Chapter VI
On the Necessity of Finding Content Before Watermark Retrieval:
Active Search Strategies for Localising Watermarked Media on the Internet

Martin Steinebach
Fraunhofer Institute for Secure Information Technology (SIT), Germany

Patrick Wolf
Fraunhofer Institute for Secure Information Technology (SIT), Germany

ABSTRACT

Digital watermarking promises to be a mechanism for copyright protection without being a technology for copy prevention. This sometimes makes it hard to convince content owners to use digital watermarking for protecting their content. It is only a passive technology adding information into the content to be protected. Therefore some active mechanism is required that completes the protection. This needs to be a search mechanism that localises potentially watermarked media on the Internet. Only then the passive information embedded in the content can help to fight illegal copies. We discuss strategies and approaches for retrieving watermarks from the Internet with the help of a media search framework. While various Internet domains like HTML pages (the Web), eBay, or FTP are discussed, the focus of this work is on content shared within peer-to-peer (P2P) networks.

INTRODUCTION

Digital watermarking has become an established media security mechanism and some see a trend towards an even wider use in the near future (Rosenblatt, 2007). Since the emergence of digital watermarking as an independent field of research more than 20 years ago (Anderson, 1996; Cox & Miller, 2002), a lot of progress has been made in imperceptibly embedding information (the *water-*

mark) into all kinds of digital media. There are algorithms for images, audio, and video but also text, 3D models, and even 3D surfaces (Belloni et al., 2006). The same holds true for retrieving (not extracting) the embedded information from watermarked media, which is naturally very closely connected to the embedding process.

All watermarking algorithms (or schemes) share basic properties like capacity, transparency, or robustness. Capacity describes how much information can be embedded. Transparency measures the (im-)perceptibility or fidelity, that is, how much (or less) does the watermark degrade the quality of the carrier medium. And robustness describes how much the embedded information changes when the carrier medium is altered (Cox, Miller, & Bloom, 2002). The type of information embedded plus the special properties of the watermarking scheme define possible application fields for the scheme.

An important field of application for digital watermarking is copyright protection. For this, information about the (copyright) owner of the medium to protect or, better, information about the receiver of the medium like customer or transaction IDs is embedded using robust watermarking schemes. The latter ensures that the source (or at least the buyer) of an illegally published medium can be identified. This is often called *transaction* or *forensic* watermarking (see also chapter XV, *Traitor Tracing for Multimedia Forensics*).

It might sound trivial, but in order to retrieve a watermark from any medium, one needs to have access to this medium, that is, the medium in question needs to be found first. This is a true challenge and this chapter discusses important issues when searching for watermarked content on the Internet.

This chapter is organised as follows: In the *Background* section, we discuss some fundamentals of the Internet and how currently copyright violations are dealt with. We will also see that watermarking is only a passive protection mechanism and needs an active component that completes the protection. This active part can be fulfilled by a search mechanism for watermarked content on the Internet as described in the *Concept* section. There, a media search framework is introduced that structures such a search and delegates it to specialised components. We show in the *Media content distribution forensics* section, how such a search framework can concretely be used to search various networks and what strategies should be taken in order to reduce the number of files that need to be checked for watermarks. Finally, some future research directions will conclude the chapter.

BACKGROUND

While many media companies see the Internet as a promising market place of the future, currently most copies of their content are transferred without causing revenues (Andrews, 2005). Several technologies aim to reduce the stream of illegitimate copies, digital watermarking being only one of them. Classical digital rights management (DRM) tries to fully control what customers can do with the media they have access to. The media are encrypted using well established cryptographic methods. Rights customers have are described in licenses that are issued by license servers along with the keys for decryption (Rosenblatt, Trippe, & Mooney, 2001). DRM systems need to be present on the customers' hardware and often bury deep into the customers' operating system in order to establish their protection—sometimes even opening the customers' system to exploits (Electronic Frontier Foundation, 2007) or refusing play-back of legitimately acquired content (Halderman, 2002). This presence is necessary since they are providing an *active* protection. All active protection mechanisms shipwreck when the item under protection leaves the domain where it can be protected (Schneier, 2001). As we humans consume media in an analogue way through rendering devices, the media necessarily