first one, even if present, are not so many. These results seem to indicate that A. Westfeld based his participation to the Contest, with particular reference to the second phase, on the sensitivity attack.

It is then possible to conclude that the results of the second phase were deeply influenced by the massive use of the sensitive attacks carried out by A. Westfeld: by excluding his attacks, in fatc, the best result was the one obtained by M. Chekrallah (EPSIL Lebanon), 36.76 dB, that is 3 dB less than Cravers result. These results seem to demonstrate then that the knowledge of the watermarking algorithm did not influence very much the outcome of the Contest.

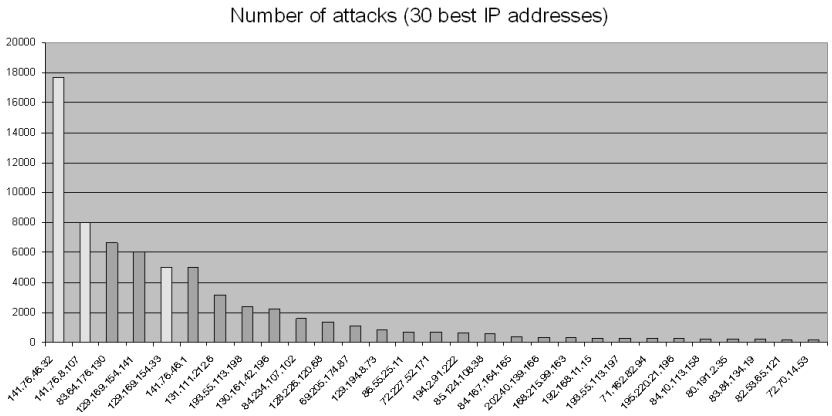

Fig. 6. Number of attacks carried out by each IP address in the First Phase. Note that the brighter columns indicate IP addresses of clients used by A. Westfeld.

2 Our view

As it is clear from the description given in the previous section, the results of the BOWS contest can not be interpreted univocally as a proof that algorithm knowledge is of no help to attackers. In the end, the best score of the second phase is much higher than that obtained in the first phase. At the same time, A. Westfeld admittedly made only a limited use of algorithm knowledge to obtain its highest score attack [5]. We should also consider that a result comparable to that obtained by A. Westfeld was obtained blindly by the group in the University of Vigo by relying on the BNSA attack. The question posed by the Wavila Challenge then is not so easy to solve. The opinion of the authors of this paper is that, all in all, knowing the watermarking algorithm is surely an help, even if the kind of improvement one can expect from such a knowledge must be carefully considered. Specifically, the following considerations hold:

– **Speeding up the attack** A problem with the BNSA attack is computational complexity. As a matter of fact removing the watermark from one of the images proposed by the BOWS contest without any prior knowledge about the watermark location, with a PSNR larger than 50dB, may take some weeks. Of course the attacker may perform some tests aiming at discovering some information about the watermark, e.g. the set of host features conveying the watermark, but this is also a time consuming exercise. A. Westfeld also admitted that the main benefit he got from the knowledge of the watermarking algorithm was a speed up of the attack. We can safely conclude, then, that the main advantage that knowing the watermarking algorithm may bring to the attacker is a (considerable) reduction of the attack complexity. On the other side, the advantage in terms of quality of the attacked document is likely to be rather limited.

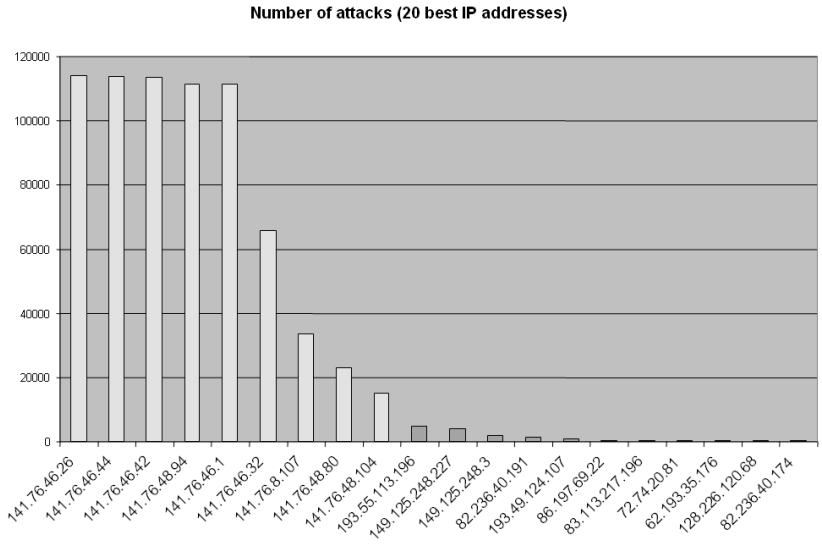

Fig. 7. Number of attacks carried out by each IP address in the Second Phase. Note that the brighter columns indicate IP addresses of clients used by A. Westfeld.

– **Watermark removal vs watermark security.** As it is becoming increasingly apparent, there is much more to watermarking security than mere watermark removal [6–8]. As clearly stated in [8] security attacks aim at acquiring full access to the watermarking channel rather then at watermark removal only. Such an ambitious goal requires that the secret key, whose secretness ensures the security of the watermarking channel, is discovered. Of course, once the secret key is disclosed, the attacker may use this knowledge to remove the watermark, but also to read it, insert a fake watermark, produce a valid forgery and so on. It is clear that security attacks are impossible if the watermarking algorithm is not known, hence making this knowledge a prerequisite of any attack against watermark security.

Acknowledgements

The work described in this paper has been supported in part by the European Commission through the IST Programme under Contract IST-2002-507932 ECRYPT. The information in this document reflects only the author's views, is provided as is and no guarantee or warranty is given that the information is fit for any particular purpose. The user thereof uses the information at its sole risk and liability.

References

1. Kerckhoffs, A.: La cryptografie militaire. Journal des Sciences Militaire **9** (1883) 5–38
2. Comesaña, P., Pérez-Freire, L., Pérez-Gonzaléz, F.: The blind newton sensitivity attack. In Wong, P.W., Delp, E.J., eds.: Security, Steganography, and Watermarking of Multimedia Contents VIII, Proc. SPIE Vol. 6072. Volume 6072., San Jose, CA, USA (2006) 149–160
3. Comesaña, P., Pérez-Freire, L., Pérez-Gonzaléz, F.: Blind newton sensitivity attack. IEE Proceedings on Information Security **153** (2006) 115–125
4. Miller, M.L., Doerr, G.J., Cox, I.J.: Applying informed coding and embedding to design a robust, high capacity watermark. IEEE Trans. on Image Processing **13** (2004) 792–807
5. Westfeld, A.: Lessons from the bows contest. In: Proceedings of VIII ACM Multimedia and Security Workshop, Geneva, Switzerland (2006) 208–213
6. Barni, M., Bartolini, F., Furon, T.: A general framework for robust watermarking security. Signal Processing **83** (2003) 2069–2084
7. Cayre, F., Fontaine, C., Furon, T.: Watermarking security: theory and practice. IEEE Trans. on Signal Processing **53** (2005) 3976–3987
8. Pérez-Freire, L., Comesaña, P., Troncoso-Pastoriza, J.R., Pérez-Gonzaléz, F.: Watermarking security: a survey. LNCS Transactions on Data Hiding and Multimedia Security (2006)

Considering Knowledge of Watermarking Algorithm and Finding the Optimal Watermark Algorithm

Kazuo Ohzeki[1], Li Cong[1], Kouhei Igarashi[1],

[1] Shibaura Institute of Technology,
14-I-30, ISE, 3-7-5 Toyosu, Koutou-ku, 135-8548 Tokyo,
Japan
{ohzeki, m105075, l03012}@sic.shibaura-it.ac.jp

Abstract. This paper discusses the problems of considering knowledge of embedding an algorithm and of exploring an acceptable watermarking system in actual applications. For these problems, a watermarking system disclosing its detector together with a watermarked image is considered. The detector hides its source code and only outputs the required minimized result after a decision of existence of the watermark. A user message part of the watermark is detached from embedded data, and is embedded in the detector. The number of parameters is so great that a sensitivity attack would have to do an astronomical number of trials to get even a single valid clue. The number of embedding keys is defined by combinatorial arithmetic. To realize this system, strict obfuscation of the detector is required. We have developed a new obfuscation concept for the specific watermark detection algorithm, which has 1:P ($P \geq 2$) transformation and which can be computationally complex at an arbitrary level. Validity of the obfuscation procedure is tentatively assumed and is not yet verified.

Keywords: sensitivity attack, asymmetric, authentication, obfuscation, computationally,

1 Introduction

This paper discusses one of the problems raised in the call for contributions of WaCha2006. For the first problem - whether knowledge of watermarking is useful or not - the sensitivity attack is considered through a new watermarking system which has a computationally complex property. The complexity should be based on a computational point of view. The sensitivity attack makes use of interim results as clues that the watermark detector produces for each small change of a watermarked image. Possible strategies against such sensitivity attacks are then to prohibit disclosure of the interim results and to increase the possibilities of changes.

To prohibit the disclosure of interim results, the detector hides the source code of its detection software program and only outputs the required minimized result after decision of existence of watermark inside the detector. To hide the detector software program, a complete obfuscation method for the detector is required. For such an obfuscation method, a transforming procedure with computationally complex features has been developed. Validity of the obfuscation procedure is tentatively assumed and is not yet verified.

The embedding process is constructed based on discrete combinatorial arithmetic with the merit of an explosive number of combinations. Combinatorial arithmetic produces a large number of embedding patterns for watermarking. Combinatorial arithmetic is not a linear operation nor does it occur in continuous space. If the embedding pattern is constructed at random, there is no direct link to effective trial testing in geometrical analysis, for example finding gradients in continuous space for the convergent point, which attackers may try to do.

Even though an embedding framework is disclosed, if the watermark has too many individual embedding parameters it is very hard to obtain a conformed set of parameters that the original owner has chosen by a computational criterion. The proposed watermarking system is an asymmetric one, which has many embedding parameters. The detector is resistant against sensitivity attacks.

Related to the second problem, the same proposed watermarking system works consistently to cope with it. As opposed to channel coding, where the noise can generally be efficiently modeled by Gaussian noise, watermarking noises include several attacks representing a wide range of noises of different natures [1]. The error rates for watermarking are larger than those of channel coding for signal communication. Error correcting codes are in general of no use for watermarking applications. This means that robustness cannot be improved by using error-correcting codes. Robustness is achieved by deepening the changes, for example by widening quantization levels. Capacity can be decided by the maximum degradation criterion. Therefore, to minimize the error rate under this restriction we must minimize embedding information to remove unimportant parts of the watermark, detaching these parts from watermarks as additional data, as auxiliary data, as backup data and as user messages. Ultimately, only the ID number of several bits is required for authentication. The proposed system, which will be described in the following section, uses the ${}_{N}C_{M}$ combinatorial embedding method, which ultimately uses a single bit ID watermark out of an astronomical number of possible embedding patterns. Because the number of embedded amount out of all candidates is limited, degradation is limited. The single ID bit is expanded to, for example, 64 bit embedded data.

2. Basic Structure of Proposed System

2.1 Motivation

This system was devised because of the need by individual people owning a homepage to keep the copyright of images on their websites without any costs or registration routines. For such a requirement, a new watermarking system was devised featuring resilience to attacks and with an authentication ability that does not require help by a trusted registration authority.

To increase robustness, the quantity of embedded information is minimized as far as possible. To establish authentication ability, a specific detector of the embedded watermark is put in the open region with the watermarked image. To disclose the detector in the open region, the detector software program must be obfuscated. Each detector reacts only to the specific image with the watermark that has been embedded and does not react to any other images without the watermark, nor to an image in

which a watermark has been embedded but subsequently lost through attacks. This can be seen as self Zero-Knowledge authentication in open region. This system may be classified into non-blind or semi-blind watermarks. However, the detector does not use the original image itself. Also, the detector is protected by obfuscation assumption. The system does not show the original image or any clues about the embedded watermark. It can be classified in a kind of blind watermarks.

2.2 Proposed System

Many papers utilize error correcting codes for improving the robustness of watermarked data [1-3]. Hernandez et al. showed experimental results using (63,36),(63,10)BCH codes, and Wang et al. did so using (8,4) extended Hamming code. However, the correcting ability depends on the Hamming distance of the codes, and BCH features did not contribute to improving the Hamming distance. Also, the class of the convolution codes is better suited to motion pictures or audio stream data, but is of no use for a single still picture, which has a limited length of data. To improve the error correction rate, one option is to reduce the source information length to be coded. In the watermarking system, we can split the watermark content into an imperative part and a subsidiary part.

A decomposition of watermark data is shown in Fig. 1. The general watermark

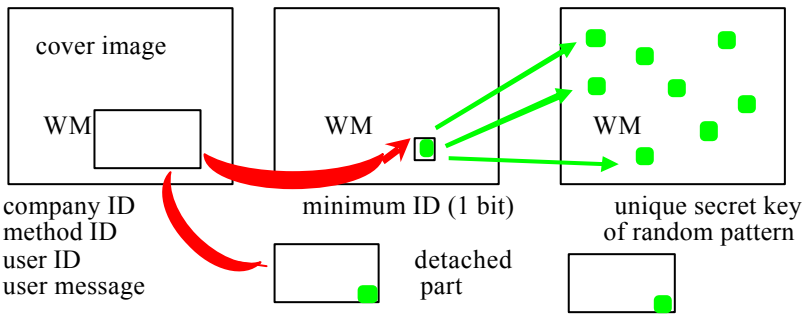

Fig. 1a Cover Image and Watermarks

Fig. 1b Reduced Watermark

Fig. 1c Unique Secret Key

layout. is shown in Fig.1a. The watermark can include all information regarding company ID, method ID, user ID, or any user messages, etc. As the capacity of the watermark is limited, we would like to reduce the amount of watermark data as much as possible. First, we decompose the watermark data into several classes. The most important part is that contributing to authentication. So, we dare to detach all watermark data except the minimum required bits for authentication. To authenticate the embedded watermark in some places in the image it is ultimately sufficient that the structure of the inspected image should accord with the detecting key operation. An example of this idea is shown in Fig.1c, whose unique secret key of random pattern will be inspected in the closed detector, which is the only one in the world, to individually match the embedded secret pattern.

The basic structure of the proposed system is shown in Fig.2. The the input picture of whole size is transformed by discrete Fourier transform (DFT) at a time. Nearly N=300x200 points around the low and middle frequency regions, except DC, are possible candidate points for embedding. The embedder selects M points (eg.M=64) from these candidates using random number generators or the like, and embeds single bit information 0 or 1 according to a secret key that the embedder only knows by quantization level shifting [4-5]. We call this embedding ${}_{N}C_{M}$, because there are N possible points and from these points the embedder selects M points, so the number of choices is the combination of N points taken M at a time.

All the M points are for the unique secret key, and other copyright information is detached from the embedded image and moved into the detector. The detector is a software program which is assumed to be computationally obfuscated. The obfuscated detector is put in an open region, paired with the corresponding watermarked image. Anyone can try to run the detector in his computer to verify whether the image has a watermark or not. The detection software checks the embedded unique pattern for the 64 points and decides "yes" if the image has more than half the significant watermarks, otherwise it decides "no". In addition, if the decision is "yes", the detector outputs related user messages which have been detached from the watermarked image and

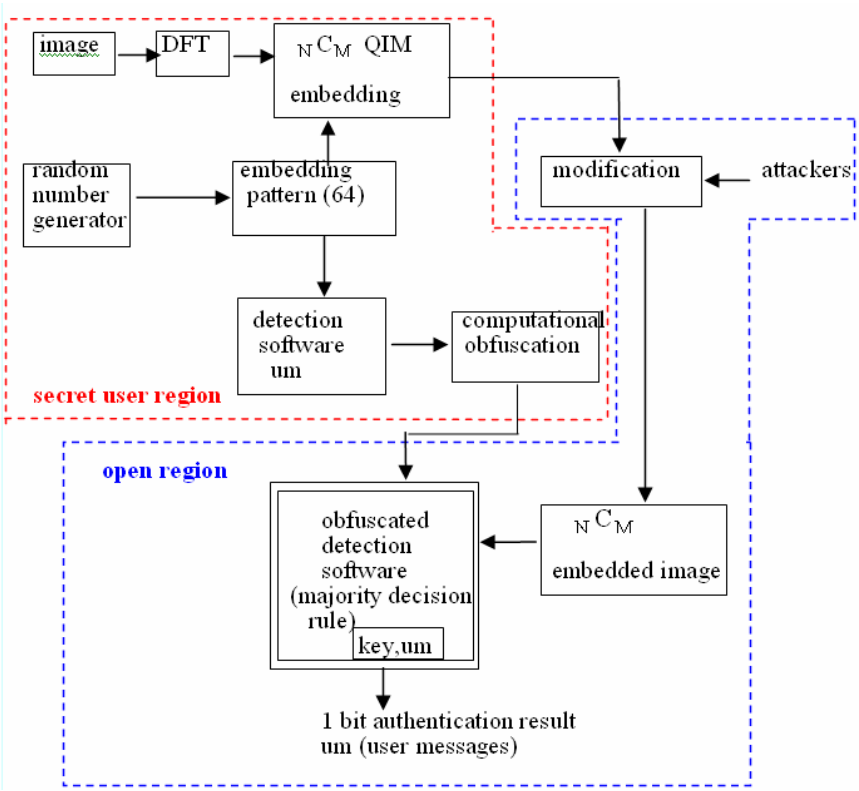

Figure 2. Proposed System

moved to the detector by the embedder. However, the detector does not output any interim results. It is very difficult to crack this detector in an open region because trial and error manipulation is not effective. The upper limit of degradation caused by attackers, who may modify the watermarked image, is assumed to be under the prescribed level in objective criteria such as signal to noise ratio (S/N), and in subjective criteria.

The conditions that the secret key, which is the embedding pattern, cannot be found by attackers in this system are as follows:

(i) It is difficult to analyze the detector software's behavior owing to obfuscation.

(ii) Even if another user tries to make another detector using the same kind of embedding program, it is probabilistically impossible for the person to find an identical key pattern owing to the large number of patterns.

The conditions above will be considered in the next section.

2.3 Estimating the Number of Embedded Patterns

In the proposed system shown in Fig.2, after transformation into the DFT domain, one of two kinds of quantization value is selected for each point [5]. It is a question of design regarding increased robustness whether to make the quantization step size large or to increase embedded points with small Q-step size and use error-correcting codes. However, the linear error correcting codes generated by polynomials do not contribute to improve correction ability beyond the Hamming distance. To achieve sufficient robustness, the quantization step size should be as large as possible. Embedded positions in the DFT domain are in low and middle frequency components, except the DC position. For a standard digital television picture in Japan with a size of 720x480, nearly 300x200=60000 points (we call this number N) can be candidates. From these N points, arbitrary M points are selected by the embedder. The method of selection is the combination of N points taken M at a time, which is written as

$${}_{N}C_{M}. \qquad \text{eq(1)}$$

The quantization is done uniformly, and the quantization error is

$$\frac{1}{12}Q^{2} \qquad \text{eq(2)}$$

for each point.

Here, we can assume that the number of embedding patterns, which is the number of keys, is much larger than the number of users. Each user randomly selects M points from N possible points, then also randomly selects 0 or 1 for each point and changes the quantization level. The set of embedded patterns that a user selects is unique. The probability of coincidence of the set of embedded patterns of two users is extremely small. In the case of N=60000, M=64, the probability of coincidence of all patterns of two users is,

$$P_{m2u} = \frac{{}_{64}C_{64}}{{}_{60000}C_{64}}\left(\frac{1}{2}\right)^{64} = 2.3\times10^{-233} \qquad \text{eq(3)}$$

because the case is the only one among the combination 60,000 points taking 64 points and for each point, two cases of bit 0 or bit 1 occur. In the case that half the selected points coincide and the other half do not,

$$P_{m2u\_h} = \frac{({}_{64}C_{32}) \times ({}_{60000-64}C_{32})}{{}_{60000}C_{64}} \left(\frac{1}{2}\right)^{32} = 5.3 \times 10^{-88} \qquad \text{eq (4)}$$

These probabilities can both be recognized as sufficiently small values for exhaustive searches.

3. Performance against Sensitivity Attack

For the sensitivity attack, as there may be many advanced attacking methods, we first consider only the direct method and expect it will work for other complicated sensitivity attacks as well.

If an attacker made a small change to a watermarked image and observed the output of the detector, the attacker could not obtain any evidence but would always obtain the same answer that the watermark exists because the changes are too small and within the threshold of quantization. If an attacker made a large change to a watermarked image, the attacker could not obtain any evidence but would always obtain the same answer that there is no watermark because the changes are too large and outside the threshold of quantization. It is very difficult to match the embedding patterns in the probability point of view. No continuous linear changes can reach the conforming line, though the differences in the watermarked image and the attacked image approach zero or diverge in the detector internally.

Next, the attacker can find out the framework of this proposed system and can attack using knowledge of the embedding algorithm. If the attacker knew the exact embedding algorithm except for the embedding parameters (in other words the attacker knew the DFT coefficients, quantization levels, and on/off differences) then a number of ${}_{N}C_{M}$ trials could yield clues to finding embedding patterns. However, the proposed system actually has more variable parameters for individual embedders, such as variable DFT coefficients, namely so-called Modified DFT (MDFT), variable quantization level width, and variable quantization level intervals. At this point, the attacker fails in sensitivity attack trials because of the astronomical number of trial times and because of the complete impossibility of obtaining effective sensitive reactions from the detector.

As long as the obfuscation works well, most of the sensitivity attack [6] trials attempt in vain to obtain a simple “yes” output because the probability of obtaining “no” output is very small, even though the attacker has nearly the same embedding program, which uses the same DFT coefficients and quantization step-sizes.

4. Obfuscation

There has been much research on obfuscation of computer software, including consideration of techniques such as layout changes, renaming, hiding memory data and changes of control flow. However, as these are superficial procedures, it is hard to evaluate their effects. Ogiso et al. proposed a method to map function pointers to array and state it is NP-hard [9].

Detection program software of the proposed watermark has the following characteristics.

(i) It is not general purpose calculation software, but very specifically designed to judge watermarks.

(ii) As usual, the QIM method is used, and data are quantized, so the precision of calculation is within a certain range of width. This implies that the software calculation is allowed to have errors, which means that the software can behave inaccurately to some extent.

A conventional obfuscated algorithm must work exactly the same as the original non-obfuscated software. Such an obfuscation transformation is called a homomorphic transformation. In consideration of this, for watermark applications, especially for the proposed one, there is no need to keep the homomorphic condition. We call such obfuscation non-homomorphic obfuscation for a specific watermarking system [10]. As an example, let consider an addition,

c=a+b

For the addition, two different sentences are produced.

$c = a + b + \varepsilon$

$c = a + b - \varepsilon$

where, ε is small value, eg. $\varepsilon = 0.000001$. Using if-close, the original addition sentence can be mapped to two meaningful sentences.

The detailed algorithm is shown in [10]. It uses 1:P transformation, which is recognized as an analog transformation. Based on the 1:P transformation, we can derive a kind of computationally complex obfuscation, which makes the software complex but would make it harder for attackers to reverse analyze it because of computational loads.

5. Experimental results

The general characteristics of the proposed system have been presented before [4,5]. In this contribution, two experiments are shown. The first one is for n times embedding. The detection rate for the first watermarked image in which the watermark is embedded by the owner himself may easily be overridden by the same embedding software. However, as the proposed system has many possible parameters as embedding keys, the detection rates remain high for n-times embedding. Figure 4 shows detection rates vs. the numbering. The embedded positions are different for each.

The signal-to-noise ratio for embedding is shown in Fig.5. The S/N constantly decreases. We assume that the watermark has significance of existence during high S/N while for lower S/N, the value of the image is small enough to be protected by the embedding watermark. From Fig.4 and Fig.5, we can find that this proposed system is valid up to the number of embedding times of 9 concerning detection ratio. On the other hand, concerning S/N, two or three processes of embedding cause slight degradation, and nine processes of embedding cause heavy degradation, which means this system has effective performance for several processes of embedding, and multiple embedding causes a loss in image value.

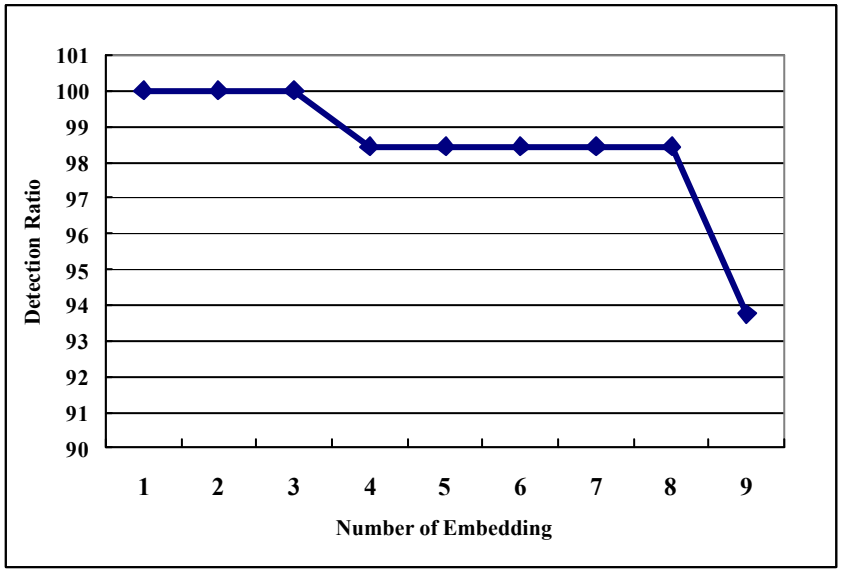

Figure 4. Detection Ratio vs Number of Embedding Times.

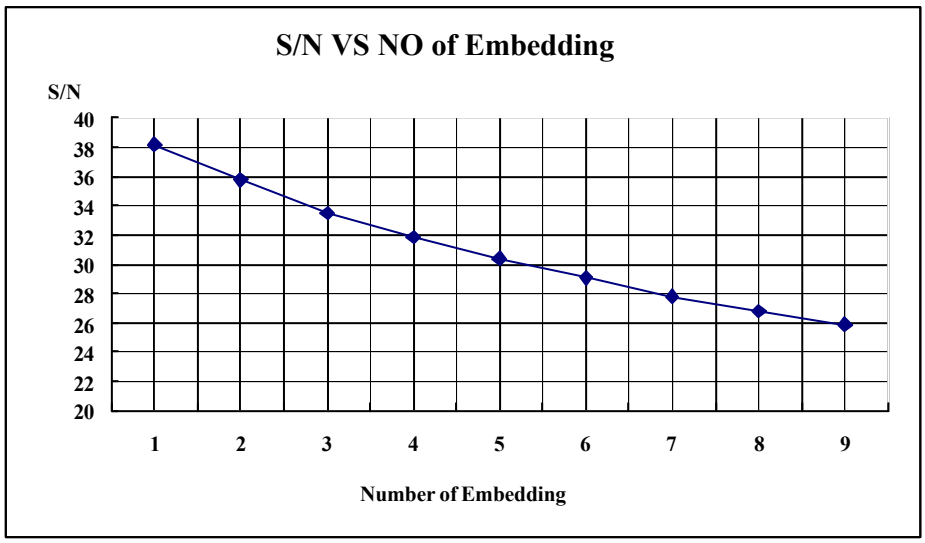

Figure 5. S/N vs. Number of Embedding Times.

6. Conclusion

An example of a system to cope with sensitivity attack is presented. The result of consideration of how this system copes with direct attacks is that the detector, which does not show any interim results, is effective in restricting sensitivity attacks. Correct small changes, which conform to the embedded patterns of a watermarked image, can contribute to removing a watermark at stages of a sensitivity attack but the probability of finding a pattern to accord with the watermark embedded pattern is much too small On the other hand, non-correct small changes, which do not conform to the embedded pattern of the watermarked image, randomly produce the output of the detector. If the change level is larger, the output of the detector will be "no watermark". If the change level is smaller, the output of the detector will be "yes, there is a watermark". The level is stronger for output of the detector than correctness of partial pattern of the watermark. The trial will not end in the attacker's lifetime.

Our watermarking system is proposed in consideration of sensitivity attacks. The embedded information is minimized to ID number only, and the other information concerning user messages is detached and embedded into the detector. The detector software program is computationally obfuscated. The obfuscation is a kind of 1:P transformation, which can be thought of as homomorphic. It can be taken as an analogous transformation, which means it is very hard to analyze in re-engineering because there is no definite inverse transformation.

The number of embedding patterns can be enlarged further. The proposed system is based on completeness of the obfuscation algorithm, which needs to be proved in the future. An analysis of coincidence for two differently embedded images in the space region rather than the frequency region is a further area of study.

Acknowledgments. The authors thank The Telecommunications Advancement Foundation (TAF) in Japan for supporting this research.

This paper is revised after WaCha2006.

References

1. S. Katzenbeisser et al. ed., "Information Hiding Techniques ..", p.135, Artech House 2000.
2. J. R. Hernandez, et al.,"The impact of channel coding on the performance of spatial watermarking for copyright protection,". ICASSP'98, vol. 5, pp. 2973--2976. May 1998.
3. James Z. Wang et al., "WaveMark: Digital Image Watermarking Using Daubechies' Wavelets and Error Correcting Coding," Proceedings of SPIE, vol. 3528, pp. 432-439, Nov. 1998.
4. K.Ohzeki, Michinori Nakajima, Kouhei Yasojima, "A Proposal of Watermarking System with Maximized Resilience", Proc. IPSJ CSEC-32 pp.61-66, Mar. 2006.(in Japanese)

5. Kazuo Ohzeki Li Cong, “Consideration on Variable Embedding Framework for Image Watermark against Collusion Attacks”, Proc. WAVILA Workshop on WaCha 2005, D.WVL.2-1.0.pdf pp.54-62, June 2005.
6. Pedro Comesaña, Luis Pérez-Freire and Fernando Pérez-González, “The Blind Newton Sensitivity Attack”, Proc. of SPIE-IS&T Electronic Imaging, SPIE Vol. 6072, 60720E, 2006.
7. .Scott Craver, Nasir Memon, Boon-Lock Yeo, Minerva M. Yeung, “Resolving Rightful Ownerships with Invisible Watermarking Techniques: Limitations, Attacks, and Implications”, IEEE Journal on Selected Areas in Communications Volume: 16, Issue: 4 pp. 573-586, May 1998
8. Stefan Katzenbeisser ed. “First Summary Report on Hybrid Systems”, ECRYPT, D.WVL.5-1.0.pdf 2005.
9. T.Ogiso et al., “Software Obfuscation on a Theoretical Basis and Its Implementation”, IEICE Trans. Vol. E86-A, No. 1, pp.176-186, Jan. 2003.
10. Li Cong and K.Ohzeki, “Asymmetric watermarking system with Disclosed Detector for Authentication in open Area”, IEICE Technical Rept. ISEC2006-72, pp.1-7, Sept.2006 .(in Japanese)
11. Andr´e Adelsbach and Ahmad-Reza Sadeghi , “Zero-Knowledge Watermark Detection and Proof of Ownership”, Proceedings of Information Hiding: 4th Inter. WS, 2001, USA, April 25-27, 2001.
12. Kuribayashi, M.; Tanaka, H., “Fingerprinting protocol for images based on additive homomorphic property”, Image Processing, IEEE Transactions on Volume 14, Issue 12, Dec. 2005 Page(s): 2129 - 2139

Practical Audio Watermarking Evaluation Tests and its Representation and Visualization in the Triangle of Robustness, Transparency and Capacity

Andreas Lang[1], Jana Dittmann[1], David, Megías[2], Jordi Herrera-Joancomartí[2]

[1]Research Group Multimedia and Security,
Department of Computer Science,
Otto-von-Guericke-University of Magdeburg, Germany
[2]Universitat Oberta de Catalunya, Spain

Abstract. Digital watermarking is a growing research area to mark digital content by embedding information into the content itself. The evaluation of watermarking algorithms provides a fair and automated analysis of specific watermarking schemes for selected application fields. In this paper, we present a intra-algorithms evaluation and analysis of a selected audio watermark scheme. Therefore, different requirements are introduced and focussed. The average, maximum and minimum values of the test results for robustness, transparency and capacity are discussed and visualized in the well known magic triangle.

1 Motivation and Introduction

Digital watermarking embeds additional information into digital medias. This technology opens and provides additional and useful features for many application fields (like DRM, annotation, integrity proof and many more). The evaluation of watermarking algorithms provides a fair and automated analysis of specific watermarking schemes for selected application fields. Currently, many researcher use their own evaluation system which does not provide comparability each other. The evaluation process can therefore be very complex and the actual research investigates into evaluation approaches with special attacks for images (see for example available tools StirMark [19], Optimark [14], Checkmark [4]) or for a specific applications like DRM, see for example [1] or so-called profiles, see for example [12, 16]. A classification of general watermarking attacks to evaluate the robustness is introduced for example in [11]. Therein attacks are classified into removal, geometrical, security and protocol attacks. [21] extends the definition by including estimation attacks or [9, 15, 2, 3] introduce attacks to gain knowledge about the secrets of the system (embedding and/or detection/retrieval) to also evaluate the security of watermarking algorithms. Besides the focus on the robustness and security evaluation for example in [10] the transparency of different steganographic and watermarking algorithms is analyzed.

Three important properties of watermarking schemes are usually applied to assess performance, namely robustness, capacity and transparency [6]. Often, an improvement in one of these properties implies a decline in some of the other ones and, thus, some trade-off solution must be attained. In general most evaluation systems do focus on robustness and transparency. Many other properties like capacity, complexity or security are neglected. In this paper, we introduce firstly a theoretical description to define these properties in a general way. This description is then used to provide exemplary a intra-algorithm evaluation and analysis of a selected n-bit audio watermarking scheme. The test results are visualized in the well known magic triangle of robustness, transparency and capacity.

This paper is organized as follows: Section 2 introduces the framework and its practical usage. Furthermore, the test scenario, test set and test goals for the evaluation of a selected watermark scheme for intra-algorithm analysis are shown. At the end of this section, the test results are presented and its visualization in the magic triangle is presented. The section 3 summarizes our approach and impacts and concludes with future work.

2 Framework and Practical Tests

In this section, we introduce our framework and set up a practical evaluation of a selected watermarking scheme to provide intra–algorithm evaluation and analysis.

If the position of a watermarking scheme is presented in the triangle of robustness, capacity and transparency, then these properties must be measured. Therefore, the following itemization introduces these three properties briefly [5].

Transparency: The transparency relates to the degree of distortion introduced by the embedding function of a watermarking scheme. It is defined in an interval $[0, 1]$ where 0 provides the worst case (the distortion of the test signal are so different that it cannot be recognized as a version of a given reference signal) and 1 is the best case (an observer does not perceive any significant difference between two given signals). Thereby, the transparency $\mathrm{tra}_{E\mathrm{rel}}$ is related to a particular audio signal. It is usually better to provide some absolute values of transparency that are not related to a particular audio signal. The absolute embedding transparency is related to a given audio test set (family of audio signals) to be marked. Then, the average ($\mathrm{tra}_{E\mathrm{ave}}$), maximum ($\mathrm{tra}_{E\mathrm{max}}$) and minimum ($\mathrm{tra}_{E\mathrm{min}}$) embedding transparency can be measured. These measure values are computed as follows: the average transparency ($\mathrm{tra}_{E\mathrm{ave}}$) is the arithmetic mean over the whole test set, the maximum transparency ($\mathrm{tra}_{E\mathrm{max}}$) identifies the best possible embedding transparency and identifies the audio signal. In contrast, the minimum transparency ($\mathrm{tra}_{E\mathrm{min}}$) is the worst embedding transparency derived by the embedding function and identifies the audio signal.

Capacity: The attacking capacity relates to the retrieved capacity of the message by using the retrieval function after performing attacks. Whereby the relative attacking capacity ($\text{cap}_{A\text{rel}}$) is related to a specific attack and specific audio signal. It is defined in an interval $[0, 1]$ whereby 0 means the worse case (the retrieved message of the retrieval function is completely different to the previous embedded message) and 1 is the best case (the retrieved message is identical to the embedded message). It is often useful to measure the average attacking capacity ($\text{cap}_{A\text{ave}}$) over a given test and attacking set. Furthermore, the maximum attacking capacity ($\text{cap}_{A\text{max}}$) identifies the attack and audio signal with the best results. In contrast, the minimum attacking capacity ($\text{cap}_{A\text{min}}$) identifies the attack and audio signal where the embedded message cannot be retrieved correctly.

Robustness: If the robustness of a watermarking scheme is measured, then two properties are important. Firstly, the retrieve ability of the embedded message after performing an attack. The second property is the attacking transparency, which should be as transparent as possible for the attacker. It means, that an attack against the watermark is successful only, if the watermark cannot be retrieved correctly and the distortion occurred by the attack are not audible. Therefore, a watermarking scheme is defined as NOT robustness, if the watermark cannot be detected in a particular audio signal whereby the attack with the best transparency return the measured robustness value. The robustness is defined in an interval $[0, 1]$, whereby 0 is the worst case (the watermark can be attacked, not able to retrieve) after performing an attacking with any audible distortions. In contrast, a robustness of 1 identifies a robust scheme, whereby the watermark can only be destroyed with extremely distortions on the audio signal. If the average robustness (rob_{ave}) is measured, then the arithmetic mean over the whole audio test set is computed. The minimum robustness (rob_{min}) identifies a particular attack and particular audio signal, where the watermark was destroyed successful. A maximum robustness makes no sense, because an attack, with the worst attacking transparency cannot be defined as successful.

There are many other properties (like complexity, security) but they and their measurement are out of focus for this publication.

2.1 Test Scenario

Our test scenario is as follows. All audio signals S provided by the test set $\mathcal{S}$ are used as cover medium. The embedding function E and its selected parameters $\boldsymbol{p}_E$ embeds the given message m into S. The average, maximal and minimal embedding transparency ($\text{tra}_{E\text{ave}}$, $\text{tra}_{E\text{max}}$ and $\text{tra}_{E\text{min}}$) of E is measured by computing the Objective Difference Grade (ODG) [8] with the implementation of [13]. Furthermore, the detection/retrieval function tries to retrieve the embedded message m' after applying the embedding function in order to identify, if the embedding was successfully or not. After a successful embedding, the marked

audio signal $\hat{S}$ is attacked by single attacks $A_{i,j}$ (whereby i is the name of the attack and j the used parameters) and its default and improved attack parameters $\boldsymbol{p}_{A_{i,j}}$ provided by StirMark for Audio (SMBA) [20]. The average, maximal and minimal attacking transparency ($\text{tra}_{A\text{ave}}$, $\text{tra}_{A\max}$ and $\text{tra}_{A\min}$) of $A_{i,j}$ with $\boldsymbol{p}_{Ai,j}$ is measured. Then the retrieval function R with its parameters $\boldsymbol{p}_R$ tries to retrieves the message m' from the attacked audio signal $\tilde{S}$. The following figure 1 shows the test scenario and introduces the simple measuring points.

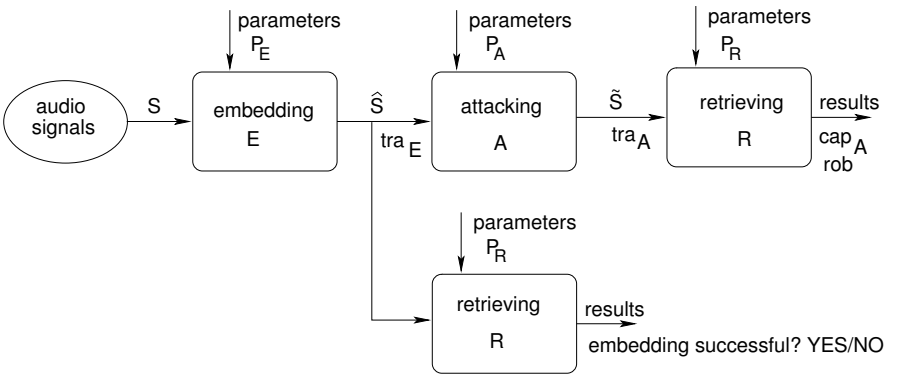

Fig. 1: Test Environment

As watermarking embedding algorithm, we select the 2A2W watermarking scheme. 2A2W works in the wavelet domain and embeds the watermark on selected zero tree nodes [17]. It does not use a secret key and can therefore categorized, from the application point of view, as an annotation watermarking scheme. An additional file is created, where the marking positions are stored to retrieve the watermark information in detection/retrieval function (non blind) [7]. By using 2A2W, the following parameters are defined for this algorithm:

- p_1: specifies the internal embedding method and at present only ZT (zerotree) is possible.
- p_2: specifies the internal coding method and at present, only binary (BIN) is possible.

2.2 Test Set

In this subsection, the test scenario and the test set used for the experiment is introduced.

The audio test set $\mathcal{S}$ contains 16 audio files of the well known SQAM [18] test set. All audio signals are in CD quality and they have a sampling rate of $44.1kHz$ with two audio channels (stereo) and $16bit$ sample resolution. The minimal length

of an audio signal is 16.3s, the maximum length 34.9s and the average length of all audio signals 21.26s. Furthermore, the audio files are categorized in three types of content, which is shown in table 1. Therefore, the first category *single instrument* contains 7 audio files, where a single music instrument is audible, the second category *speech* contains spoken text with female and male voices in the languages English, German and French. The last category *singing* contains female, male and a mixture of both singing voices.

single instruments	speech	singing
harp40_1.wav	spfe49_1.wav	bass47_1.wav
horn23_2.wav	spff51_1.wav	sopr44_1.wav
trpt21_2.wav	spfg53_1.wav	quar48_1.wav
vioo10_2.wav	spme50_1.wav	
gspi35_1.wav	spmf52_1.wav	
gspi35_2.wav	spmg54_1.wav	
frer07_1.wav		

Table 1: Audio files and its classification used for the test scenario

As message m, the string *"Tests"* is used for embedding.

2.3 Test Goals

In this subsection, our test goals are introduced. Thereby, we use the test scenario introduced above and we show exemplary the intra-algorithm analysis and evaluation. Therefore, we firstly measure the average, maximum and minimum embedding transparency, retrieved capacity after attacking and robustness for the 2A2W watermarking scheme. For embedding, the default parameters $\boldsymbol{p}_E$ are used and for attacking, the attack set $\mathcal{A}$ based on all 42 attacks provided by StirMark for Audio and their default attacking parameters [20] are used to measure the robustness and average, minimum and maximum attacking capacity. All test results are used to provide an intra-algorithm evaluation and analysis by visualizing the results in the magic triangle and to show the position of 2A2W.

2.4 Test Results

In this subsection, the test results for the intra-algorithm evaluation and analysis are presented. Thereby we show the average, minimum and maximum test results for the embedding transparency, attacking capacity and robustness and visualize them in the magic triangle.

2A2W is able to embed the message *"Tests"* into all audio files successfully. Thereby, a fixed embedding capacity is used for embedding and the retrieval function returned exactly the same message for all audio files.

The test results for the embedding, retrieval and attacking function are shown in the following table 2.

embedding E	retrieval R	attacking A
$\text{tra}_{E\text{ave}}$=0.63, $\text{tra}_{E\text{min}}$=0.02, $\text{tra}_{E\text{max}}$=0.95	retrieval is always possible	$\text{tra}_{A\text{ave}}$=0.36, $\text{tra}_{A\text{min}}$=0.02, $\text{tra}_{A\text{max}}$=1.00
		$\text{cap}_{A\text{ave}}$=0.77, $\text{cap}_{A\text{min}}$=0.00, $\text{cap}_{A\text{max}}$=1.00
		rob_{ave}=0.36, rob_{min}=0.02

Table 2: Test results for the evaluated watermarking scheme 2A2W

The best embedding transparency is measured with $\text{tra}_{E\text{max}} = 0.95$ for the audio test file *spmg54_1.wav* (which is speech) and the worst with $\text{tra}_{E\text{min}} = 0.02$ for audio test file *frer07_1.wav* (which is a single instrument). The average embedding transparency is measured with $\text{tra}_{E\text{ave}} = 0.63$. These measured results show that the embedding transparency of 2A2W depends on S and therefore the quality of embedding depends on the type of audio content. The retrieval after embedding measured shows that the given message "Tests" fits into all audio files successful. The test results for the attacking function with the following retrieval show that 2A2W is not robust against the attacks *FFT_Invert* and *Invert*, that also produce a good attacking transparency. The robustness measure shows that an average robustness of 0.36 and minimum robustness of 0.02 is measured. The minimum robustness shows that there is an attack available, which destroys the watermark and this attack has a very good attacking transparency *Invert*. If we focus only on the attacking transparency, then there are other attacks (like *Nothing* or *BitChanger*) which have also a very good attacking transparency ($\text{tra}_{A\text{max}} = 1.00$). In contrast, other attacks like *Resampling, Normalizer2* or *DynamicTimeStretch* produce a very bad attacking transparency.

If we introduce and summarize the intra–algorithm evaluation and analysis based on the attacking set and we distinguish between average, minimum and maximum values, then the following test results are obtained. The following table 3 summarized the test results needed for the visualization in the magic triangle.

	tra_E	cap_A	rob
average	0.63	0.77	0.36
maximum	0.95	0.00	n.a.
minimum	0.02	1.00	0.02

Table 3: Summarized test results for the intra–algorithm evaluation of 2A2W

For the visualization of these measured values in the magic triangle, we use the symbol ■ for embedding transparency, + for attacking capacity and × for robustness measures. Furthermore, the following symbols around the measured values show the average (●), maximum (▲) and minimum (■) location in the magic triangle. The exact position of the watermarking scheme with its average values (●) and minimum values (■) is exemplary shown. All these positions are shown in the following figure 2.

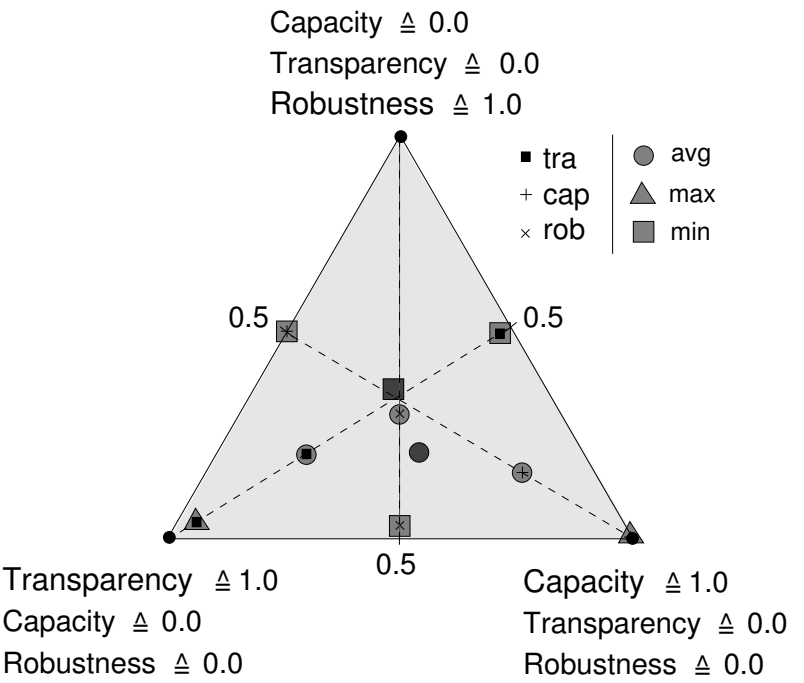

Fig. 2: Test results after attacking for 2A2W in the triangle for average, maximum and minim test results

This visualization shows the different positions of measured average, maximum and minimum values of the 2A2W watermarking scheme, evaluated with the SQAM audio file as test set and the default single attacks from StirMark for Audio.

3 Summary

In this paper, we introduces briefly and evaluation framework. Based on this methodology, the 2A2W watermarking scheme is selected for evaluation, whereby the embedding transparency, attacking capacity and the robustness is measured. These test results are introduced and its position in the magic triangle shown. The robustness evaluation showed, that 2A2W is not robust against the *Invert*

attack provided by StirMark for Audio whereby this attack has a very good attacking transparency.

With this evaluation and representation of the test result is shown, that the evaluation itself provides the measured values, whereby the representation of them is not intuitive. Therefore, other visualization form could provide a better understanding. Future work is to analyze other visualization forms and to evaluate more watermakirng schemes.

Acknowledgements

The work about single SMBA attacks described in this paper has been supported in part by the European Commission through the IST Programme under Contract IST-2002-507932 ECRYPT. The information in this document is provided as is, and no guarantee or warranty is given or implied that the information is fit for any particular purpose. The user thereof uses the information at its sole risk and liability.

Effort for transparency evaluation of the audio attacks is sponsored by the Air Force Office of Scientific Research, Air Force Materiel Command, USAF, under grant number FA8655-04-1-3010. The U.S. Government is authorized to reproduce and distribute reprints for Governmental purposes notwithstanding any copyright notation thereon. The views and conclusions contained herein are those of the authors and should not be interpreted as necessarily representing the official policies or endorsements, either expressed or implied, of the Air Force Office of Scientific Research or the U.S. Government.

UOC authors are partially supported by the Spanish MCYT and the FEDER funds under grants MULTIMARK TIC2003-08604-C04-04 and PROPRIETAS-WIRELESS SEG2004-04352-C04-04

Furthermore, we thank Jörg Wissen who helped us to set up the test environment and to analyze of the test results.

References

1. Macq, Benoit, Dittmann, Jana, Delp, Edward J., *Benchmarking of Image Watermarking Algorithms for Digital Rights Management*, Proceedings of the IEEE, Special Issue on: Enabling Security Technology for Digital Rights Management, pp. 971–984, Vol. 92 No. 6, June 2004
2. F. Cayre, C. Fontaine and T. Furon, *Watermarking security, part I: theory*, In: Security, Steganography and Watermarking of Multimedia Contents VII, Ping Wah Wong, Edward J. Delp III, Editors, Proceedings of SPIE Vol. 5681, San Jose, USA, 2005
3. F. Cayre, C. Fontaine and T. Furon, *Watermarking security, part II: practice*, In: Security, Steganography and Watermarking of Multimedia Contents VII, Ping Wah Wong, Edward J. Delp III, Editors, Proceedings of SPIE Vol. 5681, San Jose, USA, 2005

4. Checkmark Benchmarking, *http://watermarking.unige.ch/Checkmark/*, 2006
5. Jana Dittmann, David Megías, Andreas Lang, Jordi Herrera-Joancomartí, *Theoretical framework for a practical evaluation and comparison of audio watermarking schemes in the triangle of robustness, transparency and capacity*, accepted Journal in Springer LNCS Transactions on Data Hiding and Multimedia Security, 2006
6. J. Fridrich, *Applications of data hiding in digital images*, Tutorial for the ISPACS 1998 conference in Melburne, Australia, 1998
7. H. Inoue, A. Miyazaki, A. Yamamoto, T. Katsura, *A Digital Watermarking Technique Based on the Wavelet Transform and Its Robustness on Image Compression and Transformation*, IEICE Trans. Fundamentals, vol. E82-A, no. 1, 1999
8. ITU-R Recommendation BS.1387, *Method for Objective Measurements of Perceived Audio Quality*, *http://www.itu.int/rec/R-REC-bs/en*, Dec. 1998
9. T. Kalker, *Considerations on watermarking security*, In: Proceedings of the IEEE Multimedia Signal Processing MMSP01 workshop, Cannes, France, pp. 201–206, 2001
10. Christian Kraetzer, Jana Dittmann, Andreas Lang, *Transparency benchmarking on audio watermarks and steganography*, to appear in SPIE conference, at the Security, Steganography, and Watermarking of Multimedia Contents VIII, IS&T/SPIE Symposium on Electronic Imaging, 15-19th January, 2006, San Jose, USA, 2006
11. M. Kutter, S. Voloshynovskiy and A. Herrigel, *Watermark copy attack*, In Ping Wah Wong and Edward J. Delp eds., IS&T/SPIEs 12th Annual Symposium, Electronic Imaging 2000: Security and Watermarking of Multimedia Content II, Vol. 3971 of SPIE Proceedings, San Jose, California USA, 23-28 January 2000
12. Andreas Lang, Jana Dittmann, *Profiles for Evaluation - the Usage of Audio WET*, to appear in SPIE conference, at the Security, Steganography, and Watermarking of Multimedia Contents VIII, IS&T/SPIE Symposium on Electronic Imaging, 15-19th January, 2006, San Jose, USA, 2006
13. Alexander Lerch, zplane.development, *EAQUAL - Evaluation of Audio Quality*, Version: 0.1.3alpha, http://www.mp3-tech.org/programmer/misc.html, 2002
libSNDfile library, *http://www.mega-nerd.com/libsndfile/*, May, 2006
14. Optimark, *http://poseidon.csd.auth.gr/optimark/*, 2006
15. Luis Pérez-Freire, Pedro Comesaña and Fernando Pérez-González, *Information-Theoretic Analysis of Security in Side-Informed Data Hiding*, Information Hiding, pp. 131–145, 2005
16. F. A. P. Petitcolas, M. Steinebach, F. Raynal, J. Dittmann, C. Fontaine, N. Fates, *Public automated web-based evaluation service for watermarking schemes: StirMark Benchmark*, In: Security and Watermarking of Multimedia Contents III, Ping Wah Wong, Edward J. Delp III, Editors, Proceedings of SPIE Vol. 4314, Bellingham WA, USA, pp. 575–584, ISBN 0-8194-3992-4, 2001.
17. J.M. Shapiro, *Embedded image coding using zerotrees of wavelet coefficients*, IEEE Trans. Signal Processing, vol. 41, no.12, pp. 3445–3462, 1993
18. SQAM — Sound Quality Assessment Material, *http://sound.media.mit.edu/mpeg4/audio/sqam/*, 2006
19. StirMark Benchmark, *http://www.petitcolas.net/fabien/watermarking/stirmark/*, 2006
20. StirMark Benchmark for Audio, *http://amsl-smb.cs.uni-magdeburg.de/*, 2005
21. S. Voloshynovskiy et al., *Attacks on Digital Watermarks: Classification, Estimation-Based Attacks, and Benchmarks*, IEEE Communications Magazine, vol. 39(8), pp. 118–126, Aug. 2001

Visualisation of Benchmarking Results in Digital Watermarking and Steganography

Christian Krätzer

Research Group Multimedia and Security,
Department of Computer Science,
Otto-von-Guericke-University of Magdeburg, Germany

Abstract. The goal of this paper is to facilitate the discussion about fitting representation approaches for fair benchmarking and the selection and use of techniques by non-experts. To meet this goal a brief review on digital watermarking (DWM) and steganography features commonly encountered in algorithm evaluation and benchmarking is given. Then selected techniques derived from the field of information visualisation are introduced and considered for application in the visualisation of research and benchmarking results in DWM and steganography.

1 Introduction

In 1998 J. Fridrich gave an extensive overview over the state of the art on data hiding in digital imagery [1], including a definition of the data hiding term, as well as a close review on two of the most common digital data hiding techniques: digital watermarking (DWM) and steganography. In the eight years since 1998 the development of digital watermarking techniques and steganography has made large progress. In the field of digital watermarking new properties like the invertability of watermarks have to be considered as well as the shifted importance of features, which is very well illustrated by the example of complexity. This feature gained a stronger relevance with the growing importance of mobile devices and the definition of application scenarios which require real time capable watermarking algorithms. In the field of steganography of course new types of covers were considered moving from largely storage channel based approaches to more sophisticated time channel based techniques or hybrid techniques. Also here the shifting of the relevance of features shows, e.g. with the robustness of steganographic techniques which are nowadays also considering algorithms compliant with faulty communication channels [2].
Besides the actual research on DWM and steganography algorithms the comparison of benchmarking results has gained importance in scientific publications on these topics. To address this fact selected visualisation techniques are presented for discussion within the watermarking community. Some of these techniques have not been used before in DWM benchmarking and the evaluation of steganography but might prove useful in further research.

The paper is structured into the following sections: Basics on information visualisation including a notation which is used to describe the visualisation problem encountered are described in section 2. In section 3 main features of DWM and steganography algorithms are reviewed since the characteristics of these data have a strong influence on the visualisation decision. Relevance of DWM and steganography benchmarking and analysis, and therefore the appropriate visualisation of these results is emphasised in section 4. Section 5 introduces visualisation techniques, sorted by the dimensionality of the entity to be visualised. Section 6 concludes the paper.

2 Visualisation in scientific work

Robert Spence divides in [5] visualisation techniques applied in scientific work into scientific visualisation (primary related to, and representing visually, something "physical", like the flow of water in a pipe, temperature distribution in materials, etc) and information visualisation (dealing with abstract quantities e.g. baseball scores, fluctuating exchange rates between currencies, etc). Based on the nature of algorithm evaluation and benchmarking, the main focus of this document is placed therefore in what he identifies as the area of information visualisation, but we also consider selected results from scientific visualisation (like a notation for describing the visualisation task) since as a science it exceeds information visualisation in age and the maturity of theoretical research.

It is common to present data, structures and relations graphically to enable efficient analysis and communication. This presentation requires the transformation of data of different kinds into geometric information (B.H. Mc Cormick et. al [7]). The two main goals in visualisation are to present (research) results and to facilitate the analysis of the data. In [6] the importance of finding a fitting presentation for a given data set is indicated by H. Schumann and W. Müller. The application of a inappropriate presentation might easily lead to incorrect interpretations in an analysis. Therefore it is fundamental to define and describe the characteristics of a set of data (the subject of visualisation) and consider these characteristics very early in the visualisation process. Here for a description of the visualisation tasks considered a notation introduced by K.W. Brodlie et. al [8] for scientific visualisation is used and applied to information visualisation.
This notation describes the abstraction of data from a so called "underlying field" to an "entity for visualisation" E. Thereby is E an entity specified on a domain (defined by number and type of the independent variables) and yielding a range (characterised by class and dimensionality) of results. Applied to general data presentation the notation uses E_n^F for describing an entity of class F ($F \in \{S, P, V_k, T_o\}$; scalar, set of points, vector with k components or tensor field of o-th order) with an domain of order n. Also the characteristics of the domain can be described in this notation. A continuous domain is denoted with n. If the entity is defined over regions of a continuous domain the notation uses $[n]$. If the entity is defined over an enumerated set $\{n\}$ is used. It is also possible with this notation to describe the fact that multiple results are intended to be visualised

over the same domain (e.g. two scalar fields like pressure and temperature within a volume in 3D; E_3^{2S}) or to describe composite representations.
H. Schumann and W. Müller [6] and Brodlie et. al [8] introduce examples to help with the understanding of this notation. Some of these are repeated in section 5. This notation introduced for scientific visualisation now has to be applied to our needs which are mainly to be found in information visualisation. This is done by considering only what Schumann and Müller call in [6] the "abstract dimensionality" of the observed space. This includes only the data which does not contain any positional or temporal information and binding.

3 Features of steganographic applications and digital watermarking algorithms

In section 2 the importance of characteristics of the sets of data to be visualised is highlighted. Therefore the main features of steganographic applications and DWM algorithms are reviewed here to provide knowledge necessary for the application of visualisation techniques. The description of features given is based on the work of J. Fridrich introduced in 1998 in [1] for data hiding techniques in the image domain. There the most important properties of data hiding schemes were identified as robustness, undetectability, invisibility, security, complexity, and capacity. Based on the definitions given there and using the knowledge that some of the above properties (namely robustness, capacity and undetectability/transparency) are mutually competitive, clear requirements for the construction of watermarking and steganographic algorithms can be derived. In the following the features of steganographic systems and DWM approaches are reviewed briefly for their requirements in presentation techniques.

Capacity: Basically the capacity definitions in steganography and DWM are the same. The question is how much data can be embedded within one byte or one second of cover. Sometimes constrains like a predefined transparency threshold have an impact on the maximum embedding strength applicable. In steganography generally more capacity is better, in DWM the required capacity strongly depends on the chosen application scenario. For example annotation watermarking might require a large capacity at relatively small proportions of the marked object. Necesary information regarding the co-domain of functions computing the capacity is that it is commonly a non-negative, continuous value, which in most cases does not exceed the capacity of the cover.

Robustness: [1] states that the embedded information is said to be robust if its presence can be reliably detected after the image has been modified but not destroyed beyond recognition. In this definition robustness means the resistance to blind, non-targeted modifications or image operations. This image domain based description of the term robustness has been outdated by the emergence of watermarking evaluation tools like Stirmark Benchmark ([9], [10]) or Stirmark Benchmark for Audio (SMBA; e.g. [11]). Lang et. al measure the robustness of a watermarking algorithm for their SMBA in terms of robustness against a pre-

defined set of attacks (signal modifications).
This approach, which tests DWM algorithms against blind, targeted modifications, can be transferred to steganography (see [13], [2]) but here in addition to the integrity of the message the impact of the embedding on the cover(-protocol) has to be considered. If results for this approach used by Lang et. al have to be visualised, the co-domain concerned is a discrete value in the range between zero and the maximum number of attacks. Since the attacks can be grouped into classes depending on the domain they work in or the type of modification they perform, there a need to use a vector might arise to adequately describe the robustness results for the different classes identified.

Transparency (Perceptual transparency and statistical undetectability): [1] distinguishes between the two terms Undetectability (an image with an embedded message is consistent with a model of the source from which images are drawn) and Invisibility (an average human being is not capable to distinguish between carriers that do contain hidden information and those that do not). Instead of this approach to describe the transparency of a message embedding in two terms we would like to refer to a more recent and more formal approach given in [3]. There both terms used in [1] (Undetectability and Invisibility) are joined to form a more appropriate measure labelled transparency. In [3] the differences between transparency considerations in the fields of steganography and digital watermarking are considered in detail, highlighting amongst others the importance of transparency as the main feature in steganography and the strong dependance on the selected application scenario for the transparency requirements in DWM.
In the presentation of transparency results for selected algorithms scalar values or vectors containing the results from an analysis with different measurements are the most common output. Ranges differ depending on the measurement applied (e.g. ODG as defined in [12]).

Security: [1] states that an embedding algorithm is said to be secure if the embedded information cannot be removed beyond reliable detection by targeted attacks based on a full knowledge of the embedding algorithm and the detector (except the secret key), and the knowledge of at least one carrier with a hidden message. Since 1998 many publications have addressed the security of steganography and DWM algorithms respectively with attacks on their security. Examples in the field of watermarking are [14] and [15]. Steganalytic approaches (which can be considered as security attacks at this point) have been classified into groups in [16].
One important question to address is: how can security in steganography and DWM be measured? One possibility which might be applicable is the transfer of the classification paradigm from cryptographic security (a discrete scale ranging from “unconditionally secure” to “secure enough”). In this case no normalised representation can be applied for this discrete classification in an useful manner.

Invertability: The feature of invertability is a new DWM paradigm which has been developed after [1] was published. So far no application of invertability in

steganography is known to the author. Nevertheless it might be useful to research the possibility of constructing invertible steganographic protocols and algorithms and their impact on deniability (or non-repudiation) of the communication. The representation of this feature is usually a 1-Bit value (binary decision). Therefore for invertability as well as for security a representation has to be found which takes the non-continuous nature of this feature into account.

Additional features: Additional features might be identified as being necessary for a complete description of the performance of an algorithm. As a good example the complexity of the embedding and detection processes shall be mentioned which might be a necessary criterium for the decision whether an algorithm could be used on mobile devices (which normally possess limited computational capabilities). In the visualisation of the results for features not described in detail in this paper the same rule identified in section 2 applies as for the ones described here: the characteristics of a set of data (the subject of visualisation) have to be analysed and have to control the visualisation process.

Relation between characteristics: In her publication [1] J. Fridrich points out that some (namely capacity, robustness and transparency) of the characteristics mentioned above are mutually competitive when considered as requirements for information hiding techniques. Unfortunately no universal linear or functional relationship between the characteristics can be identified for the domain of information hiding techniques which would allow for a dimensional reduction in the visualisation problem.

4 Evaluation of steganography and digital watermarking approaches

What divides steganography and DWM is in most cases only the intention for which a technique is used. If steganography is seen as a means for a hidden end-to-end communication it has more in common with cryptography, which also provides privacy mechanisms for communications, than with digital watermarking. Therefore its evaluation (called steganalysis) is in many cases very similar to cryptanalysis. Benchmarking approaches for steganography algorithms or applications are uncommon (for the same reason as there is no standardisation organisation for steganography), instead steganalysis tools are benchmarked at a large scale. For obvious reasons the scientific community is more interested in creating the perfect universal, blind steganalyser than in finding the perfect steganography approach. Nevertheless the development of an advanced steganalysis tool does require the existence of advanced steganography applications. And these advanced steganography applications have to fulfill certain requirements regarding the characteristics identified in section 3. Most important is that they have to be very transparent. As an additional feature a high capacity would be significant. The robustness is neglected in most discussions about the performance of steganography algorithms, but depending on the application scenario it might be useful to sacrifice some capacity to gain robustness against format

conversions [13] or the influence of a faulty communications channel.
In contrast to steganography, where only one well-defined application scenario exists, digital watermarking has a large spectrum of possible means for application (annotation watermarking, watermarking for forensic tracking purposes, etc). Therefore in DWM the approach of benchmarking algorithms is more common than in steganography and is used to characterise selected watermarking algorithms and their fitness for one of the application scenarios. For examples on these benchmarking activities see publications concerning the WET [18] and Audio WET [17] suites. The different application scenarios have of course an impact on the relevance of certain features of the algorithm and the visualisation for research results in this area. If a problem can be considered from different angles or perspectives (application scenarios) then a graphical representation has to be as generic as possible, to cover all these angles, but at the same time it should be as intuitive as possible since it already represents a very complex problem.
As an additional factor influencing the visualisation of results for steganography and DWM algorithms many features (like capacity, transparency and robustness) might be context sensitive for selected algorithms. Therefore when testing these features on a large test-set the results of a binary or discrete decision might become "blurred" or continuous. This might result in the necessity to introduce decision thresholds, quantisation steps or the expression using probabilities, error rates or ϵ-environments in the visualisation problem.

5 Realisation in Visualisation

R. Spence implies in [5] that since we are living in a three-dimensional (3D) world one would imagine that a 3D display of data would be regarded as "natural". In practice this is limited by the capabilities of today's presentation equipment. Due to these capabilities the most commonly used forms are the textual or a two-dimensional (2D) representation of information of n-dimensional ($1 \leq n \leq \infty$) origin (also known as hypervariate or multivariate data [5]). Common techniques employed in the graphical representation are the projection of the n-dimensional space onto all pairs of axes or the usage of perspective presentations with a distorted 3rd axis (sometimes also called $2\frac{1}{2}$-D representations) for the presentation of 3D data. General problems are encountered which apply to any visualisation technique independent of the dimensionality. A good example is the question: *Which kind of scale (linear, logarithmic, etc) should be applied?*
In the following realisations for the description of (research) results in the contexts of steganography and DWM research are presented. The range of techniques introduced includes general visualisation methods applied in this field and more specific presentations taken from recent publications. First the non-graphical representation is reviewed and then visualisations are given, sorted by increasing dimensionality of the domain of the entity for visualisation using the notation of Brodlie et. al [8].

5.1 Non-graphical representation

The first form of description and comparison of (test-)results to be mentioned is one that can not be placed in Brodlies notation. Nevertheless the presentation in text form is one of the techniques most commonly used in scientific publication. A special form of the presentation in text form is the presentation in tables, allowing for a more structured presentation with possibilities for faster comparison. The following example was taken from [3] and describes, first in text form and then in table 1, the results of a transparency measurement (as the absolute value of the average ODG over a test-set of 389 files) on four selected steganography algorithms (denoted A_S with different parameterisations): *From these results it can be seen that all four A_S used with all parameters tested have a very similar embedding transparency (which in all cases is about* 0.02 *and therefore has to be considered very transparent). Differences can be foun on detail level, when considering the detection process and the context dependency of the algorithm.*

A_S	**Param.**	**avg. embed. t.** [$\lvert ODG \rvert$]
Publimark (v. 0.1.2)		0.0180
Steghide (v. 0.4.3)	Enc./ECC ON/OFF	[0.0255 .. 0.0275]
Steghide (v. 0.5.1)	Enc. std./OFF, ...	[0.0232 .. 0.0265]
LSB (v. Heutl051208)		0.01797

Table 1: Computed average $\lvert ODG \rvert$ values for all A_S and their parameters (taken from [3]).

5.2 Using entities for visualisation of dimensionality $n = 1$

The one-dimensional domain leads to (apparently) simple results in presentation. Despite most of the visualisation forms located in this domain are well known, examples are presented here for two reasons: first to facilitate the application of the notation used and second to derive knowledge for the higher dimensional representations from this class of visualisations.

1D scatter plot (E_1^P): The one-dimensional scatter plot is a simple technique projecting test results onto a single axis. Relationships between the different results are expressed in their distance. Additionaly an order is indicated. An example for a 1D scatter plot is given in figure 1 where test results from table 1 are visualised. A problem encountered in this example is the fact that some of the values given in table 1 are representing ranges (results computed using different parameterisations). The solution chosen here depicts only the minimum and maximum value of these ranges.

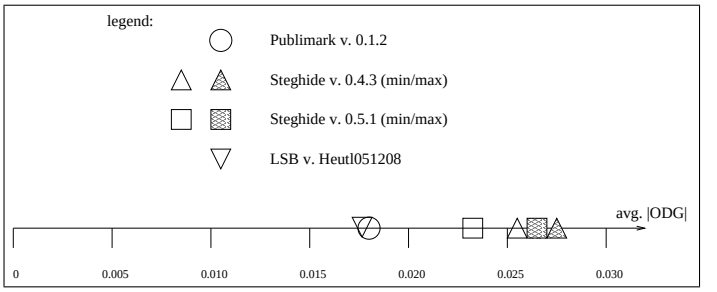

Fig. 1: Transparency results from table 1 as 1D scatter plot (ranges are given with min. and max. values).

Line graph, multiple line graphs (E_1^S, E_1^{mS}): In the line graph of a function the entity is defined pointwise over an interval of the continuous real line (input). An example for superimposed line graphs is given in figure 2. This example, taken from [21], shows the development of the standard deviation of transparency results on a DWM algorithm and different classes of audio material with a varied parameter. The functions are interpolated from a discrete set of measurements. Problems introduced by this interpolation are discussed in detail in [8].

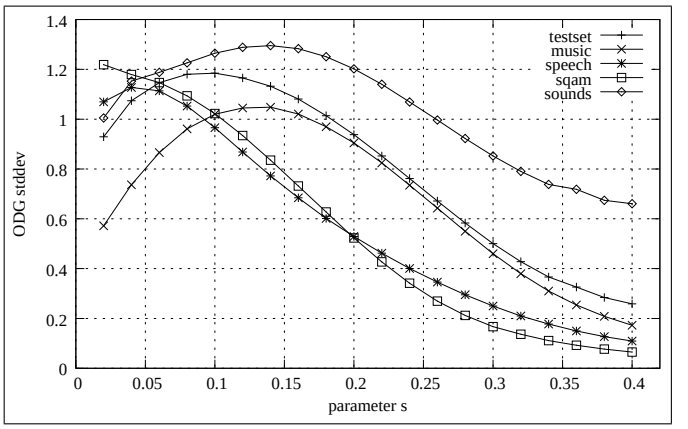

Fig. 2: Interpolated development of the standard deviation of transparency results on a DWM algorithm and different classes of audio material (taken from [21]).

Histogram and Bar chart ($E_{[1]}^S$ and $E_{\{1\}}^S$): In a histogram the entity is defined over regions of the real input. The data is aggregated into bins. The number of elements in each bin is shown in the histogram. The histogram in figure 3 (a) was taken from [19]. It shows the distribution of lengths of delays in a WLAN with and without steganography.
A bar chart depicts the values of items in an enumerated set. If the values can be seen as fractions of a whole then the results could be expressed also as a pie chart. Figure 3 (b) shows a classic example for a bar chart taken from [3].

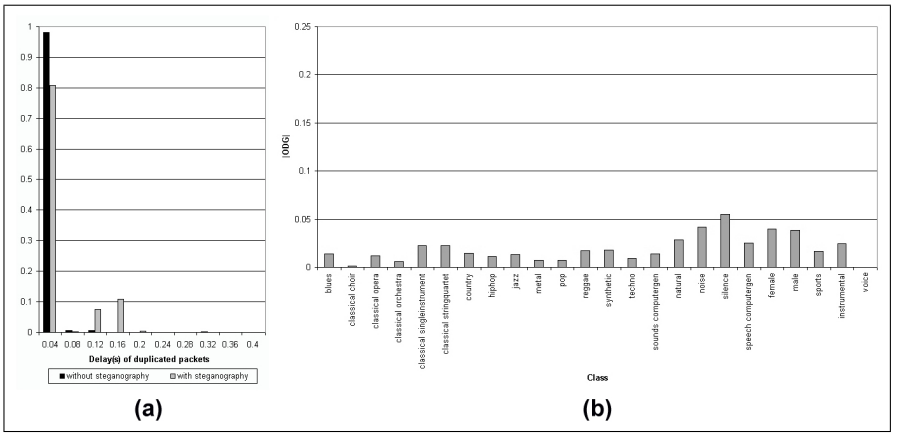

Fig. 3: (a) Results for the delays between WLAN packets with set "Retry" field and the corresponding original packets with and without WLAN steganography (taken from [19]); (b) Transparency results for a steganography algorithm on a test-set grouped into 24 classes of audio material (taken from [3]).

Pixel-based techniques (E_1^S or $E_1^{[S]}$): A technique very similar to the classical histogram is introduced in [6]. This pixel-based technique can be used to visualise results for large sets by representing each element in the set with a marker object (usually a line) with a width of one pixel and a fixed length. To encode the results for each marker object usually colour-coding is applied. In the example given in figure 4 the results from transparency benchmarks on three (two steganography and one watermarking) algorithms are represented. In this case the results were grouped into three classes (regions) and encoded with the colouring of the marker lines in white, grey and black.

Fig. 4: Results from the transparency benchmarking for two steganography and one watermarking algorithm over a test set of 389 files. The computed $|ODG|$ value is colour-coded in three classes: below 0.2 (white), between 0.2 and 1 (grey) and above 1 (black). (Values taken from [3])

The problem encountered in this visualisation form is the fact that colour-coding always has to follow certain rules reducing the maximum number of marker objects. These rules are based on the limited capabilities of the HPS (Human Perceptual System) like the limited number of colours distinguishable and possible limitations regarding individuals (e.g. colour vision deficiencies [4], [6]) or the limitations of the chosen presentation media (e.g. black-and-white print media). The consequences of these rules for the example shown above can be found in the constriction to three defined classes (regions) for the results, which results in a very low resolution for the transparency values. Nevertheless this example shows very impressively the difference between steganography and watermarking algorithms with regards to embedding transparency.

5.3 Using entities for visualisation of dimensionality $n = 2$

The natural dimensionality of print media as well as common computer displays is 2D. Therefore it would be intuitive to choose two-dimensional entities for representation in scientific work, which is most commonly communicated in print or electronic documents mimicking their printed counterparts in appearance. The fact that entities of this dimensionality are not the most common objects chosen is justified by the point that in scientific work the representation of higher dimensionality is more interesting. Nevertheless with the 2D scatter plot one example is introduced here which can be found quite often in scientific publications.

2D scatter plot (E_2^P): In this traditional scatter plot pairs of values are represented as points in the plane. The example shown in figure 5 was already used as the basis for generating figure 2 by interpolating the functions.

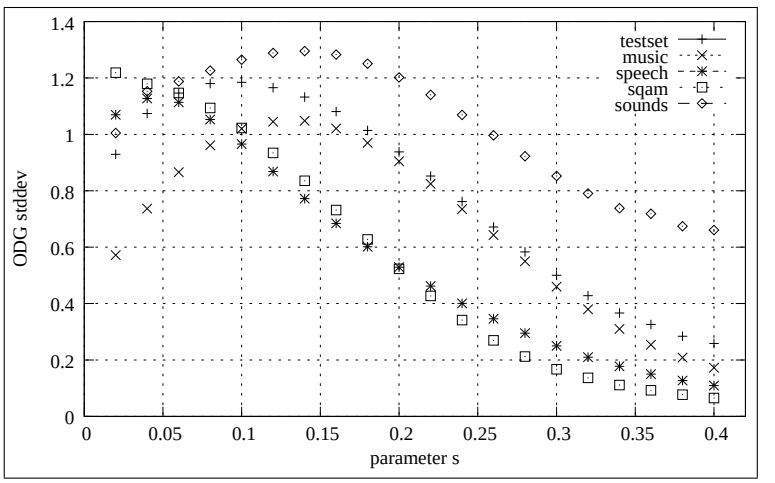

Fig. 5: Standard deviation of transparency results on a watermarking algorithm and different classes of audio material (taken from [21]).

5.4 Using entities for visualisation of dimensionality $n = 3$

The three-dimensional domain is, what Spence considers in [5] to be the "natural" domain of perception for a human audience. Therefore 3D data should be the ones most commonly chosen in presentation. The problem with this approach is that the possibilities of presentation on paper and normal computer displays are a priori limited to 2D information. Three dimensional data can be visualised naturally with appropriate hardware or by projecting them on a 2D plane. In many cases this leads to the question: Which axis should be the one which has to be scaled? Since it has to be assumed that this axis is not as precise readable as the other two, here the main characteristic with the least impact should be chosen. The information hiding paradigm concerned might decide which characteristic should be mapped on this axis (for steganography it might be robustness, while in a DWM scenario the transparency might be chosen).

3D scatter plot (E_3^P): For the 3D case of the scatter plot the result is very often projected on 2D presentation material. In this step the information presented by the 3rd dimensional component is either neglected or distorted. To prevent this techniques like colour-coding or the usage of the size of the marker glyph to indicate the value of the third component can be used.

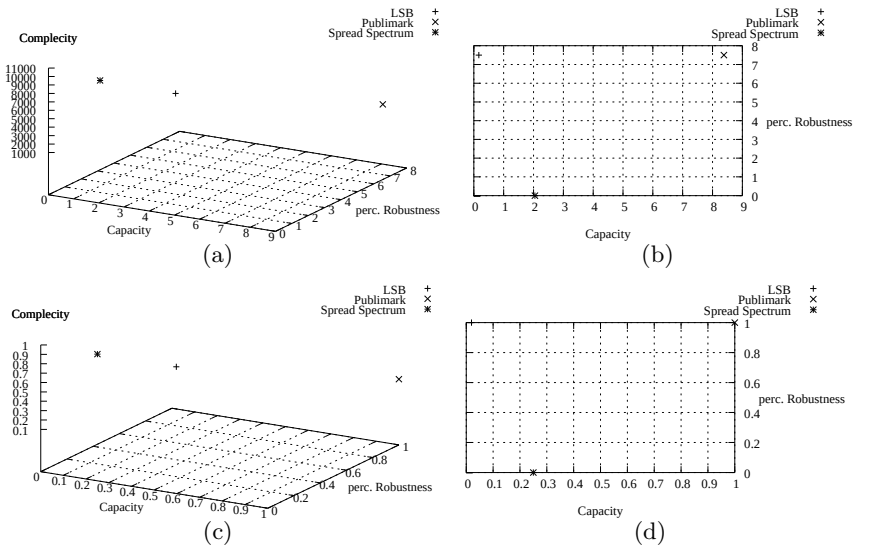

Fig. 6: Visualisation of the embedding Complexity (in seconds), Capacity (in Byte per second) and Robustness (in percent) for three algorithms in 3D scatter plots and 2D projections.

Figures 6 (a) and (c) show 3D scatter plots (non-normalised and in an unit cube) based on figures taken from [17]. Figures 6 (b) and (d) use the technique

of axial projection to generate better readable results from the 3D model. If this approach of using axial projections is applied consequently the result is called in [6] a scatter plot matrix.

Triangular representation taken from [20] (E_3^S): Exploiting the metaphor of the triangle (of Transparency, Capacity and Robustness) presented in section [1] for representation the of benchmarking results (like in [20]) leads to a complex, non-orthogonal representation of three different features in 2D. The proposed representation is shown in figure 7.

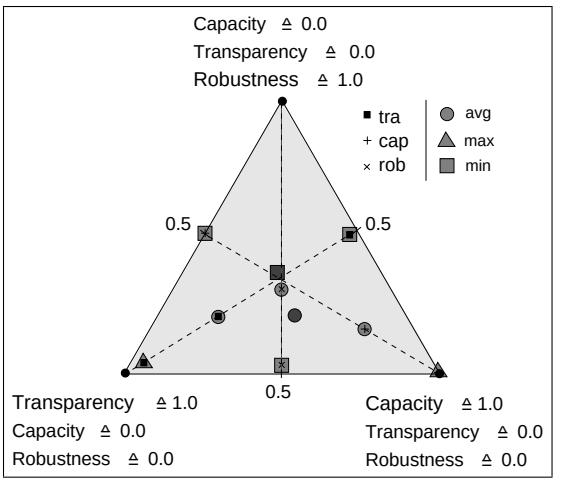

Fig. 7: Benchmarking results of the Complexity, Transparency and Robustness for different watermarking algorithms in a triangular representation (taken from [20]).

Due to the nature of this representation (three non-linear dependent values are mapped in 2D) a specific location within the triangle is not the unique representation of a single (normalised) value set for the three characteristics (if considering the distance to the corner-points as weights in the representation, a point in the centre would equally represent the sets $\{1, 1, 1\}$, $\{0.3, 0.3, 0.3\}$ and $\{0, 0, 0\}$). Other approaches advancing the idea of representation within the triangle, like exploiting area sizes or colour-coding, do not overcome the basic flaw in this representation: in many practical DWM-algorithms the three main characteristics might be dependant but not in a linear way, which means graphically that the result of placing them in a triangle will not result in a point.
Nevertheless the metaphor of the triangle is still a good approach to symbolise the fact that the three main features are dependent on each other.

5.5 Using entities for visualisation of dimensionality $n \geq 3$

Generally the number of linear independent vectors (required for an injective representation) is limited by the dimensionality of the representation system.

In 2D exist exactly two linear independent vectors, in 3D exactly three. Therefore the only way to adequately represent values of an n-dimensional functional nature it would require an n-dimensional space and an equally n-dimensional display method. If instead of functional correlations only states have to be visualised then for most n-dimensional data an adequate representation in 2D or 3D can be found. The problem here is to identify such "adequate" visualisation techniques for a well defined problem. For example the Hyperbox introduced by R. Spence in [5] gives a graphical example of a representation of 6D data in 2D. This representation form, which might be considered a very intuitive way of presentation, is not useful in the focus of this document since the introduced distortion of the data makes a perceptual comparison of results for different algorithms impossible. Another technique for the visualisation of a n-dimensional space, which was already introduced in section 5.4, is the scatter plot matrix. This concept of a projection onto all pairs of axes can easily be transferred from the 3D domain to any other dimensionality. Other representations to be introduced for example are the parallel coordinate plots and the Kiviatgraphs ([5], [6]). Both are variable in dimensionality.

Areas under a parallel coordinate plot (E_5^{mS}): The area under the curves in a parallel coordinate plot might be considered in some applications an adequate rating for the quality of an algorithm with regards to n characteristics. To use this measure in watermarking and steganography benchmarking is highly questionable since most often non-continuous values are projected and a different order of the features would result in a different area. Figure 8 displays such a parallel coordinate plot with five features for five selected algorithms. The problem in this figure is proposed by the fact that the robustness and capacity are presented by "bigger-is-better" metrics and the transparency and complexity by "lower-is-better" metrics. Nevertheless this presentation provides a good base for algorithm comparison with regards to the features identified.

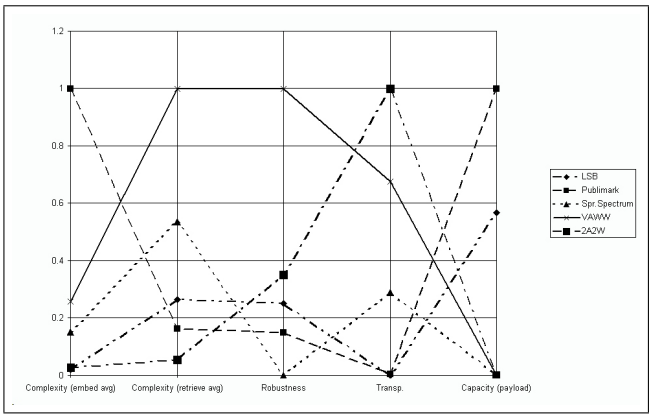

Fig. 8: Parallel coordinate plot displaying normalised benchmarking results of the Complexity (embedding and retrieval), Transparency and Robustness and Capacity for five selected algorithms (values taken from [17]).

Area(s) in a Kiviatgraph (E_5^S and E_5^{mS}): The Kiviatgraph is a presentation form very similar to the parallel coordinate plot. Here the same problems arise when considering the area within the graph as a measure for the performance of an algorithm. A simplified version in a star-shaped form is shown in image 9 (c).

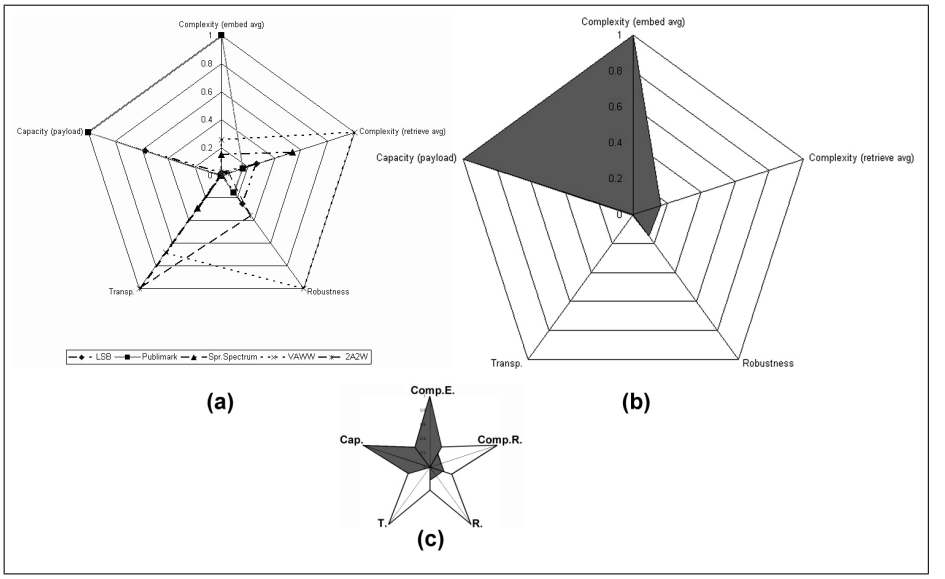

Fig. 9: Kiviat graph displaying normalised benchmarking results of the Complexity (embedding and retrieval), Transparency and Robustness and Capacity for: (a) five selected algorithms, (b) one selected algorithm, (c) simplified version of (b) (values taken from [17])

6 Summary/Conclusion

Apart from primary scientific goals like the development of universal, blind steganalysis tools, commercially exploitable watermarking algorithms or an universally accepted watermarking benchmarking approach, secondary problems like finding the appropriate representation for research results also have to be considered by the research community.
This paper basically contains an overview of features to be benchmarked in DWM and steganography as well as it provides an introduction of a number of visualisation techniques applicable to the results in this field. The goal of this paper was to facilitate the discussion about fitting representation approaches for fair benchmarking and the selection and use of techniques by non-experts. The author does not consider the introduced visualisation techniques as perfect matches for the visualisation problems at hand, but they very well show which problems can be encountered when trying to find fitting representations for complex sets of data.

Acknowledgements

I want to thank Prof. Jana Dittmann for her ideas regarding the visualisation problem discussed in this paper and Andreas Lang for providing material from his research and inspiring new ideas for creating visualisations.
The work about watermarking benchmarking described in this paper has been supported in part by the European Commission through the IST Programme under Contract IST-2002-507932 ECRYPT. The information in this document is provided as is, and no guarantee or warranty is given or implied that the information is fit for any particular purpose. The user thereof uses the information at its sole risk and liability.

References

1. J. Fridrich: *Applications of Data Hiding in Digital Images*, Tutorial for the ISPACS 1998 conference in Melburne, Australia, 1998
2. A. Westfeld: *Steganographie für den Amateurfunk*;, S. 119-130 in Jana Dittmann (Hrsg.): Sicherheit 2006, Sicherheit - Schutz und Zuverlässigkeit, Beiträge der 3. Jahrestagung des Fachbereichs Sicherheit der Gesellschaft für Informatik e.V. (GI), 20.-22. Februar 2006 in Magdeburg, LNI Vol. P-77, Bonn, 2006
3. Christian Kraetzer, Jana Dittmann and Andreas Lang: *Transparency benchmarking on audio watermarks and steganography*, SPIE conference, at the Security, Steganography, and Watermarking of Multimedia Contents VIII, IS&T/SPIE Symposium on Electronic Imaging, 15-19th January, 2006, San Jose, USA, 2006
4. ICD-10, Chapter VII H53.5, *International Statistical Classification of Diseases and Related Health Problems, 10th Revision*, World Health Organization (WHO), 1999
5. Robert Spence: *Information Visualization*, Addison Wesley, ACM Press, ISBN 0-2001-59626-1, 2001
6. Heidrun Schumann, Wolfgang Müller: *Visualisierung - Grundlagen und allgemeine Methoden*, Springer Verlag, ISBN 3-540-64944-1, 2000
7. B.H. Mc Cormick, T.A. De Fanti, M.D. Brown: *Visualization in Scientific Computing*, Computer Graphics, Vol.21 Nr.6, P. 1-14, Nov. 1987
8. K.W. Brodlie, L.A. Carpenter, R.A. Earnshaw, J.R. Gallop, R.J. Hubbold, A.M. Mumford, C.D. Osland, P. Quarendon: *Scientific Visualization - Techniques and Applications*, Springer Verlag, ISBN 3-540-54565-4, 1992
9. Fabien A. P. Petitcolas, Ross J. Anderson, Markus G. Kuhn: *Attacks on copyright marking systems*, in David Aucsmith (Ed), Information Hiding, Second International Workshop, IH98, Portland, Oregon, U.S.A., April 15-17, 1998, Proceedings, LNCS 1525, Springer-Verlag, ISBN 3-540-65386-4, pp. 219-239, 1998
10. Fabien A. P. Petitcolas: *Watermarking schemes evaluation*, I.E.E.E. Signal Processing, vol. 17, no. 5, pp. 58–64, September 2000
11. Andreas Lang, Jana Dittmann, Ryan Spring, Claus Vielhauer: *Audio watermark attacks: from single to profile attacks*, Proceedings of ACM Multimedia and Security Workshop 2005, pp. 39 - 50, ISBN 1-59593-032-9, New York, NY, USA, August 1-2, 2005
12. ITU-R Recomendation BS.1387, *Method for objective measurements of perceived audio quality*, ITU-R, 2001
13. Stefan Katzenbeisser and Fabien A.P. Petticolas: *Information Hiding Techniques for Steganography and Digital Watermarking*, Artech House Publishers, ISBN 1580530354, 2000

14. Martin Kutter, Sviatoshlav Voloshynovskiy and Alexander Herrigel: *The Watermark Copy Attack*, Electronic Imaging 2000, Security and Watermarking of Multimedia Content II, Volume 3971, 2000
15. J. Dittmann, S. Katzenbeisser, C. Schallhart and H. Veith: *Ensuring Media Integrity on Third-Party Infrastructures*, Proceedings of the SEC2005, Chiba, Japan, May, 2005
16. Neil F. Johnson, Zoran Duric, Sushil Jajodia: *Information Hiding*, Kluwer Academic Publishers, 2001
17. Andreas Lang, Jana Dittmann: *Profiles for Evaluation - the Usage of Audio WET*, SPIE conference, at the Security, Steganography, and Watermarking of Multimedia Contents VIII, IS&T/SPIE Symposium on Electronic Imaging, 15-19th January, 2006, San Jose, USA, 2006
18. Hyung Cook Kim, Eugene T. Lin, Oriol Guitart, Edward J. Delp: *Further progress in watermark evaluation testbed (WET)*, Security, Steganography, and Watermarking of Multimedia Contents 2005: pp. 241-251, 2005
19. Christian Kraetzer, Jana Dittmann, Andreas Lang, Tobias Kuehne: *WLAN Steganography: A First Practical Review*, To appear in: Proceedings of the ACM Workshop on Multimedia and Security, Geneva, Swiss, Saptember 26-17th, 2006
20. Andreas Lang, Jana Dittmann, David, Megías, Jordi Herrera-Joancomartí: *Practical Audio Watermarking Evaluation Tests and its Representation and Visualization in the Triangle of Robustness, Transparency and Capacity*, Submitted to the 2nd WaCha, Geneva, Swiss, 2006
21. Andreas Lang, Jana Dittmann: *Transparency and Complexity Benchmarking of Audio Watermarking Algorithms Issues*, to appear in Proceedings of ACM MM & Sec'06 Workshop, Geneva, Swiss, September 2006

ISBN: 978-3-929757-29-3
Printed at the Otto-von-Guericke University of Magdeburg, Germany
March 31st, 2007