SECURITY ENGINEERING FOR ZERO-KNOWLEDGE WATERMARK DETECTION

*André Adelsbach and Ahmad-Reza Sadeghi**

Horst Görtz Institute for IT Security
Ruhr-Universität Bochum, Germany
andre.adelsbach@nds.rub.de, sadeghi@crypto.rub.de

Markus Rohe

Saarland University
Saarbrücken, Germany
mail@markus-rohe.de

ABSTRACT

Many applications of watermarking schemes require to prove the presence (or absence) of a certain watermark to a potentially untrusted party. For standard watermarking schemes this poses a major security problem, since the proving party has to reveal security critical information (watermark, watermarking key, reference data) to the untrusted verifier, who could exploit this information to remove the watermark.

In this paper we review zero-knowledge watermark detection (ZKWMD) as a solution to this problem and consider its secure integration in certain applications. Furthermore, we present the first implementation results of a provably secure zero-knowledge watermark detection scheme, which shows the practicability and maturity of this technique.

1. INTRODUCTION

Digital watermarking schemes allow to embed a watermark in digital data. One may classify the watermarking schemes with respect to two main flavours: *robust* watermarking schemes allow to detect the watermark, even after the marked data has been (maliciously) modified, whereas *fragile* watermarking schemes purposely break when the digital data has been modified. Due to their contrary properties, both types of watermarking schemes have fundamentally different applications: robust watermarks are usually used to firmly link certain meta-information, e.g., regarding ownership, to digital data. The robustness property ideally guarantees that this link cannot be broken, without rendering the data useless. Prominent applications of robust watermarking schemes comprise, amongst others, *copyright protection systems* (e.g., dispute resolving [6], proofs of authorship [3] or fingerprinting [7, 15]), or *copy protection/management systems* (e.g., as proposed in the Content Protection System Architecture (CPSA) [1]). Fragile watermarking schemes are mainly used to verify integrity and authenticity of data, e.g., to ensure authenticity of images of a surveillance camera.

A common problem when applying *robust*[1] watermarking schemes in corresponding applications is as follows: On the one hand at some point in time the presence of a certain watermark has to be verified by a not fully trustworthy party. On the other hand, however, the verification requires the revealment of security critical information, such as watermark, detection key and reference data, that enables an attacker to delete the watermark.

Zero-knowledge watermark detection (ZKWMD) [11, 17, 5, 2, 12] applies cryptographic protocols to prove the presence of a certain watermark without disclosing critical information. The information required for detection, such as watermark, reference data and detection key, is concealed, preferably by using cryptographic primitives. An (interactive) zero-knowledge proof protocol is applied to prove the presence of the concealed watermark to the verifying party.[2] Concealing the inputs to detection and the zero-knowledge property of the proof protocol guarantee, that no security critical information is leaked to the verifying party.[3]

However, hiding detection information from the verifying party hinders the immediate verification of certain properties, such as the probability distribution of the watermark as desired in some applications. As a consequence, application of ZKWMD is not always straightforward. Unfortunately, this aspect was not completely understood in the past and misled some researchers to doubt the value of *strong* zero-knowledge watermark detection [12].

Outline In this paper we show that these problems, regarding the properties of hidden watermarks, can be easily solved by means of ZKWMD and we review several practical solutions to securely integrate zero-knowledge watermark detection in several applications. Furthermore, we present the first performance results of a provably secure ZKWMD protocol. We conclude that zero-knowledge watermark detection is practicable and can replace symmetric watermarking schemes in several applications.

2. ZKWMD

The basic idea of ZKWMD is to conceal the input required for detection and to apply zero-knowledge proof systems to prove that the detection criterion holds on these concealed inputs without disclosing any additional information about the security-critical data.[4] ZKWMD protocols should ideally fulfil the following main requirements:

*The information in this document reflects only the author's views, is provided as is, and no guarantee or 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. The work described in this paper has been supported in part by the European Commission through the IST Programme under Contract IST-2002-507932 ECRYPT.

[1]Here, we focus on robust watermarking, as the value of zero-knowledge watermark detection for fragile watermarks seems to be rather limited. However, the techniques outlined below are valid for fragile watermarking schemes as well.

[2]Under stronger assumptions (e.g., random oracle model) one can transform these protocols to non-interactive proofs.

[3]Another approach, trying to overcome this limitation are *asymmetric watermarking schemes* [16]. However, so far there was only little success in finding robust asymmetric watermarking schemes.

[4]Here, we only give an informal characterisation of zero-knowledge watermark detection. See [5, 2] for a more complete and formal definition.

- *Hiding secret information:* The inputs required by the ZK-WMD protocols do not reveal any information about the watermark, the detection key and the reference data.
- *Zero-knowledge:* A run of the protocol does not disclose any *new* information, i.e., information *beyond* the positive detection result and the information already leaked by the protocol inputs.
- *Soundness:* A dishonest prover should not be able to make a verifier falsely believe that the watermark, concealed in the input, is detectable in the given data.

This makes ZKWMD suitable for applications where some party has to prove the presence of a watermark to an untrusted verifier, while at the same time, the verifier does not trust the prover to tell the truth about the presence of the watermark.

2.1. Existing ZKWMD Schemes

The quality of a ZKWMD protocol can be assessed regarding the above security requirements. In particular the "hiding"-requirement is important, as it determines the information a-priori leaked by the hidden detection inputs (without even running the protocol): The ZKWMD protocols proposed in [5] hide the watermark in perfectly hiding commitments[5] (*strong* ZKWMD), while those by Craver et al. [11, 12] hide the plain-text watermark only in a list of "fake" watermarks, which is a rather weak "encryption" (*weak* ZKWMD). In order to illustrate the weakness of Craver's approach consider the following example: even when hiding the plain-text watermark in a list of 2^{40} fake watermarks, this would not achieve an adequate security level. In this case a concealed watermark, having 1000 coefficients (each 16 bit long), would have a size of more than $2^{40} \cdot 1000 \cdot 16$ bit (=2048000 GB) and the detection protocol would require to perform 2^{40} watermark detections!

Further weaknesses may lurk in the ZKWMD protocols themselves. As an example, consider again the protocol by Craver [11] which hides the correct watermark in a list of fake watermarks. This protocol is not zero-knowledge, as a cheating verifier can determine the legal watermark by repeatedly removing a watermark from the list, until the prover fails to prove legality of a watermark in the list. In a recent paper Craver et al. [12] try to alleviate this weakness by embedding several legal watermarks. However, the resulting protocol suffers from several shortcomings: (i) the protocol is still rather weak and may only be useful in very special settings, (ii) there is no security analysis given that an adversary is not capable of generating a fake watermark which fulfils the legality criteria, and (iii) crucial details regarding whitening of fake watermarks to make them indistinguishable from legal watermarks are left open[6] and it is not clear whether these issues can be fixed at all.

3. AMBIGUITY ATTACKS

Beside "low-level" robustness attacks which try to remove watermarks from protected content, there are several *protocol/application level attacks* against watermark-based applications. The most important class of protocol level attacks are the so called *ambiguity attacks*: Here, the adversary embeds additional watermarks or tries to compute watermarks, which have never been embedded in digital data, but which can, nevertheless, be detected therein.

Ambiguity attacks exploit *conceptual weaknesses of the watermarking scheme*, namely the computability of *false positives*. Since ZKWMD protocols represent zero-knowledge proof *equivalents* of the corresponding symmetric watermarking scheme, it is only natural that ZKWMD faces problems similar to those of the watermarking scheme from which it stems: this includes not only general robustness issues of the underlying symmetric watermarking scheme, but also its susceptibility to ambiguity attacks.

For standard symmetric watermarking schemes there are several means to counter ambiguity attacks on the application (protocol) level. Since ZKWMD conceals the watermark from the verifying party, countermeasures proposed for standard watermarking schemes, mostly involving complementary tests on plain-text watermarks, cannot be immediately applied. We stress that appropriate countermeasures, both for standard symmetric watermarking schemes and for zero-knowledge watermark detection, depend strongly on the respective application and its security goals. Fortunately, we will see that "zero-knowledge equivalents" of these countermeasures are quite straightforward and practicable and that ZKWMD can be securely integrated in many applications of robust watermarking schemes.

3.1. Countermeasures for ZKWMD

Countermeasures against ambiguity attacks have been particularly investigated in the context of dispute resolving applications. These countermeasures are *heuristic* approaches which basically require the watermark WM (or watermarking key) to be generated according to a special rule or *legality criteria*, such as computing the watermark signal as a one-way function of the cover-data or deriving them from a cryptographic time-stamp (see [6] for a detailed review). Performing such tests for legality of watermarks is more involved when using ZKWMD, since meaningful ZKWMD perfectly conceals the watermark. There are, however, several viable ways to overcome this hurdle:[7]

1. *Watermark Certification:* A trusted third party may certify the legality of the concealed watermark. An example for this can be found in [5], where a trusted registration center is required anyway to achieve a sufficiently strong evidence for authorship [3]. Therefore, certification of concealed watermarks involves little additional overhead, but significantly improves the overall efficiency: the registration center is only involved *once* during registration, whereas the actual authorship proofs do not involve the registration center anymore (offline proofs) due to the use of zero-knowledge watermark detection. This is a crucial improvement, due to the application of ZKWMD and was overlooked in [12].
2. *Alternative Application-Level Evidence:* In some applications the legality requirement for watermarks can be neglected, when adapting the application protocol accordingly. In dispute resolving, where ambiguity attacks may lead to authorship deadlocks, the heuristic legality requirements have been proposed to allow a dispute resolver to distinguish the fake original from the real original by means of watermarks. Alternatively, one may use timestamps (e.g., on the original

[5] Commitments are cryptographic primitives to bind a party to a value, while at the same time concealing this value from an other party [13].

[6] Note, that Craver et al. discuss whitening by pseudorandomly scrambling watermark coefficients. However, they overlook, that this requires the stego data to be pseudorandomly scrambled prior to detection as well.

[7] Here, we will only review some solutions. The interested reader can find more solutions and details in [4] and in an upcoming paper.

works) for this purpose. We refer the interested reader to [6] for a detailed discussion on dispute resolving.

3. *Zero-Knowledge Equivalents of Standard Legality Criteria:* Here we will discuss two examples of "standard" legality criteria for symmetric watermarking schemes and sketch how to verify these criteria on committed watermarks, by means of complementary zero-knowledge protocols.

   - *Dependency on cover-data* W ($WM = h(W)$): This criterion can be adapted to be used with committed watermark coefficients if we choose an appropriate hash function. Obviously, it is hard to come up with adequate zero-knowledge proofs for heuristic hash-functions based on chaotic mixing such as SHA-1 or RIPEMD-160. However, there exist provably secure hash functions based on number theoretical assumptions such as the discrete logarithm assumption. As an example consider the hash function proposed by Chaum et al. [9], which hashes a message $m = (m_1||m_2)$ as follows:[8] $h(m_1||m_2) := g_1^{m_1} \cdot g_2^{m_2} \mod p$. Using zero-knowledge proofs for the exponential and multiplicative relation from [8, 14], one can prove in zero-knowledge that a commitment contains the hash-value (watermark) of another committed value (cover data W).
   - *No correlation between watermark and cover data*: Ramkumar and Akansu [20] propose that a legal watermark should not correlate with the original data W. One can prove this criterion on committed watermarks and committed cover data running a protocol similar to the ZKWMD protocol of correlation based watermarking schemes, as introduced in [5].
   - *Condition on statistical properties*: Adelsbach, Rohe and Sadeghi [4] recently introduced cryptographic protocols for proving in zero-knowledge that a cryptographically concealed vector (e.g., a watermark) suffices certain statistical properties. These protocols can be applied in addition to zero-knowledge watermark detection to convince a verifying party that the watermark, proven to be present, suffices a certain distribution (e.g., the binary equal distribution with coefficients -1 and 1) and, as a consequence, can be considered a well-formed watermark.

4. *Interactive Watermark Generation:* Any mutually mistrusting parties may use a cryptographic protocol to interactively generate committed *legal* watermarks. Such protocols, that assure both parties that the generated watermark is drawn from a desired probability distribution, have been recently introduced by Adelsbach, Sadeghi and Rohe [4]. This solution is viable for applications, where potential verifying parties are known a-priori and their number is small. Possible applications are non-disclosing dispute resolving schemes [6]: using this protocol an author and a dispute resolver interactively construct the watermark that is to be embedded in the work, such that the watermark provably suffices a suitable distribution. Another application are fingerprinting schemes [15], where our protocol may be run between merchant and buyer to guarantee correct distribution of the fingerprint signal.

[8] Here, $||$ denotes the concatenation operation, p is a large prime number of the form $2q + 1$, q is a prime number and g_i are random generators of the unique subgroup G of $\mathbb{Z}_p^*$ where the order of G is q.

4. PERFORMANCE OF ZKWMD

To demonstrate the practicability of strong ZKWMD we implemented two strong zero-knowledge detection protocols along the lines of [5] for the well-known correlation based watermarking schemes by Piva et al. [19] (non-blind) and by Cox et al. [10] (non-blind). For this, we had to implement several cryptographic building blocks, such as the Damgård-Fujisaki commitment scheme [14] and zero-knowledge proofs for arithmetic relations and interval proofs. All sub-proofs were performed in the random oracle proof mode to make the proof non-interactive and allow for pre-computation of the proof. For our measurements, we ran the prover process and the verifier process on two workstations (Pentium 4 with 2600 MHz and 512 MB RAM) connected by 100 MBit Ethernet. For reasonable security parameters[9] our Java demonstrator reached the following results:

- Piva et al. [19] propose to use watermarks with 16000 coefficients. Computing a committed version of such a watermark takes 2:51 minutes and the committed watermark has a size of about 2 MB. After a one-time precomputation by the prover, which takes about 5:19 minutes, the prover can compute the actual proof (depending on the stego-data) in 6 seconds and the verifier can verify this proof in 27 seconds. This is due to the homomorphic property of the commitment scheme which allows a computation on committed values. The proof requires the prover to send 3.2 MB of data.
- The ZKWMD protocol for the watermarking scheme by Cox et al. [10] was performed with 1000 watermark coefficients. In this case committing to the watermark and the reference-data takes about 23 seconds. One-time precomputation by the prover requires 2:08 minutes, while the actual proof takes 2:14 minutes. The proof requires the prover to send 1.2 MB of data. Although non-blind detection requires more cryptographic sub-proofs, its practical performance is still quite good, as it requires significantly less watermark coefficients for reliable detection.

Our results show that strong ZKWMD protocols are mature and can be deployed in practice to improve the security of several applications, which have been based on symmetric watermarks so far.

5. CONCLUSION AND FUTURE WORK

Conventional watermarking schemes require to disclose *secret* information (e.g., the cover-data, watermark or the detection key) to convince a party of the presence of a watermark. A secure solution to this problem is to apply zero-knowledge proofs to prove that the detection criterion holds on concealed watermarks. The verifying party learns nothing about the secret information required for detection, whereas at the same time she is convinced that the watermark is present in the underlying data. We argued that *strong* zero-knowledge watermark detection (ZKWMD) is necessary, as previous proposals bear severe security issues.

[9] The Damgård-Fujisaki commitment scheme was instantiated on $\mathbb{Z}_n$ with 1024 bit modulus.

In this paper we showed that strong ZKWMD can replace robust symmetric watermarks in applications, such as dispute resolving, direct authorship proofs or fingerprinting. This significantly improves the overall security in case the party verifying the presence of the watermark is not fully trustworthy. The fact, that the watermark remains strongly concealed can be easily handled by adapting the application accordingly.

Finally, we presented first implementation results of strong ZKWMD protocols, which show that this technique is indeed practicable. Implementation of the advanced cryptographic building blocks took about 1 man year, but due to their modular implementation it is quite easy to assemble these building blocks to further zero-knowledge watermark detection protocols. Currently, we do joint work on implementing ZKWMD protocols for new watermarking schemes.

6. REFERENCES

[1] 4Centity. Content protection system architecture – a comprehensive framework for content protection, revision 0.81. `http://www.4centity.com`.

[2] André Adelsbach, Stefan Katzenbeisser, and Ahmad-Reza Sadeghi. Watermark detection with zero-knowledge disclosure. *ACM Multimedia Systems Journal*, 9(3):266–278, September 2003. Special Issue on Multimedia Security.

[3] André Adelsbach, Birgit Pfitzmann, and Ahmad-Reza Sadeghi. Proving ownership of digital content. In Pfitzmann [18], pages 126–141.

[4] Andre Adelsbach, Markus Rohe, and Ahmad-Reza Sadeghi. Overcoming the obstacles of zero-knowledge watermark detection. In *Proceedings of the 2004 multimedia and security workshop on Multimedia and security*, pages 46 – 55. ACM Press, 2004.

[5] André Adelsbach and Ahmad-Reza Sadeghi. Zero-knowledge watermark detection and proof of ownership. In Ira S. Moskowitz, editor, *Information Hiding—4th International Workshop, IHW 2001*, volume 2137 of *Lecture Notes in Computer Science*, pages 273–288, Pittsburgh, PA, USA, 2001. Springer-Verlag, Berlin Germany.

[6] André Adelsbach and Ahmad-Reza Sadeghi. Advanced techniques for dispute resolving and authorship proofs on digital works. In *Proceedings of SPIE Vol. 5020, Security and Watermarking of Multimedia Contents V*, 2003.

[7] Dan Boneh and James Shaw. Collusion-secure fingerprinting for digital data. In Don Coppersmith, editor, *Advances in Cryptology – CRYPTO '95*, volume 963 of *Lecture Notes in Computer Science*, pages 452–465. International Association for Cryptologic Research, Springer-Verlag, Berlin Germany, 1995.

[8] Jan Camenisch and Markus Michels. Proving in zero-knowledge that a number is the product of two safe primes. In Stern [21], pages 107–122.

[9] David Chaum, Eugène van Heijst, and Birgit Pfitzmann. Cryptographically strong undeniable signatures, unconditionally secure for the signer. In Joan Feigenbaum, editor, *Advances in Cryptology – CRYPTO '91*, volume 576 of *Lecture Notes in Computer Science*, pages 470–484. International Association for Cryptologic Research, Springer-Verlag, Berlin Germany, 1992.

[10] Ingemar Cox, Joe Kilian, Tom Leighton, and Talal Shamoon. Secure spread spectrum watermarking for multimedia. *IEEE Transactions on Image Processing*, 6(12):1673–1687, 1997.

[11] Scott Craver. Zero knowledge watermark detection. In Pfitzmann [18], pages 101–116.

[12] Scott Craver, Bede Liu, and Wayne Wolf. An implementation of, and attacks on, zero-knowledge watermarking. In J. Fridrich, editor, *Information Hiding—6th International Workshop, IHW 2004*, volume 3200 of *Lecture Notes in Computer Science*, pages 1–12. Springer-Verlag, Berlin Germany, 2004.

[13] Ivan Damgård. Commitment schemes and zero-knowledge protocols. In Ivan Damgård, editor, *Lectures on data security: modern cryptology in theory and practise*, volume 1561 of *Lecture Notes in Computer Science*, pages 63–86. Springer-Verlag, Berlin Germany, 1998.

[14] Ivan Damgård and Eiichiro Fujisaki. A statistically-hiding integer commitment scheme based on groups with hidden order. In Yuliang Zheng, editor, *Advances in Cryptology – ASIACRYPT '2002*, volume 2501 of *Lecture Notes in Computer Science*, pages 125–142. International Association for Cryptologic Research, Springer-Verlag, Berlin Germany, 2002.

[15] Funda Ergun, Joe Kilian, and Ravi Kumar. A note on the limits of collusion-resistant watermarks. In Stern [21], pages 354–371.

[16] Teddy Furon and Pierre Duhamel. An asymmetric public detection watermarking technique. In Pfitzmann [18], pages 88–100.

[17] K. Gopalakrishnan, N. Memon, and P. Vora. Protocols for watermark verification. In *Multimedia and Security, Workshop at ACM Multimedia*, pages 91–94, 1999.

[18] Andreas Pfitzmann, editor. *Information Hiding—3rd International Workshop, IH'99*, volume 1768 of *Lecture Notes in Computer Science*, Dresden, Germany, October 2000. Springer-Verlag, Berlin Germany.

[19] A. Piva, M. Barni, F. Bartolini, and V. Cappellini. DCT-based watermark recovering without resorting to the uncorrupted original image. In *Proceedings of 4th IEEE International Conference on Image Processing ICIP 97*, volume I, pages 520–523, Santa Barbara, CA, USA, October 26-29 1997. IEEE.

[20] Mahalingam Ramkumar and Ali Akansu. Image watermarks and counterfeit attacks : Some problems and solutions. In *Content Security and Data Hiding in Digital Media*, 1999.

[21] Jacques Stern, editor. *Advances in Cryptology – EUROCRYPT '99*, volume 1599 of *Lecture Notes in Computer Science*. International Association for Cryptologic Research, Springer-Verlag, Berlin Germany, 1999.