Robustness of Public Key Watermarking Schemes

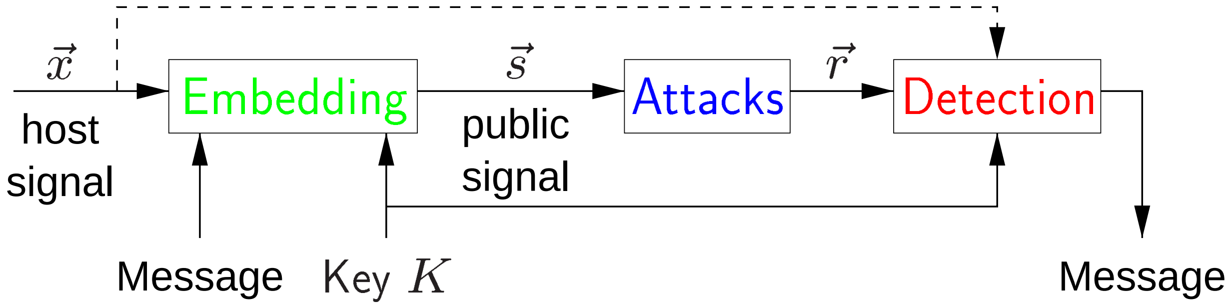
Joachim J. Eggers and Bernd Girod

Telecommunications Laboratory
University of Erlangen-Nuremberg
http://www.LNT.de

Overview

- Definition of public key watermarking
- Legendre sequence watermarking
  - Public detection based on Fourier invariance
  - Detection performance without attacks
  - Malicious attacks
- Quantization index modulation
  - Dithered scalar uniform quantization
  - 2D lattice quantization

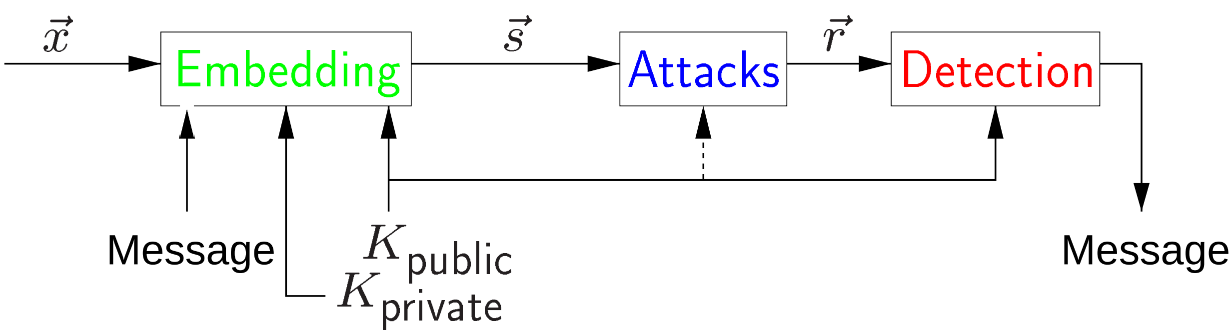
Private Key Watermarking



- Detection
  - Needs the key K
  - May need the host signal $\vec{x}$

    otherwise: blind detection
- Attacks can fully remove watermark when K is known

Public Key Watermarking

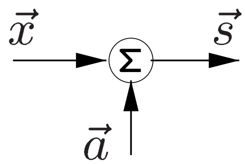


- Public detection
  - Cannot use the host signal $\vec{x}$
  - Need only the public key K_{public}
- Attacks should not work even with K_{public}

Legendre Watermarks

- Proposal by van Schyndel (ICMCS 99, Florence)

- Legendre sequence: $a_0 = 0, \quad a_n = e^{j\frac{2\pi r}{N-1}\text{ind}_g n}$



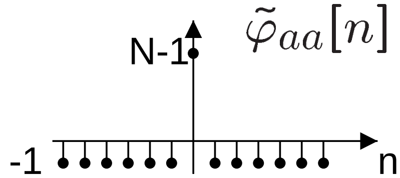
- Embedding: add Legendre sequence $\vec{a}$
  - More complex embedding schemes are possible
  - Additive scheme sufficient for analysis of detection performance

- Keep Legendre sequence $\vec{a}$ secret

Legendre Watermark Detection

- Periodic auto-correlation of Legendre sequences

$$\tilde{\varphi}_{aa}[n] = \sum_{m=0}^{N-1} a[m]a^{\star}[m + n(\text{mod } n)]$$



- Fourier invariance property $\quad G_{\mathcal{DFT}}\vec{a} = A_1\vec{a}^{\star}$

- Detection principle

$$c_p = \underbrace{(\vec{x} + \vec{a})^T}_{\vec{s}^T} G_{\mathcal{DFT}} \underbrace{(\vec{x} + \vec{a})}_{\vec{s}} / N \approx A_1$$

- Detection possible without knowing $\vec{a}$

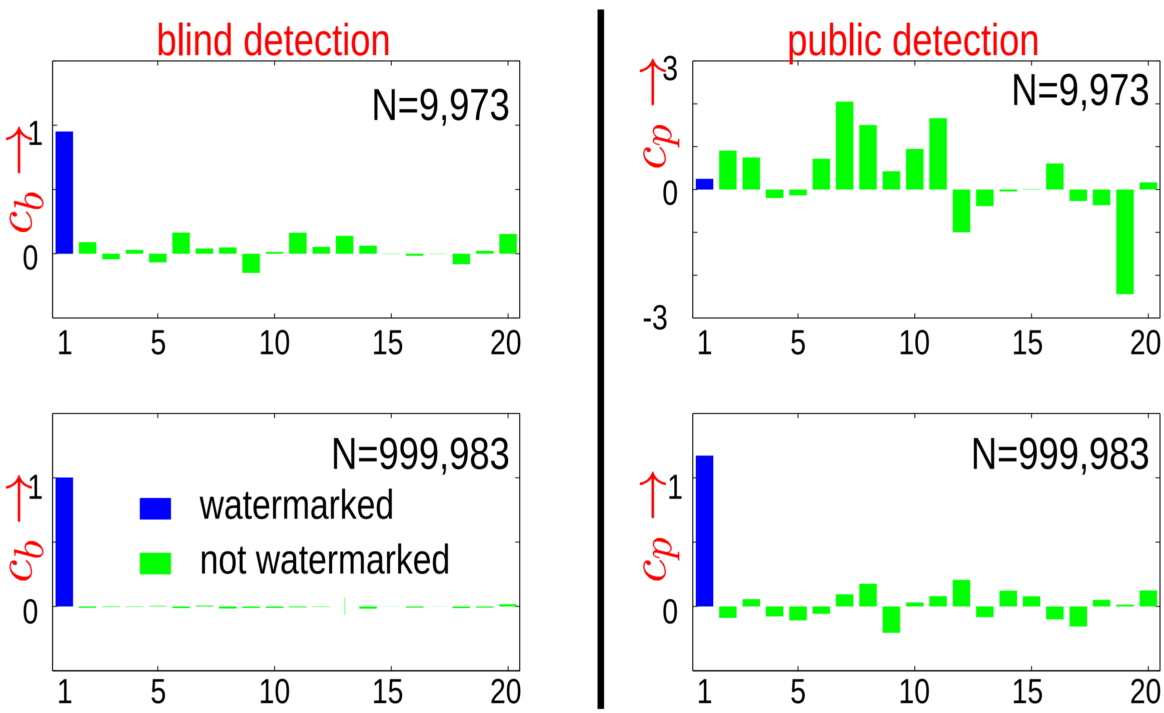
Detection Robustness

- Rough estimation
  - Assume i.i.d. host signal samples x
  - Low power watermark $\gamma = \frac{||\vec{x}||}{||\vec{a}||} >> 1$

$$\Rightarrow \text{robustness } \rho = \frac{\mathrm{E}\left\{c_p | \vec{r} \text{ is watermarked}\right\}}{\mathrm{Std}\{c_p\}} \sim \frac{\sqrt{N}}{\gamma^2}$$

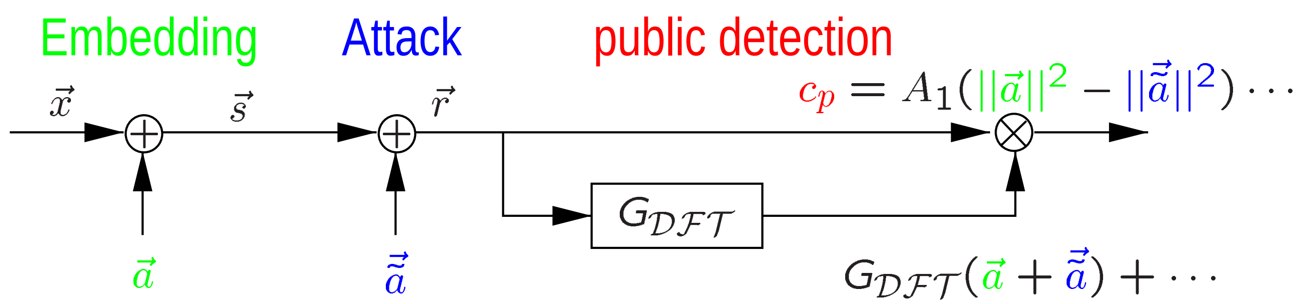
- Necessary signal length $N_{\text{public}} \approx \gamma^2 N_{\text{private}}$
- Example: $\gamma = 10$ goal: robustness $\rho \approx 10$

Detection Results



Malicious Attacks

- Remove watermark
  - Only $N-2$ Legendre sequences of length N
    $\Rightarrow$ exhaustive search feasible
- Confuse the watermark detector
  - Sequence $\vec{\tilde{a}}$ with $\quad G_{\mathcal{DFT}}\vec{\tilde{a}} = -A_1\vec{\tilde{a}}^{\star}$



Construction of Attack Sequence

- Generate random sequence $\vec{v}$

$$\vec{u} = \mathcal{R}e\left\{G_{\mathcal{DFT}} - A_1 I\right\}^{-1}\mathcal{I}m\left\{G_{\mathcal{DFT}} + A_1 I\right\}\vec{v}$$

$$\Rightarrow \vec{\tilde{a}} = \vec{u} + j\vec{v}$$

- Equation is singular for $A_1 \in \{\pm 1, \pm j\}$
  - $\{\pm 1, \pm j\}$ are eigenvalues of $G_{\mathcal{DFT}}$

  $\Rightarrow$ $\vec{\tilde{a}}$ can be constructed with eigenvectors of $G_{\mathcal{DFT}}$

$\Rightarrow$ Random sequence $\vec{\tilde{a}}$ can be found easily!

Properties of Legendre Watermarking Scheme

- Very long sequences necessary – even without attack
- Distortion penalty for confusion attack

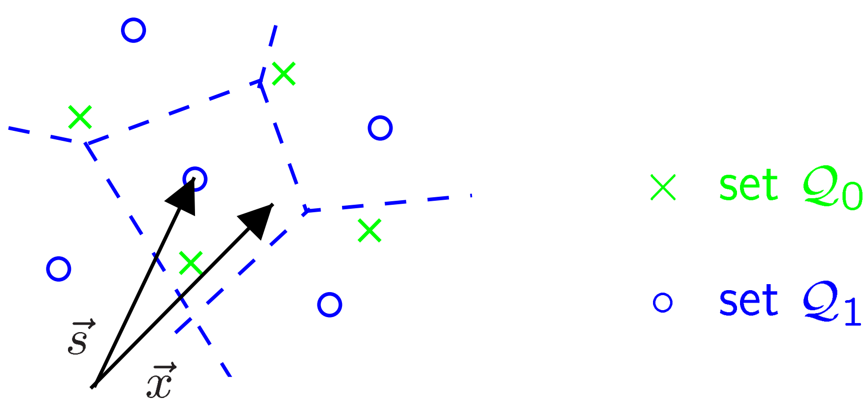
$$\frac{D_{\text{attack}}}{D_{\text{embedding}}} \approx 2 \equiv 3\text{dB}$$

- Overall rating:

⇒ Nice idea, but hardly practical!

Quantization Index Modulation

- Proposal by Chen & Wornell (1998/99)

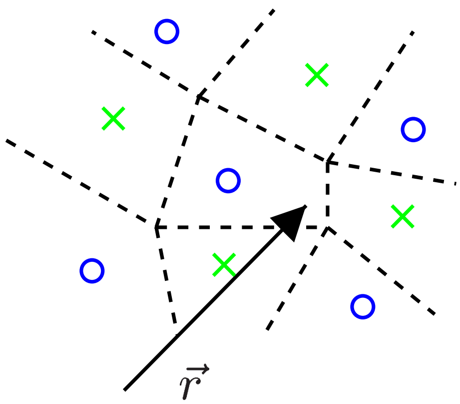


- Embedding distortion = quantization distortion

$$D_{\text{embedding}} = \mathrm{E}\left\{\frac{1}{N}||\vec{s} - \vec{x}||^2\right\}$$

Public Detection Principle

- Sets $\mathcal{Q}_0$ and $\mathcal{Q}_1$ are public



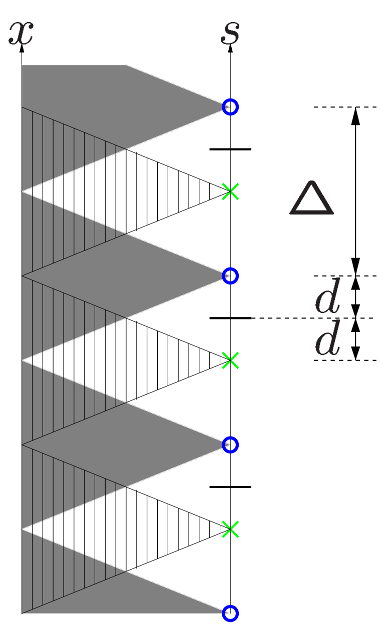
Quantize $\vec{r}$ to closest point in $\mathcal{Q}_0 \cup \mathcal{Q}_1$

- Determine watermark bit from quantizer set index

$\Rightarrow$ Watermark is publicly detectable

Dithered Uniform Quantizer

Watermark bits $\vec{b} \rightarrow$ channel coded bit sequence $\vec{z}$



- Embedding

$$s = \mathcal{Q}\{x + d\} - d$$

$$d \in \{\pm\Delta/4\}$$

$z = 0 \rightarrow d > 0 \quad s \in \mathcal{Q}_0$

$z = 1 \rightarrow d < 0 \quad s \in \mathcal{Q}_1$

- Fine quantization

$$\Rightarrow D_{\text{embedding}} = \frac{\Delta^2}{12}$$

Robustness of Scalar QIM (1)

- Chen & Wornell:

$$\frac{D_{\text{attack}}}{D_{\text{embedding}}} \geq 1 + \gamma_c \frac{3/4}{NR}$$

with

- signal length N
- rate R of watermark bits per signal sample
- γ_c denotes strength of channel code $\left(\gamma_c = d_H \frac{k_u}{k_c}\right)$

- Worst case?

Robustness of Scalar QIM (2)

- Malicious attack
  - Quantizer $\mathcal{Q}$ is public
  - Move signal points $\vec{s}$ on boundary between $\mathcal{Q}_0$ and $\mathcal{Q}_1$

  $\Rightarrow$ Public watermark detection is no longer possible

$$\frac{D_{\text{attack}}}{D_{\text{embedding}}} \geq 1 + \frac{3}{4} \equiv 2.43\text{dB}$$

- Distortion penalty too small to prevent attacks

Examples

with embedded 1D QIM watermark

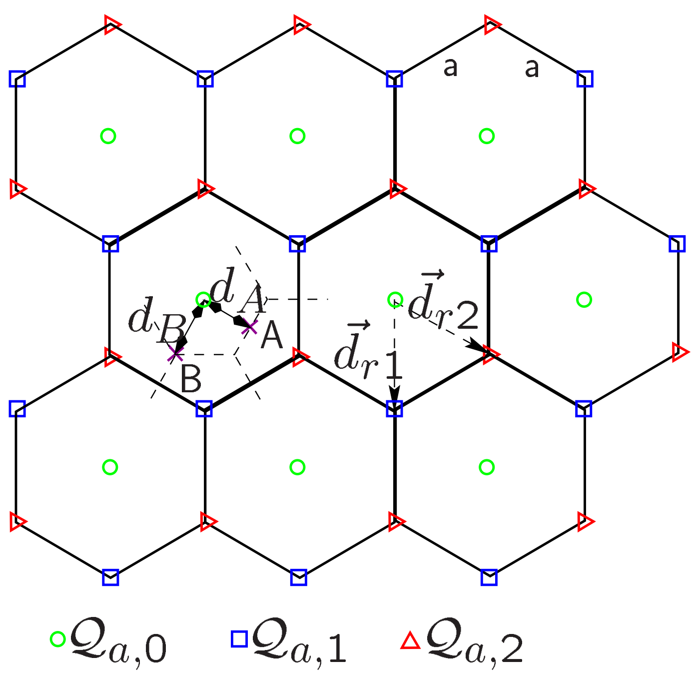
PSNR=40.16dB

after perfect attack

PSNR=37.75dB

$\frac{D_{\text{attack}}}{D_{\text{embedding}}} \equiv 2.41\text{dB}$

2-D QIM with Hexagonal Lattice



- Encode watermark bits using a ternary alphabet
- 2-D dither vectors $\vec{d}_{r1}$ and $\vec{d}_{r2}$
- Attack: move $\vec{s}$ to point A or point B with $d_A < d_B$

Robustness of 2D-QIM

- Attack A $\frac{D_{\text{attack}}}{D_{\text{embedding}}} \geq 1.6 \equiv 2.04\text{dB}$
  - Distortion penalty smaller than for 1D-QIM
  - Watermark information not completely destroyed
- Attack B $\frac{D_{\text{attack}}}{D_{\text{embedding}}} \geq 1.8 \equiv 2.55\text{dB}$
  - Distortion penalty larger than for 1D-QIM
  - Watermark information completely destroyed

$\Rightarrow$ 2D-QIM scheme is slightly more robust

Conclusion

- Public Legendre watermarking
  - Distortion penalty about 3 dB
  - Very long sequences are necessary
  - Scheme is not practical
- Quantization index modulation
  - Easy to implement even in practical applications
  - 1D-QIM distortion penalty of 2.43 dB
  - 2D-QIM distortion penalty of 2.55 dB
- Public scheme with sufficiently large distortion penalty?