The use of Watermarks in the Protection of Digital Multimedia Products

G.Voyatzis and I.Pitas

Department of Informatics

University of Thessaloniki, Thessaloniki, 54006 GREECE

fax: +3031-996304, email: {voyatzis,pitas}@zeus.csd.auth.gr

Contact author : I.Pitas

Abstract

The watermarking of digital images, audio, video and multimedia products in general has been proposed for resolving copyright ownership and verifying originality of content. This paper studies the contribution of watermarking for developing protection schemes. A general watermarking framework (GWF) is studied and the fundamental demands are listed. The watermarking algorithms, namely watermark generation, embedding and detection, are analyzed and necessary conditions for a reliable and efficient protection are stated. Although the GWF satisfies the majority of requirements for copyright protection and content verification, there are unsolved problems inside a pure watermarking framework. Particular solutions, based on product registration and related network services, are suggested to overcome such problems.

1 Introduction

The digital form of photographs, paintings, speech, music, video etc. became very popular in the last decade. Digital facilities for creating, processing and storing multimedia products have been found very convenient by creators, providers, editors and customers. At the same time, digital network communications have grown rapidly. In such an environment, digital products can be easily copied, processed for various purposes, broadcasted and/or publicly exposed. However, these revolutionary capabilities are also available to pirates who use them illegally for their personal interest by violating the legal rights of the providers and customers. Subsequently, security issues should be accounted for in the digital networked distribution systems for multimedia products.

Digital piracy, dealing with multimedia products, generally, includes the following cases :

- *Illegal access.* A pirate tries to receive a digital product from a network site without permission.
- *Intentional tampering.* A pirate modifies a digital product in order to extract/insert features for malicious reasons and then proceeds to its retransmission. The authenticity of the original product is lost.
- *Copyright violation.* A pirate receives a product and resells it without getting the permission to do so from the copyright owner.

Techniques based on cryptography, digital signatures and digital watermarks can be used for countering digital piracy [1].

Private or public key cryptography [2] can be used for data access control. Encrypted products are accessible, and decryption is possible only by someone who possesses a proper key. Well established algorithms (e.g. RSA [3] and DES [4]) can be used for this purpose. The encryption/decryption techniques should manipulate large amounts of digital data and should achieve real-time encryption/decryption e.g. for video and digital TV applications [1]. The properties of chaotic systems seem quite useful and suitable for such purposes [5].

Digital signatures are based on cryptographic algorithms and they have been proposed for checking authenticity of digital short messages [2]. A digital signature standard (DSS [6]) has been officially adopted. By using a private key, the original creator produces a digital signature for each product. A public verification algorithm checks if the contents of the product comply to the corresponding signature. The application of such signatures to digital images, video or audio is proven to be inconvenient and impractical because of the large size of the signature which is appended to the original data. Hash functions applied to proper data subsampling have been proposed to solve such problems [7, 8].

Digital watermarking is a rather new technique [9, 10, 11, 12]. It is associated with the ancient technique of information hiding known as *steganography* or "covered writing" [13]. In contrast to cryptography, steganography does not raise suspicions that an important message can be possibly carried inside a harmless medium (e.g. a digital image). Watermarking aims at hiding a secret and personal message in order to protect the copyright of a product [14, 15, 16] or to demonstrate its authenticity, namely, its content originality, also referred as content verification, data integrity or tamper proofing [10, 17]. The decoded watermark information may be just a binary decision indicating watermark existence or absense in the product. An important difference between steganography and watermarking is related to the attackers' goal. A pirate tries to reveal the information carried by a steganographic message. In the case of watermarking, a pirate either tries to remove the watermark in order to violate copyright, or to reproduce it after product tampering in order to achieve a false positive content verification.

This paper describes a general watermarking framework for digital products and analyzes its possible contribution to develop an overall system for copyright protection and content verification. Watermarking has been applied successfully to many multimedia modalities. A large portion of the watermarking literature deals with the copyright protection of still digital images (e.g. [18, 19, 20, 21, 22, 23, 24]). Watermarking algorithms for digital video [25, 26] and digital audio [27, 28, 29] have been developed as well. The proposed watermarking framework is general and abstract and can cover all the above mentioned modalities.

The structure of this paper is the following. In the next section we introduce the general watermarking framework and its basic properties. Section 3 demonstrates the main steps of a watermarking algorithm and the most important techniques that can be used for their implementation. Sections 4 and 5 deal with the capability of watermarks to provide efficient copyright protection and content verification respectively. Unsolved problems and reliability problems are discussed. Finally, in section 6, we propose special digital services and product registration, which can act complementary to watermarking, in order to accomplish an efficient and reliable protection system.

2 The general watermarking framework (GWF)

2.1 The basic digital product distribution model

A detailed description of a real distribution mechanism of digital products is quite complex and includes entities like the original creators, editors, multimedia integrators, resellers, state/regulatory authorities etc. This paper refers to a simplified distribution model presented in Figure 1. The "provider" represents collectively the copyright owner, the editor or the reseller. The user (also called customer) receives the digital product through a network distribution channel. Pirates are unauthorized providers who retransmit products without having the permission of the legal copyright owner or tamper intentionally some original products and retransmit non-authentic versions of them. The customer cannot be protected directly from receiving piratical copies. The watermarking framework, that we shall describe in the following, is applicable in the basic distribution environment shown in Figure 1.

2.2 The watermarking framework

Various forms of digital watermarks can be found in the literature. Watermarks can have the form of LSB manipulations (e.g. [30]), hidden mark codes (e.g. [31, 32]), invisible textures [27], secret constraints in transform domains [14, 33] etc. Generally, we can define as watermark a digital signal W:

$$W = \{w(\mathbf{k}); \mid w(\mathbf{k}) \in U \ , \ \mathbf{k} \in \hat{W}^d\} \tag{1}$$

that is superimposed on digital products through an embedding procedure [9, 10, 34, 35]. $\hat{W}^d$ denotes the watermark domain of dimensions $d = 1, 2, 3$ for audio, still images and video respectively. The watermark signal may have a binary form ($U = \{0, 1\}$ or $U = \{-1, 1\}$) [19, 36, 27, 23] or the form of Gaussian noise ($U = (-1, 1) \subset I\!R$) [37, 28, 38]. Sometimes, W is called "original watermark" in order to distinguish it from transformed watermark versions $\mathcal{F}(W)$ that may also appear during the watermark embedding/detection procedure.

The general watermarking framework (GWF) is defined as the *six-tuple* $(\mathbf{X}\,, \mathbf{W}\,, \mathbf{K}\,, \mathcal{G}\,, \mathcal{E}\,, \mathcal{D})$ related to the system of Figure (1):

1. $\mathbf{X}$ denotes the set of digital products X to be protected.

2. $\mathbf{W}$ is the set of the possible watermark signals defined by Equation (1).

3. $\mathbf{K}$ is a set of ID numbers (e.g. sets of integer parameters) that are called *watermark keys.*

4. $\mathcal{G}$ denotes the algorithm that generates the watermarks by using a key and the digital product to be watermarked:

$$\mathcal{G} \ : \mathbf{X} \times \mathbf{K} \rightarrow \mathbf{W} \ \ , \ \ W \ = \ \mathcal{G}(X, K) \tag{2}$$

5. $\mathcal{E}$ is the embedding algorithm that casts a watermark W in a digital product X_0:

$$\mathcal{E} : \mathbf{X} \times \mathbf{W} \times I\!R \rightarrow \mathbf{X} \ \ , \ \ X_w \ = \ \mathcal{E}\ (X_0\ , \ W) \tag{3}$$

X_w denotes the watermarked version of X_0.

6. Finally, $\mathcal{D}$ denotes the detection algorithm defined as follows:

$$\mathcal{D} \;:\; \mathbf{X} \times \mathbf{K} \rightarrow \{0,1\} \tag{4}$$

$$\mathcal{D}(X,W) = \begin{cases} 1 & \text{if } W \text{ exists in } X \ (H_1) \\ 0 & \text{otherwise } (H_0) \end{cases}$$

where H_0 and H_1 denote the null and the alternative hypothesis respectively.

2.3 Basic definitions

Usually, we search for watermarks in digital products that have been somehow modified intentionally or unintentionally. Therefore, we introduce the notion for perceptual similarity of products defined as follows:

Perceptual similarity: if $X, Y \in \mathbf{X}$ then the notation $X \sim Y$ denotes that the digital products X and Y have the same perceptual appearance. $X \not\sim Y$ denotes that either X and Y are completely different products or Y shows quality reduction with respect to X.

In general, perceptual similarity is based on subjective criteria. However, objective error measures (e.g. [39], [40]) or concepts from content-based search in databases (e.g. [41]) may be used in order to decide about perceptual similarity.

The capability of the detector $\mathcal{D}$ to distinguish watermarks that are not exactly identical is generally limited. Two watermarks are assumed to be different when positive detection of the first watermark does not imply positive detection of the second one. Thus, we introduce the following definition:

Watermark equivalence: The watermark W_1 is equivalent to W_2 ($W_1 \simeq W_2$) when:

$$\mathcal{D}(X,W_1) = 1 \;\Rightarrow\; \mathcal{D}(X,W_2) = 1$$

In many watermarking paradigms, watermark equivalence essentially refers to high watermark correlation. Obviously, identical watermarks are equivalent but the inverse does not hold in general. Equivalent watermarks may differ significantly. The operators $\sim$ and $\simeq$ are commutative and transitive. We remark that perceptual similarity has a different meaning for copyright protection than for content verification.

2.4 Basic properties and necessary conditions for the GWF

The GWF should satisfy specific conditions in order to form a trustworthy basis for copyright protection or for content verification of digital products:

1. *Perceptual invisibility.* The watermark embedding should not produce perceivable data alterations. X_w should not contain any perceptual distortion that reduces the quality of the original product X_0. This property implies easily that:

$$X_0 \sim X_w$$

2. *Key uniqueness.* Different keys should not produce similar watermarks, i.e.:

$$K_1 \neq K_2 \;\Rightarrow\; W_1 \not\simeq W_2$$

   for any product $X \in \mathbf{X}$ and $W_i = \mathcal{G}(X, K_i)$.

3. *Watermark validity.* Only *valid* watermarks should be used in the watermarking scheme. A watermark $W \in \mathbf{W}$ is valid for a particular product $X \in \mathbf{X}$ if and only if:

$$\exists K \in \mathbf{K} \;\text{ such that }\; \mathcal{G}(X, K) = W$$

4. *Non-invertibility.* The function $G_{X^*}(K) = \mathcal{G}(X^*, K)$ should not be invertible i.e. $G_{X^*}^{-1} : \mathbf{W} \to \mathbf{K}$ does not exist or cannot be estimated or approximated from W. By considering a non-surjective mapping G_{X^*}, this condition is directly satisfied. However, this may not be sufficient for the watermarking scheme. In practice, non-invertibility means that for any watermark signal W it is practically impossible to find another *valid* watermark similar to W.

5. *Product dependency.* When $\mathcal{G}$ is applied on different products with the same key, different watermarks should be produced, i.e., for any particular key $K \in \mathbf{K}$ and for any $X_1, X_2 \in \mathbf{X}$:

$$X_1 \not\sim X_2 \;\Rightarrow\; W_1 \not\simeq W_2$$

where is $W_i \;=\; \mathcal{G}(X_i, K)$.

6. *Multiple watermarking.* In general, watermarking of an already watermarked product X_w by using a different key is possible. This feature can be exploited by a pirate. However, it is also desirable in certain cases, e.g. for product stamping and tracing in the distribution channels when several resellers exist. If $X_{w_i} = \mathcal{E}(X_{w_{i-1}}, W_i), i = 1, 2, ...$, then the original watermark should be still detectable in X_{w_i}:

$$\mathcal{D}(X_{w_i}, W_1) = 1 \quad \forall\, i \leq n$$

where n is a sufficient number of coexisting watermarks such that $X_{w_n} \sim X_0$ and $X_{w_{n+1}} \not\sim X_0$.

7. *Reliable detection*: The positive detector output should have an acceptable minimal degree of certainty. If P_{fa} is the detector probability of false alarm then

$$P_{fa} < P_{thres} \tag{5}$$

where P_{thres} is a proper probability threshold which is chosen by the provider.

8. *Robustness.* Let X_0 be a digital product and X_w its watermarked version ($\mathcal{D}(X_w, W) = 1$). We denote by $\mathcal{M}$ a multimedia data processing operator that processes the digital products $X \in \mathbf{X}$. Then the following condition should hold:

$$\mathcal{D}(Y, W) = 1 \;, \;\; \forall Y \sim X_w \;\text{ and } Y = \mathcal{M}(X_w)$$

For any $Y' = \mathcal{M}(X_0)$, the condition $\mathcal{D}(Y, W) = 0$ holds by definition.

9. *Computational efficiency.* The watermarking algorithm should be effectively implemented by hardware or software. Especially, the watermark detection algorithm should be fast enough for multimedia data monitoring in the distribution network.

3 Watermark casting and detection using the GWF

Watermarking comprises of two main operations, namely *casting* $(\mathcal{G}, \mathcal{E})$ and *detection* $(\mathcal{G}, \mathcal{D})$ that are presented schematically in Figure 2.

3.1 Watermark generation

Watermark signals of the form (1) are usually based on pseudo-random number generators or chaotic systems. m-sequences or Gausian pseudonoise signals can be easily produced providing a large set of uncorellated signals and sufficient security (non-predictability and non-invertibility). The produced watermark W may require further transformation to become suitable for embedding. For analytical reasons, it is convenient to decompose the procedure $\mathcal{G}$ in two parts:

$$\mathcal{G} = \mathcal{T} \circ \mathcal{R} \ , \ \mathcal{R} : \mathbf{K} \to \mathbf{W} \ , \ \mathcal{T} : \mathbf{W} \times \mathbf{X} \times \mathbf{K} \to \mathbf{W} \tag{6}$$

The first component $\mathcal{R}$ outputs the original watermark $\tilde{W} \in \mathbf{W}$ that depends exclusively on the key K. When $\mathcal{R}$ is based on a pseudo-random number generator, the key K is mapped directly to the seed of the random number generator (e.g. [16, 37]). When chaotic systems are used, the key set is formed by a convenient transformation of the initial conditions [42]. In both cases the key set is sufficiently large, $\mathcal{R}$ satisfies key uniqueness and, obviously, $\tilde{W} = \mathcal{R}(K)$ is a valid watermark for the key K. Also, the inversion of $\mathcal{R}$ is practically impossible.

The second component $\mathcal{T}$ modifies the original watermark to obtain the watermark W that is product dependent. $\mathcal{T}$ should take into account only salient data features, e.g. data characteristics that are robust to manipulations:

$$\mathcal{T}(\tilde{W}, X_0) \simeq \mathcal{T}(\tilde{W}, X_w) \simeq \mathcal{T}(\tilde{W}, X'_w) \tag{7}$$

where X_0 denotes an original product, X_w a watermarked one and $X'_w = \mathcal{M}(X_w)$, such that $X'_w \sim X_w$.

3.2 Watermark embedding

The embedding procedure is defined as a superposition of the digital watermark signal $W = \{w(\mathbf{k})\}$ onto the original product $X_0 = \{x_0(\mathbf{k})\}$. The most common and simple watermark embedding rules are the following [37]:

$$x_w(\mathbf{k}) = x_0(\mathbf{k}) + \alpha w(\mathbf{k}) \quad \textbf{(additive rule)} \tag{8}$$

$$x_w(\mathbf{k}) = x_0(\mathbf{k}) + \alpha x_0(\mathbf{k}) w(\mathbf{k}) \quad \textbf{(multiplicative rule)} \tag{9}$$

The variable x refers either to the sample intensity/amplitude (spatial/temporal domain) or to a transform coefficient magnitude (transform domain). Parameter α may be different in various data samples (e.g. across image regions) for reasons of perceptual watermark masking. The spatial domain and the additive rule have been used in many algorithms for watermark embedding [27, 16, 19, 36, 23, 21]. However, transform domains may be proven very useful for watermarking. The DFT phase [27, 18] and magnitude [43] have been used for watermark hiding. Watermark embedding in the DFT magnitude is robust to some elementary geometrical modifications (namely rotation and scaling). The DCT domain provides watermarks robust to compression, filtering and other digital processing operations [37, 44]. Recently watermarking based on Discrete Wavelet Transform has been proposed, providing remarkable robustness to JPEG and JPEG2000 [45, 46]. Watermarking techniques for digital images/video, based on $n \times n$ block partitioning, that embed the watermarks by creating special constraint relations of the DCT coefficients of the image blocks have been proposed in [33, 47].

We can describe the embedding procedure $\mathcal{E}$ by considering a generalized superposition operator $\oplus$. The data samples (pixels, audio samples) of the watermarked product are produced as follows:

$$x_w(\mathbf{k}) \;=\; x_0(\mathbf{k}) \;\oplus H(\mathbf{k}) w(\mathbf{k}) \tag{10}$$

where $h(\mathbf{k}) \in I\!R$. The matrix $H = \{h(\mathbf{k})\}$ is d-dimensional and is called *watermark embedding mask*. The operator $\oplus$ includes appropriate data truncation and quantization if needed e.g. due to pixel/sample bit length. $H(\mathbf{k})$ should depend on the product data and is selected so as to provide perceptual invisibility and, at the same time, adequate watermark resistance to product modifications.

3.3 Watermark detection

Watermark detection should be performed on any product X, preferably without the use of the original product. Some watermarking techniques use the original product in order to address efficiently some watermark robustness issues (e.g. [20]). However, the use of the original is a big disadvantage when watermarking is used for product monitoring in network distribution or broadcasting. Therefore, we consider detection without resorting to the original product.

We proceed to watermark detection, by generating first the watermark using $\mathcal{G}$. $\mathcal{G}$ is based exclusively on the product X to be checked and on the key K. However, since K is a constant, we may generate the fixed watermark $\tilde{W}$ once and use it for detection in many products.

After watermark generation the detector $\mathcal{D}$ is applied. The realization of $\mathcal{D}$ implies the following errors:

Type I error: Watermark is detected although it does not exist in the data (false positives).

Type II error : Watermark is not detected in the data although it exists (false negatives).

The above errors occur with specified false alarm (P_{fa}) and rejection (P_{rej}) probabilities respectively. Figure 3 presents schematically the detection errors when a statistical test based on normal distribution is performed. Let $c = 1 - P_{fa}$ denote the certainty of a positive detection, then:

$$c \geq c_{thres} \implies \text{watermark exists} \tag{11}$$

The parameter c_{thres} is the *certainty level* for detection and is chosen by the provider who performs the detection test. (11) is directly related to the reliability condition (5). Hypothesis testing can be used for statistical certainty estimation and detection error minimization [48]. Generally, when false positives become insignificant ($P_{fa} \to 0$) the probability to reject a watermark increases ($P_{rej} \to 1$) and vice versa.

In many cases detection is based on the correlation between the watermark signal and the watermarked product [9, 23, 37, 28]. Statistical tests on the "difference of means" can be also performed for watermark detection [27, 16, 22].

3.4 Satisfaction of basic demands

The GWF inherits the properties of watermark uniqueness, watermark validity, non-invertibility and image dependency from the algorithm $\mathcal{G}$. Embedding $\mathcal{E}$ is responsible for watermark invisibility and robustness. Finally, detection $\mathcal{D}$ should provide high detection certainty and computational efficiency. We remark that $\mathcal{E}$ is to be applied to a relatively small number of digital products whereas $\mathcal{D}$ should be tested on a very large number of accessible products located anywhere in the distribution network. Generally, in the proposed framework, we consider public watermarking techniques that are applied using private keys. The protection is provided by the key not by the casting algorithm. However, alternatively, private (secret) watermark embedding techniques could be applied as well. In this case, the validity of the watermark should be proven.

3.5 Hiding and detecting information streams

The watermark detection, as analysed in section 3.3, extracts just one bit of information (yes/no) at a given certainty level. Such a binary answer is directly associated with the fundamental question: 'Does the watermark $W(X, K)$ exists in the product X?' In the the proposed watermarking framework only the legal owner, who possesses the correct key K, can perform watermark detection. In this case, the hiding of information bit streams is redundant. Bit stream hiding is useful e.g. when a trusted authority (e.g. an authors' collector society) watermarks the products of various copyright owners with the same key. In this case, this authority can use bit stream information hiding for storing and tracing author's identity in watermarked products for monitoring applications. Furthermore, embedding and extracting m-bit watermarks could be proven useful in many cases, e.g., watermarks may present trade-marks that demonstrate an elegant indication for rightful ownership [22, 47]. However, from a mathematical point of view, the power of a single-bit or multiple-bit watermarking schemes can be compared only in terms of the overall detection error. In other cases, multiple bits may be used in order to determine geometrical modifications of the product and, subsequently, to resynchronize the embedded and the reference watermark during detection [21]. Secure embedding and extraction of such additional bits may be obtained, without any restrictions in the watermarking framework, by using private algorithms. Embedding of many bits is necessary when watermarks are used for content verification [49]. Error correction techniques (e.g. Reed-Solomon codes) can be used for recovering bit errors [43].

4 Copyright protection in the general watermarking framework

Copyright violation primarily harms the interests of the providers rather than those of the customers. By considering the basic distribution model of Figure 1 and any distributed product X, the GWF aims to answer effectively the question made by a particular provider: "Am I the copyright owner of X?". By casting a watermark in the products before their distribution to customers or to publicly accessible networks, the above mentioned question becomes "Does my own watermark exist in X?". Consequently, copyright protection is directly related to the copyright owner's ability to detect the embedded watermarks. It must be noted that the GWF can not directly answer the questions "Is this product protected?" or "Whose is this product?". This is because GWF is based on watermarking detection using strictly private keys. Such questions can be replied only when a trusted authority has its own watermarking key, casts watermarks on behalf of the copyright owners and monitors the multimedia product distribution.

In the GWF, all watermarks for products belonging to the same owner are generated by the same key K, which is private. This key is also used for detecting the watermark in any suspicious product existing in the distribution network. Such products should be provided by an automated search procedure $\mathcal{S}$:

$$X = \mathcal{S}(NetworkDomain)\ , \quad X \in \mathbf{X} \tag{12}$$

Watermark detection is applied to any product X supplied by $\mathcal{S}$ and a positive result indicates potential copyright ownership. The certainty level of the detection is chosen by the provider. We distinguish the following cases:

- **Detection with low certainty**. In this case false alarms are frequent. However, the probability to miss a watermarked product is very small. In the case of positive detection, further actions for ascertaining the watermark existence or for copyright proof are required.
- **Detection with high certainty**. In this case $P_{fa} \to 0$ and the detector provides very reliable positive detection. Watermark detection may stand as a strong evidence of legal ownership in a court of law. However, high certainty level increases the rejection probability and the watermarks are proven less robust to intentional or unintentional attacks.

4.1 Direct intentional attacks

A pirate should attack GWF in order to undermine its capability to indicate copyright ownership. The watermark should be robust to unintentional attacks, i.e. to digital processing operations that preserve perceptual similarity. Direct intentional attacks to a watermarking system, i.e. attacks without product modifications, are possible as well. We always assume that the original product is unavailable during detection. We present the following possible attacks and ways for defence:

Extraction of counterfeit watermarks [50]. A pirate forms a signal W', for a particular product X, which forces the detector $\mathcal{D}$ to output a positive answer. Therefore W' is a watermark signal that had never been embedded in X and the pirate uses it as his/her own watermark. However, $\mathcal{G}$ is not invertible and W' cannot be associated to a key, i.e. counterfeit watermarks W' are not valid for the GWF. Craver *et al* [51] have claimed that the watermarking techniques of Cox *et al* [35] and Pitas and Kaskalis [16] are vulnerable to this attack. However, the condition of watermark validity and non-invertibility, which can be easily included in the above-mentioned techniques, renders the extraction of valid counterfeit watermarks impossible.

Detection of false positives [50]. A pirate, after a trial and error procedure, finds a key K^* that provides a positive detection and claims that K^* indicates his/her ownership on a product. However, a watermark can be used as a proof for ownership when it is detected with very high certainty. In this case, false alarms are extremely rare and, subsequently, this attack is unfeasible.

Statistical watermark extraction [28, 36]. The possession of a great number of digital images, all watermarked with the same key, should not dispose the watermark by applying statistical estimation methods (e.g. averaging). Such statistical recoverability is prevented by using product-dependent watermarks.

Multiple Watermarking [16]. An attacker may use this property of the GWF to embed his/her own watermark. Both watermarks (the original and the piratical) can be detected by using the corresponding key. The *original owner* of the product under question is the one who can dispose a copy containing *only* his/her watermark. As a result, the owner must have an archive of his/her watermarked products in order to counter this attack.

Watermarking by using arbitrary keys. Malevolent users or providers may apply watermarking by using arbitrary keys to any accessible product in public network domains that allow data uploading ("watermark bombing"). If these products are publicly exposed, a great confusion may arise when the automated watermark search procedure $\mathcal{S}$ is applied. The negative consequences of such an attack are restricted by the fact that it demands an enormous set of watermark keys. On the other hand, the available network domains, where malevolent users may put such "bombs", are rather few in comparison to the total accessible (read only) domains.

This list of direct watermark attacks is not complete, since new attacks are devised by the pirates or the scientists that study the security provided by the watermarks.

4.2 Unsolved problems for the GWF

Although, watermarking technologies had an impressive growth in the last years, they are still not mature enough in order to provide full fledged copyright protection. In the following, we present some watermarking shortcomings:

Robustness to data processing operations. Watermark robustness under JPEG or MPEG compression for high compression ratios has captured the attention of many researchers so far. However, such robustness is not necessary because a high compression ratio generally implies low product quality. Watermarks ought to be trustworthy and robust to all digital processing algorithms that do not reduce significantly product quality. Without resorting to the original image, the watermark detection in a watermarked product that has been modified by a cascade of elementary geometric transformations (scaling, cropping, rotation, reflection) or by the jitter and mosaicing attacks [52] is not an easy task. Invariant watermarks under rotation, scaling and translation have been proposed recently by Ruanaidh and Pun [43]. However, as far as we know, all current methods fail in cases of combined geometrical attacks.

Development of new techniques for compression and filtering. Watermarking and compression are competing techniques. The first aims to add imperceivable information in the product, whereas the second attempts to remove redundant information. Although, very robust watermarks under compression have been developed for the current compression technologies, this may not be the case for forthcoming compression techniques, e.g. content-based ones. Furthermore, new filters or techniques may be developed by pirates in order to remove watermarks (e.g. [53]). The most important point is that, ideally, the watermark must be robust to all current and future processing techniques. Once the watermarked product is out in the distribution, it is vulnerable to any future attack.

Private key loss/theft. The pirates have strong incentives in stealing the key of large copyright owners (e.g. news agencies, multimedia data libraries, movie makers). This incentive becomes even stronger, if copyright protection schemes based on one or a few trusted authorities, each having a unique key, will prevail. The theft or the reverse engineering of the private key and its possession by a pirate may cause the watermark removal from all the products that belong to the particular provider. A solution to this serious problem might be given by combining secure *time-stamping* and *time-dependent* watermarks. However, no such efficient techniques have been developed yet.

Although a large set of intentional attacks can be efficiently countered, all watermarking techniques proposed so far are vulnerable to piracy in some way. The GWF can be proven dangerously unstable under the above mentioned conditions. Subsequently, to our opinion, digital protection schemes that are based exclusively on watermarking cannot provide complete solutions to the problem of copyright protection at the current technological level.

5 Content Verification using the GWF

Content verification of a product X is relative to an original product X_0 (authentic). Protection against malevolent tampering is required either by the providers or the users. Namely:

- A pirate can break an access control mechanism and replace the digital products on a server with tampered versions [49]. The provider should have a mechanism to check the originality of his/her products in the server.
- A distributor can modify the multimedia data content intentionally or not, thus violating the moral (fraternity) rights of a copyright owner. Thus, the copyright owners want to have the capability to check the integrity of their work in the distribution channels.
- The users want to check, using a trustworthy and efficient technique, the originality of the products that they receive through distribution networks.

The first case is usually addressed by methods based on access control. But, also, the first and second case can be dealt with private key watermarking schemes. Content verification is then equivalent to the detection of a watermark which should be very sensitive to any modification. In the following, we focus our study to the third case case, due to its financial significance.

5.1 Basic protection scheme and forgery

We consider the distribution model of Figure 1. The general watermarking framework, described in section 2, can contribute to content verification as follows:

- The original provider (creator), using a private key K_{pr}, performs watermark embedding to the original product X_0 and produces X_w. It is assumed that watermark embedding does not alter significantly the contents of X_0 and, thus, X_w is considered to be an authentic copy of X_0.
- Customers proceed to the detection of the particular watermark that characterizes exclusively the original provider and guarantees the content integrity of the product. Positive watermark detection denotes that the product is original.

In contrast to watermarking for copyright protection, the pirates do not want to remove watermarks from the products X_w. They have two goals:

1. They try to preserve the original authenticity proof for a product, after a malicious tampering.
2. They insert forged authenticity proofs to other products in order to affect their significance or value.

5.2 Basic demands and protection efficiency

The content verification scheme presented above, imposes the following requirements on the watermarking algorithm:

Public key detection. Since watermark detection is performed by the users, a public detection key for this purpose should be available for each provider. Namely, watermark casting should be based on a private key K_{pr}, but detection should employ a public key K_{pub} that accompanies the product and is associated exclusively to the original provider.

Effective watermark fragility/robustness. Fragile watermarks can be used for content integrity [54]. Fragility is effective when, at the same time, provides sufficient sensitivity to alterations and security. For example, modifications of the LSB are extremely sensitive but not secure, because the product may be altered significantly without changing the LSB. Although fragility is a basic watermark property for content verification, robustness should be also considered in some special cases that do not affect significantly the content integrity (e.g. high quality compression, or other insignificant modifications that are needed to insert the product in a multimedia environment). Generally, local modifications (e.g. object insertion/extraction, montage) should cause verification failure.

Information about tampering. Watermark detection provides a binary (true/false) decision on content integrity. When, content originality is not verified, it would be very useful to extract product tampering information (e.g. which region has been falsified). Therefore, multiple bit watermarking may be useful.

Security against forgery. Watermark creation without the knowledge of private key K_{pr} and, thus, subsequent creation of forged authenticity proofs on other products should be impossible.

Watermarks with fragility to certain data processing or tampering operations have been proposed. Zhu *et al* [55] proposed watermarks for proving authentication of compressed images. Yeung and Mintzer [49] proposed a technique for determining intentional modifications performed in some image regions. Both methods are based on private key watermark detection. The significance of security issues in the watermarking framework creates serious difficulties in developing public key watermarking. Hartung and Girod [56] proposed a technique for public key watermarking on video data.

6 Complementary protection schemes and watermark contribution

We showed in the previous sections that, at the current technological level, serious problems may be encountered when a copyright protection or content verification scheme is based exclusively on watermarking. Therefore, the GWF may be incorporated as a part of a more complete digital product protection system.

6.1 Product storage/registration for copyright protection

Product registration to a trusted authority is a well established way for protecting intellectual property rights (IPR) for various conventional intellectual products, e.g. books, movies etc. Registration information can be used to form indisputable proofs of original ownership. An effective copyright protection system can be designed by considering, additionally to GWF, the following requirements:

- The provider possesses a personal archieve $\mathcal{L}$ of his/her digital products and a "matching procedure" $\hat{m}$ such that:

$$\hat{m}(X, \mathcal{L}) = \begin{cases} 1 & \text{if } \tilde{X} \in \mathcal{L} \text{ and } \tilde{X} \sim X \\ 0 & \text{otherwise} \end{cases} \tag{13}$$

- The provider can register his/her watermark key K_{pr} to a trusted authority. Thus, he/she ensures its uniqueness and increases the reliability of the watermarking scheme. Instead of a private key, a private watermark casting software package may be registered.

- The provider registers his/her watermarked products X_w for any $X \in \mathcal{L}$ to a trusted authority to ensure and time-stamp his/her ownership on X_w.

A protection system based on product registration includes the following actions:

1. Watermark casting is applied by using the registered key or the private software.

2. The original product is included in the provider's archive $\mathcal{L}$ and a watermarked copy is registered to a trusted authority before distribution to customers.

3. The provider proceeds to an automated distribution network monitoring by using $\mathcal{S}$ (see Eq. 12). When low certainty searching level is used, the reliability of a positive result can be reinstated by searching the library $\mathcal{L}$ using the procedure $\hat{m}$.

4. The demonstration of the registered copy X_w in a court of law is the proof for copyright ownership.

This copyright protection scheme is illustrated in Figure 4. We note that, by including registration schemes in the protection system, the watermark contribution is restricted to a signaling mechanism for tracking/monitoring products used illegally in the network. Therefore, watermarks should be resistant to the greatest variety of product modifications. After a positive watermark detection, the registration authority must provide the final and reliable proof about legal ownership.

6.2 Content verification through network servers and authenticity headers

The public key watermark detection requirements and/or insufficient watermark fragility create problems that should be tackled in order to develop a stable and trustworthy content verification system. An efficient solution can be developed along the following lines:

- The provider creates a content verification system (CVS), e.g. a network server, accessible to all customers that provides content verification for his products by using private key watermarking.

- Any allowable modification/processing of the product, which does not destroy its authenticity but obstructs the watermark detection, is declared in an appropriate product "authenticity header" (A_H). The header includes also the site of the corresponding CVS.

Content verification is performed by the following watermark casting and detection procedures:

Watermark casting. The owner uses a private key K_{pr} to watermark a product X. By using a private and secure algorithm, the owner generates a public key $K_{pub} = \mathcal{K}(X, K_{pr})$ which is available to all customers via the authenticity header of the product.

Content verification Procedure. Any customer can verify the originality of a particular product X by sending it to the corresponding CVS. The CVS determines the private key $K_{pr} = \mathcal{K}^*(X, K_{pub})$ and performs watermark detection after taking into account the data modifications that are declared in the product authenticity header. Positive watermark detection with sufficient certainty proves authenticity. The detection result is relayed back to the interested customer. Public key cryptography can be used to have a secure customer/provider communication channel.

The content verification scheme is shown in Figure 5. It is clear that the proposed system does not require public key watermarking. Also, watermarking fragility can be effectively manipulated through the declared product modifications. We remark that a pirate does not aim to remove the watermark and, therefore, should not remove or change the information of the publicly accessible authenticity header. Forgery is not easy, because private watermarking can provide sufficient security.

7 Conclusions

In this paper we presented some fundamental concepts about watermarking and its contribution for developing multimedia data copyright protection and content verification systems. Product distribution usually occurs in public distribution networks where piracy is possible. The providers demand protection from illegal copying and retransmission of their products (copyright violation). On the other hand, the customers (users) and providers demand content verification as a protection against the distribution of non-authentic products.

We have defined and analyzed a general watermarking framework (GWF) applied to a simple product distribution model that includes the provider, the customer, a network distribution channel and, possibly, pirates. GWF consists of three main procedures: watermark generation, embedding and detection. Basic properties should be satisfied, so that GWF contributes efficiently to copyright protection and content verification systems.

Copyright protection, which is based exclusively on watermarking, seems insufficient, at least at the current technological level. The demand of watermark robustness is not easily satisfied and the possibility of future watermark removal by using new intentional or unintentional attacks can not be ignored. An effective copyright protection scheme can be developed by employing product and key registration to trusted authorities. In this case, watermark detection is essentially a signaling mechanism for copyright violations. Watermark robustness to a wide class of possible product modifications is the crucial feature of any watermarking scheme.

In the case of content verification we have different requirements. In this case, watermarks should be fragile to data modifications. Public key watermarking is very important for content verification. However, the use of public watermark detection algorithms poses security problems for the watermarking scheme. We propose the use of secure publicly accessible authenticity servers, controlled by the providers or by trusted authorities. In this content protection scheme, private watermarking can be applied efficiently for content verification.

References

[1] B. M. Macq and J. J. Quisquater. Cryptology for digital TV broadcasting. *Proceeding of the IEEE*, 83:944–957, June 1995.

[2] D. R. Stinson. *Cryptography, Theory and Practice.* CRC Press, New York, 1995.

[3] R. L. Rivest, A. Shamir, and L. Adleman. A method for obtaining signatures and public-key cryptosystems. *Commun. ACM*, 21:120–126, 1978.

[4] FIPS PUB 46. Data encryption standard, 1977.

[5] J. Scharinger. Fast encryption of image data using chaotic kolmogorov flows. *Journal of Electronic Imaging*, 7(2):318–325, 1998.

[6] FIPS 186. Digital signature standard, 1994.

[7] M. Naor and M. Yung. Universal one-way hash functions and their cryptographic applications. In *Proceedings of 21st STOC*, pages 33–43, 1989.

[8] M. Schneider and S. F. Chang. A robust content based digital signature for image authentication. In *Proceedings of ICIP'96*, volume III, pages 227–230, Lausanne, Switzerland, September 1996.

[9] A. Z. Tirkel R. G. Schyndel and C. F. Osborne. A digital watermark. In *Proceedings of ICIP'94*, volume II, pages 86–90, 1994.

[10] W. Bender, D. Gruhl, and N. Morimoto. Techniques for data hiding. In *Proceedings of SPIE*, volume 2420, page 40, 1995.

[11] K. Matsui and K. Tanaka. Video steganography: How to secretely embed a signature in a picture. In *Proceedings of IMA Intellectual Property Project*, volume 1, pages 187–206, 1994.

[12] H. Berghel and L. Ó Gorman. Protecting ownership rights through digital watermarking. *IEEE Computer*, 29:101–103, July 1996.

[13] N. F. Johnson and S. Jacodia. Exploring steganography: Seeing the unseen. *IEEE Computer*, pages 26–34, February 1998.

[14] E. Koch and J. Zhao. Towards robust and hidden image copyright labeling. In *Proceedings of 1995 IEEE Workshop on Nonlinear Signal and Image Processing*, pages 452–455, N. Marmaras, Greece, 20-22 June 1995.

[15] J-J. Quisquater O. Bruyndonckx and B. Macq. Spatial method for copyright labeling of digital images. In *Proceedings of 1995 IEEE Workshop on Nonlinear Signal and Image Processing*, pages 456–459, N. Marmaras, Greece, 20-22 June 1995.

[16] I. Pitas and T.H. Kaskalis. Applying signatures on digital images. In *Proceedings of 1995 IEEE Workshop on Nonlinear Signal and Image Processing*, pages 460–463, N. Marmaras, Greece, 20-22 June 1995.

[17] G. L. Friedman. The trustworthy digital camera: Restoring gredibility to the photographic images. *IEEE Transactions on Consumer Electronics*, 39(4):905–910, 1993.

[18] J. Ó Ruanaidh, W. J. Dowling, and F. M. Boland. Phase watermarking of digital images. In *Proceedings of ICIP'96*, volume III, pages 239–242, Lausanne, Switzerland, September 1996.

[19] R. B. Wolfgang and E.J. Delp. A watermark for digital images. In *Proceedings of ICIP'96*, volume III, pages 219–223, Lausanne, Switzerland, September 1996.

[20] I. J. Cox and J. P. Linnartz. Some general methods for tampering with watermarks. *IEEE Journal of Selected Areas of Communications*, 16(4):587–593, 1998.

[21] M. Kutter, F. Jordan, and F. Bossen. Digital watermarking of color images using amplitude modulation. *Journal of Electronic Imaging*, 7(2):326–332, 1998.

[22] G. Voyatzis and I. Pitas. Digital image watermarking using mixing systems. *Computer & Graphics*, 22(3), 1998.

[23] A. Z. Tirkel, C. F. Osborne, and T. E. Hall. Image and watermark registration. *Signal Processing*, 66(3):373–383, 1998.

[24] J. F. Delaigle, C. De Vleeschouwer, and B. Macq. Watermarking algorithm based on a human visual model. *Signal Processing*, 66(3):319–335, 1998.

[25] F. Hartung and B. Girod. Copyright protection in video delivery networks by watermarking of pre-compressed video. Springer Lecture Notes in Computer Science, 1997.

[26] F. Hartung, P. Eisert, and B. Girod. Digital watermarking of MPEG-4 facial animation parameters. *Computer & Graphics*, 22(3), 1998.

[27] W. Bender, D. Gruhl, N. Morimoto, and A. Lu. Techniques for data hiding. *IBM Systems Journal*, 35(3&4):313–335, 1996.

[28] M.D.Swanson, B.Zhu, A.H. Tewfik, and L. Boney. Robust audio watermarking using perceptual masking. *Signal Processing*, 66(3):337–355, 1998.

[29] P.Bassia and I.Pitas. Robust audio watermarking in the time domain. In *Proceedings of EUSIPCO'98*, volume 1, pages 25–28, Rodos, Greece, September 1998.

[30] G. Voyatzis and I. Pitas. Chaotic mixing of digital images and applications to wateramrking. In *Proceedings of ECMAST'96*, pages 687–694, Louvain-la-Neuve, Belgium, 28-30 May 1996.

[31] D. L. Hecht. Embedded data clyph technology for hardcopy digital documents. In *Proceedings of SPIE*, volume 2171, 1995.

[32] J. Brassil, S. Low, N.Maxemchuk, and L. Ó Gorman. Electronic marking and identification techniques to discourage document copying. In *Proceedings of Infocom'94*, pages 1278–1287, June 1994.

[33] A. G. Bors and I. Pitas. Image watermarking using dct domain constraints. In *Proceedings of ICIP'96*, volume III, pages 231–234, Lausanne, Switzerland, September 1996.

[34] I. Pitas. A method for signature casting on digital images. In *Proceedings of ICIP'96*, volume III, pages 215–218, Lausanne, Switzerland, September 1996.

[35] T. Leighton I. J. Cox, J. Kiliant and T. Shamoon. Sequre spread spectrum watermarking for images, audio and video. In *Proceedings of ICIP'96*, volume III, pages 243–247, Lausanne, Switzerland, September 1996.

[36] N. Nikolaidis and I. Pitas. Robust image watermarking in the spatial domain. *Signal Processing*, 66(3):385–403, 1998.

[37] I. J. Cox, J. Kilian, F. T. Leighton, and T. Shamoon. Secure spread spectrum watermarking for multimedia. *IEEE Transactions on Image Processing*, 6(12):1673–1687, 12 1997.

[38] C. I. Podilchuk and W. Zeng. Perceptual watermarking of still images. Electronic Proc. of IEEE Signal Processing Society 1997, Workshop on Multimedia Signal Processing, June 23-25 1997.

[39] A. B. Watson, R. Borthwick, and M. Taylor. Image quality and entropy masking. In *Proceedings of SPIE*, volume 3016, 1997.

[40] Scott Daly. The visible differences predictor: An algorithm for the assessment of image fidelity. in Digital Images and Human Vision, 1992.

[41] P. Lewis, J. Kuan, S. Perry, M. Dobie, H. Davis, and W. Hall. Content based navigation from images. *Journal of Electronic Imaging*, 7(2):275–281, 1998.

[42] G. Voyatzis and I.Pitas. Chaotic watermarks for embedding in the spatial domain. In *Proceedings of ICIP'98*, Chicago, USA, October 1997.

[43] J. J. K. Ruanaidh and T.Pun. Rotation, scale and translation invariant spread spectrum digital image watermarking. *Signal Processing*, 66(3):303–317, 1998.

[44] A. Piva, M.Barni, and F. Bartolini. Dct-based watermark recovering without resorting to the uncorrupted original image. In *Proceedings of ICIP'97*, volume I, pages 520–523, Atlanta, USA, October 1997.

[45] X. G. Xia, C. G. Boncelet, and G. R. Arce. A multiresolution watermark for digital images. In *Proceedings of ICIP'97*, volume I, pages 548–551, Atlanta, USA, October 1997.

[46] D. Kundur and D. Hatzinakos. A robust digital image watermarking method using wavelet-based fusion. In *Proceedings of ICIP'97*, volume I, pages 544–547, Atlanta, USA, Octomber 1997.

[47] C. T. Hsu and J.L. Wu. Hidden signatures in images. In *Proceedings of ICIP'96*, volume III, pages 223–226, Lausanne, Switzerland, September 1996.

[48] A. Papoulis. *Probability & Statistics.* Prentice Hall, 1991.

[49] M. M. Yeung and F. Mintzer. An invisible watermarking technique for image verification. In *Proceedings of ICIP'97*, volume II, pages 680–683, Atlanta, USA, October 1997.

[50] 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*, 16(4), 1998.

[51] S. Craver, N. Memon, BL. Yeo, and M. Yeung. On the invertibility of invisible watermarking techniques. In *Proceedings of ICIP'97*, volume I, pages 540–544, Atlanta, USA, October 1997.

[52] F. Petitcolas, R. J. Anderson, and M.G. Kuhn. Attacks on copyright marking systems. In *Proceedings of 2nd Workshop on Information Hiding*, Oregon, USA, April 14-17 1998.

[53] G.C.Langelaar, R.L.Lagendijk, and J. Biemond. Removing spatial spread spectrum watermarks by non-linear filtering. In *to appear in Proceedings of EUSIPCO'98*, Rodos, Greece, September 1998.

[54] F. Mintzer, G. W. Braudaway, and M. M. Yeung. Effective and ineffective digital watermarks. In *Proceedings of ICIP'97*, volume III, pages 9–12, Atlanta, USA, October 1997.

[55] B. Zhu, M. D. Swanson, and A. H. Tewfik. Transparent robust authentication and distortion measurement technique for images. In *Proceedings of Conf. on Digital Signal Processing*, pages 45–48, Loen, Norway, September 1996.

[56] F. Hartung and B. Girod. Fast public-key watermarking of compressed video. In *Proceedings of ICIP'97*, volume I, pages 528–531, Atlanta, USA, October 1997.

Figure Captions

1. The basic model for digital product network distribution.
2. Watermarking algorithms for a) casting and b) detection.
3. Hypothesis testing for watermark detection based on normal distributions.
4. Copyright protection scheme based on watermarking and product registration.
5. Content verification system based on private watermarking and public accessibility to authentication servers. DCh, QCh, VCh denote the distribution, verification request and reply channels respectively.

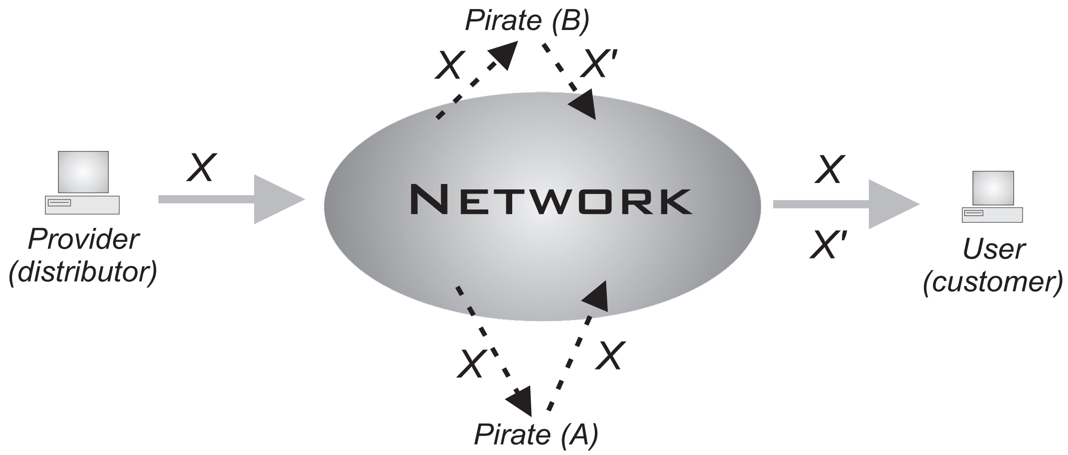

Figure 1: The basic model for digital product network distribution.

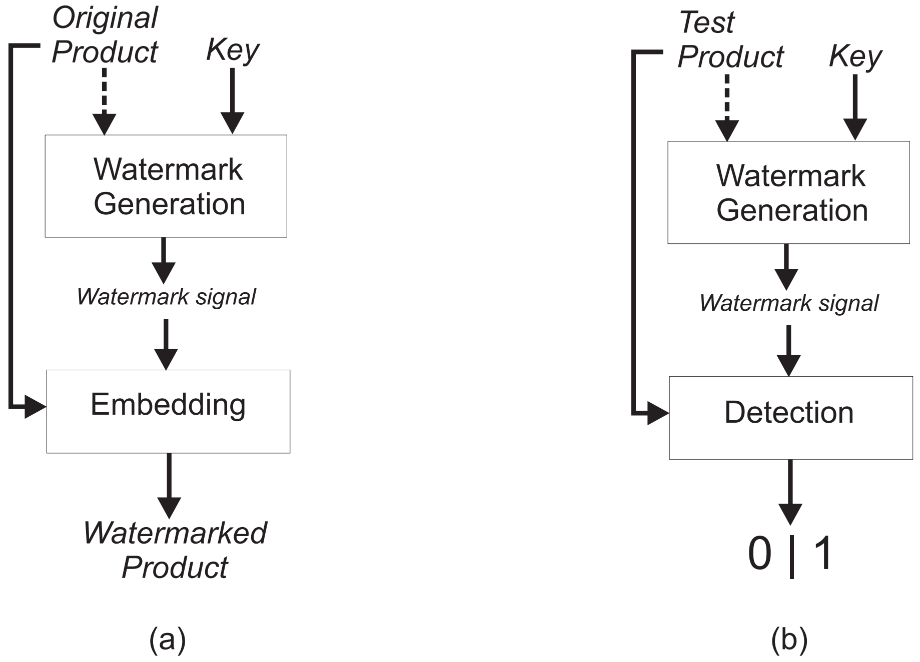

Figure 2: Watermarking algorithms for a) casting and b) detection.

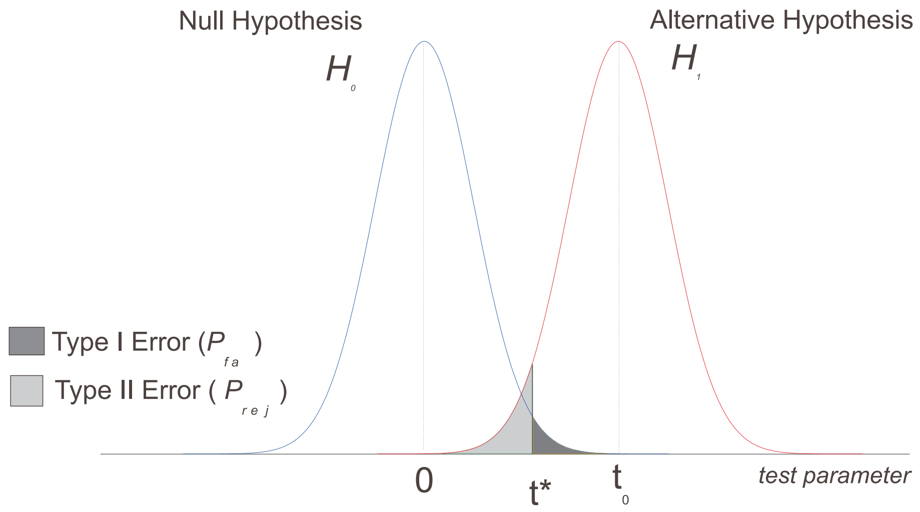

Figure 3: Hypothesis testing for watermark detection based on normal distributions.

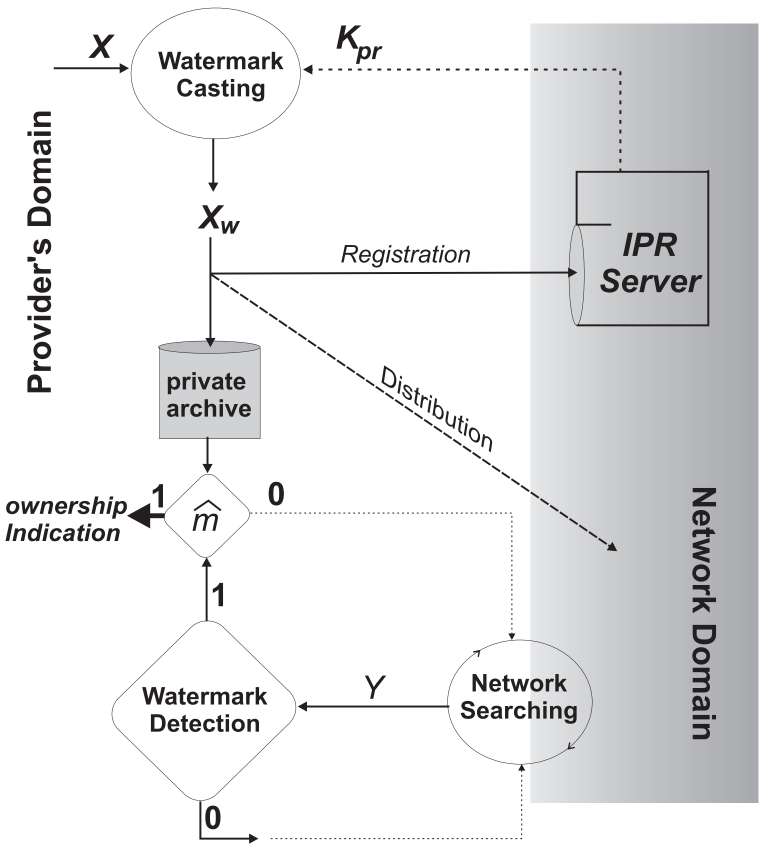

Figure 4: Copyright protection scheme based on watermarking and product registration.

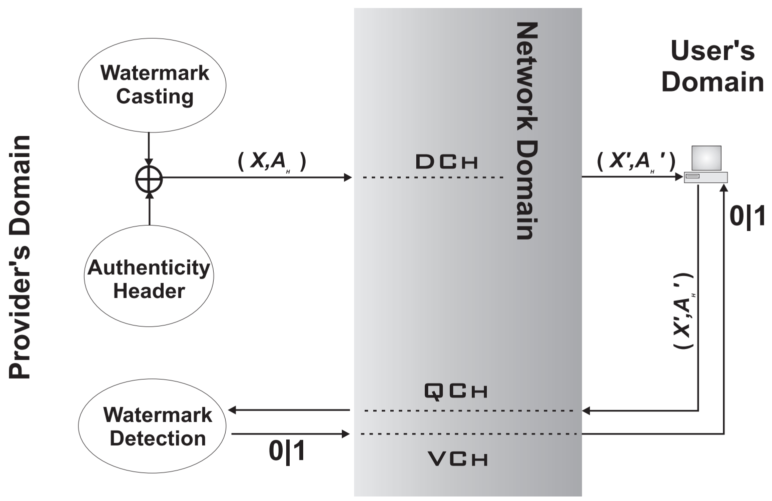

Figure 5: Content verification system based on private watermarking and public accessibility to authentication servers. DCh, QCh, VCh denote the distribution, verification request and reply channels respectively.