

---

# Security & Watermarking

## The example of Copy Protection

**Teddy FURON**

**teddy.furon@inria.fr**



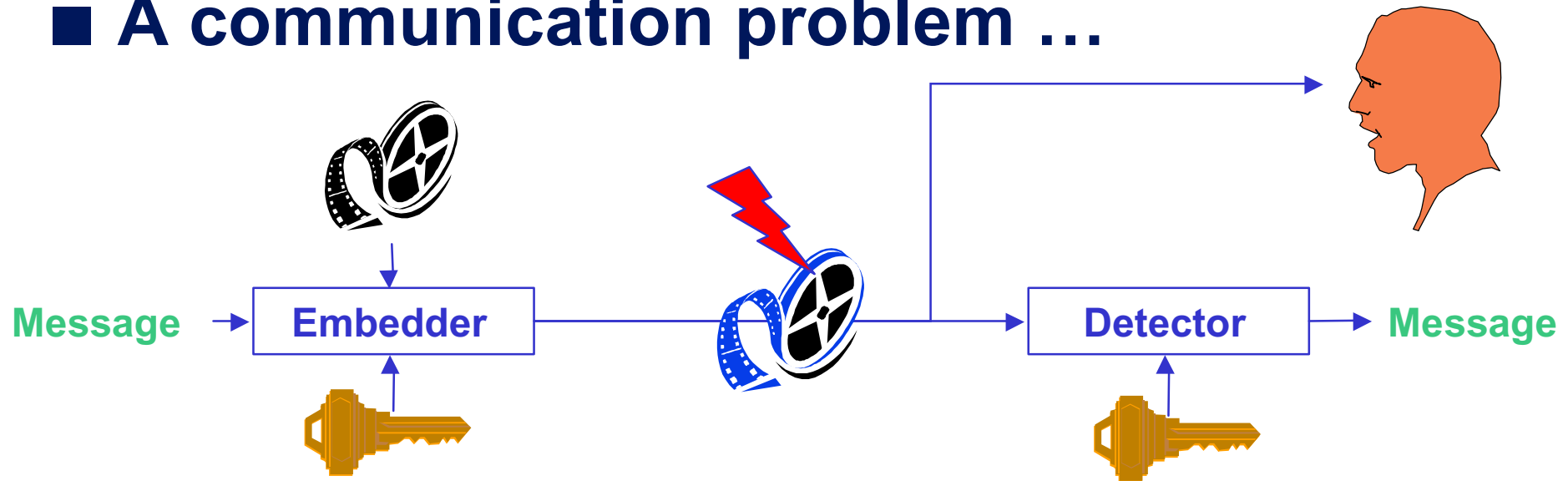
# Presentation outline

---

- **Short Overview of Watermarking.**
- **The Copy Protection Framework.**
- **Watermarking & Security**
- **Asymmetric Watermarking Scheme**

# Watermarking Overview

## ■ A communication problem ...



... under constraints:  
non perceptibility (*filigrane*)  
robustness (*tatouage*)  
capacity

# Region of Interest

---

## ■ Non perceptibility

- Use of Human Visual System to control locally the watermarking power.
- Optimal: to embed a maximum of energy without perceptual distortion.

## ■ Capacity

- To be maximized or fixed by the application.

## ■ Robustness

- It has really increased during the last 5 years.
- Recipes with no cook book.

## ■ Security:

- The role of the watermark in the global system.
- What level of security does it provide to the system?

# Overview of digital watermarking - II

---

- 'Smart Image' concept (invited talk SPIE EI'99 - DIGIMARC's CEO).

digital signature      International Image Number  
copy control  
access control      author's name      URL: [www.lenna.org](http://www.lenna.org)  
date of creation



Should I do this?



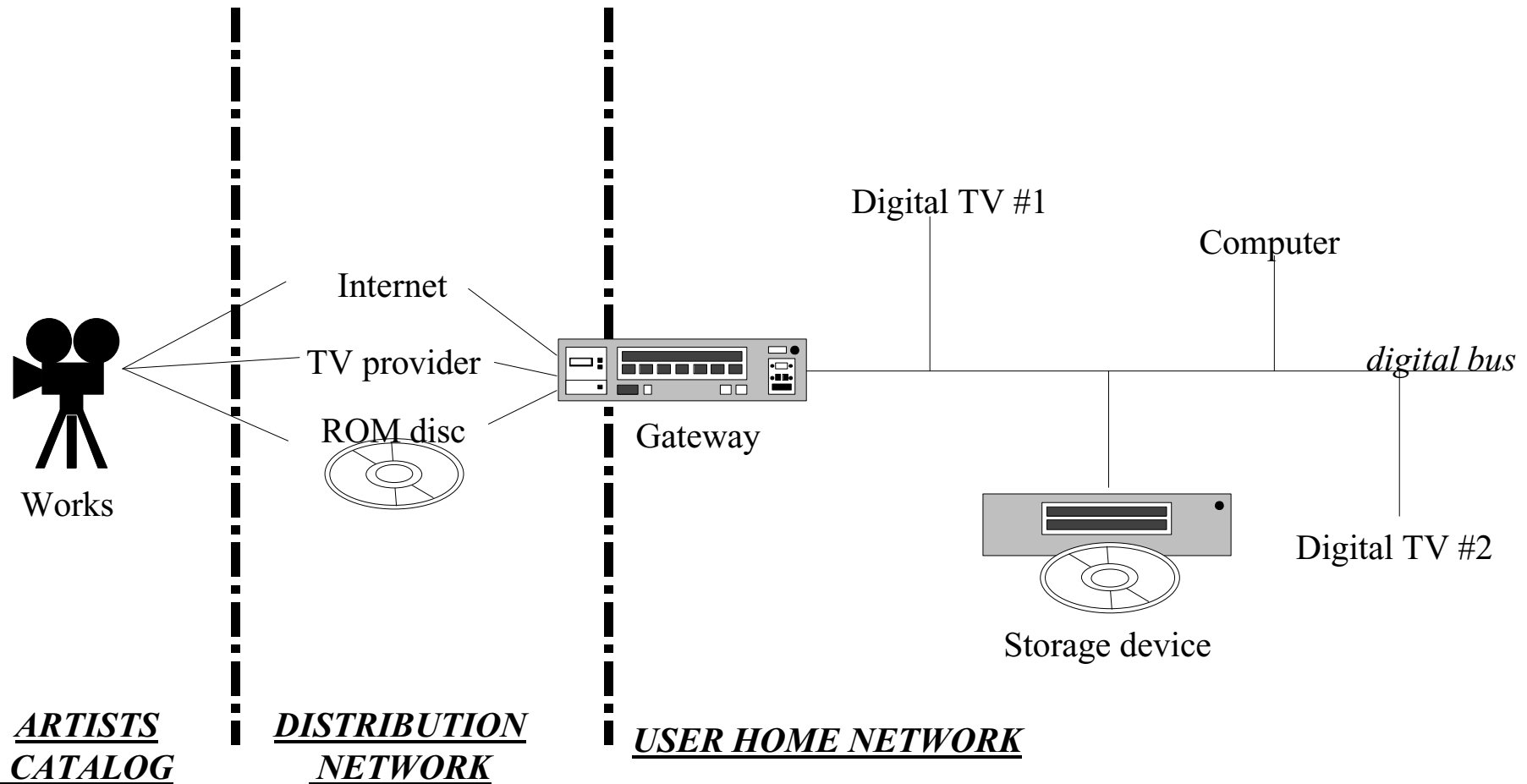
It is secure?

# Presentation outline

---

- Short Overview of Watermarking
- **The Copy Protection Framework**
- Watermarking & Security
- Asymmetric Watermarking Scheme

# The copy protection framework

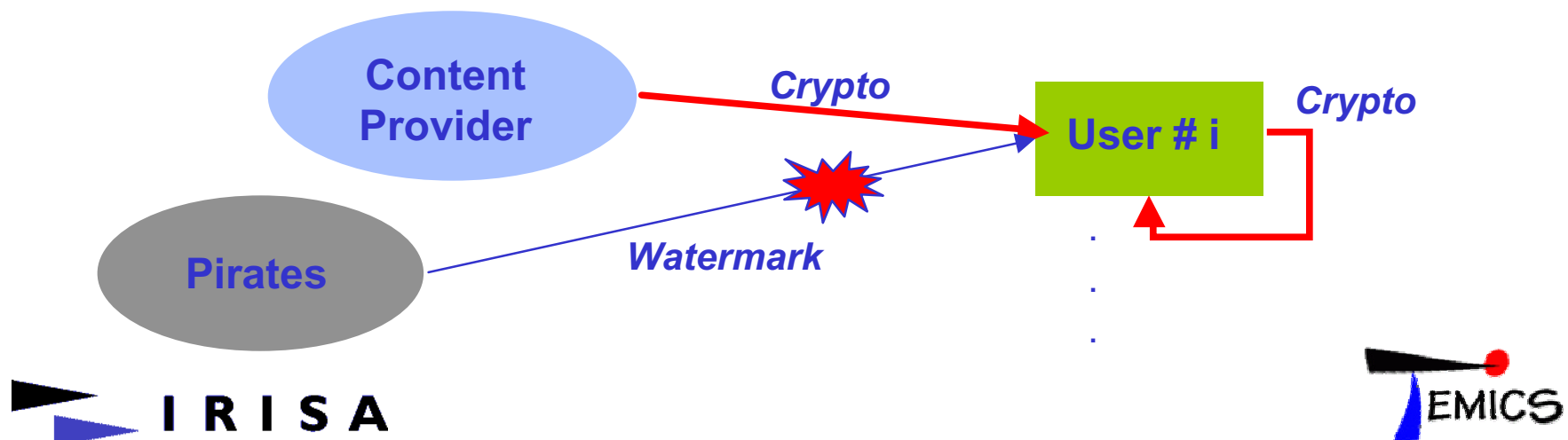


# Role of the WM in the CP system

---

## ■ Quick summary:

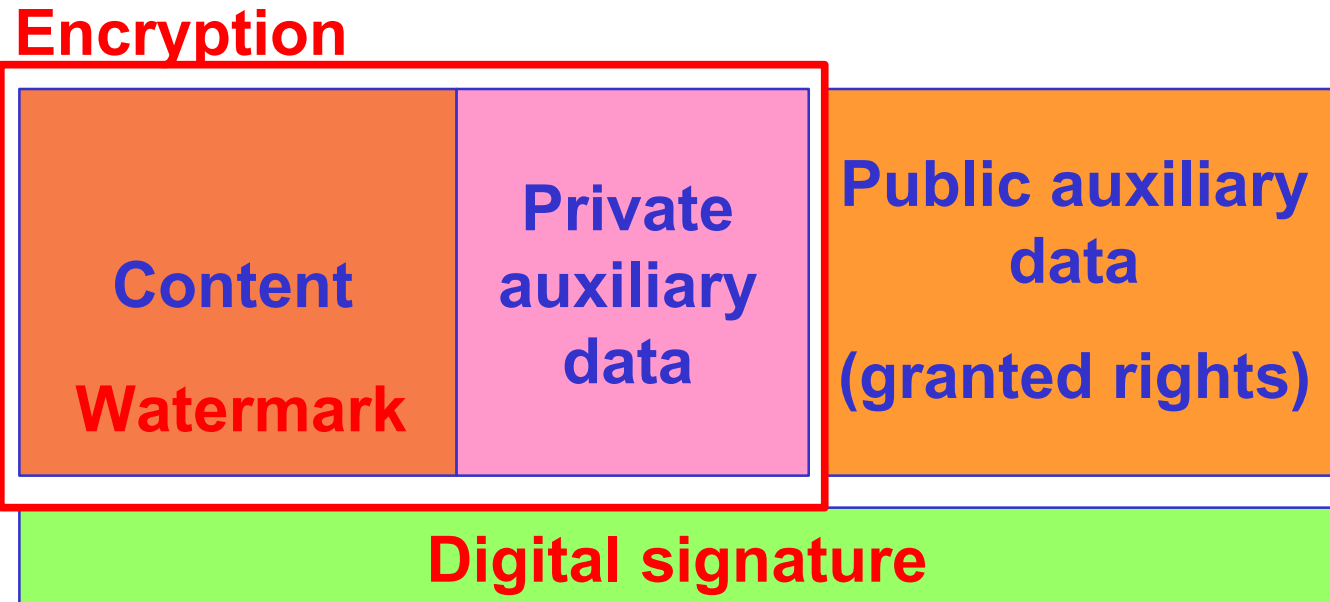
- ❑ The management of the CA and the CP realized only by **cryptographic** functions.
- ❑ The role of the watermarking is **reduced to the minimum**.
- ❑ It targets **only** “analog path” or “bit to bit” pirated copy.



# Role of the watermarking

---

- WM = flag warning compliant devices (" 'smart images'").



- Content is either protected (**WM+Encrypted+DS**) either in the clear.
- BUT, compliant devices reject watermarked content in the clear.

# Threats Analysis

---

- All protected contents are watermarked in the same way.
  - *Watermarked contents only attack.*
- The old contents were not protected.
  - *Known original contents attack.*
- All the compliant devices have a WM detector.
  - *Oracle attack.*

# Examples

---

- Attack on <clear/watermarked> pairs.
  - SDMI Challenge



- Attacks on chosen contents:



# Presentation outline

---

- Short Overview of Watermarking
- The Copy Protection Framework
- **Watermarking & Security**
- Asymmetric Watermarking Scheme

# Cryptography & Watermarking

---

- Usually, nothing in common.
- Except the way cryptography analyses security.
  - Kerckhoffs' Principle: **Algorithms are public except some parameters, called secret key.**
  - Shannon's definition of a security level: **Amount of observations required to get all the information available about the secret key.**
$$H(W|O)=H(W)-I(O;W)$$
  - Diffie-Hellman classification **according to the type of observations.**

# WM structure: the embedding stage

---

## ■ The watermark signal creation.

- From a secret key, create the watermark signal to be embedded.

## ■ The extraction function:

- It extracts perceptually important features from the content.

$$r_o = X ( C_o )$$

## ■ The mixing function:

- It mixes the extracted vector and the WM signal.

$$r_w = F ( r_o , w )$$

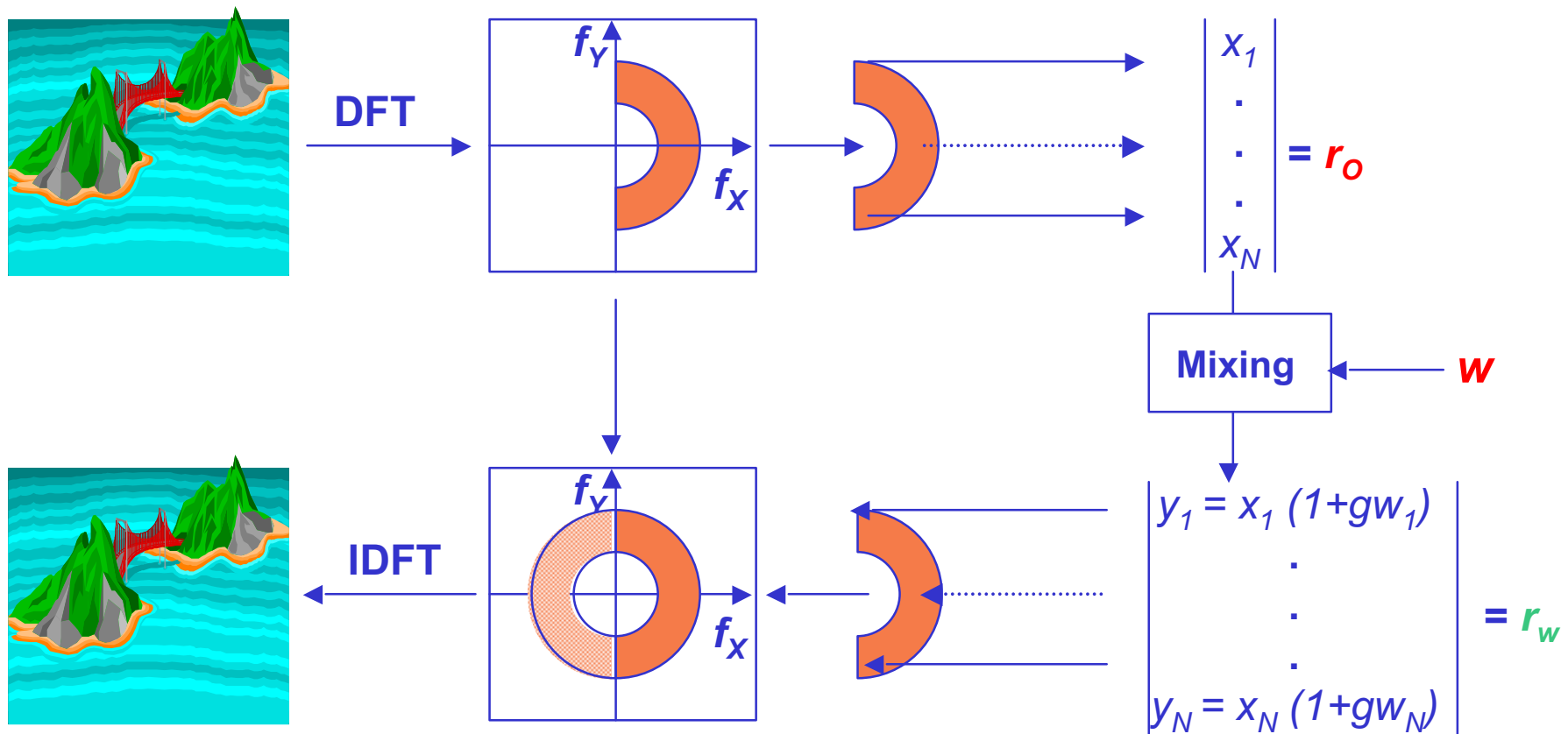
## ■ The inverse extraction function:

- It creates a watermarked content from the new extracted vector and the original content.

$$C_w = X^{-1} ( r_w , C_o )$$

# Example

- **Multiplicative** embedding on some **modules of DFT** coefficients.



# WM structure: the detection stage

---

## ■ The extraction function:

$$r_u = X ( C_u )$$

## ■ The detection process:

- Likelihood that the received content is watermarked.

$$d = \text{detect} ( r_u , \textit{secret key} )$$

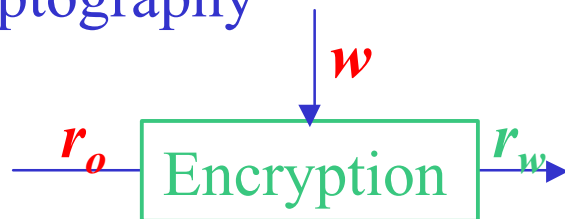
## ■ Comparison to a threshold:

$$D = \begin{cases} \text{Yes if } d > T \\ \text{No if } d < T \end{cases}$$

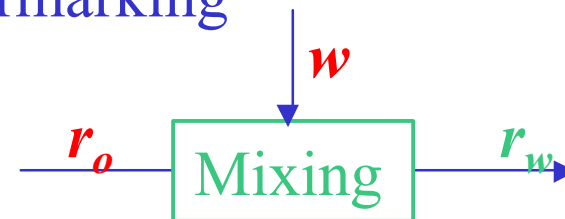
# Shannon's theorem - I

## Watermarked contents only attack

Cryptography



Watermarking



- Initialisation:  $r_o$  and  $w$  independent random processes:  $(R_o, W)$   
 $\Rightarrow$  The ignorance of the adversary is measured by the entropy  $H(W)$ .

- Run: A key is chosen and watermarked contents are produced.

$\Rightarrow$  The ignorance of the adversary is measured by  $H(W / R_w)$ .

- Definition: A system is perfect if  $H(W) = H(W / R_w)$ .

- Theorem 3 [Shannon48]: If a crypto-system is perfect, then

$$H(R_o) = H(W)$$

- Theorem 4 : If a data hiding system is perfect, then

$$H(W) = H(R_o)$$

# Shannon's theorem - II

- Difference #1:

$$H(R_o) < H(W)$$

κρυπτω: I hide

$$H(R_o) > H(W)$$

στεγανω: I cover with

- Difference #2:

Cryptography:  $H(W)$  = Length of the key in bits.

Data hiding:  $H(W)$  does not mean anything for real signals.

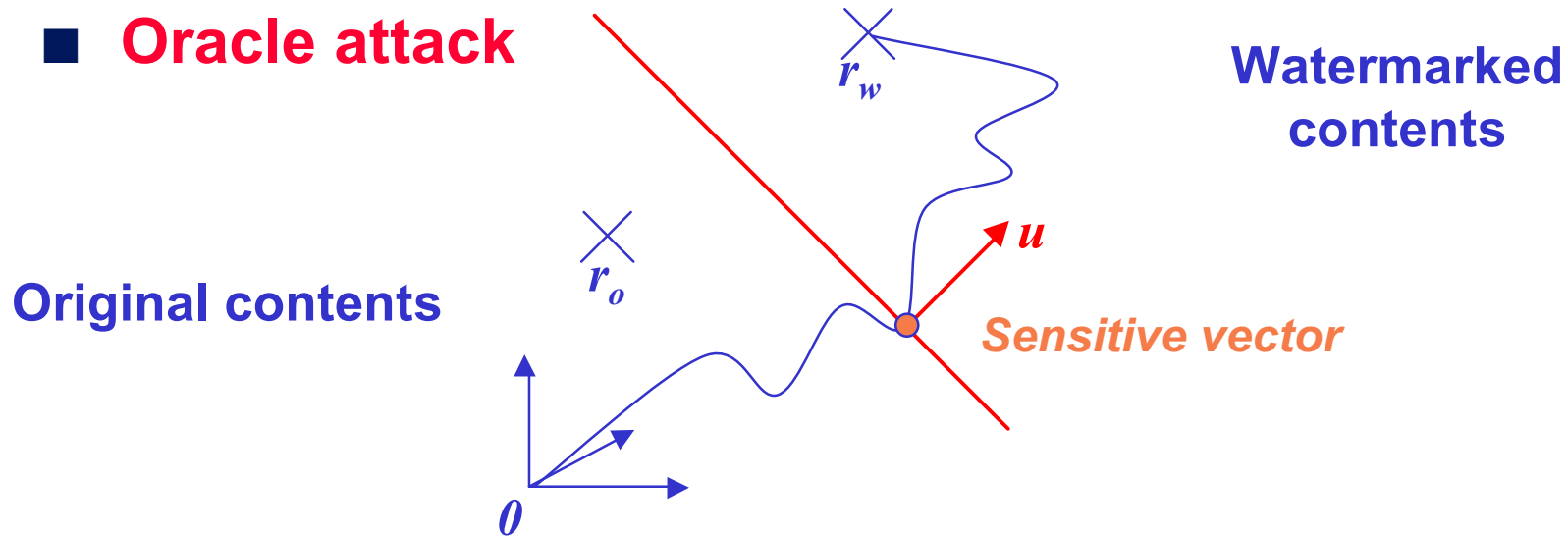
- Corollary: A DSSS is not a perfect data hiding system if the watermark signal is repeated in more than  $T_{lim}$  independent vectors.

$$r_w' = gw + \frac{1}{T} \sum_{k=0}^{T-1} r_{o,k} = gw + r_o' \quad T_{lim} \leq \frac{\sigma^2}{g^2} \left( \frac{\det C_o}{\det C_w} \right)^{1/N} = G^{-1}$$

- The *Power Spectrum Condition* [Su01] maximizes this bound.

# Sensitivity attack

## ■ Oracle attack

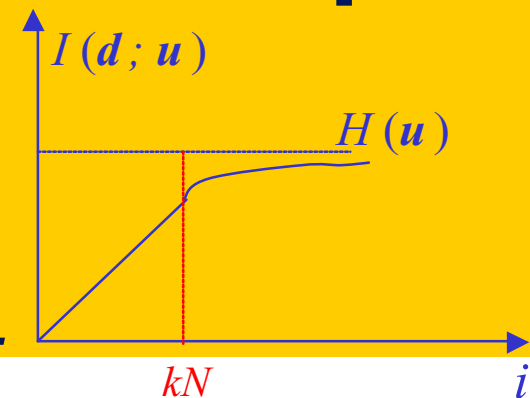


## ■ Theorem 5: sensitivity attack [Kalker&Linnartz98]

Denote  $d_i = \{d_1, \dots, d_j\}$ , then

$$I(d_i; u) \propto i$$

So that an estimation of  $u$  needs  $O(N)$  tries.



# Diffie Hellman terminology

---

- Terminology inspired by the article [DiffieHellman76].

**Watermarked contents only attack:** the pirate has only access to watermarked contents.

**Contents pairs attack:** the pirate has access to watermarked contents and their original version.

**Chosen watermarked contents attack (oracle attack):** the pirate has access to a detector (as black sealed box). He chooses the contents to be tested.

**Chosen original contents attack:** the pirate has access to an embedder (as black sealed box). He chooses the contents to be watermarked.

# Symmetric schemes

---

## ■ WM signal creation:

- $w$  is Gaussian white noise.
- Embedding:  $r_w = r_o + w$

## ■ Detection = correlation:

$$d = r_u^t \cdot w$$

## ■ Advantages: Direct Seq. Spread Spectrum => SIGSALY

- Non perceptibility
- Impossibility to decode without the key
- Good performances against jamming

## ■ Threat :

- Contents are watermarked in the same way.
- Detection is a linear process

# Attacks on symmetric schemes

---

- ***Watermarked contents only attack:*** the “average attack”.
  - An average of 10 sec. of movie is sufficient.
  
- ***Known original content attack:***
  - One good image is enough.
  
- ***Oracle attack:***
  - This attack needs  $O(N)$  tries. (cf. J.P. Linnartz / Ton Kalker)

# Presentation outline

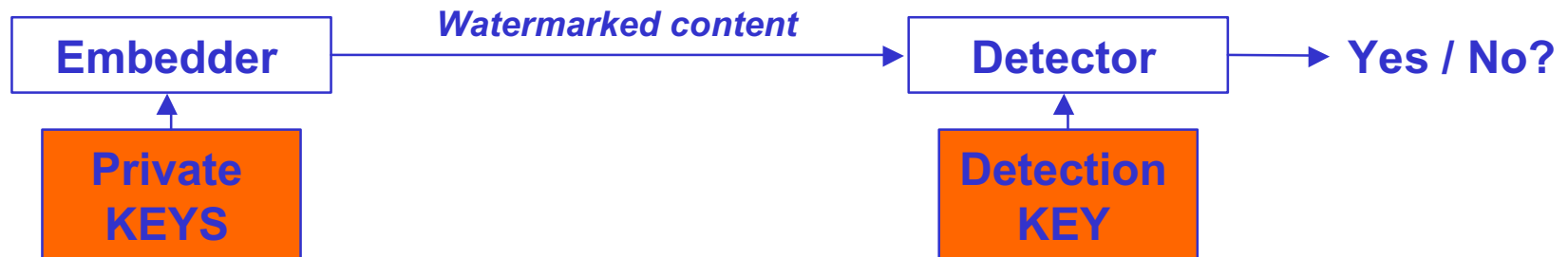
---

- **Short Overview of Watermarking.**
- **The Copy Protection Framework.**
- **Watermarking & Security**
- **Asymmetric Watermarking Scheme**

# Asymmetric schemes

---

## ■ Definition



## ■ Desired advantages

- Even the detector does not know.
- There is a big amount of suitable private keys.

## ■ Better security against key estimation attack.

- Re-newability,
- One time pad (Th. Shannon).

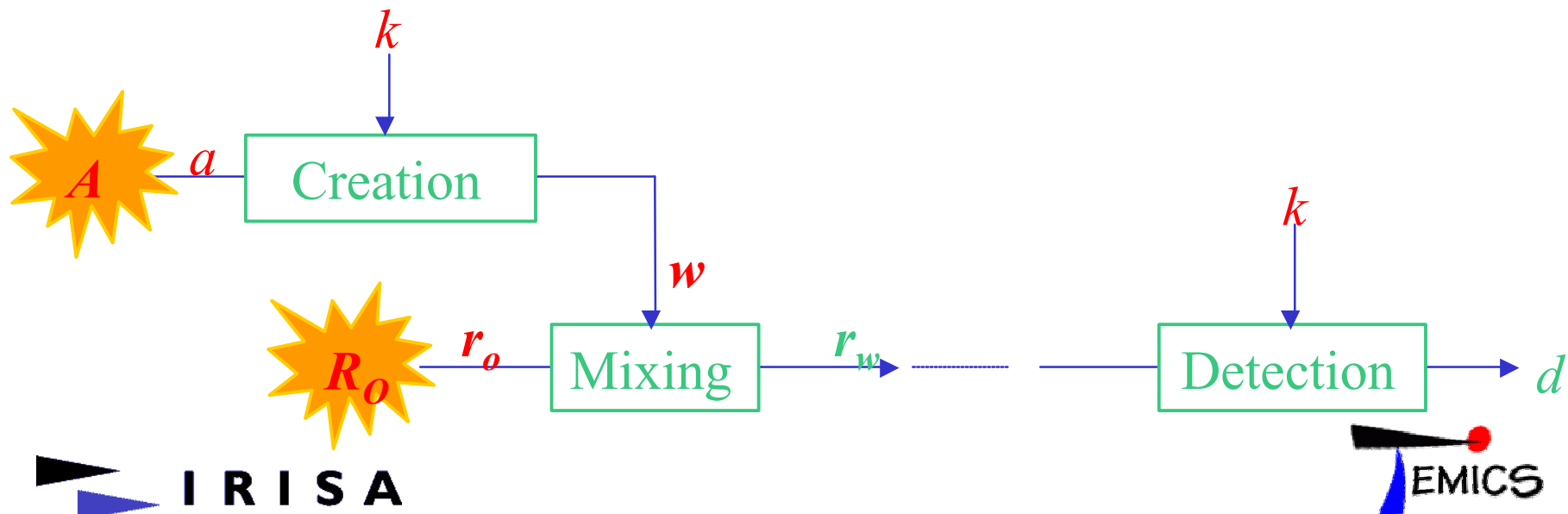
# Inspired from cryptography

## ■ How to encrypt one bit?

- Probabilistic encryption schemes (Goldwasser-Micali) [Menezes96]

## ■ We randomize the embedding.

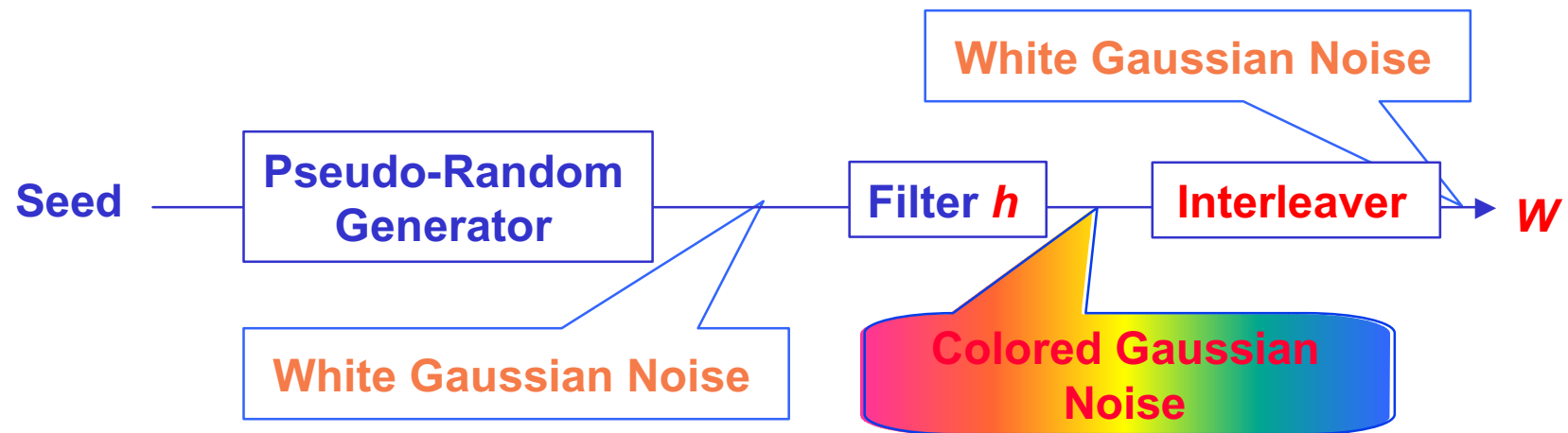
- The random  $a$  is not needed at the detector.
- Hence, a certain asymmetry.



# Our method: the embedding stage

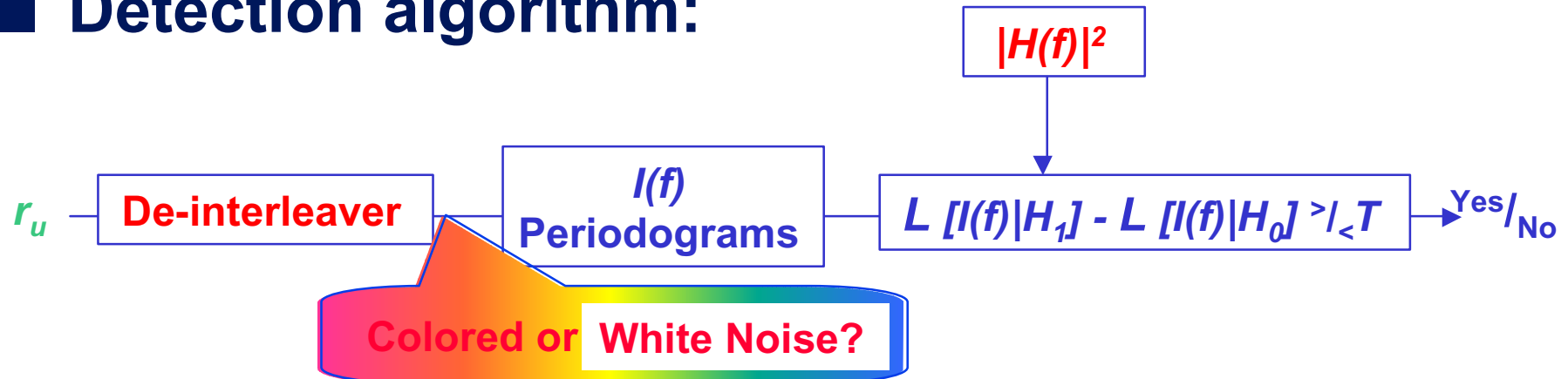
---

## ■ WM signal creation:



# Our method: the detection stage

## ■ Detection algorithm:



## ■ Original contents and watermark signals are independent processes.

$$S_u(f) = S_w(f) + S_o(f)$$
$$S_u(f) = g^2 |H(f)|^2 + cte$$

# Attacks on this asymmetric scheme

---

## ■ *Watermarked contents only attack:*

- The “average attack” is useless.
- The goal is to retrieve some information about the secret key:  
**interleaver**
  
- Strategy:
  - *Brute force attack*: try all them until one finds it.
  
- Issue:
  - Pirates can not perform the detection ( $|H(f)|^2$  is missing).
  
- Solution:
  - Sphericity test           ⇒ Drouiche & Fay test.
  - It performs poorly in our application.
  
- Not possible in practice.

# Attacks on asymmetric scheme (II)

---

## ■ *Known original content attack:*

- The difference gives only one WM signal.
- The goal is to retrieve some information about the secret key: **interleaver**
- Drouiche & Fay test performs well in this case.

- *Brute force attack:*  $N!$  possible permutations.

$N=2048 \Rightarrow$  Sterling Formula  $N! \sim 2^{19000} \gg 2^{300}$  part. in the Universe

- Assume the creation of the interleaver is public.



# Attacks on asymmetric scheme (III)

---

## ■ Oracle attack:

- The pirate tests faked contents:

$$v_{i,j}[k] = \delta_i[k] + \delta_j[k]$$

$$\underline{v}_{i,j}[k] = \delta_i[k] - \delta_j[k]$$

- There are  $O(N^2)$  faked contents. **This is not a good security level in cryptography!**
- BUT, detectors are designed for a decision rate of 10s.

Example:

$N=2^{14}$

- symmetric scheme:  $O(N) \Rightarrow$  2 days.
- asymmetric scheme:  $O(N^2) \Rightarrow$  85 years.

# Attacks on asymmetric scheme (IV)

---

## ■ Oracle attack (II):

- The pirate succeeded to discover a tilting vector  $v_{i,j}$ .

$$r_u' = r_u + e \cdot v_{i,j}$$

- He expects:

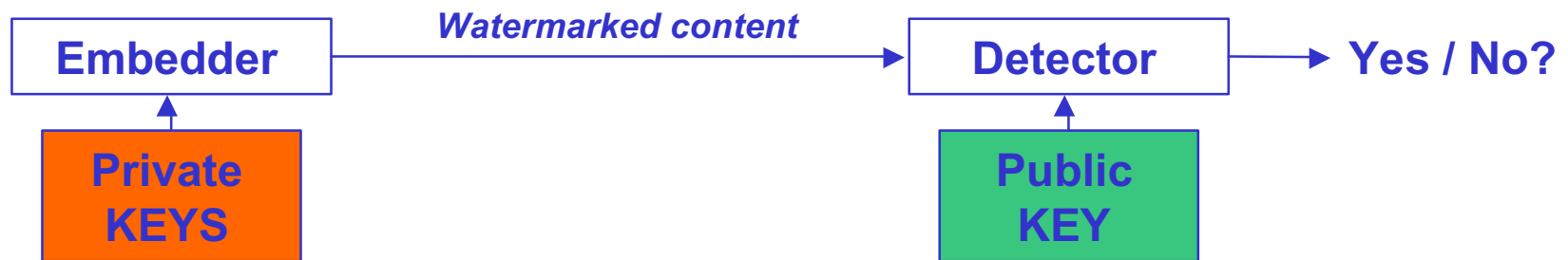
$$d' = d - e^2 C_u / N + d''$$

- Because the periodogram is not a linear process!
- This attack does not always work.

# Towards public key schemes

---

- Last attack: *Reverse engineering* of the detection chip.
- Definition of a public key scheme:



- Desired Advantages
  - Disclosure of the detection key give no advantage to pirate.
  - Detection implementation in software possible.
- This would provide a “perfect security” level.

# Conclusion about asymmetry

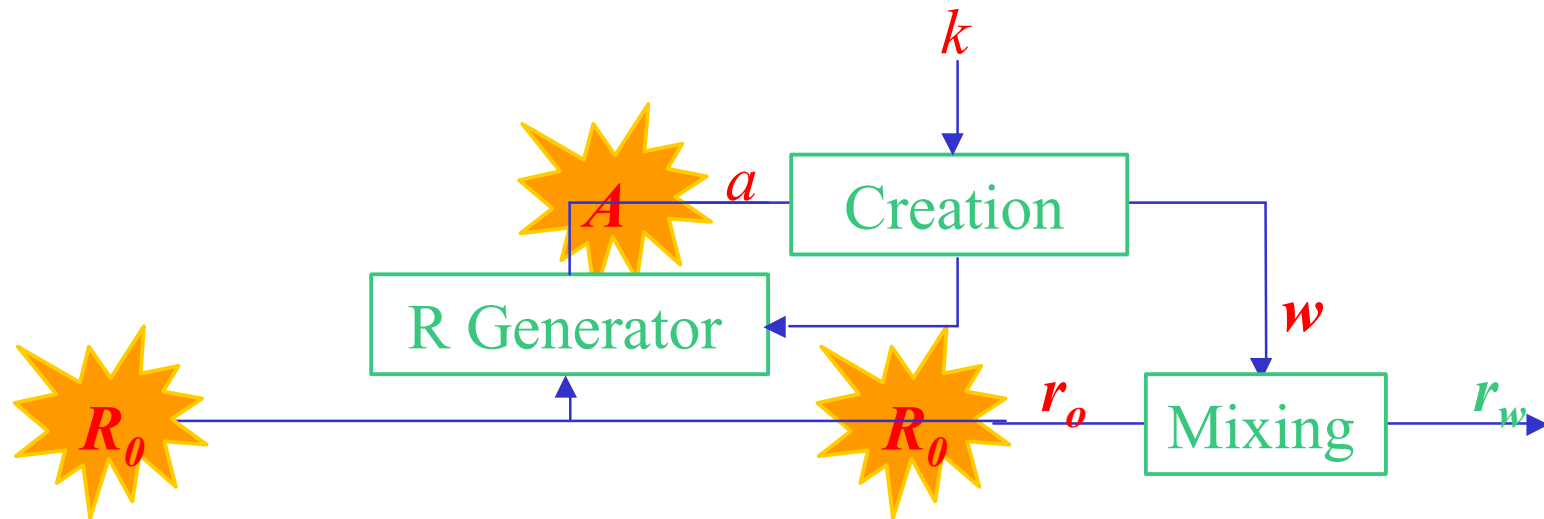
---

- 😊 **Asymmetric schemes provide a better security level against malicious attacks:**
  - ❑ *Watermarked contents only attack, Known original content attack, Oracle attack,*
- 😐 **Is this method is versatile?**
  - ❑ Whatever DSSS technique can be derived into an asymmetric scheme.
  - ❑ Only suitable for the copy protection framework! (capacity of 1 bit).
- 😞 **Asymmetric detectors less efficient:**
  - ❑ Efficacy comparison asym. Vs. sym. : a decrease 1:10
  - ❑ Hence, extracted vectors are longer.
  - ❑ Complexity increases (amount of content, memory, computer power).

# Perspective: SI at the embedding

---

- Idea: Derive one source from the other one.



- Theory:

Choose the best  $a$  to maximize detectability

# Conclusion (II)

---

- **Security is an important feature.**
  - Describe your targeted application. Especially the role of the WM in the global system.
  - Threat analysis: what the pirate can do & what he can not do.
  - Estimate the complexity of the attacks and their impact on the system.
  
- **At last, Watermarkers have understood security (Kerckhoff, Shannon, Diffie-Hellman).**
  
- **Does public key WM scheme exist?**