A Computational Model for Watermark Robustness

André Adelsbach[1], Stefan Katzenbeisser[2], and Ahmad-Reza Sadeghi[3]

[1] Horst Görtz Institute for IT Security, Ruhr-Universität Bochum
andre.adelsbach@nds.rub.de
[2] Institut für Informatik, TU München
katzenbe@in.tum.de
[3] Horst Görtz Institute for IT Security, Ruhr-Universität Bochum
sadeghi@crypto.rub.de

Abstract. Multimedia security schemes often combine cryptographic schemes with information hiding techniques such as steganography or watermarking. Example applications are dispute resolving, proof of ownership, (asymmetric/anonymous) fingerprinting and zero-knowledge watermark detection. The need for formal security definitions of watermarking schemes is manifold, whereby the core need is to provide suitable abstractions to construct, analyse and prove the security of applications on top of watermarking schemes. Although there exist formal models and definitions for information-theoretic and computational security of cryptographic and steganographic schemes, they cannot simply be adapted to watermarking schemes due to the fundamental differences among these approaches. Moreover, the existing formal definitions for watermark security still suffer from conceptual deficiencies.

In this paper we make the first essential steps towards an appropriate formal definition of watermark robustness, the core security property of watermarking schemes: We point out and discuss the shortcomings of the existing proposals and present a formal framework and corresponding definitions that cover those subtle aspects not considered in the existing literature. Our definitions provide suitable abstractions that are compatible to cryptographic definitions allowing security proofs of composed schemes.

1 Motivation

Multimedia applications deploy various cryptographic and watermarking techniques to maintain security. Typical application scenarios are dispute resolving, proof of authorship and asymmetric and anonymous fingerprinting.

In this context, the security analysis and security proofs for the resulting composed schemes require a suitable formal framework and reasonable security definitions. Modern cryptography already uses established formal models and definitions for information-theoretic and computational security. Inspired by cryptographic methodology similar approaches have been proposed for

J. Camenisch et al. (Eds.): IH 2006, LNCS 4437, pp. 145–160, 2007.

steganography [1,2,3,4]. In contrast, less investigation has been done with this regard for watermarking schemes, and the existing approaches do not cover the subtle aspects essential for reasonable formal security definitions, analysis and abstraction of watermarking schemes.

The need for formal definitions of watermarking schemes, and their most notable properties, such as robustness, false-positive and false-negative probabilities, is manifold: first, one requires formal definitions as suitable abstractions to build, analyse and prove the security of applications on top of watermarking schemes. Second, one requires suitable formal definitions to prove the robustness of watermarking schemes. Furthermore, such definitions provide valuable guidance and basis in the development of provably robust watermarking schemes.

One should note that steganography, although likewise watermarking a means for information hiding, differs from watermarking with respect to various aspects. The most important difference concerns their requirements: In steganography there is a strong hiding requirement, stating that an adversary cannot even detect the *presence* of some stego-message in stego-data. In watermarking, however, one usually does want to prevent watermarks to be detectable by an adversary.[1] Instead, the challenging core property, distinguishing watermarking schemes from other cryptographic or data-hiding primitives, is the robustness property, which guarantees that a watermark cannot be removed without significantly distorting the stego-data and making it useless.[2] Due to the fundamental difference between steganography and digital watermarking, one cannot simply adapt recent definitions of steganographic security [3,4].

In this paper, we point out and discuss the shortcomings of the existing proposals for watermark security definitions as well as the subtle aspects/parameters that these proposals do not cover. In fact, our review shows that even the meaning of watermark security is still not well understood, mainly because many authors do not focus on the main, distinguishing security property of watermarking schemes, which cannot be achieved by applying complementary cryptographic measures: *robustness*. We propose formal (and intuitive) definitions for watermarking schemes, including robustness, that (i) incorporate these aspects/parameters and (ii) can be used as a suitable abstraction for security proofs of composed schemes in the context of various applications.

2 Related Work

In recent years, there has been a remarkable body of literature on definitions for robustness and security of watermarking schemes. Most of the existing proposals

[1] One may require watermarking schemes to provide an optional secrecy property, requiring that adversary cannot obtain any information about the concrete watermark embedded in the stego-data. This requirement is very different and much easier to achieve (e.g., by using standard encryption schemes) than the strong hiding property which is at heart of steganographic systems.

[2] Note, that we do not consider fragile watermarking schemes, because fragility can be achieved quite easily, using cryptographic primitives.