IST-2002-507932

ECRYPT

European Network of Excellence in Cryptology

Network of Excellence
Information Society Technologies

D.WVL.20

Final Report on the Security of Practical Systems

Due date of deliverable: 31st July 2008
Actual submission date: July 29, 2008

Start date of project: 1 February 2004
Duration: 4 years

Lead contractor: Katholieke Universiteit Leuven (KUL)

<table>
<thead>
<tr>
<th>Dissemination Level</th>
<th></th>
</tr>
</thead>
<tbody>
<tr>
<td>PU Public</td>
<td>X</td>
</tr>
<tr>
<td>PP Restricted to other programme participants (including the Commission services)</td>
<td></td>
</tr>
<tr>
<td>RE Restricted to a group specified by the consortium (including the Commission services)</td>
<td></td>
</tr>
<tr>
<td>CO Confidential, only for members of the consortium (including the Commission services)</td>
<td></td>
</tr>
</tbody>
</table>
Final Report on the Security of Practical Systems

Editors
Mauro Barni (CNIT), Sviatoslav Voloshynovskiy (UNIGE)

Contributors: All partners involved in WVL2

July 29, 2008
Final

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. The information in this document is provided as is, and no 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.
Contents

1 Introduction

2 Security analysis of practical systems
   2.1 Spread Spectrum Watermarking Security [PFPG08b] .......................... 2
   2.2 Practical Security Analysis of Dirty Paper Trellis Watermarking [BD07] ... 3
   2.3 Security of lattice-based data hiding against the Watermarked Only Attack [PFPG07a],[PFPG08a] .................................................... 3
   2.4 On The Tradeoff Between Security And Robustness Of The Trellis Coded Quantization Scheme .................................................... 4

3 Design of secure and robust watermarking systems
   3.1 Practical Performance Analysis of Secure Modulations for WOA Spread-Spectrum based Image Watermarking [MBC07] ............................ 5
   3.2 Zero-Knowledge Watermark Detector Robust to Sensitivity Attacks [TPG07] 6
   3.3 On a watermarking scheme in the logarithmic domain and its perceptual advantages [CPG07c] and Dither Modulation in the Logarithmic Domain [CPG07b] 7

4 The BOWS contests
   4.1 First BOWS (Break Our Watermarking System) Contest ........................ 8
   4.2 BOWS attack package ...................................................................... 9
   4.3 “Breaking the BOWS Watermarking System: Key Guessing and Sensitivity Attacks” [CPG07a] (extended version of “Two different approaches for attacking BOWS” [CPG07e]) ........................................ 9
   4.4 Broken Arrows (a watermarking technique designed for the BOWS-2 challenge) [FB] ................................................................. 10
   4.5 A Regression-Based Restoration Technique For Automated Watermark Removal ................................................................. 11

5 Steganography - Steganalysis
   5.1 Publimark [Pub08] .......................................................................... 12
   5.2 Benford’s law in image processing [PGHA07] ..................................... 13

6 Authentication
   6.1 Tamper-proofing of Electronic and Printed Text Documents via Robust Hashing and Data-Hiding [VVK+07] ............................................. 14
   6.2 Authentication of Biometric Identification Documents via Mobile Devices [VKV+08] ................................................................. 15
6.3 Invertible chaotic fragile watermarking for image authentication [SNP08] . . . 15
6.4 A Novel Interpretation of Content Authentication [CBPG07] . . . . . . . 16

7 Miscellaneous works
7.1 Comparative study of chaotic and white noise signals in digital watermarking [MKP08] . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . 17
7.2 Watermarking 3-D Volumes in the Discrete Fourier Transform Domain [SP07] 18
7.3 Review of image watermarking techniques [TNP08] . . . . . . . . . . . . . . 19
7.4 On Distortion Measures Associated to Random Desynchronization Attacks [BPG07] . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . 20
7.5 “Modeling gabor coefficients via generalized gaussian distributions for face recognition” [GJP07] . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . 20
7.6 Secure Surface Identification Codes [BVKV08] . . . . . . . . . . . . . . . . 21
7.7 Detection of malevolent changes in digital video for forensic applications . . 22
7.8 A Secure Multidimensional Point Inclusion Protocol [TPKCL07] . . . . . . 22
7.9 Privacy Preserving Error Resilient DNA Searching through Oblivious Automata [TPKC07] . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . 23

8 Main contributions of WVL2 across the whole project
8.1 Watermark robustness . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . 24
8.2 Watermark security . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . 25
8.3 The BOWS contests . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . 25
8.4 Miscellaneous works . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . 26
1 Introduction

In this report the activity carried out within WVL2 during the last year (Y4) of the ECRYPT project is presented, at the same time a brief summary of the results obtained throughout the duration of the project is given.

The activity carried out in Y4 has followed the same path of the previous years, i.e. i) put into practice the theoretical insights obtained in WVL1 to analyze the security of practical watermarking systems, ii) design new kind of attacks that get as close as possible to the theoretical limits found in WVL1, and iii) develop new watermarking algorithms that improve the state of the art with regard to security and robustness.

Within the above framework, the first part of this report is split into 6 parts (section 2 through section 7). In section 2, the theoretical security framework developed in WVL1 is used to evaluate the security of some popular watermarking systems, thus continuing a similar effort carried out in the previous years. In section 3, the efforts made to design new watermarking systems with improved security and robustness are presented. Section 4 is dedicated to the two BOWS contests organized by Wavila. The design criteria behind the two contests are discussed together with the most important scientific results the contests resulted into.

Though the main focus of WVL2 was more on watermarking, the research has touched steganography and steganalysis, with some interesting results that are summarized in section 5. The next section (section 6) is devoted to watermarking-based authentication of digital (and in some cases analog) contents, including text documents and biometric data.

Section 7 describes a bunch of interesting results obtained by Wavila partners that can not be directly brought back to the activity plan of WVL2. The existence and abundance of such results is a direct consequence of the great number of partners active within WVL2 and by the practically-oriented nature of this workpackage. It is, in fact, natural that whenever practical problems are touched new research issues come out that could not be predicted in advance.

In the last part of the deliverable (section 8), a brief summery of the activity carried out within WVL2 throughout the duration of the project is given.

2 Security analysis of practical systems

Following a similar effort made in the previous years, and in line with the activity plan of WVL2, the core of the activity during the last year of the project has been the analysis of the security of some of the most popular, and best performing, watermarking systems available today. In particular, in [PFPG08b] the analysis of the security of SS watermarking is moved one step further with respect to previous works, permitting to clearly quantify the trade-off between security and robustness. In [BD07], the security of dirty trellis watermarking is analyzed for the first time. This is a very important step, since it is one of the first times that the security of a complete (and practical) watermarking system is analyzed, being the great majority of previous works focused on the analysis of general classes of watermarking algorithms framed in highly idealized scenarios. In [PFPG07a],[PFPG08a], the security of lattice-based watermarking is investigated under the WOA (Watermarked Only Attack) model,
thus extending a previous work carried out within WVL2 where security was addressed under the KMA (Known Message Attack) scenario.

The section ends with the presentation of the results of an unconventional analysis where statistical watermark invisibility (a typically steganographic concept) is studied together with robustness (a requirement which is typical of watermarking applications rather than steganography).

\section{Spread Spectrum Watermarking Security [PFPG08b]}

\textit{Addressed problem:} This work considers the security of the most popular spread spectrum watermarking techniques both from theoretical and practical points of view, according to the framework established in [CPFPG05]. Security is understood as the difficulty of estimating the secret parameters of the embedding function based on the observation of watermarked signals.

\textit{Motivations:} Although the security of spread spectrum techniques has been already addressed before by other authors, some problems which are worth addressing still remain open. These problems are the tradeoff robustness-security for spread spectrum techniques, the evaluation of security in asymptotic conditions, the derivation of performance bounds, and the theoretical analysis of the practical estimators already proposed.

\textit{Technical approach:} This research comprises both theoretical and practical analysis. In the theoretical part, the security is quantified by means of the Shannon’s equivocation function. This measure allows to compute the information about the secret spreading vector that leaks from the observation of watermarked signals, providing the fundamental security bounds. In the practical part, the cost function of previous estimators of the spreading vector, which are based in Principal Component Analysis (PCA) and Independent Component Analysis (ICA), are analyzed from a statistical point of view. Two new cost functions for implementing practical estimators of the spreading vector are proposed. A practical framework for the implementation of the estimators, based on optimization over manifolds, is proposed and evaluated in practical images.

\textit{Obtained results:} The theoretical results allow to quantify: 1) the security of plain spread spectrum watermarking [CKLS97]; 2) the tradeoff robustness-security when using methods with host rejection [MI03],[MF03]; 3) fundamental bounds on the estimation of the spreading vector, in terms of MMSE and normalized correlation. The practical results reveal the conditions for which the PCA and ICA approaches are successful, and the new proposed estimators (Informed ICA and the Constant Modulus Criterion) are shown to work better than ICA and PCA in a number of situations. A new measure (the chordal distance) for quantifying the estimation accuracy of subspaces is proposed.
2.2 Practical Security Analysis of Dirty Paper Trellis Watermarking [BD07]

Addressed problem: This work presents a practical method to estimate the secret key used by Dirty Paper Trellis watermarking schemes. The set-up used here is the WOA (Watermarked content Only Attack) setup: the secret key is estimated from a collection of content watermarked with the same secret key.

Motivations: The use of trellis for watermarking is a practical way to perform dirty paper coding [MDC04]. Dirty Paper Trellis watermarking makes use of both informed coding and informed embedding and it is considered as a good choice in order to obtain both robust and high-capacity watermarking. Consequently a security analysis of these scheme is very relevant.

Technical approach: In the watermarked contents only attack (WOA) setup, the watermarked data-set exhibits clusters corresponding to the different patterns attached to the arcs of the trellis. The $K$-means clustering algorithm is used to estimate these patterns and a co-occurrence analysis is performed to retrieve the connectivity of the trellis.

Obtained results: Experimental results demonstrate that it is possible to accurately estimate the trellis configuration, which enables to devise an optimal tampering attack which is much more efficient than simple additive white Gaussian noise (AWGN).

2.3 Security of lattice-based data hiding against the Watermarked Only Attack [PFPG07a],[PFPG08a]

Addressed problem: The security of nested lattice codes for data hiding is considered both from information-theoretic and practical perspectives, according to the framework proposed in [CPFPG05]. This work shows that, if the embedding parameters are not properly adjusted, the secret parameters of the embedding function in lattice data hiding can be estimated and used for performing attacks to the system.

Motivations: All the signals watermarked by the same user convey information about his/her secret key. Hence, if the system is not perfectly secure then an attacker can estimate the secret key (or the secret parameters of the algorithm) if he manages to gather several contents watermarked with the same secret key. In a subsequent stage, he/she can use this knowledge to attack other images.

Technical approach: The followed approach extends the scope of the work carried out in [PFPGFC06] (focused on the “Known Message Attack” scenario) to a more general scenario, known as “Watermarked Only Attack”, where no knowledge about the embedded messages is assumed. For the theoretical part, the information about the secret key (in this case, a secret dither vector) provided by the observations is computed, providing fundamental security bounds (paying special attention to the impact of the coding rate in security), and
showing under what conditions it is possible to achieve perfect secrecy. In the practical part, a joint set-membership and Bayesian approach is devised in order to estimate the secret dither vector in practical scenarios.

Obtained results: Results are given both at the theoretical and practical level. In the theoretical part, several lemmas and theorems formally state: 1) asymptotically, security and robustness are not conflicting requirements; 2) there exist practical constructions for achieving perfect secrecy; 3) if the distortion compensation parameter is sufficiently large, the asymptotic information leakage (for a large number of observations) is simply given by that of the KMA case plus one term which is the embedding rate. The information leakage for a practical nested lattice code (Distortion Compensated Dither Modulation [CPGB06]) is also computed, in order to illustrate the most important results. In the practical part, the results of applying the proposed estimator to several scenarios using nested lattice codes are shown, considering the accuracy in secret dither estimation for different lattice codes and embedding rates. Finally, a “reversibility attack” (for watermarked images) using the estimated dither vector is proposed, achieving complete removal of the embedded information and a virtually perfect host recovery, confirming the threat represented by the security attacks.

2.4 On The Tradeoff Between Security And Robustness Of The Trellis Coded Quantization Scheme

Addressed problem: The Scalar Costa Scheme (SCS) is robust to Additive White Gaussian Noise (AWGN) attack but is drastically insecure since its probability density function for Gaussian host signal is severely discontinuous. An improved scheme has been proposed by Guillon et al. which increases the security to the detriment of the robustness. We propose a new watermarking scheme, based on the combination of the Spread Transform (ST) and the Trellis Coded Quantization (TCQ) which is secure and robust to AWGN attack.

Motivations: In the context of the steganography, Cachin has defined the notion of secure scheme by the closeness of the Probability Density Function (PDF) of the host and marked signals. The distance criterion is the Kullback-Leibler Distance (KLD) or also called relative entropy. A steganographic analysis of the SCS shows that this scheme is not secure according to the Cachin’s criterion. Indeed, the SCS introduces many artifacts in the PDF of the watermarked signal.

Technical approach: A simple inspection of the statistics of the SCS watermarked signal gives away the presence of the watermark. Guillon et al propose a modification of the SCS which leads to considerably improve the steganographical security of the scheme. But, as a price to pay, the proposed scheme imposes many constraints in terms of robustness. The aim of this work is to find a quantization based watermarking scheme which is steganographically secure and robust. For this work, we propose to combine a highly statistically invisible trellis coded quantization scheme (TCQ) with the robust and well known Spread Transform (ST) based SCS. For the new system proposed (STTCQ) we study the best compromise in terms of robustness and invisibility. A validation of the obtained results on real images is also proposed.
Obtained results: Steganography and robust watermarking are often animated by contradictory goals since for steganography, the prime goal is to have a high level of security, defined as the closeness of the PDF of the host and marked signals, often to the detriment of the robustness to an attack. For robust watermarking, this is precisely the inverse. So, it is interesting to design a watermarking scheme which is secure and robust. Toward this end, we have proposed a new quantization-based watermarking scheme based on the Trellis Coded Quantization on a spread transform domain which allows a good tradeoff between security and robustness with respect to other standard quantization based watermarking schemes.

3 Design of secure and robust watermarking systems

The first two papers of this section are a clear example of how the theoretical and practical analysis of the security of existing watermarking schemes may lead to the design of new systems with improved security. This is the case of the systems presented in [MBC07], where the natural and circular watermarking concepts are used to design two practical image watermarking schemes. In [TPG07], an attempt to couple security and robustness against sensitivity attacks is made by applying zero knowledge watermark detection to a watermarking system that is particularly suited to prevent sensitivity attacks (or at least make them as complex as possible).

The last result reported in this section is more oriented towards watermark robustness (rather than security). In particular, the research described in [CPG07c] and [CPG07b] appears to be particularly interesting since it manages to deal with valuemetric attacks in a very elegant way, while at the same time taking into account perceptual issues.

3.1 Practical Performance Analysis of Secure Modulations for WOA Spread-Spectrum based Image Watermarking [MBC07]

Addressed problem: This research line proposes a practical implementation of secure spread-spectrum watermarking schemes for digital images. Before these schemes (call natural watermarking and circular watermarking) were only applied on synthetic signals.

Motivations: Natural and Circular watermarking schemes enables to achieve a given degree of security regarding Watermark Only Attacks; using these techniques it is impossible to estimate the secret subspace for Natural Watermarking or the secret key for Circular Watermarking. This work has investigated the possibility to apply these schemes on digital images.

Technical approach: The watermark is embedded in a secret subspace that is generated from the wavelet domain. The power of the watermark is tuned in order to preserve both imperceptibility and robustness regarding JPEG compression. Multiplicative embedding is performed to address the problem of visual distortion and the embedding parameters are modified in order to not jeopardizing the security of the scheme. For the implementa-
tion of Natural Watermarking, a special care is devoted to the choice of the wavelet subbands.

*Obtained results:* The proposed implementation enables to have both secure and robust watermarking schemes dedicated to digital images. Regarding JPEG compression, with an embedding distortion of 45dB, the implementation of Circular and Natural watermarking exhibits a BER of 0.1 for a quality faction of 10%.

### 3.2 Zero-Knowledge Watermark Detector Robust to Sensitivity Attacks

**Addressed problem:** This work copes with the problem of achieving a watermarking protocol able to withstand blind sensitivity attacks, while concealing the secret parameters using zero-knowledge protocols in order to perform a public detection.

**Motivations:** Current zero-knowledge watermark detectors are based on a linear correlation between the asset features and a given secret sequence. This detection function is susceptible of being attacked by sensitivity attacks, for which zero-knowledge does not provide protection. Furthermore, many complex functions that could be used to get more involved detection boundaries cannot be implemented through zero-knowledge protocols.

**Technical approach:** A new zero-knowledge blind watermark detection protocol is presented, based on the maximum likelihood detection function for generalized Gaussian distributed host features. For some choices of parameters, the detection boundary is very complicated, what makes it more robust to sensitivity attacks. As a consequence, the combination of this detector with zero-knowledge techniques results in a secure and robust protocol. In order to implement it, two zero-knowledge proofs for calculation of modulus and square root are presented. The new protocol is compared in terms of performance and efficiency with the previous zero-knowledge protocols based on correlation detector for additive spread-spectrum and Spread-Transform Dither Modulation (ST-DM).

*Obtained results:* Two zero-knowledge proofs for modulus and square root have been developed; they serve as building blocks for the zero-knowledge implementation of the generalized Gaussian maximum likelihood detector, and also open new possibilities in the design of high level protocols. The developed zero-knowledge detector has shown improved resistance against blind sensitivity attacks with respect to the previous protocols, with a better performance than correlation based detection; Moreover, when the employed spreading sequence is binary antipodal, the resulting communication complexity of the protocol can be greatly reduced, making it even more efficient than the previous ones.
3.3 On a watermarking scheme in the logarithmic domain and its perceptual advantages [CPG07c] and Dither Modulation in the Logarithmic Domain [CPG07b]

*Addressed problem:* In these works a new family of quantization-based methods is proposed, based on both Dither Modulation and Spread Transform Dither Modulation, aiming at dealing with scaling attacks, and which presents another outstanding property: they produce perceptually shaped watermarks.

*Motivations:* Scaling attacks are well-known to be some of the most harmful strategies against quantization-based watermarking methods, as they desynchronize the decoder, completely ruining the performance of the watermarking system with almost non perceptually altering the watermarked signal.

*Technical approach:* Although some proposals are already available in the literature, some of them based on a non-linear transformation (e.g., A-law compansion) previous to the embedding, this is still an open topic that we study in this paper from an innovative approach: the watermark is embedded in the logarithmic domain using a quantization based system; the cases where a projection is performed previously to the quantization, and where the logarithmic transform of the host signal is not projected are compared, as well as their differential and non-differential versions. A performance analysis is introduced, obtaining theoretically the probability of decoding error.

*Obtained results:* The intuitive idea that the differential schemes are more sensitive than non-differential ones to additive noise attacks has been quantified; nevertheless, one should also consider that the differential schemes are invulnerable to valumetric attacks.

Furthermore, a set of techniques that perform a projection before quantization have been considered, as well as those techniques that do not consider that projection, obtaining the interesting result that, under some reasonable assumptions on the projecting matrix, the performance of the latter is better than that of the former.

The usefulness of the proposed techniques is also proved by some empirical results that show the perceptual advantages of the logarithmic schemes. This goodness is based on the fact that the logarithmic schemes proposed in this paper are perceptually shaping the watermark, i.e. embedding a larger amplitude watermark in those coefficients where the original host signal is larger, so they take advantage of contrast masking.

4 The BOWS contests

The organization of the two BOWS contests turned out to be among the most successful initiatives of the Wavila virtual lab. As it has already been discussed in previous reports, the organization of such contests has proven to be a valuable research tool for a number of reasons, among which the most important ones can be summarized as follows: 1) the contests served
to pinpoint the weaknesses of state-of-the-art methods, and promote new research aimed at improving those methods; 2) the inherent applicability of the attacks served as a benchmark to test results developed under more theoretical conditions; 3) the existence of independent attackers acted in a way as a “Monte Carlo” testing of the algorithms.

The first three subsections below, refer to the first BOWS contest (already discussed in the previous deliverable of WVL2). Specifically, the design criteria and the objectives of the first contest are analyzed in [PB07]. In the subsequent section, the development of a package for the automatic application of a subset of the attack strategies developed for the contest is described. The package can be a useful teaching tool and can be used both for subsequent contests and to benchmark the security/robustness of any watermarking system under the same attack model adopted within BOWS I. Finally, in [CPG07a], [CPG07e], two new attacking strategies expressly developed for the BOWS contest and proved to be particularly effective against any watermarking algorithm in a wide variety of contexts are described.

The subsequent subsections, refer to the second BOWS contest, whose results and impact on the watermarking research community are still to be fully appreciated. First the design criteria used to develop the watermarking algorithm that has been used during the second contest are reviewed. Interestingly, the results of the first contest has been widely used to design a powerful watermarking system with improved performance especially from a robustness perspective. At the same time, the attacking strategies used by contenders that took part in BOWS II have been refined leading to several interesting results, that start appearing in the watermarking literature as demonstrated in [Wes08].

4.1 First BOWS (Break Our Watermarking System) Contest

Addressed problem: To investigate how and when an image watermarking system can be broken while preserving the highest possible quality of the content, in the case the watermarking system is subject to a massive world-wide attack. The great number of participants and the echo that the Contest has had in the watermarking community, contributed to make BOWS a great success.

Motivations: The BOWS Contest was not intended 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.

Technical approach: contenders were allowed to try to erase the embedded watermark from 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 were asked 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 to ask to run the detection process; finally they obtained as answer the result of the detection and the PSNR achieved [PB07].
Obtained results: From a scientific point of view, many insights into the problems attackers have to face with when operating in a practical scenario have been obtained, confirming the threat posed by the sensitivity attack, which turned out to be the most successful attack. At the same time, several interesting modifications of such an attack have been proposed to make it work in a real scenario under limited communication and time resources.

4.2 BOWS attack package

Addressed problem: Which functions are necessary in order to attack digital watermarking methods in a public challenge? Which steps are suitable to determine weaknesses of an unknown watermark?

Motivations: Unfortunately, when we started to participate in the BOWS contest, there was no framework to base on. Especially for attacks that require a large number of calls to the online detector, an automated execution of suitable steps is necessary [Wes07a].

Technical approach: We propose an environment that provides basic functions for analysis of digital watermarking and the processing of spatial images (transformations, filters, quality assessment).

Obtained results: The BOWS package, a workbench for attacks on the particular watermarking system used in the first BOWS challenge [ECR06], along with examples for using this experimental environment, is available for education or subsequent BOWS events. The BOWS package provides an offline detector for several platforms. With our high level implementation of the trellis-based code in R [R D07] we can quickly find single coefficient attacks and derive high-quality images (62.6 dB PSNR) with full knowledge of the key [Wes07b].

4.3 “Breaking the BOWS Watermarking System: Key Guessing and Sensitivity Attacks” [CPG07a] (extended version of “Two different approaches for attacking BOWS” [CPG07e])

Addressed problem: From December 15, 2005 to June 15, 2006 the watermarking community was challenged to remove the watermark from 3 different $512 \times 512$ watermarked images while maximizing the Peak Signal to Noise Ratio (PSNR) measured by comparing the watermarked signals with their attacked counterparts.

Motivations: Probably challenging the watermarking community (and the public in general) to break a certain watermarking system is valuable for a number of reasons: 1) the contest serves to pinpoint the weaknesses of state-of-the-art methods, and likely, promote new research aimed at improving those methods; 2) the inherent applicability of the attacks serves as a benchmark to test results developed under more theoretical conditions; 3) the existence of independent attackers acts in a way as a “Monte Carlo” testing of the algorithms.
Technical approach: Within this framework we tried to remove the watermark from the provided images in two different circumstances:

1. The attacker completely lacks any knowledge of the used watermarking method and only has access to a detector, that he feeds with an image, and provides a binary output. This situation corresponds to the first stage of the BOWS challenge.

2. The attacker knows all the details about the watermarking scheme, except for a secret parameter, the secret key, which is only shared by embedder and detector.

For the first case we used the blind sensitivity attack previously published (BNSA), whereas for the second one we followed a strategy based on an exhaustive search on secret key space.

Obtained results: The PSNRs obtained for the 3 considered images were 53.5051 dB, 56.1106 dB, and 59.9275 dB for the exhaustive search technique, and 57.5496 dB, 57.8056 dB, and 60.0081 dB for the BNSA.

An important conclusion regarding the BNSA is that it is possible to trade-off the final PSNR and the computational load; this compromise is achieved by reducing the number of gradient components that are actually computed.

4.4 Broken Arrows (a watermarking technique designed for the BOWS-2 challenge) [FB]

Addressed problem: The watermarking technique “Broken Arrows” has been designed especially for the BOWS-2 (Break Our Watermarking Scheme 2nd Edition) contest [BF07]. The BOWS-2 challenge is divided into three episodes with different contexts. The first episode aims at benchmarking the robustness of the technique against common image processing tools (compression, denoising, filtering...). The second episode is dedicated to oracle attacks. The third episode focuses on threats when many contents watermarked with the same secret key are released.

Motivations: Regarding the first episode, the scheme must be efficient so that it strongly multiplexes the original content and the watermarking signal in a non-reversible way when the secret key is not known. Moreover, no robustness against geometrical attacks is needed because they yield low PSNR values unacceptable in the contest. Considering the second episode, the technique must be sufficiently simple so that the software implementation of the detector runs very fast because we expect a huge number of trials during this second episode. Counterattacks should be included if possible in the design. The third episode implies that the contenders are expected to deduce some knowledge about the secret key in order to better hack the pictures. Consequently “Broken Arrows” must not be trivially hacked.

Technical approach: The practical choices of the algorithm are motivated by the general constraints in image watermarking, e.g. security, robustness and distortion. The visual distortion has been taken into account by choosing the medium and high frequencies
of the image thanks to the wavelet transform and applying a proportional embedding. The PSNR of the watermarked images is controlled during the embedding, resorting to norm conservation property of some orthogonal transforms and by taking into account the proportional embedding step. Moreover the algorithm relies on two techniques in order to have a decent robustness. The first one is commonly known as informed embedding: the watermarked vector is generated in order to be as far as possible from the border of the detection region. Furthermore, proportional embedding in a transform domain enables to shape the spectrum of the watermark like the spectrum of the host. Regarding the constraint of security, the original content is projected successively onto lower dimension subspaces in order to ease the creation of the watermark signal. However, the first projection is private and depends on the secret key. This prevents the pirate from tracing the contents in the successive subspaces and it restricts his play-ground to a very high dimensional space. The dimension is almost as big as the number of pixels in the image. The detection region is composed of several regions introducing some diversity in the embedding because the host contents are pushed towards many different regions. Finally, at the detection side, the security is also strengthened by randomizing the decision of the detector when the signal is near the frontier and by introducing notches in the detection region.

Obtained results: The algorithm provides a pretty good robustness regarding common robustness attacks with a embedding PSNR of 43dB and a probability of false alarm between $10^{-6}$ and $10^{-7}$. Regarding complexity, for a $512 \times 512$ grey-scale image, the computational time for an embedding is of approximately 1.0s for the embedding and 0.8s for the detection on the BOWS-2 server (a 3-ghz Intel Xeon). Consequently the BOWS-2 server, with 2 dual-core processors, has the possibility to detect around 350 000 images per day.

4.5 A Regression-Based Restoration Technique For Automated Watermark Removal

Addressed problem: The self-similarities attack [RDCD02] is a restoration technique for image de-watermarking. In this this attack is automated and made independent on its image database.

Motivations: The performance of the self-similarities attack depends on a database of image blocks by which the actual watermarked blocks are replaced. For a generic attack the image database has to be from a generic source. But its performance could be improved if the database comes from the same source as the image under attack.

Technical approach: We propose to employ regression to model the elements of the feature space (wavelet coefficients in the case of the BOWS-2 challenge [ECR07]) from its surrounding. Because spatially close elements in images strongly depend on each other, but the elements of the watermark do not, the image can be preserved by estimation from the surrounding while the watermark is completely removed.
Obtained results: In the experiments, an average PSNR of 30 dB for successfully broken images using the proposed regression-based technique [Wes08] was determined. This technique can also work as an estimator of the watermark if the difference to the marked original image is calculated. 10,000 of such estimated watermarks have been successfully combined in Episode 3 of BOWS-2 to remove the watermark with 46 dB PSNR on average.

5 Steganography - Steganalysis

The activity of WVL2 also touched steganography and steganalysis, as demonstrated by the two subsections below. The first one regards the development of a public tool for the benchmarking of steganographic algorithms and its use for the evaluation of the security of a particular audio steganographic scheme. The second one discusses the extension of the well-known Benford’s law to image steganalysis with very interesting results both from a theoretical (the extension of Benford’s law to digital images) and practical (use of such an extension of steganalysis) points of view (see [PGHA07] for further details).

5.1 Publimark [Pub08]

Addressed problem: Publimark is a tool, which embeds a steganographic message into selected LSBs of audio signals by using a scalar Costa’s scheme. The symmetric key required by the scalar costar scheme is asymmetric encrypted and also embedded by using trellis code quantization. In our work, we address the security evaluation of an embedded steganographic message with the tool Publimark [Pub08].

Motivations: A secure embedded digital steganographic message should be non detectable for non authorized users. Motivated from this assumption, we focus on the black box, gray box and white box security evaluation of the embedding technique implemented as tool Publimark [Pub08].

Technical approach: Our evaluation concept evaluates with black box evaluation techniques on one hand the embedding positions of the steganographic message and on the other hand, it estimates the used capacity. With gray box testing, statistical analysis are used to differ between a cover and a marked object. Thereby, correlation, normal distribution and chi-square test as well as the sreganography/watermark copy attack are performed to evaluate the security of the embedded steganographic message. Our white box evaluation is based on source code analysis, whereby the used key space and content depending embedding are in focus.

Obtained results: The security evaluation results can be summarized as follows [Neu08]: Our black box evaluation tests, performed with digital silence audio content identified, that Publimark embeds the steganographic message into each third LSB by increasing it. A key depended random generator generates the bit sequence used for embedding. In addition, there is an exception between sample value 1557 and 1661, where no LBSs are modified. The first 1557 sample values (remind, only each third of it) are used to embed the asymmetric
encrypted random selected symmetric key used for the scalar Costa’s scheme. After sample value 1661 the embedding message is embedded into each third sample value and if digital silence, or audio signals, where at the beginning is silence used as cover signal, then it is easy to classify between a cover and a marked object. Furthermore, the length of the embedded message can be easily estimated by counting the modified LSBs. The statistical analysis of marked audio signals did not always distinguish between marked and cover signals, whereas we identified no significant statistical characteristics. Our white box evaluation identified the two security leaks of a small key space for the internal randomly selected symmetric key (32 bit) and the content independent embedding (embedding into digital silence gives a lot of information about the embedding technique, message size, etc.).

5.2 Benford’s law in image processing [PGHA07]

Addressed problem: In this work a generalization of Benford’s law for the first significant digit is developed. This generalization is based on keeping two terms of the Fourier expansion of the probability density function of the data in the modular logarithmic domain. We prove that images in the Discrete Cosine Transform domain closely follow this generalization.

Motivations: Benford’s law of “anomalous digits” was enunciated by General Electric’s physicist Frank L. Benford in 1938, and predicts the frequency of appearance of the most significant digit (MSD) for a broad range of natural and artificial data. Since Benford’s paper, many works have made significant contributions at both the fundamental and the application levels.

Technical approach: It is shown that while images in the “pixel” domain seem not to obey Benford’s law, the situation changes quite dramatically when they are transformed using the Discrete Cosine Transform (DCT). Furthermore, we present a generalization of Benford’s law, based on Fourier analysis, that leads to a much closer fit to the observed digits frequencies. We also give a theoretical explanation of why images in the DCT domain satisfy the generalized law; such explanation heavily relies on well known and thoroughly tested statistical properties of DCT coefficients. Finally, we hint at some possible applications in forensics, by showing how the Fourier-based formulation can be used to detect whether an image has been watermarked.

Obtained results: It is shown how a generalization of Benford’s law can be employed for steganalytic purposes in images, that is, for detecting whether a certain natural image contains a hidden message. We have done so by proving for the first time that Generalized Gaussian distributions follow a generalized form of Benford’s law and, furthermore, that this extends to combinations of GGs, opening the gate to video forensic applications.
6 Authentication

A core activity of WVL2 during the whole duration of the project regarded the use of digital watermarking for authentication purposes. In this framework, during the last year of the project, the activity focused on the authentication of different types of media, including printed and digital text documents [VVK+07], biometric data [VKV+08] and digital images [SNP08]. As a further, more general, result, a novel theoretical framework for watermarking-based authentication has been developed, and used to find a good trade-off between the different requirements imposed by the authentication scenarios, namely the probability of successful attack, security and embedding distortion [CBPG07].

6.1 Tamper-proofing of Electronic and Printed Text Documents via Robust Hashing and Data-Hiding [VVK+07]

Addressed problem: We deal with the problem of authentication and tamper-proofing of text documents that can be distributed in electronic or printed forms. We advocate the combination of robust text hashing and text data-hiding technologies as an efficient solution to this problem. However, there are two main concerns in this approach, namely the limited data storage capacity offered by current text data-hiding methods and the lack of reliable and secure robust text hashing functions.

Motivations: The advocated approach for text document authentication is very attractive for various reasons. First, the authentication of the document is performed directly without accessing a hash database. Second, the hash cannot be easily separated from the document. Finally, this approach can be easily implemented into any modern text editing tool and the resulting document can be stored using a suitable electronic format.

Technical approach: First, we consider the problem of text data-hiding in the scope of the Gel’fand-Pinsker data-hiding framework. Second, we study two approaches to robust text hashing that are well suited for the considered problem. The experimental work compares both approaches and shows their robustness against typical intentional/unintentional document distortions.

Obtained results: Firstly, we addressed the problem of limited data storage capacity of current text data-hiding methods by considering their combination in the scope of the Gel’fand-Pinsker text data-hiding framework. In this context, two modern text data-hiding methods, namely color index modulation and location index modulation, and their combination, were studied. Secondly, we studied two text hashing algorithms, namely OCR + MAC text hashing and random tiling text hashing, that are particularly well suited for the considered problem. In particular, we showed by experimentation that OCR + MAC text hashing shows better applicability than random tiling text hashing. However, we have also observed that the OCR + MAC text hashing method highly relies on the accuracy of the OCR tool. Moreover, the

---

1Actually WVL2 is the fusion of the two WPs of the initial activity plan, one of which was devoted to watermarking-based authentication.
experimental work also confirms that both text hashing algorithms are robust against typical legitimate document distortions that include electronic format conversion, printing, scanning, photocopying, and faxing.

6.2 Authentication of Biometric Identification Documents via Mobile Devices [VKV+08]

Addressed problem: We consider the problem of authentication of biometric identification documents via mobile devices such as mobile phones or PDAs. We assume that the biometric identification document holds biometric data (e.g., face or fingerprint) in the form of text, and personal data in the form of text, both being printed directly onto the identification document.

Motivations: The security of person identification requires a theoretic investigation of several issues. The theoretic aspects of biometric fusion have been investigated thoroughly in the past. However, any secure solution also needs to address the issue of document security. The current state-of-the-art in the domain of document authentication is lacking a thorough theoretical framework which guides the design of practical secure authentication systems with the required accuracy of performance, usually measured in terms of probability of error in decision making.

Technical approach: We propose a novel low-cost approach to document security, and particularly authentication, based on data-hiding and perceptual hashing, biometrics and portable devices with optic equipment and present an information-theoretic analysis of the proposed approach. The proposed solution uses digital data-hiding in order to cross-store the biometric data inside the personal data and vice versa.

Obtained results: A theoretical framework was developed which enables the analysis of future authentication systems based on robust visual hashing and data hiding and guides their design. In particular, we advocate the separation approach which uses robust visual hashing techniques in order to match the information rates of biometric and personal data to the rates offered by current image and text data-hiding technologies. We also described practical schemes for robust visual hashing and digital data-hiding that can be used as building blocks for the proposed authentication system. These schemes share a common requirement, namely resistance to legitimate distortions to which the identification document may be subjected.

6.3 Invertible chaotic fragile watermarking for image authentication [SNP08]

Addressed problem: This work, which is a continuation and finalization of work on the same topic presented in deliverable D.WVL.9, addresses the problem of fragile, blind watermarking of still images for authentication purposes. More specifically, the method aims at achieving data-dependent invertible authentication. Any manipulation that takes place in an authenticated image can be localized in a relatively small region of it.
Motivations: The primary motivation behind this work is the fact that the majority of image fragile watermarking methods cannot achieve invertibility, i.e., there is no way to retrieve the original image. In terms of motivations for the technical approach that was followed these were provided by the sensitivity of chaotic signals to initial conditions as well as the robustness provided by image dependent watermarks.

Technical approach: The binary watermark generation process takes place in the spatial domain. A non-linear iterative system, more specifically a chaotic Chebyshev function with properly selected parameters, is applied on the pixel intensities (data-dependence). Non-linearity in watermark embedding allows us to exploit the chaos property of extreme sensitivity to initial conditions and chaotic function parameters, while data-dependence enhances the system’s security performance. During detection, the inverse procedure is applied. If the correct key and chaotic function parameters are used and the image is not manipulated, the output of the detection function will lie in a specific range of values. In this case, watermark can be precisely reproduced and erased, and thus the authentic image can be retrieved. Otherwise, the detector’s output value will eventually fall in an "illegal" interval thus signalling a violation of the image integrity. Moreover, three layers of watermarks (consisting of blocks of different sizes) are superimposed to the authentic image. Any malicious modification of the image content is detected and localized in the block where it is identified. Within the reporting period the problem of chaotic synchronization, which may result in two different watermarks converging towards the same values when applied on the same image has been tackled. Moreover a theoretical study of the tampered region localization properties of the proposed method has been performed. Robustness and security issues of the method have been also considered. Finally, a study on the selection of appropriate chaotic function parameters has been performed. The method has been also experimentally tested (see next section).

Obtained results: Experiments with a large number of watermarks and different images verified that the method has an extremely low false positive detection rate, zero false negative detection rate and very good localization accuracy (especially when the layered scheme is used). A journal paper has been authored and submitted for publication in Chaos, Solitons and Fractals [SNP08]

6.4 A Novel Interpretation of Content Authentication [CBPG07]

Addressed problem: This work deals with practical and theoretical issues raised by the information-theoretical framework for authentication with distortion constraints proposed by Martinian et al. .

Motivations: In the last years multimedia editing tools have undergone an impressive evolution, putting powerful capabilities within reach of average unskilled users. This seeming advantage constitutes at the same time a serious threat. Indeed, using those advanced tools, the authenticity of multimedia contents can be effectively compromised by a much larger number of people than ever before. This new trend stresses the importance of developing multimedia authentication techniques aimed at solving this critical issue.
**Technical approach:** The optimal schemes proposed by Martinian et al. rely on random codes which bear close resemblance to the dirty-paper random codes which show up in data hiding problems. On the one hand, this would suggest to implement practical authentication methods employing lattice codes, but these are too easy to tamper with within authentication scenarios. Lattice codes must be randomized in order to hide their structure. One particular multimedia authentication method based on randomizing the scalar lattice was recently proposed by Fei et al.. We reexamine here this method under the light of the aforementioned information-theoretical study, and we extend it to general lattices thus providing a more general performance analysis for lattice-based authentication. We also propose improvements to Fei et al.’s method based on the analysis by Martinian et al., and we discuss some weaknesses of these methods and their solutions.

**Obtained results:** The obtained results show the trade-off between figures-of-merit as probability of successful attack (closely related to the reliable achievable rate of the data hiding system), security, embedding distortion and computational cost. Whereas Martinian’s approach shows very good performance from the probability of successful attack, security and embedding distortion point of view, it is computationally unfeasible; on the other hand, Fei’s approach is computationally cheap, but at the cost of needing a significantly larger embedding distortion for a given probability of successful attack. The proposed scheme tries to fill the gap between these two extreme approaches, providing a range of intermediate strategies. The performed analysis outlines the equilibrium that a system designer should take into account when devising an authentication scheme.

### 7 Miscellaneous works

This section includes any research activity that, though not directly mentioned in the WAV-ILA workplan, concerns contiguous topics that for their nature play an important role from a practical perspective.

#### 7.1 Comparative study of chaotic and white noise signals in digital watermarking [MKP08]

**Addressed problem:** The problem addressed in this work was that of experimentally comparing the performance of a certain family of chaotic watermark signals against other signals used as watermarks namely white noise and colored highpass noise signals. Performance evaluation was done in a correlation detection framework, under the influence of attacks.

**Motivations:** The motivation for this work came from previous theoretical studies that showed the superior performance of chaotic watermark signals in such a framework but mainly in a attack-free setup.
Technical approach: White noise, coloured highpass noise, lowpass chaotic signals and highpass chaotic signals were used as watermarks and embedded into cover images. Chaotic signals were generated using the skew tent map and their frequency characteristics were controlled by the parameter $\alpha$. The watermark was embedded in the three detail bands of DWT and blind detection was performed by correlation. The presence of a watermark was determined in these watermarked images after they were subjected to cropping, JPEG compression and additive noise.

Obtained results: It was found that the highpass chaotic watermarks perform steadily better than the highpass noise signals in the presence of the imposed attacks. It was also found that lowpass chaotic signals have the best overall performance for the attacks discussed, with these signals performing best in six out of 12 experiments (mainly JPEG compression and noise addition). Highpass chaotic signals perform next best with best results in five out of the 12 experiments (mainly cropping). In general, chaotic signals performed better than the corresponding noise signals in the presence of the attacks presented. This work has been performed in cooperation with researchers from National University of Ireland, Maynooth. A journal paper has been published in Chaos, Solitons and Fractals [MKP08].

7.2 Watermarking 3-D Volumes in the Discrete Fourier Transform Domain [SP07]

Addressed problem: The problem addressed in this work is that of robust blind watermarking of 3D volumetric (voxel-based) data like those resulting from medical imaging devices such as MRI, CT etc. Watermarks for such data should be robust to a wide range of data manipulations that include filtering, compression, rotation around the three axes, 3D translation and scaling.

Motivations: The limited literature for watermarking of such data along with the increased use of volumetric data provided sufficient motivation for dealing with this topic. In terms of the approach that was selected (Fourier domain watermarking and utilization of watermarks of symmetric nature), the motivation came from the properties of 3D Fourier transform as well as the need to perform watermark detection with limited computational complexity in case of attacks.

Technical approach: The watermark is a binary signal that is embedded in a multiplicative way on the magnitude of the middle frequency transform coefficients. More specifically, the watermark is embedded in a spherical shell of inner radius $R_1$ and outer radius $R_2$. In addition, the watermark is constructed so as to have a degree of rotational symmetry i.e. so that its values are identical between the 20 pyramids that form an icosahedron. Detection is performed by correlation, in a blind way. Due to its symmetric nature, watermark detection in rotated volumes has to be checked only for a reduced set of potential rotation angles e.g. for z-axis rotation angles lying in the interval $[0, \frac{2\pi}{5}]$ (instead of $[0, 2\pi]$) thus reducing the detection complexity. The watermark is also invariant to translation due to the properties of the Fourier transform and behaves well in scaling.
Obtained results: The method has been tested on grayscale medical volumetric data. Experiments showed that due to the middle frequencies where the watermark is embedded, the method is robust to 3D mean and median filtering compression. It is also robust to histogram equalization and sufficiently robust to isotropic scaling but not robust to anisotropic scaling. Moreover, the experimental results verified the invariance to translation and the fact that a reduced search space can be used to cope with rotations. A journal paper has been authored and published in the IEEE Transactions on Multimedia [SP07].

7.3 Review of image watermarking techniques [TNP08]

Addressed problem: A review of the basic issues and recent developments in image watermarking was conducted. This resulted in the authoring of a book chapter which is actually an updated version of the book chapter [TNP05] authored within WVL.2. The book chapter provides a critical view on the performance, pros and cons of methods and approaches proposed up to now in the literature.

Motivations: The motivation behind authoring this book chapter was to provide readers with an up-to-date comprehensive review of the area of image watermarking and highlight the basic principles, challenges, and techniques that have been proposed in order to cope with robustness and security issues.

Technical approach: Being a review chapter, this work does not propose any new solutions but rather reviews existing ones. Issues that are reviewed in this chapter are (not exhaustive list):

- Basic principles, applications and major categories of image watermarking algorithms.
- Requirements of copyright protection watermarking methods.
- Performance metrics and benchmarking systems.
- Attacks against copyright protection watermarking systems.
- Technical approaches towards coping with attacks and achieving good detection performance.
- Watermarking techniques for image content integrity and authentication.

Emphasis has been placed on covering the developments in the field that appeared since the authoring of the predecessor [TNP05] of this chapter.
7.4 On Distortion Measures Associated to Random Desynchronization Attacks [BPG07]

**Addressed problem:** In this work, we investigate geometric distortion measures which aim at being simultaneously general, related to human perception, and easy to compute in stochastic contexts.

**Motivations:** Desynchronization attacks based on fine resampling of a watermarked signal can be very effective from the point of view of degrading decoding performance. Nevertheless, the actual perceptual impact brought about by these attacks has not been considered in enough depth in previous research.

**Technical approach:** Our approach is based on combining the stochastic characterization of the sampling grid jitter applied by the attacker with empirically relevant perceptual measures. Using this procedure, we show that the variance of the sampling grid, which is a customary geometric distortion measure, has to be weighted in order to carry more accurate perceptual meaning. Indeed, the spectral characteristics of the geometric jitter signal have to be relevant from a perceptual point of view, as intuitively seen when comparing constant shift resampling and white jitter resampling. Finally, as the geometric jitter signal does not describe in full the resampled signal, we investigate more accurate approaches to producing a geometric distortion measure that takes into account the amplitude modifications due to resampling.

**Obtained results:** This paper tried to establish a formal framework for the study of the statistical properties of distortion measures for the difference between a geometrically distorted signal (stochastically resampled) and the corresponding original. We have given a way to obtain a version of the MSGT measure, which is boundable for stochastic sampling. This bound shows that the variance of the sampling grid, used in different prior works as the distortion parameter, may skew the performance results. Also, we have discussed distortion measures that encompass not only the distortion of the sampling grid but also the interpolation used.

7.5 “Modeling gabor coefficients via generalized gaussian distributions for face recognition” [GJP07]

**Addressed problem:** In this research, we tackle the problem of proposing statistical models for Gabor coefficients in a face recognition scenario.

**Motivations:** The main motivations behind this work are the following:

- Although statistical models for image-derived features have been widely employed in image processing applications, no one has proposed or used statistical priors for Gabor features in face recognition.
Gabor-based templates require a large amount of memory to be stored. Once statistical models have been proposed, one can take advantage of these underlying statistics to quantize coefficients (data compression).

**Technical approach:** Classical Gabor-based face recognition systems make use of 40 Gabor filters for feature extraction. For each of these filters, we extract features from face images and store the obtained coefficients in separate bunches. Each bunch of coefficients is modeled using an univariate Generalized Gaussian (GG) density, whose parameters are chosen via a ML approach. In order to assess the accuracy of the fitted models, we measured the Kullback-Leibler divergence between each of the histograms and the corresponding fitted densities.

To achieve data compression, we took advantage of the underlying statistics by means of the Lloyd-Max algorithm, and carried out face authentication experiments on a standard database to evaluate whether degradation exists or not when coefficients are compressed.

**Obtained results:** The main results obtained are listed below:

- GGs provide a good model for Gabor coefficients, with parameters depending on the frequency and orientation of the original Gabor filter.
- Statistical-based data quantization allows to greatly reduce the amount of data to be stored without degrading the performance of the system (even with only 8 quantization levels (3 bits), performance is not decreased).

### 7.6 Secure Surface Identification Codes [BVKV08]

**Addressed problem:** We introduce an identification framework for random microstructures of material surfaces.

**Motivations:** These microstructures represent a kind of unique fingerprints that can be used to track and trace an item as well as for anti-counterfeiting.

**Technical approach:** We first consider the architecture for mobile phone-based item identification and then introduce a practical identification algorithm enabling fast searching in large databases. The proposed algorithm is based on reference list decoding. The link to digital communications and robust perceptual hashing is shown. We consider a practical construction of reference list decoding, which comprises computational complexity, security, memory storage and performance requirements. The efficiency of the proposed algorithm is demonstrated on experimental data obtained from natural paper surfaces.

**Obtained results:** We have proposed a mobile architecture and a secure protocol based on microstructures to address the issue of counterfeiting branded goods. The secure protocol has been compared with existing approaches, namely digital communications and robust perceptual hashing. It has been shown that the proposed protocol satisfies the requirements in
terms of performance, security, complexity, memory storage. These results have been verified experimentally. The mobile architecture has been implemented, and two proposed approaches, reference list decoding and minimum reference distance decoding, have been shown to offer very good results in terms of probability of error whilst offering a huge reduction in complexity when compared to classic minimum distance. These results have been experimentally verified. Thus, we have shown that the practical results support the theory and the proposed protocol is a fruitful approach to the stated problem.

7.7 Detection of malevolent changes in digital video for forensic applications

*Addressed problem:* Is it possible to identify if all the scenes of a video sequence have been taken with the same camera? Are we able to distinguish if a video sequence has been altered (e.g. time duration)?

*Motivations:* It is crucial to provide reliable, inexpensive, and fast tools able to assess digital video and images authenticity. In particular it would be useful a reliable identification of the device used to acquire a particular digital content and the verification of its integrity.

*Technical approach:* We have developed a new methodology to detect forgeries in digital videos, starting from a method for still images proposed by Fridrich *et al.* in [LFG06] based on PRNU. The system is able to create a self-building reference pattern by using the initial frames of the video sequence. With this method we are able to assert if all the scenes of a video have been taken with the same camera and if some images have been malevolently inserted; moreover we are also able to determine if modifications have been made within the frames of the video.

*Obtained results:* Experimental results show that this method is able to assert if all the scenes of a video have been taken with the same device or some frames have been fraudulently inserted to change video sequence semantic. The system still continues to work even in presence of MPEG compression.

7.8 A Secure Multidimensional Point Inclusion Protocol [TPKCL07]

*Addressed problem:* The point inclusion problem refers to deciding whether a point lies in a certain spatial region. In a multiparty scenario, two participants decide whether a point known to the first lies inside a region specified by the second. In a secure solution, neither party must gain knowledge about the other’s input.

*Motivations:* Point inclusion involving multidimensional signals is an underlying problem in many common signal processing applications that must be run in untrusted environments. There are many examples of these applications, as biometric authentication, classification, database queries, secure positioning, or watermarking/fingerprinting.
Technical approach: In this work, an efficient and provably secure two-party protocol for solving the point inclusion problem in a convex region bounded by hyperplanes in $N$-dimensional space is presented. In this construction, the public key encryption scheme of Paillier [Pai99] is used for concealing the input coordinates of the point, the relative position of this point and each of the hyperplanes is computed under encryption, the BITREP gate [ST06] is used to extract the result for each hyperplane, and they are merged again into the binary decision. The presented primitive can also be extended to multi-party scenarios, as well as to non-convex regions, as every non-convex region can be expressed as the disjoint union of convex regions.

For dealing with unconnected regions, the construction is extended to regions specified as the union of several hyperellipsoids, with the same privacy properties as the former one.

Obtained results: In this work, a new primitive for securely solving the $N$-dimensional point inclusion problem in polytopes and in hyperelliptic regions is presented. The primitive is useful in many applications, including biometrics, classification, database queries, positioning and watermarking. The presented primitive is analyzed in terms of communication, round and computation complexity, including a full complexity analysis of the BITREP subblock, and input packing is proposed as a complexity reduction strategy when the number of dimensions is high.

The protocol is presented for the semi-honest model, but a sketch is provided for extending it to malicious parties. Extending the protocol for use with more than two parties is straightforward and requires a convexity proof when the polytope is shared among several parties.

7.9 Privacy Preserving Error Resilient DNA Searching through Oblivious Automata [TPKC07]

Addressed problem: The problem of privacy-preserving DNA searching consists in checking if a short template (e.g., a string that describes a mutation leading to a disease), known to one party, is present inside a DNA sequence owned by another party, accounting for possible errors (produced by mutations or by the sequencing process) and without disclosing to each party the other party’s input.

Motivations: Human Desoxyribo-Nucleic Acid (DNA) sequences offer a wealth of information that reveal, among others, predisposition to various diseases and paternity relations. Privacy concerns about DNA information have traditionally been addressed through laws and procedures, but the breadth and personalized nature of this information highlights the need for privacy-preserving protocols.

Technical approach: The main technical contribution of this work is a protocol that allows to execute any finite state machine in an asymmetric two-party setting in an oblivious manner, requiring a communication complexity which is linear both in the number of states and
the length of the input string. This protocol can be used to solve the problem of oblivious approximate string matching (searching), in which two parties collectively determine whether their strings differ (one is approximately a substring of the other) by a given number of edit errors, without disclosing their strings to each other. The proposed solution translates one of the strings into a finite-state machine and executes it obliviously. The protocol is also extended to automata with non-binary output.

**Obtained results:** A protocol for the secure evaluation of finite state machines is presented. Besides showing its security in the semi-honest setting, the protocol has been proven to be efficient in terms of communication complexity, the latter being linear in the size of the input alphabet and in the number of states of the FSM.

Secure DNA matching is presented as the main application field, thereby overcoming the performance deficits of previous approaches, and constituting the first efficient privacy-preserving solution for error-resilient DNA searching. Furthermore, the presented protocol can also be used for privately solving any problem that involves matching a string against a regular expression or running a sequential transducer, such as searching a DNA database with incomplete definitions, file parsing (oblivious spam checkers or virus analyzers) or lexical analysis.

8 Main contributions of WVL2 across the whole project

Before summarizing the main results that have been obtained in the 4 years of the project, we recall that after the reshaping of the Wavila virtual lab in Y2 (in this process the new WVL2 incorporated the initial WVL4 "Hybrid systems" and WVL5 "Authentication"), the objectives of WVL2 focused on bridging the theoretical security definitions and developments obtained in the scope of WVL1 with real practical systems in the chosen application domains.

8.1 Watermark robustness

In the starting period of WVL2 the research activity aimed at closing the gap between existing practical watermarking methods based on the spread spectrum framework and theoretically achievable rates according to Gel’fand-Pinsker framework. The main problem originated from the interference of host data, which severely degraded the communication rate. The development of a comprehensive framework and theoretical foundations of quantization- and lattice-based schemes, where WAVILA researchers greatly contributed to, has allowed closing the above gap using low-complexity practical methods. Moreover, a careful analysis of the error probability was performed for fixed rates. A particular focus of this line of research was the investigation of host statistics impact on the performance of practical data-hiding systems under additive attacks.

A second important line of activity regarded the robustness of watermarking techniques against specific content modifications such as amplitude scaling and geometrical synchronization that appeared to be an essential problem for practical schemes. The WAVILA researchers
proposed several solutions to these problems by using a number of elaborated techniques including spherical codes, rational dither modulation, non-uniform quantization techniques and self-reference watermarking are found. Moreover, the information-theoretic and detection-theoretic consideration of synchronization techniques in the context of digital watermarking created the necessary theoretical basis for the analysis of security leakages provided by the synchronization mechanisms. The above activity culminated in the first WAVILA Challenge (WaCha’05) workshop organized by UVIGO, CNIT and GAUSS in Barcelona, Spain, (June, 2005), part of which was explicitly targeted to the analysis of the possible approaches to cope with watermark de-synchronization induced by geometric attacks. Useful theoretical and practical hints were obtained.

8.2 Watermark security

While (within WVL2) the initial period of project was more focused on watermark robustness, the research activity soon focused on the core issue of the project, i.e. watermarking security. By following the theoretical framework developed within WVL1 (related to the theoretical definition of security of digital watermarking systems), significant progresses have been achieved in WVL2 with respect to the evaluation of the security of various practical watermarking techniques based on both spread-spectrum and side-informed embedding. The analyzed systems include

- Additive Spread Spectrum [CPFPG05, PFPG08b];
- Improved Spread Spectrum [PFPG08b];
- Lattice-based Data Hiding [PFPG08a, PFPG07a, PFPG07b, PFPGFC06, BH05];
- Costa’s scheme [PFCPG05b];
- Quantization-based Detection Techniques [PFCPG05a];
- Dirty Paper Trellis watermarking [BD07];
- Content-Based watermarking techniques [BG05].

In order to emphasize the practical nature of WP2, practical methods for estimating the secret parameters of the embedding functions have been devised and tested for all the above methods. The performance of these estimators have been compared to the theoretical bounds provided by the measures proposed in WVL1.

8.3 The BOWS contests

To practically validate the power of the developed theoretical frameworks, a new initiative called Break Our Watermarking System (BOWS) has been launched. The idea behind this contest is to disable the detection of watermarks embedded into 3 images by using a state-of-the-art watermarking method whose algorithm was not disclosed during the first stage of the contest. According to the WVL2 objectives, such kind of contest should provide useful hints about the selection of attacking strategies by uninformed attackers and the way to extract
security leaks from the detector answer. This contest clearly revealed the power of sensitivity attacks that studied the unknown decision boundary of used detector. The contest has raised a significant interest in the watermarking community.

Taking into account the theoretical and practical achievements as well as extending very important feedback obtained from the first web-based contest, a second contest (BOWS2) was organized, which was divided into three episodes targeting: (a) Robustness against image processing (episode 1); (b) Sensitivity against oracle attacks (episode 2) and (c) Key estimation attacks (episode 3). The results of this contest will be presented at a special session during some reputable international conference as well as reported in one of the issues of EURASIP Journal on Information Security (EURASIP JIS).

### 8.4 Miscellaneous works

In addition to the above well defined and coherent research activities, a huge amount of work has been devoted to the development of new watermarking schemes that permit to get closer to the theoretical bounds on watermark robustness and security. Such activity encompassed all possible media contents, including: audio, images, video, text and 3D objects; and touched a wide variety of applications including: copyright protection, fingerprinting, authentication, steganography. A detailed description of the various algorithms developed by all the partners during the project duration can be found in the deliverables produced by WVL1 (including this one for the results obtained during the last reporting period).
References

T. Furon, F. Cayre and P. Bas, editors, Information Hiding: 9th international workshop, Lecture
notes in computer science, Saint-Malo, France, 2007. Springer Verlag, Berlin, Germany.


[BG05] Patrick Bas and Alejandro Lobo Guerrero. First wavila challenge: several considerations on
the security of a feature-based synchronisation scheme for digital image watermarking. In 1st
WAVILA CHALLENGE (WaCha 2005), Barcelona, Catalonia (Spain), June 2005.

[BH05] P. Bas and J. Hurri. Security of dm quantization watermarking scheme: a practical study for
digital images. In International Workshop on Digital Watermarking, Lecture notes in computer
science, Sienna, Italy, 2005. Springer Verlag, Berlin, Germany.

[BPG07] Félix Balado and Fernando Pérez-González. On distortion measures associated to random desyn-
chronization attacks. In Proceedings of the conference Security, Steganography, and Watermark-
ing of Multimedia Contents IX, San Jose, CA, January 2007. SPIE.


[CBPG07] Pedro Comesaña, Félix Balado, and Fernando Pérez-González. A novel interpretation of content
authentication. In Proceedings of the conference Security, Steganography, and Watermarking of
Multimedia Contents IX, San Jose, CA, January 2007. SPIE.


[CPFPG05] Pedro Comesaña, Luis Pérez-Freire, and Fernando Pérez-González. Fundamentals of data hiding
security and their application to Spread-Spectrum analysis. In 7th Information Hiding Work-
shop, IH05, Lecture Notes in Computer Science, Barcelona, Spain, June 2005. Springer Verlag.

[CPG06] Pedro Comesaña and Fernando Pérez-González. The impact of the cropping attack on scalar

[CPG07a] Pedro Comesaña and Fernando Pérez-González. Breaking the bows watermarking system: Key

[CPG07b] Pedro Comesaña and Fernando Pérez-González. Dither modulation in the logarithmic domain.
In Proceedings of the the International Workshop on Digital Watermarking, Guangzhou, China,
December 2007.

[CPG07c] Pedro Comesaña and Fernando Pérez-González. On a watermarking scheme in the logarithmic
domain and its perceptual advantages. In Proceedings of the IEEE International Conference on

of the ACM Multimedia and Security Workshop (MMSEC’07), Dallas, TX, September 2007.

[CPG07e] Pedro Comesaña and Fernando Pérez-González. Two different approaches for attacking bows. In
Proceedings of the conference Security, Steganography, and Watermarking of Multimedia Con-
tents IX, San Jose, CA, January 2007. SPIE.

[CPGB06] Pedro Comesaña, Fernando Pérez-González, and Félix Balado. On distortion-compensated
dither modulation data-hiding with repetition coding. IEEE Transactions on Signal Processing,


[GJPGC+07] Daniel González-Jiménez, Fernando Pérez-González, Pedro Comesaña, Luis Pérez-Freire, and
José Luis Alba Castro. Modeling gabor coefficients via generalized gaussian distributions for
face recognition. In Proceedings of the IEEE International Conference on Image Processing,
volume 4, pages 485–488, San Antonio, TX, September 2007. IEEE.


