A SURVEY ON WATERMARKING APPLICATION SCENARIOS AND RELATED ATTACKS

A. Nikolaidis S. Tsekeridou A. Tefas V. Solachidis

Department of Informatics
Aristotle University of Thessaloniki
Box 451, Thessaloniki 540 06, GREECE
e-mail: nikola@zeus.csd.auth.gr

ABSTRACT

A thorough investigation on all possible scenarios where digital imperceptible watermarking is applicable is presented in this paper. All previously proposed watermarking schemes fall to at least one of the referenced application categories. Possible attacks are divided into categories and application scenarios are presented, always referring to the watermarking parameters involved.

1. INTRODUCTION

A variety of imperceptible watermarking schemes have been proposed over the last few years [1]-[6]. In general, publications on the subject tend to focus on the technical details of the specific scheme (increase of robustness, improvement of imperceptibility, etc.), paying little attention on the application scenarios where the proposed method could fit in. Most of the methods are said to be suitable for either copyright protection or authentication, i.e. for a single specific application with no investigation is done on the possibility of applying the same scheme to other applications as well. The main reason for this is that no attempt for a detailed and systematic listing and categorization of the existing application scenarios took place so far.

The current paper presents a detailed analysis of all possible watermarking application scenarios. A similar analysis dealing with specific scenarios has been performed in [7]. The aim of the paper is to help watermarking scheme developers realize the range of applications that an existing scheme may address, or help them in shaping the characteristics of a new method according to the target application.

An overview of the paper follows. Section 2 presents a coarse classification of watermarking applications, section 3 reviews a proposed attack categorization scheme, and section 4 analyzes various watermarking application scenarios, referring in parallel to the types of attacks that may be encountered in each scenario. Finally, conclusions are drawn in section 5.

2. WATERMARKING APPLICATIONS CLASSIFICATION

Several attempts have been made to discriminate the various classes into which each watermarking application falls [8]-[11]. The most popular classification scheme seems to be the one that is based on the kind of information conveyed by the watermark:

- *IPR protection applications*: In this class watermarking is used as a means to convey information about content ownership and intellectual property rights. This class includes applications such as copyright protection, fingerprinting, usage control and piracy tracking.
- *Content verification applications*: In this case, the watermark indicates whether the multimedia content has undergone any alterations, and in certain cases, pinpoints the type and location of alterations. Typical applications of this class are authentication and integrity checking.
- *Information hiding applications*: In this class watermarks are used as information carriers. The information might be relevant or irrelevant to the product on which they are embedded and may be intended for a specific class of users or a specific use. Applications of this category include people metering and secure communications (including steganography).

3. ATTACK CATEGORIES

Before introducing the various watermarking applications scenarios it might be useful to classify all possible types of attacks on the basis of their effect on the watermark and the way the watermark is interpreted by the detector [12]-[14]. In this way, four broad categories can be formed [15]:

- *Removal attacks*: This category includes attacks that aim at removing the watermark without degrading the perceptual quality of the product. These can be unintentional attacks that occur during common processing operations by the user/system (compression, filtering, resizing, printing, scanning, etc.), or malicious ones like noise addition to weaken the strength of the watermark, or the collusion attack which tries to combine different watermarked versions of the same image to generate an average image that is very close to the original, thus reducing the watermark strength or totally removing the watermark.
- *Presentation attacks*: Instead of removing the watermark, these attacks aim at manipulating the content in such a way that the detector cannot find the watermark. The intention is essentially the same as in the previous category, but the techniques employed to achieve it are different. One example of such attacks is the mosaic attack in which the watermarked image is divided into parts and reassembled using proper HTML tags in order to fool web-based agents. Thus the watermark cannot be detected in any of the individual

The work presented in this paper has been financed by the IST-1999-10987 CERTIMARK project.

image parts which the web crawler accesses. Other examples of such attacks are rotation, enlargement, and affine transformations in general.

- *Interpretation (protocol) attacks*: In this case, the intention of the attacker is to render the watermarking scheme unreliable. This can be done for example by producing a counterfeit original after subtracting a counterfeit watermark from a watermarked image. The attacker can then claim that the watermarked image contains his own watermark and also that he has the original product, thus creating an ownership deadlock [16].
- *Legal attacks*: This category is quite different from the ones presented above, since it implies all the actions that can be taken in a law court in order to damage the credibility of watermarks as proofs of ownership/authenticity in case of disputes. In other words, it does not include manipulations of the watermarked product, but attempts to take advantage of the lack of legal foundation on watermarking as a proof of ownership (i.e. gaps in the legislation on copyright laws), and challenging the credibility of the owner.

4. APPLICATION SPECIFIC SCENARIOS

In this section a detailed explanation of all possible application scenarios involving a watermarking scheme is carried out. The various entities that are involved, the entity that mainly benefits from the use of the watermarking scheme, the types of possible attacks, the watermark characteristics, and the kind of information that may be conveyed by the watermark, should all be defined, for a clear and concise definition of each application scenario. This can help watermarking system developers in understanding what is involved in the efficient implementation of watermarking methods targeted at such an application. Such an approach is attempted in the subsequent categorization:

4.1. Copyright protection

4.1.1. Copyright protection without distribution network

This is a scenario in which only two entities are involved, the copyright owner and the user. Figure 1 shows the entities in this case. The copyright owner, who is also the content owner, is concerned whether he can protect his intellectual property by proving that a certain product is copyrighted by him. The product might have been manipulated by a user in an illegal way in an effort to remove or destroy the watermark or replace it with his own one. This addresses the need for a scheme that is robust to all types of attacks that were defined in the previous section. A zero-bit embedding scheme is adequate enough, since the owner wants to make a decision about the presence or absence of his watermark. This means that we only have to use a detector. Private-key technology should be employed because the copyright holder sells his product only to specific users and he is the one that, at a later time, performs detection. Use of the original image could sufficiently increase detection performance and control false alarms, although blind schemes are usually preferred, since the original is not present or hard to retrieve at the detector's location. A typical example of this scenario occurs when the content owner wants to prove his ownership in a law court against an attacker who has redistributed the product as his own.

Fig. 1. Transaction without a distribution network.

4.1.2. Copyright protection through distribution network without TTP (trusted third party)

This scenario introduces a new entity, the distribution network, through which the product is delivered to the intended users. This is illustrated in Figure 2. Robustness to all categories of attacks is required. A private key can be used in this case, however the need for introducing a public-key watermarking scheme is obvious, since the product is to be made available to a quite large group of people, and, thus, watermark detection may not be centralized, a fact that increases the possibility of key theft. Again, zero-bit watermarks should be constructed and the detector should again be able to answer the question whether the product belongs to the copyright owner, who is the interested entity. A sample case complying to this scenario is that of a web crawler or intelligent web agent that searches the Internet for material copyrighted by a certain owner.

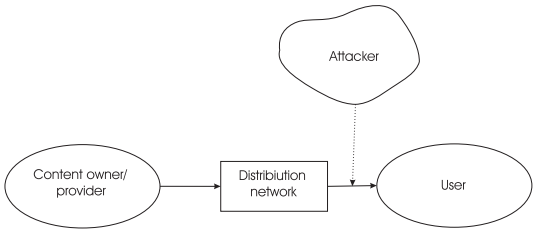

Fig. 2. Transaction through a distribution network.

4.1.3. Copyright protection through distribution network with TTP

In addition to the entities referenced in the previous scenarios, a TTP (trusted third party) arises. The TTP which is responsible for performing detection on behalf of the content owner, provided that the owner has already registered his watermarked products to the TTP. The detector has to answer the question who the registered user of the product under investigation is, thus addressing the need for a single / multiple-bit watermarking scheme, which means that a decoder should be employed after the detector. Private keys should be used because of the centralized control the TTP offers. Again, in a case of copyright dispute, the TTP can resolve rightful ownership. All types of attacks are encountered here as well.

4.2. Broadcast monitoring

This scenario involves, apart from the product owner and the user, a distributor, which in this case is the broadcaster. There can be two alternative aims for broadcast monitoring. The first one refers to piracy tracking done by monitoring stations at the side of the receivers. Robust watermarks should be incorporated to ensure protection against potential attackers. All categories of attacks are possible. Public keys should be used since the detection is not done

by the content owner himself but by the monitoring stations. A single/multiple-bit watermark is needed to identify the owner and protect his rights during transmission of the corresponding material. Use of the original is not feasible since the scenario concerns real time applications. Typical examples of this scenario are TV, Internet and radio providers sending out broadcasts that are monitored at various stations in order to decide if a digital medium is legally transmitted and by whom. Usually, the watermark contains just the index to the entry in a central database, which corresponds to a specific broadcaster and aids in its identification.

The second alternative concerns people metering. In this case, the interested party is the broadcaster who wants to get information about his broadcasts ratings. This is accomplished again by monitoring stations that decode the watermark which contains information about the identification of the broadcaster and of the broadcast content, as well as the time of broadcast and sometimes the receiver's location. This multitude of information requires the employment of a multiple-bit scheme. The fact that the detection of the watermark is done by the monitoring stations calls for use of public keys. Robustness should be provided against intentional removal or presentation attacks inflicted by competing broadcasters that want to make the watermark unreadable and generate lower than expected rating levels. Unintentional removal attacks (e.g. transmission noise) should also be coped with.

4.3. Fingerprinting

This is the typical scenario in which the watermark conveys identification information of the user instead of the owner. This information is inserted by the distributor and is different for each copy of the same product, thus characterizing each single transaction. The distributor is the one interested in tracing illegal copies in order to protect his IPR. Either private key watermarks or public key watermarks, in the case that web crawlers perform detection, can be used. Multiple-bit watermarks should be employed to enable user identification. The potential attackers aim at removing the watermark or misleading the detector, so that the owner cannot prove in a law court that the copy does not belong to the attacker. Alternatively, they may even insert their own watermark so that the interpretation becomes ambiguous. This means that all four categories of attacks should be coped with. Since the same copy may fall in the hands of different users in the distribution line, it is evident that the watermarking schemes should attain robustness to multiple watermarking, i.e. many embedded fingerprints.

4.4. Authentication/Integrity checking

This could be considered, once again, a two-fold scenario, depending mainly on which the interested party is. In the first case, the copyright owner wants to check whether the content has been altered, how and to what extent, either by the distributor or by the user. This is done using fragile or semi-fragile watermarks. Once again, private keys should be used since detection is done by the owner, along with multiple bit watermarks, because they should reflect the alterations that the content has undergone. Possible attackers are concerned about forging authenticity or destroying integrity without affecting the watermark. This means that robustness to interpretation attacks should be considered, as well as to unintentional removal attacks (e.g. compression). A usual application is authentication for multimedia distributed by news agencies. The copyright holder wants to know if the agency has used their source in a misleading way.

In the second case, the user is interested in verifying whether the product he has purchased is authentic or not. Since the embedding is done by the owner, whereas the detection is performed by the buyer, public keys should be employed for security. Either robust or fragile watermarks can be used, depending on whether the user simply wants to check authenticity, or wants to know what changes occurred in the content, respectively.

4.5. Usage control

The content provider is interested in constraining the level of control the end user may have on the purchased product. A single/multiple-bit watermark is required, since the owner wants to be able to check, for example, how many copies of the content have been made. Special purpose devices at the user's end are used for reading the content and detecting the watermark. These devices should be able to change the watermark bits that convey access permission information. Depending on the information carried by the watermark, the devices can allow or prohibit certain operations on the content. Since detection and decoding is done by remote equipment, public keys are employed. Removal attacks should be coped with, since the user may try to make illegal use of the product, as well as presentation attacks, that aim at making the watermark unreadable, although it is still present. An example for this scenario is the use of devices with hardware watermark detectors/embedders for reading/writing, respectively, CDs and DVDs to enable detection/embedding of copy control information. Their control mechanisms prohibit illegal copying of the original discs [17].

4.6. Information hiding

In this scenario the watermarked media is considered as the main channel conveying side-channel information to the end user. Side channel information of three different types exists: public, private and hidden. Their common characteristic is that the watermark contains useful information that may or may not be related to the cover media. For this reason, multiple-bit watermarks are needed. A public side channel watermark contains information about the content in which it is embedded, meant to be accessed by any legal buyer or user of the product. Only unintentional removal attacks should be faced, since an attacker has no benefit from misleading the detector or rendering the watermark undetectable. The interested party is the user who wants to be further informed about the content that he has access to. Therefore, the watermarks should be detected using public keys. An example of this scenario is embedded image annotations or captioning.

In the case of private side channel watermarks the information is intended only for specific authorized users, i.e. a small group of people and a private-key watermarking scheme is adequate, since no intermediate distributor is involved. In addition to removal attacks, also presentation and interpretation attacks may be expected, since malicious attackers may want to remove or destroy the side information. An example of this scenario is audio information hidden inside streaming video, which is only to be revealed to the subscribers of this service.

Finally, hidden side channel information is related to steganography. In this case, the watermark is the only important information and the cover media is just the carrier i.e. it is of no interest to the end user. Obviously, multiple-bit watermarking is necessary and private keys are again to be used by the authorized users.

High robustness against attacks of all categories but the legal ones is required. Typical examples are covert communications such as transactions between military and state offices, intelligence agencies, banks and other financial organizations.

5. CONCLUSIONS

In this paper we have been concerned with the description and categorization of all possible watermarking applications scenarios. The entities taking part, the interested party, the properties of the watermarks and the possible attacks encountered, are all addressed in each scenario. It is clear that although various classifications of these applications can be derived, one should always examine the specific scenario to which his scheme could be applied, in order to appropriately define the parameters involved in the implementation of his method.

6. ACKNOWLEDGEMENTS

This work was based in part in [7]. The following persons should be considered as co-authors of this paper: N. Nikolaidis and I. Pitas (Aristotle University of Thessaloniki), P. Nguyen (Thales Communications, France), C. Busch (FhG/IGD, Germany), F. Balado (University of Vigo, Spain), J.-L. Dugelay (Institut Eurecom, France), F. Cayre (Catholic University of Louvain, Belgium), J. C. Vorbrüggen (MediaSec Technologies, Germany), F. Dumas (INA, France), T. Kalker (Philips Electronics, the Netherlands), R. L. Lagendijk (Delft University of Technology, the Netherlands), J. Barda (Netimage, France), C. Rollin (SACD, France).

7. REFERENCES

[1] I.J. Cox, J. Kilian, F.T. Leighton, and T. Shamoon, "Secure spread spectrum watermarking for multimedia," *IEEE Trans. on Image Processing*, vol. 6, no. 12, pp. 1673–1687, December 1997.

[2] H. Berghel and L. Ó Gorman, "Protecting ownership rights through digital watermarking," *IEEE Computer Magazine*, vol. 29, pp. 101–103, July 1996.

[3] B. Macq and I. Pitas (Editors), "Special issue on watermarking," *Elsevier Signal Processing*, vol. 66, no. 3, 1998.

[4] F. Hartung, P. Eisert, and B. Girod, "Digital watermarking of MPEG-4 facial animation parameters," *Computer & Graphics*, vol. 22, no. 3, 1998.

[5] M. Kutter, F. Jordan, and F. Bossen, "Digital watermarking of color images using amplitude modulation," *SPIE Journal of Electronic Imaging*, vol. 7, no. 2, pp. 326–332, 1998.

[6] C.-T. Hsu and J.-L. Wu, "Hidden digital watermarks in images," *IEEE Trans. on Image Processing*, vol. 8, no. 1, pp. 58–68, January 1999.

[7] "Watermarking applications and requirements for benchmarking," IST project CERTIMARK (IST-1999-10987), Deliverable 2.1.

[8] G. Voyatzis and I. Pitas, "The use of watermark in the protection of digital multimedia products," *Proceedings of the IEEE, Special Issue on Identification and Protection of Multimedia Information*, vol. 87, no. 7, pp. 1197–1207, July 1999.

[9] R. Barnett, "Digital watermarking: applications, techniques and challenges," *IEE Electronics and Communication Engineering Journal*, pp. 173–183, August 1999.

[10] I. J. Cox, M. L. Miller, and J. A. Bloom, "Watermarking applications and their properties," in *Proc. of Int. Conf. on Information Technology: Coding and Computing 2000*, 27-29 March 2000, pp. 6–10.

[11] N. Memon and P. W. Wong, "Protecting digital media content," *Communications of the ACM*, vol. 41, no. 7, July 1998.

[12] I.J. Cox and J.P. Linnartz, "Some general methods for tampering with watermarks," *IEEE Journal of Selected Areas of Communications*, vol. 16, no. 4, May 1998.

[13] F. Hartung and B. Girod, "Watermarking of uncompressed and compressed video," *Elsevier Signal Processing, Sp. Issue on Copyright Protection and Access control*, vol. 66, no. 3, pp. 283–301, 1998.

[14] S. Craver, N. Memon, B-L. Yeo, and M. Yeung, "Resolving rightful ownerships with invisible watermarking techniques: Limitations, attacks and implications," *IEEE Journal of Selected Areas in Communications*, vol. 16, no. 4, May 1998.

[15] S. Craver, B-L. Yeo, and M. Yeung, "Technical trials and legal tribulations," *Communications of the ACM*, vol. 41, no. 7, July 1998.

[16] S. Katzenbeisser and F.A.P. Petitcolas, *Information Hiding Techniques for Steganography and Digital Watermarking*, Artech House, Boston - London, 2000.

[17] J.A. Bloom, I.J. Cox, T. Kalker, J.-P.M.G. Linnartz, M.L. Miller, and C.B.S. Traw, "Copy protection for dvd video," *Proceedings of the IEEE*, vol. 87, no. 7, pp. 1267–1276, 1999.