Lecture 12 (Continuation). System attacks.

Unauthorized WM embedding

1. Brute force attack

With the knowledge of SG algorithm and stegokey simply embed WM in new CO .

Protection: Before WM embedding to perform authentication by digital signature with the use of secret key unknown for an attacker.

2*. WM copying*

Attacker copies WM from some CO and embed it in another CO that is needed to be watermarked .

How to copy WM :

$$C_{w1}(n) = C_1(n) + w(n)$$

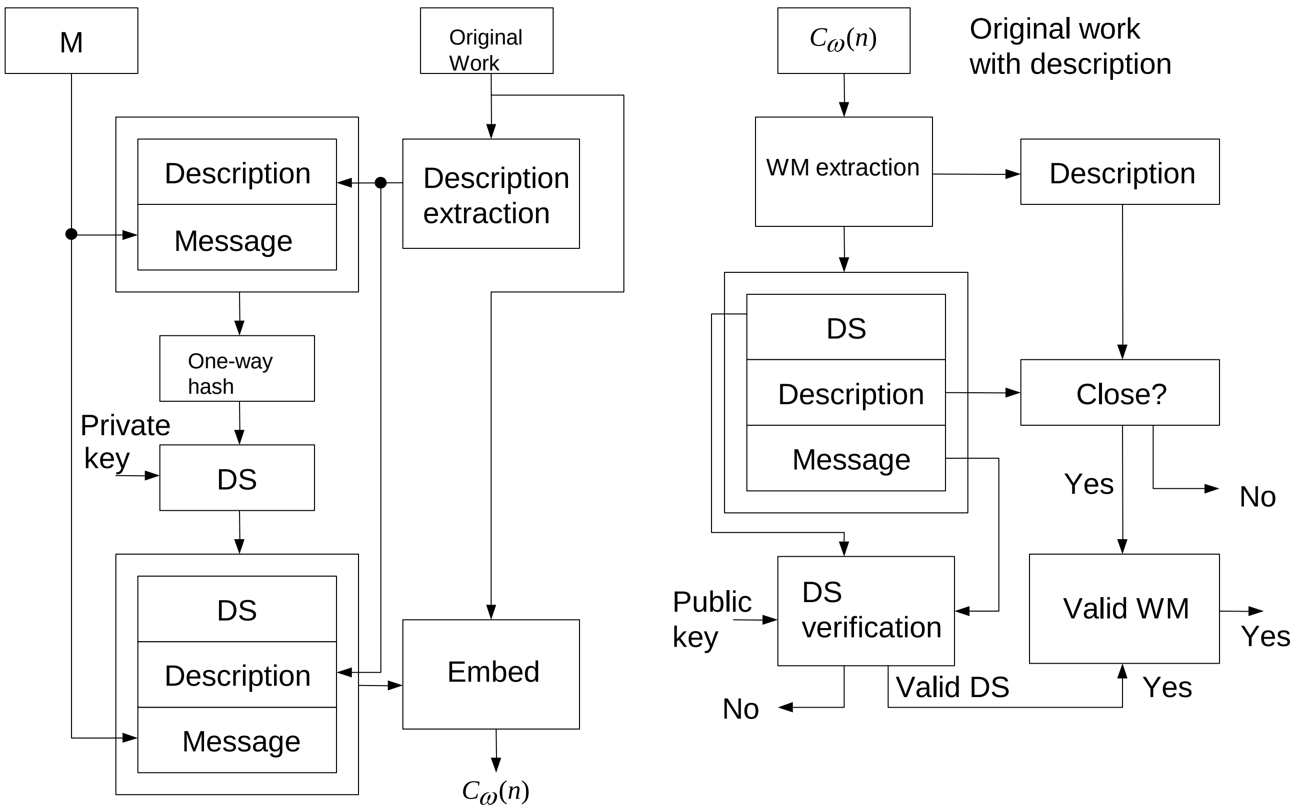
$$C_{w2}(n) = C_2(n) + \tilde{w}(n),$$

where $\tilde{w}(n)$ - is estimation of $w(n)$ (see in Lecture 11 “Estimation attack”)

The simplest case: If it is LSB-based WM then one can copy these LSB to LSB of another CO.

Protection against copy attack

The main idea: linking WM to CO.



a) WM embedding b) WM extraction

*(*M – the WM message; OWF one way hash function; a description of CO is based on information unlikely to change, such as the lowest-frequency component ; comparison means an inexact comparison , for example a calculation of correlation between the embedded and a description of the received CO and comparison it against a threshold .

The feature of method: Valid WM is confirmed even under slight distortion of CO.

3. *Ambiguity attack*

Ambiguity attacks create the appearance that a WM has been embedded in CO when in fact no such embedding has taken place. An adversary can use this attack to claim ownership of distributed CO.

Attack's technique:
a) With the use of informed decoder

$$\left.\begin{aligned} C_w(n) &= C(n) + w(n) \\ C'(n) &= C_w(n) - w'(n) \end{aligned}\right\},$$

w(n) – original WM
w`(n) – fake WM
C`(n) – fake CO (close to original *C(n)*)

For an attacker with informed decoder we get:

$$\Lambda_a = \sum_{n=1}^{N} (C(n) + w(n) - C(n) - w(n) + w'(n))w'(n) = \sum_{n=1}^{N} w'(n)w'(n) = N\alpha^2 > \lambda,$$

where λ – threshold. In fact:

$$C(n) + w(n) - C'(n) = C(n) + w(n) - (C_w(n) - w'(n)) = C(n) + w(n) - (C(n) + w(n) - w'(n)) =$$
$$= C(n) + w(n) - C(n) - w(n) + w'(n) = w'(n)$$

b) With the use of blind decoder

The main idea: To construct a fake WM $w'(n)$, that appear to be a noise signal but has a high correlation with distributed Work $C_w(n)$.

Variant of solution

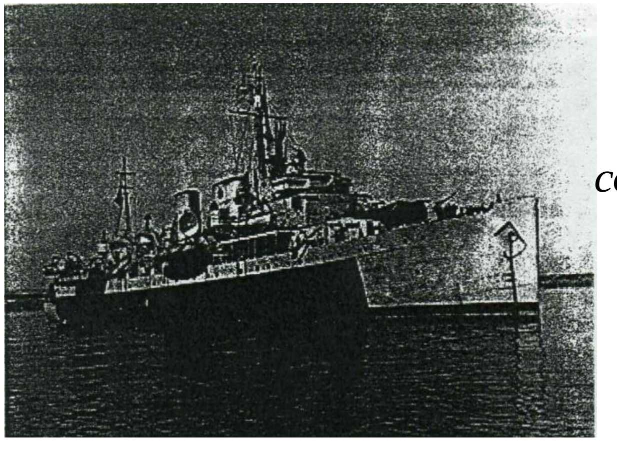
$1.C_w(n) = C(n) + w(n), n = 1,2...N$ (original WM-ed message)

$2.C'_w(n) = C(n) + \varepsilon(n), \varepsilon(n) -$ small noise

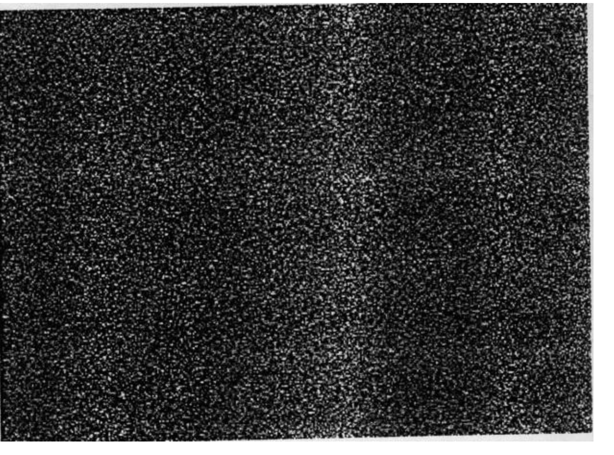
$3.DCT(C'_w(n))$

$4.DCT' = Rand\,|\,DCT(C'_w(n))\,| -$ a randomization of $DCT(C'_w(n))$ amplitude

$5.IDCT(DCT') = w'(n) -$ is looking as noise but has a large correlation coefficient with $C_w(n)$



$cor.coef.((C_w(n), w'(n)) = 0.968$



a) The original image $C_w(n)$ b) Fake WM $w'(n)$

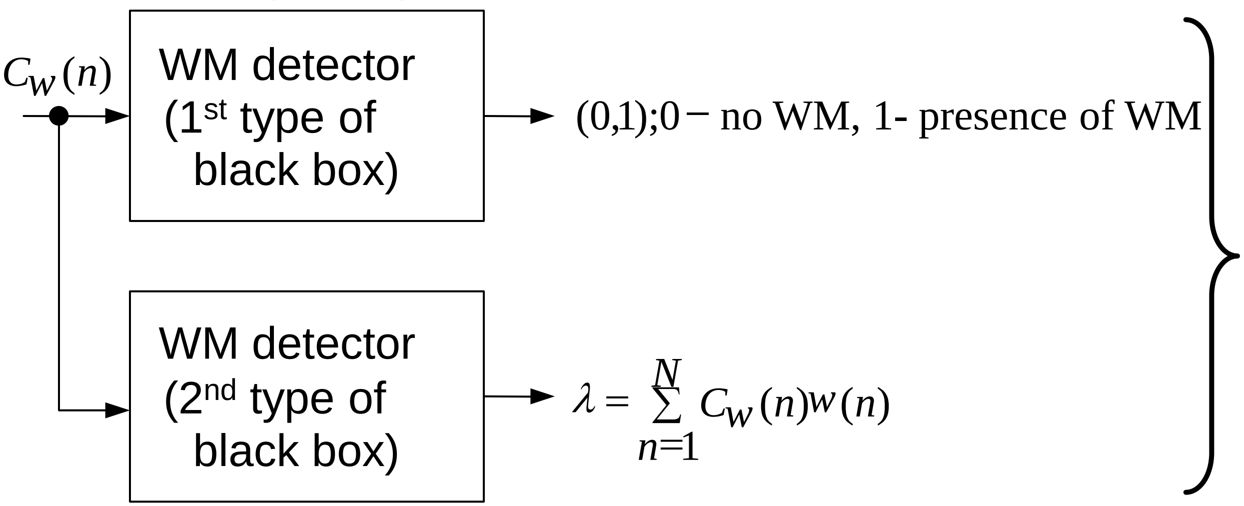
6. $C^{'}(n) = C_W(n) + \alpha w^{'}(n), \alpha \leq 1$

$\sum_{n=1}^{N} C^{'}(n) w^{'}(n) > \lambda$ and arises a dispute – both legal and illegal user can make equal claims of ownership

Protection against ambiguity attack:

-use another embedding technique rather than additive embedding; in particular with the use of one-way hash functions [19].

4. Sensitivity analysis attacks



Attacker changes $C_W(n)$ to $C_W^{'}(n)$ until WM is not already detected