Chapter 1

Introduction

1.1 Situation and objectives

In the last years, the great development of the Internet has generated an exponential growth of the amount of available data in the network. These data are easily shared and can be accessed from any computer connected to the network.

Obviously, not all data can be freely distributed, so they must be protected against unauthorized uses. The common way of protecting data consists of restricting their accessibility through encryption. In this way, data stay protected when stored in insecure places or transmitted through insecure channels. Unencrypted data can then be recovered by authorized users who know the decryption key. The aforementioned solution cannot be applied when authorized users are not completely trusted as, once data are decrypted, they are not protected anymore. So, other methods must be used to protect data when they have to be made available to possibly dishonest users. These methods protect data in such a way that protection stays imperceptible, *i.e. transparent*, to users. Thus such methods aim at *transparent protection of data.*

In this thesis we focus on the transparent protection of two kinds of data:

Multimedia content: In electronic commerce of multimedia content, merchants sell their products to possibly dishonest buyers that may redistribute them. In this case, data are imperceptibly protected against illegal redistribution by using *steganography* [1] to embed copyright information in them.

Very often, multimedia contents are published in untrusted places where they could suffer malicious alterations. In an environment where only a few users (not all of them) need to make sure the published contents have not been altered, transparent protection of such contents is a good solution. Steganography can be used to provide authentication while keeping protection imperceptible.

Statistical microdata: Policy makers and researchers often request statistical offices to release data to perform statistical studies. When these data contain information about individual entities (microdata), privacy becomes a top priority issue. Data must be released in a way that combines statistical utility and protection of the privacy of entities concerned.

One of the properties of products sold in electronic format is that they can be copied very cheaply and without quality loss. Although this reduces the production costs, it facilitates illegal redistribution. This illegal redistribution can be rather efficient and fast when taking advantage of the Internet potential. Current research in electronic copyright protection is

[1] *Steganography is the art of hiding a secret message within a larger one in such a way that others cannot discern the presence or contents of the hidden message [KP00].*

focused on copy detection. In it, steganography is used to embed copyright information in the product before being sold. In case of discovery of illegal copies, this additional information will be retrieved and will allow ownership of the content to be proven or the identity of the dishonest buyer who began such distribution to be determined.

An important part of our research is focused on the development of new copyright protection schemes offering new properties.

Traditionally, authentication and integrity of data being published and available through computer networks has been ensured by digital signatures. A digital signature consists of a message attached to the content to be protected that allows to detect any alteration that the product may suffer. It also allows the signature respondent to be authenticated. In the case of multimedia contents, dealing with an attached message is not a feasible option, especially if we want third parties to stay unaware of such protection. Rather than an attached message, a better solution is to imperceptibly embed the authentication message inside the content. In this way, content can be authenticated while overcoming the drawbacks of attached messages and achieving transparency.

Part of our research is focused on the study of reversible steganography for lossless authentication and integrity protection of multimedia contents.

Increased corporate, government and academic demand has prompted official statistics to release individual respondent data (microdata). Obviously, dissemination of such microdata must be done in a way that ensures protection of the confidentiality of survey respondents. Protecting confidentiality necessitates perturbing the data, so that individual respondent cannot be identified, while preserving the analytical properties of the data file. Methods to pertub data in this way are called *statistical disclosure control* methods (also called *masking* methods when referred to microdata). Methods offering low disclosure risk usually cause a greater data perturbation leading to high information loss. On the other side, low information loss leads to high disclosure risk. Thus, there is a tradeoff between information loss and disclosure risk in masking methods.

Some masking methods presented in the literature offer good masking properties while keeping information loss low. Our research in this area has focused on the development of post-processing methods to post-process the output of current well-performing masking methods in order to decrease information loss while keeping disclosure risk low.

When protecting data in the way addressed in this thesis, the protection method introduces some amount of noise into the data. Generally, this noise is not a matter of concern as it remains imperceptible (or transparent) to users. But when dealing with precision-critical data, this noise may turn data useless for some applications.

To solve the above drawback, we have developed methods that make it possible for trusted users to remove part of the protection to obtain a clearer version of data. In this sense, untrusted users just see the completely protected data while, the more trusted a user is, the more noise she can remove.

1.2 Structure of this thesis

This thesis is organized as follows.

Chapter 2 presents a state of the art on transparent protection of multimedia contents and statistical microdata. It is divided in three main sections. The first two deal with the application of steganography to multimedia contents protection, focusing on digital images. More precisely, the first chapter section is about copyright protection while the second section is about data authentication. The third section deals with statistical disclosure control methods for statistical microdata.

Chapter 3 presents our contributions to watermarking for digital images. It is composed of five sections. The first section presents an algorithm to provide imperceptibility to mark embedding algorithms for images. The next two sections present two new watermarking schemes. Details on the properties of each scheme as well as examples are given. The fourth section proposes mixing watermarked digital objects to increase the robustness of current watermarking schemes by combining their properties. Specific examples involving image watermarking systems are given to prove the effectiveness of the approach. Finally, the last chapter section provides a study on the invertibility of a well known spread-spectrum watermarking scheme which proves suitable for lossless image authentication.

Chapter 4 presents our contributions to binary collusion-secure fingerprinting codes. More precisely, we present a construction that generates codes secure against collusions of up to three dishonest buyers. For a moderate number of possible buyers, our construction results in shorter codewords than the current general proposal by Boneh and Shaw.

In Chapter 5, our contributions to statistical disclosure control of statistical microdata are presented. The first chapter section presents a new score to compare different masking methods that allows consideration of masked data sets with a number of records not equal to the number of records of the original data set. Next, we propose a post-masking optimization procedure which enhances current best-performing masking methods by decreasing information loss while keeping disclosure risk low.

Chapter 6 presents a novel application of watermarking to providing multilevel access to precision-critical data. This chapter is divided in two sections proposing solutions for statistical microdata and multimedia contents, respectively.

The concluding remarks and a summary of the results presented in this thesis can be found in Chapter 7. Some guidelines for future research are given in this chapter as well.