Database Watermarking for Copyright Protection

Radu Sion

Network Security and Applied Cryptography Lab
Computer Science, Stony Brook University
sion@cs.stonybrook.edu

Summary. As increasing amounts of data are produced, packaged and delivered in
digital form, in a fast, networked environment, one of its main features threatens to
become its worst enemy: zero-cost verbatim copies. The ability to produce duplicates
of digital Works at almost no cost can now be misused for illicit profit. This mandates
mechanisms for effective rights assessment and protection.

One such mechanism is based on Information Hiding. By concealing a resilient
rights holder identity “signature” (watermark) within the digital Work(s) to be pro-
tected, Information Hiding for Rights Assessment (Watermarking) enables ulterior
court-time proofs associating particular Works with their respective rights holders.

One main challenge is the fact that altering the Work in the process of hid-
ing information could possibly destroy its value. At the same time one has to be
concerned with a malicious adversary, with major incentives to remove or alter the
watermark beyond detection – thus disabling the ability for court-time proofs –
without destroying the value of the Work – to preserve its potential for illicit profit.

In this chapter we explore how Information Hiding can be deployed as an effective
tool for Rights Assessment for discrete digital data. More specifically, we discuss
numeric and categorical relational data.

1 Introduction

Mechanisms for privacy assurances (e.g., queries over encrypted data) are
essential to a viable and secure management solution for outsourced data.
On a somewhat orthogonal dimension but equally important, we find the
requirement to be able to assert and protect rights over such data.

Different avenues are available, each with its advantages and drawbacks.
Enforcement by legal means is usually ineffective, unless augmented by a
digital counterpart such as Information Hiding. Digital Watermarking as a
method of Rights Assessment deploys Information Hiding to conceal an in-
delible “rights witness” (“rights signature”, watermark) within the digital
Work to be protected (see Figure 1). The soundness of such a method relies
on the assumption that altering the Work in the process of hiding the mark
does not destroy the value of the Work, while it is difficult for a malicious adversary ("Mallory") to remove or alter the mark beyond detection without doing so. The ability to resist attacks from such an adversary, mostly aimed at removing the watermark, is one of the major concerns in the design of a sound solution.

![Diagram of Watermarking and Extraction Process]

**Fig. 1.** Introduction: (a) *Digital Watermarking* conceals an indelible "rights witness" ("rights signature", watermark) within the digital Work to be protected. (b) In court, a detection process is deployed to prove the existence of this "witness" beyond reasonable doubt (confidence level) and thus assess ownership.

There exists a multitude of semantic frameworks for discrete information processing and distribution. Each distinct data domain would benefit from the availability of a suitable watermarking solution.

Significant research efforts [2] [3] [8] [11] [14] [15] [22] [24] have been invested in the frameworks of signal processing and multimedia Works (e.g., images, video and audio).

Here we explore Information Hiding as a rights assessment tool for *discrete* data types i.e., in a relational database context. We explore existing watermarking solutions for numeric and categorical data types.

The Chapter is organized as follows. In Section 2 we explore the broader issues and challenges pertaining to steganography for rights protection. Then, in Sections 3 and 4 solutions for numeric respectively categorical data types are introduced. Related work is discussed in Section 5. Section 6 briefly discusses the current state of the art and Section 7 concludes.

## 2 Model

Before we proceed however, let us first understand how the ability to prove rights in court relates to the final desiderata, namely to *protect* those rights. After all, doesn’t simply publishing a summary or digest of the Work to be protected – e.g., in a newspaper, just before releasing the Work – do the job?