Lecture 11. «Sophisticated» attacks on WM.

We considered before the following (“non-sophisticated”) attacks:

- randomization of LSB,
- randomization of quantization levels,
- additive noise.

The first and the second attack are able to remove WM completely from LSB and QIM based WM , respectively.

The third attack is inefficient if SS-based WM with the large length N of pseudo random sequence (PRS) is used.

Consider the following «sophisticated » attacks to which are vulnerable SS-based WM even with large N:

1. Filtering.
2. Subtraction of WM followed by estimation of WM embedded in CO.
3. Tandem of D/A and A/D transforms .
4. Compression/decompression.
5. Synchronization attacks.
6. Collusion attacks.
7. System attacks.

Consider these attacks one by one.

1. Filtering. (The idea is to reduce WM amplitude keeping good CO quality)
Attack model against 0-bit WM.

$$C'_W = (C_W * h)(n) + \varepsilon(n), n = 1, 2, \ldots N$$

where $C_W(n) = C(n) + \alpha\pi(n),\ \pi(n) \in \{+1, -1\},\ \pi(n) \in \text{i.i.d.}$

$\varepsilon(n)$ - additive noise $E\{\varepsilon(n)\} = 0,\ Var\{\varepsilon(n)\} = \sigma_\varepsilon^2$
$h(n)$ - attack filter pulse response
"*" - convolution symbol

Decoder

$\Lambda \geq \lambda \Rightarrow$ presence of WM (1)
$\Lambda < \lambda \Rightarrow$ absent of WM

$$\Lambda = \sum_{n=1}^{N} (C'_W(n) - (C * h)(n))(\pi * h)(n) \quad (2)$$

λ - some threshold
(Informed decoder under known attack filter.)

Evaluation of probabilities P_m *and* P_{fa} .

$$P_m = 1 - Q(\lambda' - \mu),\ P_{fa} = Q(\lambda'), \tag{3}$$

where $\lambda' = \dfrac{\lambda}{\alpha\sigma_\varepsilon \sqrt{\sum_{n=1}^{N} \left(\tilde{h}(k)\right)^2}}$, $\mu = \dfrac{\alpha}{\sigma_\varepsilon} \sqrt{\sum_{n=1}^{N} \left(\tilde{h}(k)\right)^2}$

$\tilde{h}(k)$ - attack filter frequency response.

Frequency response of attack filter as ideal LF filter:

$$\tilde{h}(k) = \begin{cases} 1, \text{ if } 0 \le k \le \dfrac{K_h}{2}, \\ 0, \text{ for else } k \end{cases} \tag{4}$$

where $K_h : 0 \le K_h \le N - 1$

Performance evaluation of LF filter attack[25]

For the same reliability as it was without filtering attack it is necessary to increase N, in $\left(\dfrac{K_h}{N}\right)^{-1}$ times.

Remark 1. Filtering attack is acceptable with such K_h, that maintains a good quality of CO (the last condition is verified by simulation).

Remark 2. Filtering attack can be reduced significantly if PRS is chosen not as «white» (i.i.d.-sequence), but as «colored» (correlated sequence).

But in turn for colored PRS one can improve filtering attack choosing additive attack noise as colored noise .

Simulation results under filtering and additive noise attacks [25]:

Image 1 «Fish boat»																	
Case O				Case A					Case B					Case C				
α	σ_ε	η	N	δ	α	σ_ε	η	N	δ	α	σ_ε	η	N	δ	α	σ_ε	η	N
5	5	2,000	38	0,75	5	6	1,787	101	0,75	8	6	1,796	68	0,75	8	11	2,006	124
				1,00	5	7	2,124	220	1,00	12	7	2,159	87	1,00	12	17	2,087	169
				1,25	5	7	2,062	320	1,25	16	7	1,927	82	1,25	16	22	1,942	175
Image 2 «Lena»																		
Case O				Case A					Case B					Case C				
α	σ_ε	η	N	δ	α	σ_ε	η	N	δ	α	σ_ε	η	N	δ	α	σ_ε	η	N
7	6	1,735	29	0,75	7	8	1,653	89	0,75	12	9	1,796	70	0,75	12	15	1,738	102
				1,00	7	9	1,817	190	1,00	17	9	1,792	77	1,00	17	22	1,755	143
				1,25	4	9	1,755	281	1,25	22	9	1,691	73	1,25	22	29	1,798	156

Table 1. The minimal values of PRS lengths N, which provide $P_m = P_{fa} = 10^{-3}$ given some parameters and the following types of attacks:

O – no filtering attack

A – filtering and additive white noise attack for white PRS

B – filtering and additive white noise attack for colored PRS –

C – filtering and additive colored noise attack for colored PRS

δ – two demensional filter parameter

$$h(k) = \frac{1}{A} e^{\frac{-n_1^2+n_2^2}{2\delta^2}}, \quad A = \sum_{n=1}^{N} e^{\frac{-n_1^2+n_2^2}{2\delta^2}}, \eta = \frac{\eta_w}{\eta_a}$$

Remark 1. Simulation showed that CO quality keeps good if filter parameter δ is less then 1.

Remark 2. The use of colored PRS is equivalently to embedding PRS in intermediate frequency domain (see Lecture 9), whereas the use of colored additive attack noise instead of white additive noise is equivalently to removal of noise frequency domain in which is not embedded PRS frequency component .

Remark 3. In place of colored PRS one can use so called *tile-based WM (TWM)*, where $\pi = const$ on m adjacent samples. Then «m» can be chosen in such a way that filtering attack is useless but the embedding rate reduces in m times.

2. **Estimation attack.**

WM embedding.

$$C_W(n) = C(n) + \alpha(-1)^b \pi(n),\ n = 1, 2, \ldots N \quad (5)$$

where $\pi(n)$ – takes the same values at «m» $= \dfrac{N}{N_0}$ adjacent samples.

$$C'_W(n) = C_W(n) - \alpha' \pi \tilde{\omega}(n),\ n = 1, 2, \ldots N \quad (6)$$

where α' – some coefficient (in general case $\alpha' \neq \alpha$)

$\tilde{\omega}(n)$ – some estimate of $(-1)^b \pi(n)$.

Correlation decoder.

$$b = \begin{cases} 0, \text{ if } \Lambda \geq 0 \\ 1, \text{ if } \Lambda < 0 \end{cases}, \quad (7)$$

a) Blind decoder:

$$\Lambda = \sum_{n=1}^{N} (C'_W(n) - m_c)\pi(n),\ m_c = E\{C(n)\} \quad (8)$$

b) Informed decoder:

$$\Lambda = \sum_{n=1}^{N} (C'_W(n) - C(n))\pi(n)$$

Remark. The use of PRS $\pi(n)$ that is constant at m samples allows to improve robustness to filtering attack and also to some geometric attacks. Such PRS is called usually the *block repetition code* of the length m .

Calculation of the error probability [26].

$$P \le Q\left(\frac{N\left(\alpha - \alpha'\left(1-2P_{es}\right)\right)}{\sqrt{N_0\left(A + 4\left(\alpha'\right)^2 m^2 P_{es}\left(1-P_{es}\right)\right)}} \right) \quad \text{(for blind decoder)} \qquad (10)$$

$$P \le Q\left(\frac{\left(\alpha - \alpha'\left(1-2P_{es}\right)\right)}{\sqrt{4\left(\alpha'\right)^2 P_{es}\left(1-P_{es}\right)}} \sqrt{N_0} \right) \quad \text{(for informed decoder)} \qquad (11)$$

where $A = Var\left\{ \sum_{n=1}^{N} C(n) \right\}$ (12)

P_{es} – the probability of incorrect estimation $\pi(n)$ at «m» adjacent samples, e.g.

$P_{es} = \Pr\left\{ \tilde{\omega}(n) \ne (-1)^b \pi(n) \right\}$.

Signal-to –noise ratio after attack:

$$\eta_a = \frac{\sigma_c^2}{Var\left\{\alpha\pi(n) - \alpha'\tilde{\omega}(n)\right\}} = \frac{\sigma_c^2}{\alpha^2 - 2\alpha\alpha'\left(1-2P_{es}\right) + \left(\alpha'\right)^2} \qquad (13)$$

If $C(n) = const$ at «m» adjacent samples, then $A = m^2\sigma_c^2$ (14)

If $C(n)$ are independent on «m» adjacent samples , then $A = m\sigma_c^2$ (15)

In general case

$$m\sigma_c^2 \le A \le m^2\sigma_c^2 \quad (16)$$

(It is assumed that $C(n)$ are independent between blocks of the length m).

Comparison of estimation attack and additive noise attack.

$$C'_W(n) = C_W(n) + \varepsilon(n),\ Var\{\varepsilon(n)\} = \sigma_\varepsilon^2,\ E\{\varepsilon(n)\} = 0 \quad (17)$$

$$P' \approx Q\left(\sqrt{\frac{\alpha^2 N^2}{(A_c + A_\varepsilon)N_0}}\right),\ \text{where}\ A_\varepsilon = Var\left\{\sum_{n=1}^{N}\varepsilon(n)\right\} \quad (18)$$

Additive noise optimization for TBW: $\varepsilon(n) = \varepsilon$ for m adjacent samples.

Then

$$P' = Q\left(\sqrt{\frac{\alpha^2 N^2}{(A_c + m^2\sigma_\varepsilon^2)N_0}}\right) \quad (19)$$

Signal-to-noise ratio after attack:

$$\eta'_a = \frac{\sigma_c^2}{\alpha^2 + \sigma_\varepsilon^2} \quad (20)$$

Optimization of parameter α'*, that results in maximization of* P*, given parameters* $N_0,\ m,\ \alpha,\ \sigma_c^2,\ P_{es},\ \eta$.

We use numerical calculations by formulas (10), (11), (13), (19), (20) to solve this problem and compare estimation attack with additive noise attack.

Example. Let $A_c = m^2\sigma_c^2,\ \alpha = 5,\ m = 10,\ N_0 = 950,\ \sigma_c = 50.$

The results of optimization procedure (*P,α`*) given different η and P_{es} are shown in Table below.

Estimation attack					Additive noise attack	
$\eta \setminus P_{es}$	0.499	0.4	0.3	0.2		σ_ε^2
1		0.00230; 2	0.01800; 4	0.19500; 6		
0.95	0.00107; 1	0.00230; 2	0.01800; 4	0.19500; 6	0.00103	1
0.9	0.00107; 1	0.00230; 2	0.01800; 4	0.19500; 6	0.00103	1
0.75	0.00118; 3	0.00492; 4	0.03300; 5	0.31200; 7	0.00103	4
0.5	0.00132; 5	0.01000; 6	0.08900; 7	0.45100; 8	0.00108	25

Conclusion. Even for large probability of incorrect estimation of PRS bits ($P_{es} \approx 0{,}2$) estimation attack is better than additive noise attack.

Estimation of PRS bits.

Correlation decoder:

$$\tilde{\pi}(n)=\begin{cases}1, \text{ if } \Lambda \geq 0\\ -1, \text{ if } \Lambda < 0\end{cases} \quad (21)$$

where $\Lambda=\sum_{n=1}^{m}\left(C_{\omega}(n)-m_c\right),\ m_c=\frac{1}{N}\sum_{n=1}^{N}C_{\omega}(n)$ (22)

Improved correlation decoder:

$$\Lambda'=\sum_{n=1}^{m}\left(C_{\omega}(n)-\bar{m}_c\right),\ \text{where}\ \bar{m}_c=\frac{1}{N_0}\sum_{n=1}^{N_0}C_{\omega}(n),\ \text{where}\ N_0<N$$

Remark.Further improvement of estimation is an implementation of Wiener filter[26].

Decoder of «jumps»:

$$\tilde{\pi}(n+1,n+m)=\begin{cases}1, \text{ if } \Lambda \geq \lambda\\ -1, \text{ if } \Lambda < -\lambda\\ \tilde{\pi}(n-m+1,n), \text{ if } |\Lambda|<\lambda\end{cases} \quad (23)$$

where $\Lambda=C_{\omega}(n+1)-C_{\omega}(n)$

Theoretical formulas to find P_{es} are not explicit and therefore it is commonly to find this value by simulation.

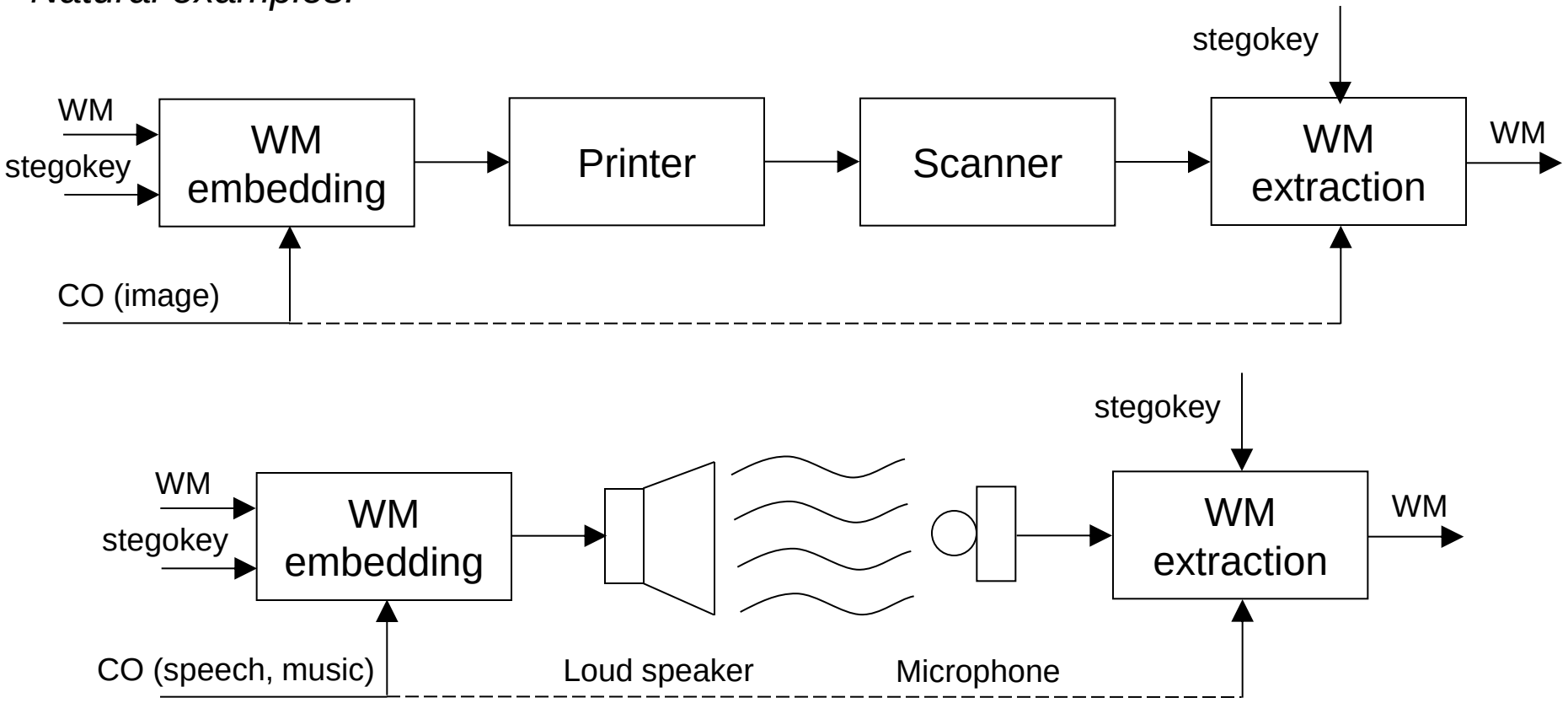
Conclusion on simulation results. Decoder of jumps is superior to other decoders.

Example.

$$\alpha \geq 3,\ m \leq 10,\ \sigma_c \approx 50.\ \text{Then}\ P_{es}=0.3.$$

3. Transforms D/A and A/D.

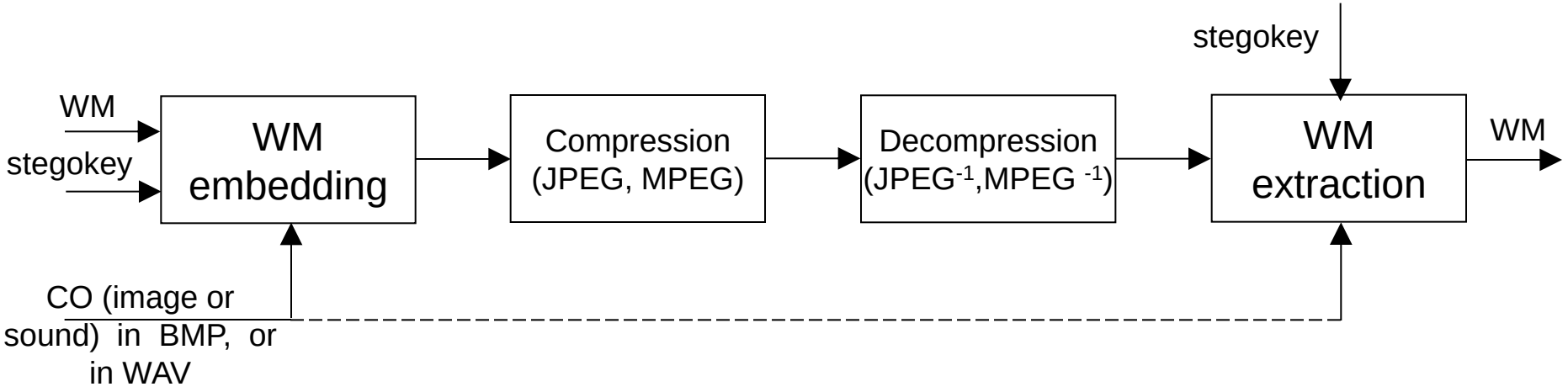
Natural examples:



Robustness of WM can be provided by embedding WM in DCT or DWT domains with the use either SS or FH signals .

4. Compression/decompression transforms.

Natural examples:



Robustness of WM can be provided if SS or FH signals are used in DCT or DFT domains with appropriated matching with JPEG or MPEG formats.