Digital watermarking techniques safeguard electronic content.

making a MARK

By
Mahalingam Ramkumar
and Nasir Memon,
Polytechnic University

An important issue that continues to arise in the use of digital content on the Internet is the protection of the rights of content owners. One approach to developing content-protection mechanisms is digital watermarking. A digital watermark is essentially an auxiliary signal embedded in the content in a process known as data hiding. The mechanism for embedding and extracting the auxiliary signal uses a secret key known only to the transmitter and receiver.[1] The process must allow access by authorized users while foiling any active adversary who would try to distort, remove, or even forge an auxiliary signal (the digital watermark).

A variety of signal-processing tools and algorithms can be applied to digital watermarking. Such algorithms are based on aspects of the human visual system, properties of signal transforms (e.g., Fourier and discrete cosine transforms), noise characteristics, properties of various signal-processing approaches, and so on. Although issues such as visual quality, robustness, and real-time constraints can be accommodated, it is still not clear whether all the desirable properties for digital watermarking can be achieved by any single algorithm. In most cases these properties have an inherent tradeoff, and typically the application dictates the optimal balance between competing properties.

Cryptographic issues lie at the core of many applications of information hiding, but unfortunately, they have received little attention. It is often assumed that simply using appropriate cryptographic primitives such as encryption, time-stamps, digital signatures, and hash functions would result in secure digital-watermarking applications. We believe this is far from the truth; in fact, the design of secure digital-watermarking techniques requires an intricate blend of cryptography, information theory, and signal processing.[2]

types of watermarks

In a conventional communications scenario, an information signal modulates a carrier signal. In data hiding, the auxiliary signal (the digital watermark) that is inextricably tied to the information signal modulates the latter. From the point of view of a communications engineer, data hiding can be seen as modulation of the information signal by the auxiliary signal. Extraction of the auxiliary signal from the information signal does not depend on how the information signal is transmitted, as long as the information signal is recovered with reasonable fidelity at the receiver.

The fundamental difference between a conventional communication scenario and data hiding is that while modulation of the carrier signal by the information signal typically changes the carrier signal drastically, modulation of the information signal by the auxiliary signal should only introduce imperceptible changes to the information signal. Let *c(t)* be some information signal and the auxiliary *s(t)* signal. A modulator *E* yields the composite signal $\hat{c}(t)$

$$\hat{c}(t) = E(c(t), s(t)),\ for\ d(c(t), \hat{c}(t)) \leq \varepsilon \qquad [1]$$

where *d* (., .) is a suitable distortion metric for the information signal, and ε is a measure of the permitted distortion of the information signal (see figure 1).

At the receiver we typically have a noisy version $\tilde{c}(t) = \hat{c}(t) + n(t)$ of the composite information signal $\hat{c}(t)$, where *n(t)* represents noise from the transmission channel. The receiver should be able to obtain an estimate $\tilde{s}(t)$ of the auxiliary signal *s(t)*. The goal is to make $\tilde{s}(t)$ as faithful as possible to *s*(*t*), in the presence of channel noise.

Watermarking techniques that do not require the original information signal image during the extraction process are called oblivious (or public or blind) watermarking techniques. Watermarking schemes can be classified as either robust or fragile. Robust watermarks are often used to prove ownership claims and so are generally designed to withstand common signal-processing tasks such as compression, cropping, scaling, filtering, contrast enhancement, printing/scanning, etc., in addition to malicious attacks aimed at removing or forging the watermark. In contrast, fragile watermarks are designed to detect and localize small changes to the image data.

There are three basic types of robust watermarks:

Type 1—Additive watermarks

Recall that data hiding can be seen as modulation of an information signal by an auxiliary signal. An obvious way to perform this modulation, as is done in Type I methods, is to add the auxiliary signal to the information signal. This approach is not a very efficient way of communicating the auxiliary signal, however. Typically the information signal has much larger amplitude than the auxiliary signal as we have to satisfy the distortion constraint. As far as the detector of the auxiliary signal is concerned, the information signal is noise. Even if there is no real noise in the channel, it is still very difficult to detect the low-power auxiliary signal buried in a substantially larger information signal.

However, if the original information signal *c*(*t*) is available at the receiver (non-oblivious watermarking), the "noise" due to *c*(*t*) can be completely eliminated. Under such a scenario, Type I is optimal. Type I is also optimal when *c*(*t*) is not available at the receiver, for very low signal-to-noise ratios (SNRs) (SNR → 0, or the channel noise *n*(*t*) is substantially stronger than *c*(*t*)).

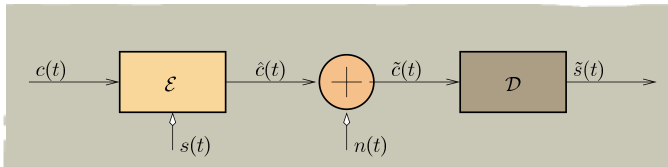

Figure 1 In data hiding, the information signal *c(t)* is modulated by an auxiliary signal *s(t)* to yield $\hat{c}(t)$. *E* is the modulator. During transmission, noise *n(t)* corrupts the signal $\hat{c}(t)$, to yield $\tilde{c}(t)$. The decoder *D* gives an estimated value of the auxiliary signal $\hat{c}(t)$.

Type II—Quantization-based watermarks

The drawback of Type I methods is that they consider *c(t)* as noise, which is not true. By definition, noise is something about which we have no information. Clearly, in this case we do have some information about *c(t)* as the detector captures $\tilde{c}(t)$, which is not very different from *c(t)*, assuming the channel noise is within reasonable limits. Methods that leverage this fact are called Type II methods and can be characterized by embedder *E* and detector *D*, which are exact inverses. Mathematically,

$$\hat{c}(t) = E\ \ (c(t), s(t))\ and\ s(t) = D\ \ (\hat{c}(t)) \qquad [2]$$

Another interesting characteristic of Type II methods is that it is impossible to recover the original *c(t)* from $\hat{c}(t)$, as compared to Type I methods, for which *c(t)* can be recovered as $\hat{c}(t) - s(t)$.

Most Type II methods proposed in the literature use some form of quantizer (with step size Δ) to implement *E* and *D*. For a quantizer with step size Δ, the distortion (ε) introduced by the embedder is a function of Δ. As long as the power of the channel noise is substantially less than the distortion ε, Type II methods perform reasonably well. However, as channel noise becomes comparable to ε, their performance deteriorates rapidly.

Type III

Though the choice of invertible *E*, *D* for Type II watermarking is obvious when there is no noise in the channel, it is neither intuitive nor obvious that *E*, *D* should be invertible for any finite channel noise.

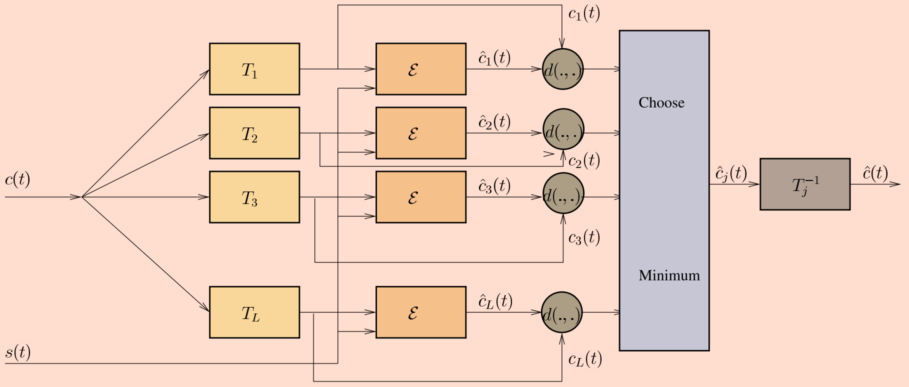

Figure 2 Embedding in extended Type III involves embedding *s(t)* in L homomorphisms $c_i(t), 1 \leq i \leq L$ of *c(t)*. The homomorphism that yields the minimum embedding distortion (homomorphism j in the figure) is inverted to obtain $\hat{c}(t)$,the composite signal. The inputs to the block labeled "Choose Minimum" are $d(c_i(t), \hat{c}_i(t))$ $1 \leq i \leq L$.

It is intuitive that we should increase step size Δ as channel noise increases. However, we have to do this without increasing distortion. We compensate for distortion by adding back the negative of some part of the introduced distortion to the composite signal. This yields Type III embedders, which are actually a composite of a Type II embedder and a truncation operator (see figure 2).

While Type II embedders are characterized by a single parameter Δ, Type III embedders are characterized by an additional parameter *k*, which determines the extent of compensation.[3,4] Type II embedders may be seen as a special case of the more generic Type III embedders, with no compensation. What is more interesting is that Type I embedders are actually Type III embedders with Δ→∞. While Type I methods are ideal for very low SNRs, and Type II methods are ideal for high SNRs, Type III methods fall conveniently between Types I and II, depending on the SNR.

a new approach—extented Type III

Costa showed that the interfering noise due to the information signal *c*(*t*) can be completely eliminated for any channel noise (and thereby achieve theoretical capacity limit).[5] The primary disadvantage of Costa's formulation is that in order to achieve complete elimination of the noise due to *c*(*t*), one needs to employ "codebooks" of impractical sizes. Additionally, the decoder needs to know the channel noise power.[6]

Even though Type III methods share some similarities to Costa's formulation, they are conceptually very different. Our group has developed an extension of Type III watermarking that is closer to Costa's formulation. While they do not have the two disadvantages of Costa's formulation mentioned above, they do not achieve complete suppression of the noise due to *c*(*t*), and hence, capacity.

For data-hiding applications, an auxiliary message is typically a sequence of bits implemented as a symbol $1 \leq m \leq M$ that is mapped to auxiliary signal $s_m(t)$. The auxiliary signal is then embedded in the information signal. At the receiver, the auxiliary message symbol is extracted from $\tilde{s}_m(t)$, the noisy version of $s_m(t)$. The mapping of each symbol *m* to a sequence $s_m(t)$ (and the inverse mapping from $\tilde{s}_m(t)$ to *m*) may be performed by a codebook lookup at the encoder/decoder.

Typically, there is a one-to-one correlation between the number of symbols used and the size of the codebook. In Costa's formulation, however, each symbol has multiple representations, resulting in a huge codebook shared by the encoder and the receiver. Having multiple representations for each symbol provides the embedder the freedom to choose an auxiliary signal that satisfies the distortion criterion for embedding. The receiver, however, does not know which representation of the symbol was actually embedded and therefore has to search the entire codebook for the best fit.

Costa's scheme is impractical for two reasons. First, each symbol can have on the order of 2^{40} representations, resulting in codebook size of $M \times 2^{40}$. Second, the codebook is designed for a particular noise power, or the receiver must know the channel noise power. A decoding error may result even if the channel noise is less than the noise the system (or codebook) was designed for.

Limitations notwithstanding, Costa's formulation achieves capacity, which Type III embedding does not. Therefore, an obvious question arises: If we increase the codebook complexity of a Type III system by raising codebook size by a factor *L* (say 2 to 20), will we achieve a reasonable capacity improvement over Type III systems, which use a single codebook of size *M*?

Let $c_1(t) = T_1(c(t))$ represent a homomorphism of *c*(*t*). One can derive other *L* – 1 homomorphisms $T_2 \ldots T_L$ such that

$$d(c_1(t); c_j(t)) >> \varepsilon \text{ for all } i \neq j \qquad [3]$$

where $c_i(t) = T_i(c(t)), 1 \leq i \leq L$. [4]

Ideally, the homomorphisms should be such that $d(c_1(t); c_j(t))$ should be maximized for all $i \neq j$. The auxiliary signal $s_m(t)$ (corresponding to the symbol m) modulates each $c_i(t)$, to yield $\hat{c}_i(t)$, $1 \leq i \leq L$. After calculation of the embedding distortion for each homomorphism, we choose the homomorphism that yields the minimum distortion. Let k be the index (of the homomorphism) that yields the minimum distortion. The composite signal is now obtained as (see figure 2)

$$\hat{c}(t) = T_k^{-1}(\hat{c}_k(t)) \qquad [6]$$

The receiver does not know which of the L homomorphisms was actually used by the encoder. Therefore the receiver tries every homomorphism (see figure 3). For every homomorphism, the receiver obtains an estimate of the embedded auxiliary signal $\hat{s}_i(t)$, $1 \leq i \leq L$, which is compared with the M codebook entries $s_1(t)$ $s_m(t)$ to obtain ρ_{ij}, the correlation of $\tilde{s}_i(t)$, $1 \leq i \leq L$ with $s_j(t)$, $1 \leq j \leq M$. The index j of the maximum of ρ_{ij} is deemed the estimate of the auxiliary symbol m.

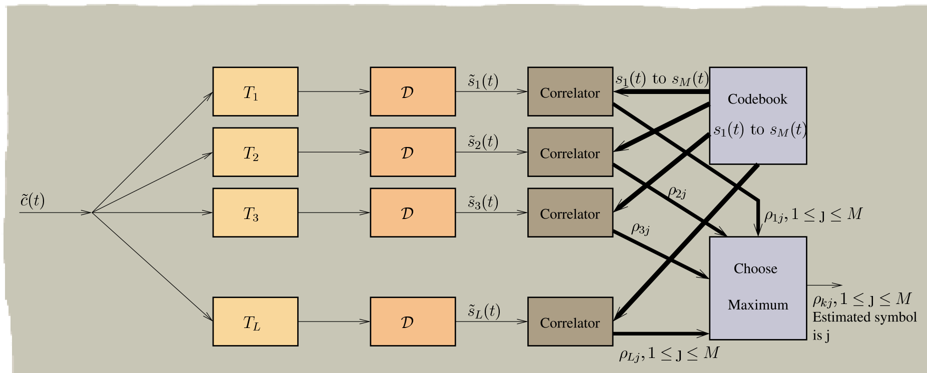

Figure 3 Detection in extended Type III. The detector D yields L estimates $\tilde{s}_i(t)$, for each homomorphism i, $1 \leq i \leq L$. Each correlator correlates $\tilde{s}_i( )$ with every entry $s_1(t)$ to $s_m(t)$ in the codebook, and outputs $\rho_{ij}, 1 \leq i \leq L, 1 \leq j \leq M$. The index j maximum of ρ_{ij} is deemed the estimate of m.

For Type II and Type III systems designed for nonzero SNR, capacities rapidly fall to zero even for finite noise power (unlike Type I methods where the fall in capacity is asymptotic). This behavior implies that, statistically, if a composite signal carrying some auxiliary signal $s(t)$ is subjected to some high but finite amount of noise, we cannot expect to find even a trace of $s(t)$ in the composite signal.

We use this to our advantage, however. Note that if a modulated information signal embedded with $s_m(t)$ in the k^{th} homomorphism reaches the receiver, statistically, the chances that $s_1(t)$ for $l \neq$ m could have been embedded in the other $L - 1$ homomorphisms in the receiver, interfering with detection process, is vanishingly small as long as $d(c_1(t); c_j(t)) >> \varepsilon$ for all $i \neq j$.

The freedom in choosing the homomorphism that yields minimum distortion for embedding conserves the power of the embedded signal (or permits one to increase the power of the embedded signal while keeping the distortion constant). An increase in signal power reduces the probability of false-alarm exponentially. The penalty paid is the small probability of additional decoding error introduced because the detector has to choose between $L \times M$ possible signals instead of M possible signals. The probability of false detection due to the $L - 1$ homomorphisms that were discarded is a constant, independent of the channel noise.

During detection, the discarded homomorphisms manifest themselves as a uniformly distributed noise between $\pm\Delta/4$. If p is the probability that the correlation between the uniformly distributed noise $\pm\Delta/4$ and some $s_m(t)$ for $1 \leq m \leq M$ is greater than a threshold, then the probability of false alarm, compared to Type III methods using a codebook with one entry per symbol is increased by a constant factor $p \times (L - 1) \times M$. An increase in signal power, however, reduces the probability of false alarm exponentially.

As long as the exponential decrease in false alarm probability due to increase in signal power is able to compensate for the constant term $p \times (L - 1) \times M$ in the total false-alarm probability, extended Type III can do better than Type III. Simulations indicate that extended Type III does indeed perform better than Type III. However, more quantitative analysis of the tradeoff between the two false alarm probabilities is a subject of our current research.

Mahalingam Ramkumar is a postdoctoral researcher and Nasir Memon is an associate professor at the Department of Computer and Information Science, Polytechnic University, Brooklyn, NY. For questions, contact Memon at phone: 718-260-3970; fax: 718-260-3609; e-mail: memon@poly.edu.

References

1. I. Cox, J. Kilian, et al., IEEE Transactions on Image Processing, vol. 6 (12) pp. 1673-1687, (1997).

2. S. Craver, N. Memon, et al., IEEE Journal on Selected Areas in Communications, 16(4):573-586, May 1998.

3. M.Ramkumar, A. Akansu, Proc. of SPIE 3845, pp. 55-65, Boston, MA (1999).

4. B. Chen and G. Wornell, Proc. of SPIE 3971, pp. 48-59, San Jose, CA (2000).

5. M. Costa, IEEE Trans. on Information Theory, IT-29, pp. 439-441, May 1983.

6. M. Ramkumar, A. Akansu, "On the Optimality of Signaling Methods for Oblivious Data Hiding and their Performance," submitted to the IEEE Transactions of Signal Processing, Jan. 2001.