Assurer la sécurité des contenus multimédia, de leur création à leur diffusion
How to protect multimedia pieces of content, from their creation to their distribution

HDR defense

Caroline Fontaine

November, 28th 2011

Some security issues

Confidentiality
Integrity
Authentication
Cryptology

Errors
Error Correcting Codes

Alice

Bob

Subjective Integrity
Fragile watermarking

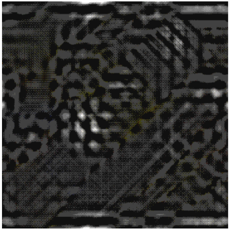
Stealthiness
Steganography

Copyright
Cryptology (DRM)
Robust watermarking (identify the copyright owner)
Fingerprinting (identify the legitimate end user)

Cryptology [1917-] Error Correcting Codes [1947-] Information Hiding [1990-]

Information Hiding in a nutshell

+

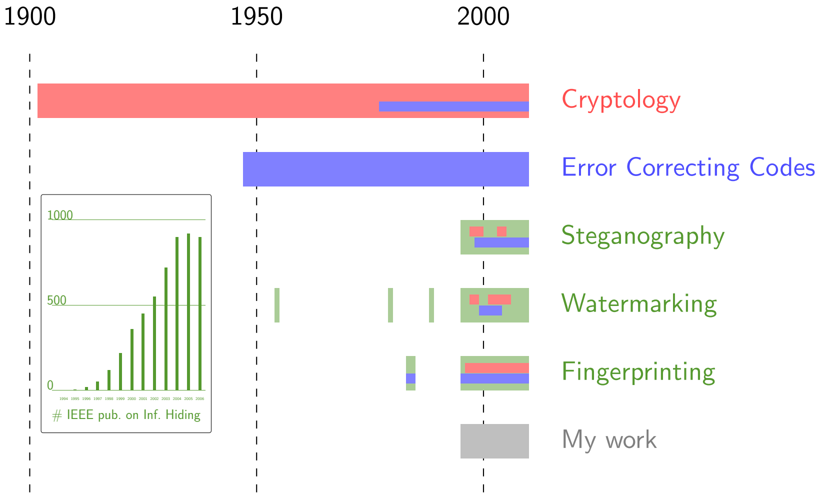


=

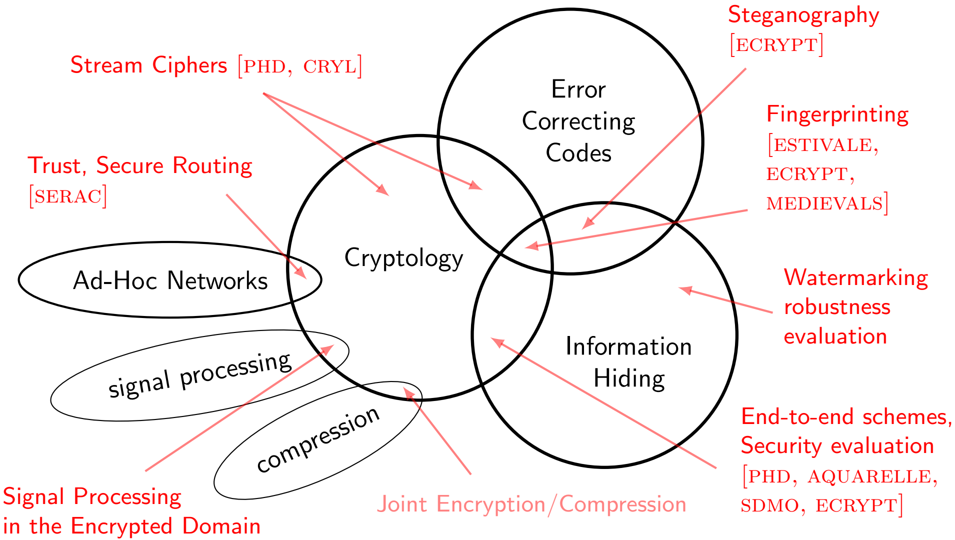
Trade-offs between :	capacity,	imperceptibility,	robustness,	security
Steganography : (stealth com.)	+	++	-	++
Robust Watermarking : (id. cop. owner)	+	+	++	++
Fingerprinting : (id. end user)	+	+	++	++

+ important, - not important

Modern Evolutions and Crossings



Crossings and Contributions



Some contributions

Design (and attack) of stream ciphers, based on Highly Nonlinear Boolean Functions obtained with the help of Error Correcting Codes [FF98,Fon99,CCCF00,CCCF01,FFJ04,BRWF05] PhD V. Bénony [02-06]

Design of steganographic schemes based on Error Correcting Codes [FG07,FG09,ABF11] PhD M. Barbier [08-11]

Design of content protection architectures mixing cryptographic and watermarking primitives [AFD98,ABD+99,ABTD+06,FDD+08]

Transposed cryptanalysis methodology to the study of the security of watermarking schemes [FR02,CFF05b,CFF05e]

Improvement of the robustness and security of `Broken Arrows` watermarking technique [CXFF09,XFF10a,XFF10b] PhD F. Xie [07-10]

Design of fingerprinting schemes based on a watermarking layer and an anti-collusion code [XFF08,CXFF09,CFF10,CFFC11]
PhD F. Xie [07-10], PhD A. Charpentier [08-11]

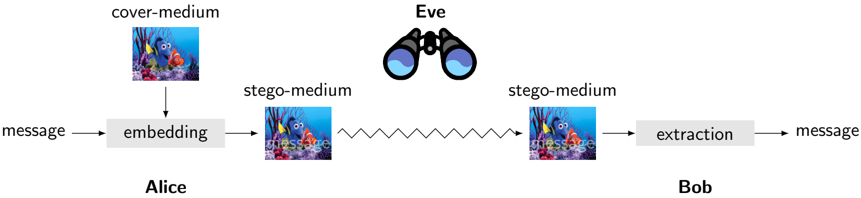
Outline

The Warden is watching . . .

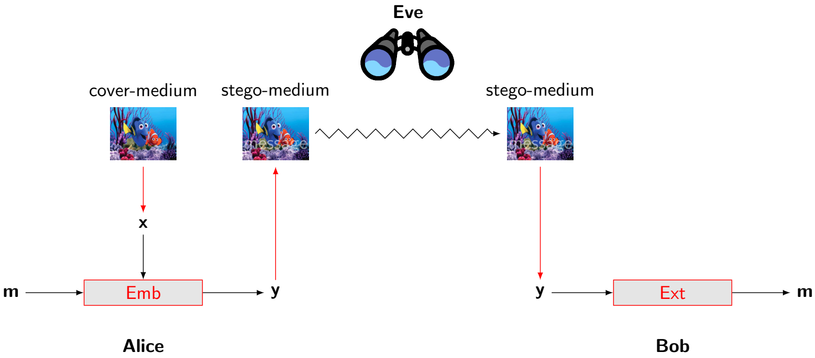
The Prisoners and the Passive Warden [Sim83] :

Alice and Bob want to send each other some important secret messages. Eve keeps a watch on. If she suspects something is going wrong, she interrupts the communication.

⇒ Alice and Bob must exchange only innocuous looking documents! They cannot rely only on cryptograhy, they need a steganographic scheme.



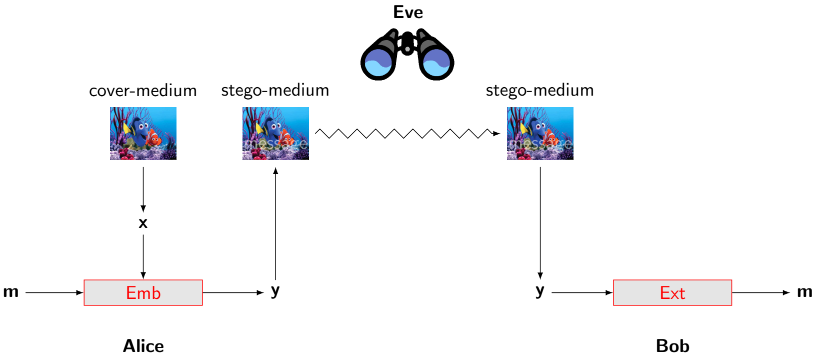
Steganographic schemes : design issues



Critical choices to prevent steganalysis (no perfect security currently achievable) :

- Which vectors to derive from the medium ?
- How to process them with Emb and Ext ?

Steganographic schemes : design issues

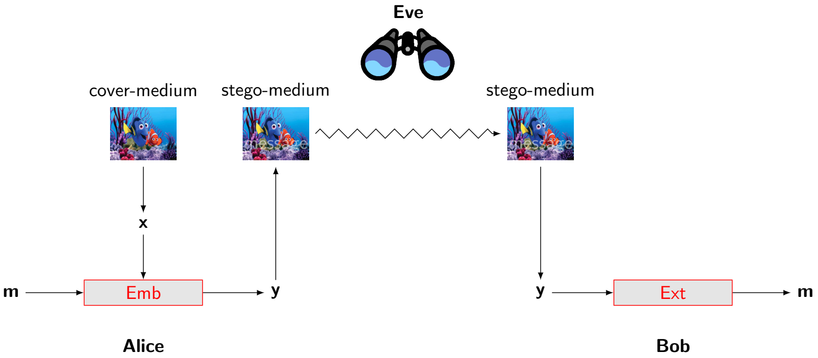


Critical choices to prevent steganalysis (no perfect security currently achievable) :

- Which vectors to derive from the medium ?
- How to process them with Emb and Ext ?

$\Rightarrow$ One strategy : to minimize distorsion, one way : syndrome coding

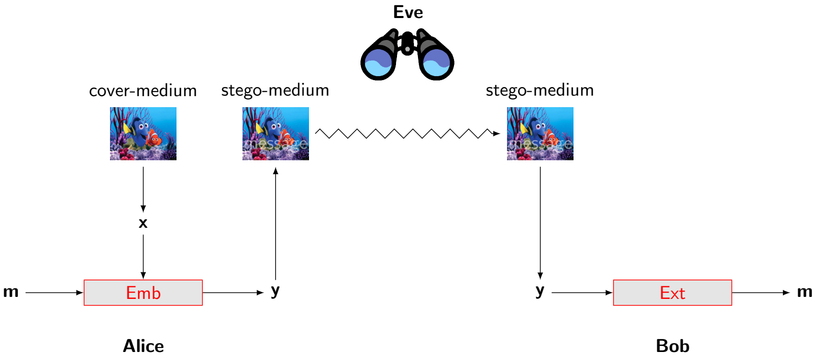
Minimizing distortion with syndrome coding



One strategy : to minimize distortion

$$\mathrm{Ext}(\mathrm{Emb}(\mathbf{x}, \mathbf{m})) = \mathbf{m}$$
$$d_H(\mathbf{x}, \mathrm{Emb}(\mathbf{x}, \mathbf{m})) \leq T$$

Minimizing distorsion with syndrome coding



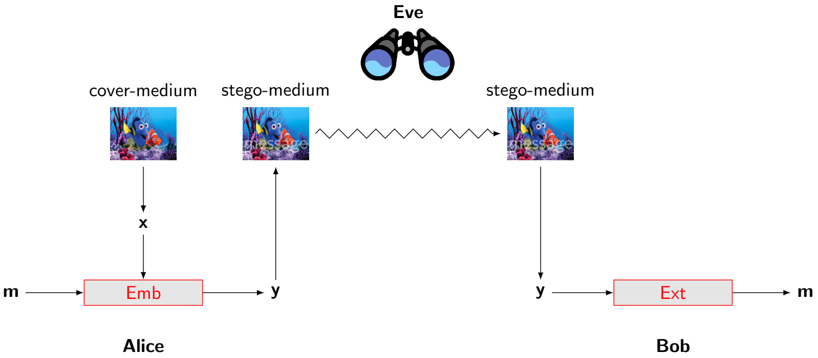
One strategy : to minimize distorsion

$$\mathrm{Ext}(\mathrm{Emb}(\mathbf{x},\mathbf{m})) = \mathbf{m}$$

$$d_H(\mathbf{x},\mathrm{Emb}(\mathbf{x},\mathbf{m})) \leq T$$

/* $\mathrm{Emb}(\mathbf{x},\mathbf{m})_i = \mathbf{x}_i \ \forall i \in \mathcal{W}$ wet paper [FGLS05] */

Minimizing distortion with syndrome coding



One strategy : to minimize distortion , one way : syndrome coding (e.g. F5 [Wes01])

$$\mathrm{Ext}(\mathrm{Emb}(\mathbf{x}, \mathbf{m})) = \mathbf{m}$$

$$d_H(\mathbf{x}, \mathrm{Emb}(\mathbf{x}, \mathbf{m})) \leq T$$

$\mathrm{Emb}(\mathbf{x}, \mathbf{m})$ of syndrome $\mathbf{m}$
$\mathrm{Ext}(\mathbf{y})$ = syndrome of $\mathbf{y}$

/* $\mathrm{Emb}(\mathbf{x}, \mathbf{m})_i = \mathbf{x}_i \; \forall i \in \mathcal{W}$ wet paper [FGLS05] */

Syndrome coding, more formally

Syndrome coding has been introduced and discussed in [Cra98,Bie01], and properly formalized in [GK03,GK09]. It has been widely studied.

Let $\mathcal{C}$ be a q-ary linear code of length n, dimension k and parity check matrix $\mathbf{H}$: $\mathcal{C} = \{\mathbf{c} \mid \mathbf{c} \cdot \mathbf{H}^t = 0\}$ is a vector subspace of $\mathbb{F}_q^n$ of dimension k.

$$\begin{array}{rcll} \mathrm{Emb}(\mathbf{x}, \mathbf{m}) & = & \mathbf{x} + D(\mathbf{m} - \mathbf{x} \cdot \mathbf{H}^t) & \leftarrow \mathrm{Ext}(\mathrm{Emb}(\mathbf{x}, \mathbf{m})) = \mathbf{m}, d_H() \leq T \\ \mathrm{Ext}(\mathbf{y}) & = & \mathbf{y} \cdot \mathbf{H}^t & \leftarrow \text{syndrome of } \mathbf{y} \end{array}$$

$D()$ must return $\mathbf{e}$, $d_H(\mathbf{e}, 0) \leq T$, of syndrome $\mathbf{e} \cdot \mathbf{H}^t = \mathbf{m} - \mathbf{x} \cdot \mathbf{H}^t$.

Syndrome coding, more formally

Syndrome coding has been introduced and discussed in [Cra98,Bie01], and properly formalized in [GK03,GK09]. It has been widely studied.

Let $\mathcal{C}$ be a q-ary linear code of length n, dimension k and parity check matrix $\mathbf{H}$: $\mathcal{C} = \{\mathbf{c} \mid \mathbf{c} \cdot \mathbf{H}^t = 0\}$ is a vector subspace of $\mathbb{F}_q^n$ of dimension k.

$$\begin{array}{rcll} \mathrm{Emb}(\mathbf{x},\mathbf{m}) & = & \mathbf{x} + D(\mathbf{m} - \mathbf{x} \cdot \mathbf{H}^t) & \leftarrow \mathrm{Ext}(\mathrm{Emb}(\mathbf{x},\mathbf{m})) = \mathbf{m}, d_H() \leq T \\ \mathrm{Ext}(\mathbf{y}) & = & \mathbf{y} \cdot \mathbf{H}^t & \leftarrow \text{syndrome of } \mathbf{y} \end{array}$$

$D()$ must return $\mathbf{e}$, $d_H(\mathbf{e},0) \leq T$, of syndrome $\mathbf{e} \cdot \mathbf{H}^t = \mathbf{m} - \mathbf{x} \cdot \mathbf{H}^t$.

- Does such a vector $\mathbf{e}$ exist ?
- How to find it ?
- May we choose between several vectors $\mathbf{e}$?

Syndrome coding, more formally

Syndrome coding has been introduced and discussed in [Cra98,Bie01], and properly formalized in [GK03,GK09]. It has been widely studied.

Let $\mathcal{C}$ be a q-ary linear code of length n, dimension k and parity check matrix $\mathbf{H}$: $\mathcal{C} = \{\mathbf{c} \mid \mathbf{c} \cdot \mathbf{H}^t = 0\}$ is a vector subspace of $\mathbb{F}_q^n$ of dimension k.

$$\begin{array}{rcll} \mathrm{Emb}(\mathbf{x},\mathbf{m}) & = & \mathbf{x} + D(\mathbf{m} - \mathbf{x}\cdot\mathbf{H}^t) & \leftarrow \mathrm{Ext}(\mathrm{Emb}(\mathbf{x},\mathbf{m})) = \mathbf{m}, d_H() \leq T \\ \mathrm{Ext}(\mathbf{y}) & = & \mathbf{y}\cdot\mathbf{H}^t & \leftarrow \text{syndrome of } \mathbf{y} \end{array}$$

$D()$ must return $\mathbf{e}$, $d_H(\mathbf{e}, 0) \leq T$, of syndrome $\mathbf{e}\cdot\mathbf{H}^t = \mathbf{m} - \mathbf{x}\cdot\mathbf{H}^t$.

- Does such a vector $\mathbf{e}$ exist ?	Not always (depends on $\mathcal{C}$)
- How to find it ?	A matter of decoding
- May we choose between several vectors $\mathbf{e}$?	If list decoding is possible

Syndrome coding, more formally

Syndrome coding has been introduced and discussed in [Cra98,Bie01], and properly formalized in [GK03,GK09]. It has been widely studied.

Let $\mathcal{C}$ be a q-ary linear code of length n, dimension k and parity check matrix $\mathbf{H}$: $\mathcal{C} = \{\mathbf{c} \mid \mathbf{c} \cdot \mathbf{H}^t = 0\}$ is a vector subspace of $\mathbb{F}_q^n$ of dimension k.

$$\begin{array}{rcll} \mathrm{Emb}(\mathbf{x}, \mathbf{m}) & = & \mathbf{x} + D(\mathbf{m} - \mathbf{x} \cdot \mathbf{H}^t) & \leftarrow \mathrm{Ext}(\mathrm{Emb}(\mathbf{x}, \mathbf{m})) = \mathbf{m}, d_H() \leq T \\ \mathrm{Ext}(\mathbf{y}) & = & \mathbf{y} \cdot \mathbf{H}^t & \leftarrow \text{syndrome of } \mathbf{y} \end{array}$$

$D()$ must return $\mathbf{e}$, $d_H(\mathbf{e}, 0) \leq T$, of syndrome $\mathbf{e} \cdot \mathbf{H}^t = \mathbf{m} - \mathbf{x} \cdot \mathbf{H}^t$.

- Does such a vector $\mathbf{e}$ exist ?	Not always (depends on $\mathcal{C}$)
- How to find it ?	A matter of decoding
- May we choose between several vectors $\mathbf{e}$?	If list decoding is possible

$\Rightarrow$ $\mathcal{C}$ must be chosen really carefully

Syndrome coding, more formally

Syndrome coding has been introduced and discussed in [Cra98,Bie01], and properly formalized in [GK03,GK09]. It has been widely studied.

Let $\mathcal{C}$ be a q-ary linear code of length n, dimension k and parity check matrix $\mathbf{H}$: $\mathcal{C} = \{\mathbf{c} \mid \mathbf{c} \cdot \mathbf{H}^t = 0\}$ is a vector subspace of $\mathbb{F}_q^n$ of dimension k.

$$
\begin{array}{rcll}
\mathrm{Emb}(\mathbf{x}, \mathbf{m}) & = & \mathbf{x} + D(\mathbf{m} - \mathbf{x} \cdot \mathbf{H}^t) & \leftarrow \mathrm{Ext}(\mathrm{Emb}(\mathbf{x}, \mathbf{m})) = \mathbf{m}, d_H() \leq T \\
\mathrm{Ext}(\mathbf{y}) & = & \mathbf{y} \cdot \mathbf{H}^t & \leftarrow \text{syndrome of } \mathbf{y} \\
/* \quad \mathbf{y}_i & = & \mathbf{x}_i \qquad \forall i \in \mathcal{W} & \text{wet paper } */ \qquad \text{even harder!}
\end{array}
$$

$D()$ must return $\mathbf{e}$, $d_H(\mathbf{e}, 0) \leq T$, of syndrome $\mathbf{e} \cdot \mathbf{H}^t = \mathbf{m} - \mathbf{x} \cdot \mathbf{H}^t$.

- Does such a vector $\mathbf{e}$ exist ? — Not always (depends on $\mathcal{C}$)
- How to find it ? — A matter of decoding
- May we choose between several vectors $\mathbf{e}$? — If list decoding is possible

$\Rightarrow$ $\mathcal{C}$ must be chosen really carefully

Which code should we use ?

A lot of codes have been studied : Hamming, BCH, Convolutional, etc

Which criteria have been addressed ?

- Embedding efficiency (heavily studied, optimal codes)

Which code should we use ?

A lot of codes have been studied : Hamming, BCH, Convolutional, etc

Which criteria have been addressed ?

- Embedding efficiency (heavily studied, optimal codes)
- Probability of success (almost never addressed)

Which code should we use ?

A lot of codes have been studied : Hamming, BCH, Convolutional, etc

Which criteria have been addressed ?

- Embedding efficiency (heavily studied, optimal codes)
- Probability of success (almost never addressed)
  - Dry paper : success is ensured only for perfect codes (Hamming and Golay, but their embedding efficiency is not good)
  - Wet paper : success is ensured only for MDS [q-ary] codes

When success is not ensured, the probability of success decreases exponentially with the message length !

Which code should we use ?

A lot of codes have been studied : Hamming, BCH, Convolutional, etc

Which criteria have been addressed ?

- Embedding efficiency (heavily studied, optimal codes)
- Probability of success (almost never addressed)
  - Dry paper : success is ensured only for perfect codes (Hamming and Golay, but their embedding efficiency is not good)
  - Wet paper : success is ensured only for MDS [q-ary] codes

  When success is not ensured, the probability of success decreases exponentially with the message length !

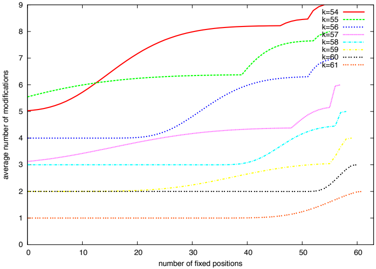
Contributions :

- Reed-Solomon codes (MDS, list decoding) [FG07,FG09]
- A variant of syndrome coding, that ensures embedding success [ABF11] PhD M. Barbier [08-11]

How Reed-Solomon codes can help

With F. Galand [FG07,FG09] (IH 2007)

- ✓ Good parameters (*e.g.* covering radius)
- ✓ MDS (embedding is ensured in the wet paper context)
- ✓ Unique decoding (Lagrange) + List decoding (Guruswami-Sudan)



Estimated Gain of List decoding. $q = 64$, $n = 63$, Plot only $\Delta\omega \geq 0.3$

- ✓ List decoding → gain in average embedding efficiency
- ☹ Guruswami-Sudan is hard to implement
- ϟ implementation of Guruswami-Sudan
- ϟ must derive q-ary vectors $\mathbf{x}$ from the media

Randomized Syndrome Coding

With D. Augot and M. Barbier [ABF11] (IMACC 2011)

"How can we design a scheme that ensures embedding ?"

Our idea : randomize a part of the syndrome,

$$\text{replacing} \quad \mathbf{y} \cdot \mathbf{H}^t = \mathbf{m}$$
$$\text{by} \quad \mathbf{y} \cdot \mathbf{H}^t = (\mathbf{m} || \mathbf{R})$$

- ✓ Embedding success, even in the wet paper context
- ✓ We provided a way to send the length of **R** to the recipient
- ☹ Loss in embedding efficiency (vs. traditional synd. coding)
- ✓ $[\frac{q^p-1}{q-1}, n-p, 3]$ Hamming codes :
  the relative loss in embedding efficiency is only $\frac{\lceil \log_q((q-1)\#\mathcal{W}+1) \rceil}{p}$
- ↯ Must be studied further

Steganography : conclusion and further work

We focused on the success on the embedding,
while preserving a good embedding efficiency.

Reed-Solomon codes :

- ✓ RS could (should ?) be used in practical schemes
- ↯ Native q-ary steganography should be studied
- ↯ Implementation of Guruswami-Sudan list decoding

Randomized Syndrome Coding :

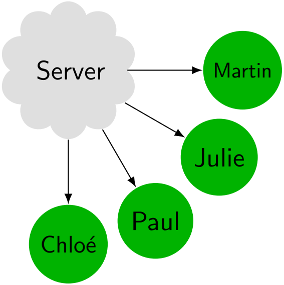
- ↯ Needs to be further studied

Other tracks :

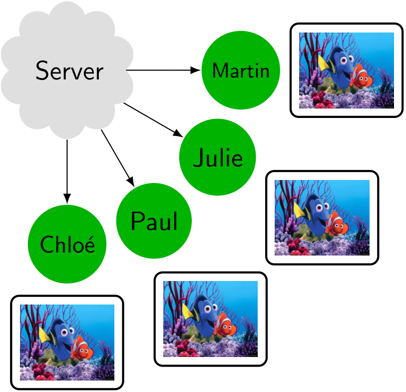
- ↯ Active Warden

Outline

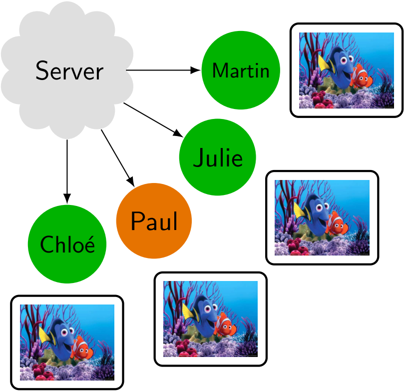
How to prevent illegal redistribution ?



How to prevent illegal redistribution ?

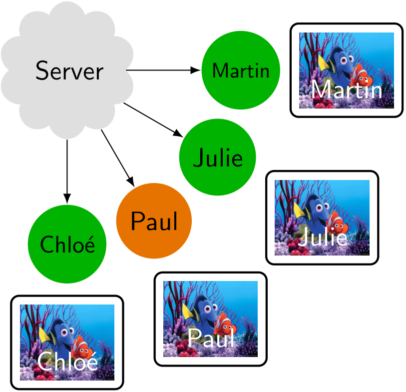


How to prevent illegal redistribution ?



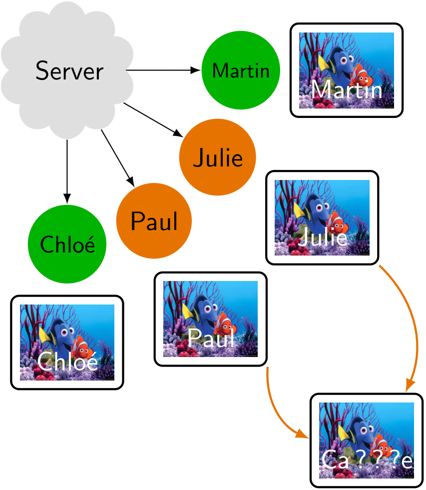
- Cryptography is not sufficient

How to prevent illegal redistribution ?



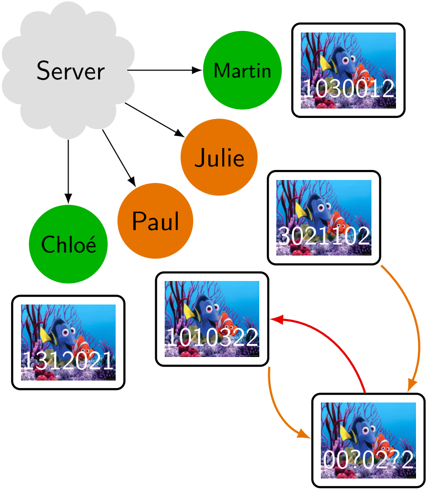
- Cryptography is not sufficient
- A need for watermarking

How to prevent illegal redistribution ?

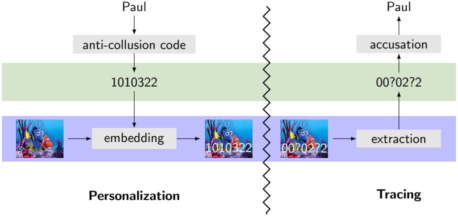


- Cryptography is not sufficient
- A need for watermarking

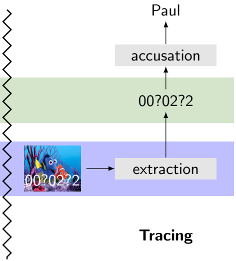
How to prevent illegal redistribution ?



- Cryptography is not sufficient
- A need for watermarking
- A need for an anti-collusion code with a structure enabling tracing



How to link models with reality



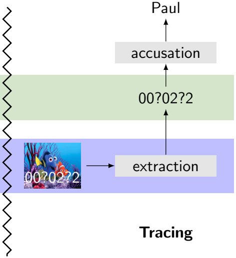
message layer

signal layer

Boneh & Shaw introduced in 1995 a model which remains the most used today (with its extensions)

- ✓ Simple to express
- ✓ Has been intensively studied
- ☹ Not so realistic

How to link models with reality



message layer

signal layer

Boneh & Shaw introduced in 1995 a model which remains the most used today (with its extensions)

- ✓ Simple to express
- ✓ Has been intensively studied
- ☹ Not so realistic
- ✓ BUT we can force reality to fit the model :

off-line block-based watermarking and on-line switching

0	0	0	0	0	0	0
1	1	1	1	1	1	1
2	2	2	2	2	2	2
3	3	3	3	3	3	3

Chloé	1	3	1	2	0	2	1
Paul	1	0	1	0	3	2	2
Julie	3	0	2	1	1	0	2
Martin	1	0	3	0	0	1	2

Attacks and assumptions

 | 1 | 0 | 1 | 0 | 3 | 2 | 2 | $\mathbf{X}_{\text{Paul}}$

 | 3 | 0 | 2 | 1 | 1 | 0 | 2 | $\mathbf{X}_{\text{Julie}}$

 | 1 | 0 | 3 | 0 | 0 | 1 | 2 | $\mathbf{X}_{\text{Martin}}$

Collusion (c users among n)

 | 1 | 0 | 3 | 1 | 0 | 2 | 2 | $\mathbf{Y}$

Copy/paste blocks (*e.g.* random, maj., etc)

 | 0 | 0 | ? | 0 | 2 | ? | 2 | $\mathbf{Y}$

Fusion of blocks (*e.g.* averaging)

Boneh & Shaw : Marking Assumption $X_{j_1 i} = \ldots = X_{j_c i} = a \Rightarrow Y_i = a$.

 | ? | 0 | 1 | 2 | ? | 1 | ? | $\mathbf{Y}$

Individual signal processing (*e.g.* compression)

To prevent errors and erasures, watermarking must be as robust as possible.

In the steps of Boneh & Shaw

Strong traceability : $\mathbb{P}$(**accuse an innocent user**) $= 0$

- ☉ error correcting codes
- ☹ $n \geq 3, c \geq 2$: only copy/paste attacks
- ☹ codes too long, on huge alphabets [HvLLT98,BCE$^+$01,SSW01]

Weak traceability : $\mathbb{P}$(**accuse an innocent user**) $< \varepsilon$

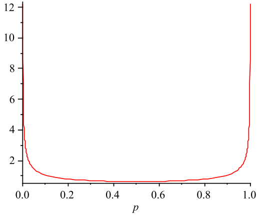
- ☉ error correcting codes + probabilistic codes
- ✓ copy/paste + fusion attacks
- ☉ Peikert's bound [PSS03][Tar03,Tar08] : $m \geq \mathcal{O}(c^2 \ln(n/\varepsilon))$
- ✓ first codes to meet the bound : Tardos codes

Binary Tardos code [Tar03]+[SKC08]

n, $1 << c$, $\varepsilon_1 << \varepsilon_2$, $m = 2\pi^2 c^2 \lceil \ln(1/\varepsilon_1) \rceil$, $Z = 2\pi c \lceil \ln(1/\varepsilon_1) \rceil$.
m secret probabilities p_i drawn according to the pdf $f(p) = \frac{1}{\pi\sqrt{p(1-p)}}$.

$$\begin{array}{cccccc} & p_1 & p_2 & p_3 & \dots & p_m \\ X_1 & 1 & 0 & 1 & \dots & 0 \\ X_2 & 0 & 1 & 0 & \dots & 1 \\ X_3 & 1 & 1 & 0 & \dots & 1 \\ \vdots & \vdots & \vdots & \vdots & \vdots & \vdots \\ X_n & 0 & 0 & 0 & \dots & 1 \end{array}$$

with $\mathbb{P}(X_{ij} = 1) = p_i$

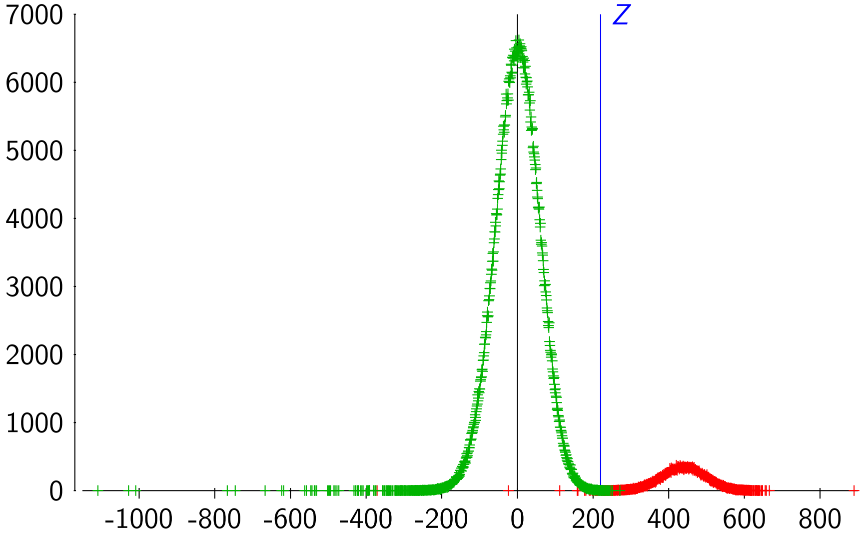


Accusation : User j's score $S_j = \Sigma_{i=1}^{m} g(Y_i, X_{ji}, p_i) >^? Z$

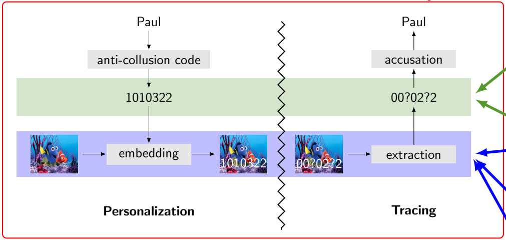
$$\begin{array}{ccccccccccc} Y & & 1 & & 1 & & 0 & & \dots & & 1 \\ X_j & & 1 & & 0 & & 1 & & \dots & & 0 \\ S_j & = & g(1,1,p_1) & + & g(1,0,p_2) & + & g(0,1,p_3) & & \dots & + & g(1,0,p_m) \end{array}$$

$g(1,1,p) = g(0,0,1-p) = \sqrt{(1-p)/p}$ $\qquad$ $g(1,0,p) = g(0,1,1-p) = -\sqrt{p/(1-p)}$

Tardos codes have been studied a lot



Contributions on fingerprinting



Design of an asymmetric fingerprinting protocol dedicated to Tardos codes (IH 2011)

Estimation of the pirates' strategy, and optimization of Tardos' scores computation (EI 2009 + TS 2010)

How to provoque multiple detections with `Broken Arrows`, and manage them with Tardos codes (MM&Sec 2008)

Improvement of `Broken Arrows` robustness (EI 2009)

Improvement of `Broken Arrows` security (MM&Sec 2010)

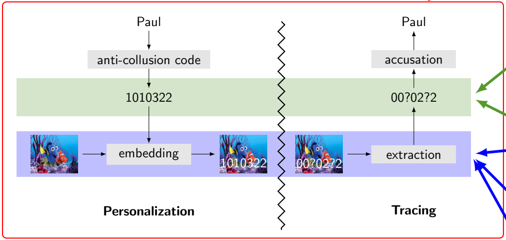
PhD A. Charpentier [08-11]

PhD F. Xie [07-10]

+ Implementation of FANTOMAS platform

M. Desoubeaux [07-09]

Contributions on fingerprinting



Design of an asymmetric fingerprinting protocol dedicated to Tardos codes (IH 2011)

Estimation of the pirates' strategy, and optimization of Tardos' scores computation (EI 2009 + TS 2010)

How to provoque multiple detections with `Broken Arrows`, and manage them with Tardos codes (MM&Sec 2008)

Improvement of `Broken Arrows` robustness (EI 2009)

Improvement of `Broken Arrows` security (MM&Sec 2010)

PhD A. Charpentier [08-11]

PhD F. Xie [07-10]

+ Implementation of FANTOMAS platform

M. Desoubeaux [07-09]

`Broken Arrows` + Tardos : a good match

With F. Xie and T. Furon [XFF08] (MM&Sec 2008)

Problem : "Fusion attacks are critical, and easy to perform."

Our idea : if the embedding technique is sufficiently robust, one can be able to detect multiple symbols in case of a fusion attack (e.g. averaging).

- `Broken Arrows` is a very robust zero-bit watermarking technique, designed in 2007 for BOWS-2 contest.
- We adapted it to embed q-ary symbols, and combined it with a q-ary Tardos code ($q = 4$).

$\Rightarrow$ Fusion attacks really lead to multiple symbols detections.

But Tardos codes were not designed to take them into account ...

- We modified the score computation to take them into account.

$\Rightarrow$ It worked really well (and even better than we thought).

Broken Arrows + Tardos : a good match

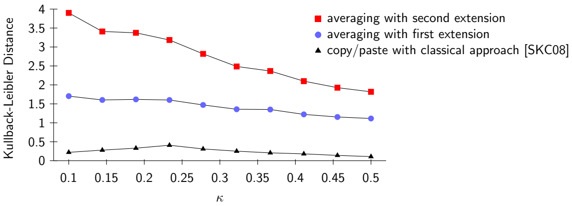
q-ary "Tardos" codes [SKC08] :

- Each $\mathbf{p}_i = (p_i^0, \ldots, p_i^{q-1}) \sim$ Dirichlet distribution of shape parameter κ
- Generation $\mathbb{P}(X_{ji} = a) = p_i^a$
- Score $S_j = \sum_{i=1}^m \delta_{Y_i = X_{ji}} g_1(p_i^{Y_i}) + (1 - \delta_{Y_i \neq X_{ji}}) g_0(p_i^{Y_i})$

We proposed two different extensions to take advantage of $\mathcal{Y}_i = \{Y_i^1, \ldots, Y_i^{L_i}\}$:

$$S_j = \sum_{i=1}^m \sum_{\ell=1}^{L_i} \delta_{Y_i = X_{ji}} g_1(p_i^{Y_i}) + (1 - \delta_{Y_i \neq X_{ji}}) g_0(p_i^{Y_i})$$

$$S_j = \sum_{i=1}^m \delta_{X_{ji} \in \mathcal{Y}_i} g_1\left(\sum_{\ell=1}^{L_i} p_i^{Y_i^\ell}\right) + (1 - \delta_{X_{ji} \notin \mathcal{Y}_i}) g_0\left(\sum_{\ell=1}^{L_i} p_i^{Y_i^\ell}\right)$$



Broken Arrows + Tardos : a good match

q-ary "Tardos" codes [SKC08] :

Each $\mathbf{p}_i = (p_i^0, \ldots, p_i^{q-1}) \sim$ Dirichlet distribution of shape parameter κ

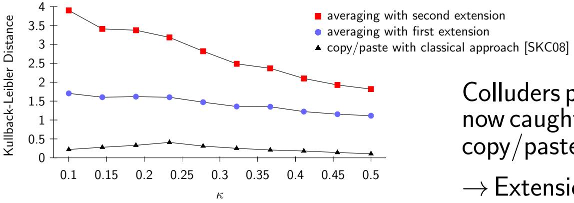
Generation $\mathbb{P}(X_{ji} = a) = p_i^a$

Score $S_j = \sum_{i=1}^m \delta_{Y_i = X_{ji}} g_1(p_i^{Y_i}) + (1 - \delta_{Y_i \neq X_{ji}}) g_0(p_i^{Y_i})$

We proposed two different extensions to take advantage of $\mathcal{Y}_i = \{Y_i^1, \ldots, Y_i^{L_i}\}$:

$$S_j = \sum_{i=1}^m \sum_{\ell=1}^{L_i} \delta_{Y_i = X_{ji}} g_1(p_i^{Y_i}) + (1 - \delta_{Y_i \neq X_{ji}}) g_0(p_i^{Y_i})$$

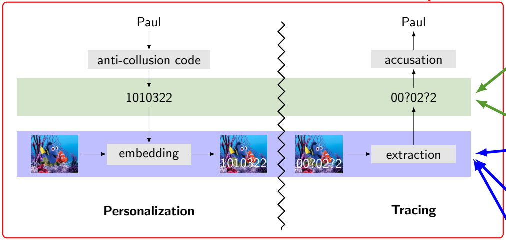
$$S_j = \sum_{i=1}^m \delta_{X_{ji} \in \mathcal{Y}_i} g_1\left(\sum_{\ell=1}^{L_i} p_i^{Y_i^\ell}\right) + (1 - \delta_{X_{ji} \notin \mathcal{Y}_i}) g_0\left(\sum_{\ell=1}^{L_i} p_i^{Y_i^\ell}\right)$$



Colluders performing an averaging are now caught more efficiently than for a copy/paste attack!

$\rightarrow$ Extension of this work in [SKSC11].

Contributions on fingerprinting



Design of an asymmetric fingerprinting protocol dedicated to Tardos codes (IH 2011)

Estimation of the pirates' strategy, and optimization of Tardos' scores computation (EI 2009 + TS 2010)

How to provoque multiple detections with `Broken Arrows`, and manage them with Tardos codes (MM&Sec 2008)

Improvement of `Broken Arrows` robustness (EI 2009)

Improvement of `Broken Arrows` security (MM&Sec 2010)

PhD A. Charpentier [08-11]

PhD F. Xie [07-10]

+ Implementation of FANTOMAS platform

M. Desoubeaux [07-09]

Dynamical optimization of Tardos' scores

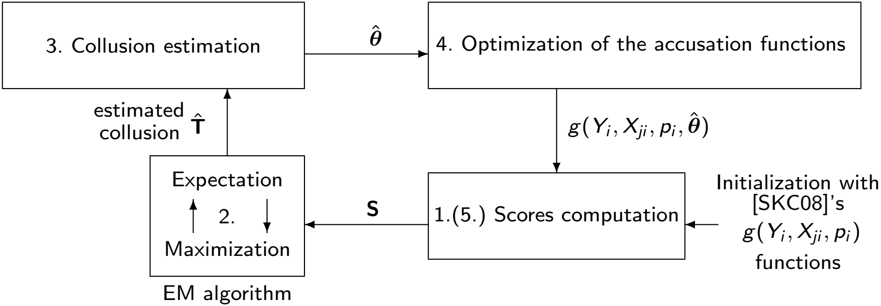
With A. Charpentier and T. Furon [CXFF09] (EI 2009), [CFF10] (TS 2010)

"Are [Tar03,Tar08,SKC08]'s parameters and functions the best ones ?"

- **Tardos [Tar03,Tar08] and Škorić et al. [SKC08]** : for a given c, the scores distributions, $\mathcal{N}_I = \mathcal{N}(0, \sigma_I^2)$ and $\mathcal{N}_C = \mathcal{N}(\mu_C, \sigma_C^2)$, remain the same whatever the colluders' strategy.
- **Furon et al. [FGC08]** :
  - When the colluders' strategy is not known, [Tar03,Tar08,SKC08]'s choices lead to the maximal Kullbach-Leibler Distance between $\mathcal{N}_I$ and $\mathcal{N}_C$. (binary case)
  - BUT if we know the colluders' strategy, we can derive functions $g(Y_i, X_{ji}, p_i)$ leading to a higher Kullbach-Leibler Distance between $\mathcal{N}_I$ and $\mathcal{N}_C$. (binary case)

Dynamical optimization of Tardos' scores

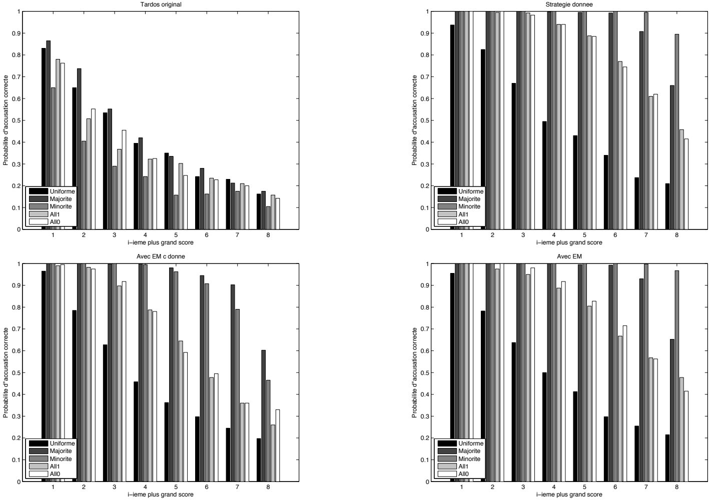
We pushed further, providing a better optimization of the scores, and a way to estimate the colluders' strategy $\theta = \{\mathbb{P}(Y_i = 1|\Sigma_i = \sigma_i), \sigma_i = 0..c\}_{i=1..m}$.
Assumption : the strategy is the same for all the components



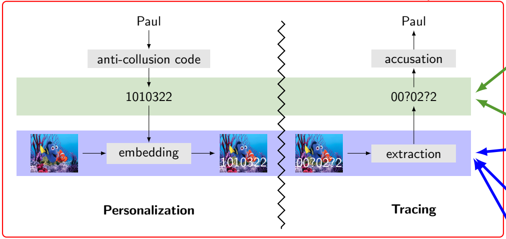
Dynamical optimization of Tardos' scores

$\mathbb{P}(k\text{-th highest score is guilty})$ $\quad k \in \{1, \ldots, 8\}$, $m = 1000$, $c = 8$, $n = 5000$.

[SKC08] / known c, θ / given c, estimated $\hat{\boldsymbol{\theta}}$ / all are estimated



Contributions on fingerprinting



Design of an asymmetric fingerprinting protocol dedicated to Tardos codes (IH 2011)

Estimation of the pirates' strategy, and optimization of Tardos' scores computation (EI 2009 + TS 2010)

How to provoque multiple detections with `Broken Arrows`, and manage them with Tardos codes (MM&Sec 2008)

Improvement of `Broken Arrows` robustness (EI 2009)

Improvement of `Broken Arrows` security (MM&Sec 2010)

PhD A. Charpentier [08-11]

PhD F. Xie [07-10]

+ Implementation of FANTOMAS platform

M. Desoubeaux [07-09]

How to prevent parties from cheating

With A. Charpentier, T. Furon and I. Cox [CFFC11] (IH 2011)

"Can we trust the provider who delivers the pieces of content ?"

In the usual (symmetric) scenario . . .

- An untrustworthy provider may frame an innocent buyer !
- Any accused buyer can argue he/she has been framed by an untrustworthy provider !

Asymmetric fingerprinting protocols have been introduced in [PS96].

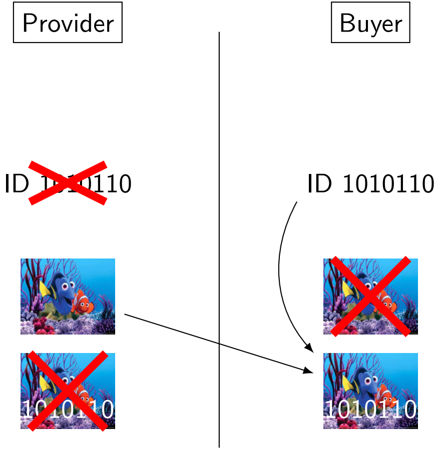
- Most of them (not ours) also provide anonymity of the Buyer.
- Very few also (not ours) provide privacy on the delivered content.

✓ Embedding and tracing techniques are sufficiently mature today to provide complete specifications for such protocols.
☹ No existing protocol is compliant with Tardos codes.

How to prevent parties from cheating

Designing a protocol based on Tardos codes : rules and challenges.

1. Generation of the ID (fingerprint)
2. Embedding

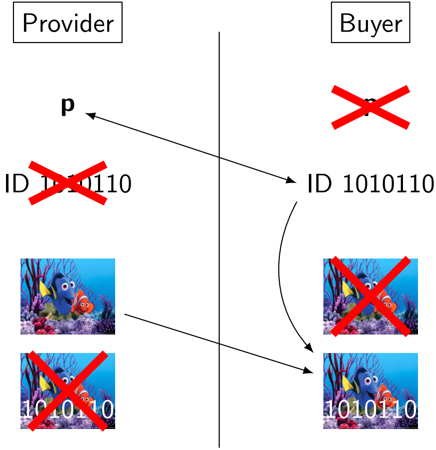


How to prevent parties from cheating

Designing a protocol based on Tardos codes : rules and challenges.

1. Generation of the ID (fingerprint)

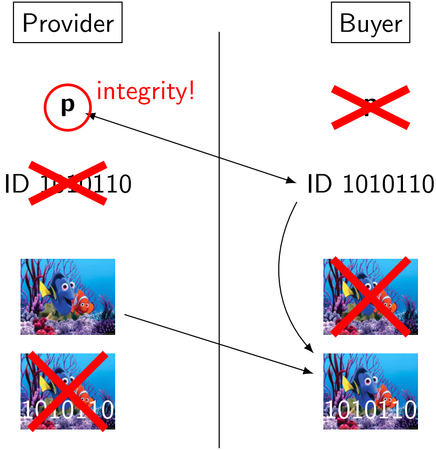
2. Embedding



How to prevent parties from cheating

Designing a protocol based on Tardos codes : rules and challenges.

1. Generation of the ID (fingerprint)
2. Embedding

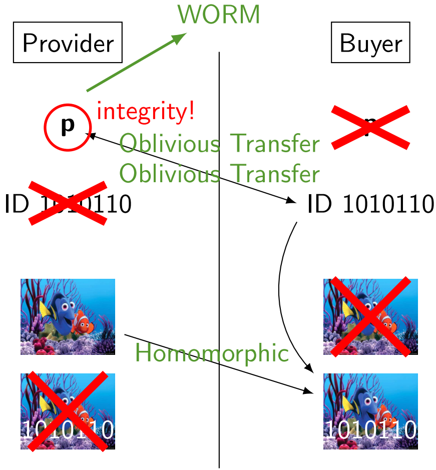


How to prevent parties from cheating

Designing a protocol based on Tardos codes : rules and challenges.

1. Generation of the ID (fingerprint)

2. Embedding



Fingerprinting : conclusion and further work

All aspects addressed : watermarking, anti-collusion code, protocols.

Tardos codes :

- ✓ Multiple detection for q-ary Tardos codes.
- ✓ Estimation of the colluders' strategy, and optimization of the accusation for binary Tardos codes.
- ϟ Optimization should be extended to q-ary Tardos codes (really hard).

Asymmetric fingerprinting protocol :

- ✓ First protocol compliant with binary Tardos codes.
- ϟ Proofs ?
- ϟ Anonymity and Privacy ?

Outline

General conclusion and further work

Crossings offer new points of views, new ideas, and a complete overview.

- ✓ Cryptanalysis methodology applied to the definition and study of watermarking security.
- ✓ Syndrome Coding in Steganography.
- ✓ An asymmetric fingerprinting protocol, with all primitives detailed.

My favorite prospects :

- ϟ Syndrome Coding in Steganography.
- ϟ Asymmetric fingerprinting protocol : proofs, commutative encryption, anonymity/privacy.
- ϟ Anonymity issues in general.
- ϟ Implementation of homomorphic encryption schemes.
  PhD S. Fau [11-14]

Dynamical optimization of Tardos' scores

Optimization step (4.) : dominating term in the K-L Distance

		Colluders' strategy				
	Accusation strategy	Uniform	Majority	Minority	All1	All0
c=3	Uniform	**98 (71)**	106 (80)	100 (53)	97 (66)	97 (66)
	Majority	96 (67)	**110 (84)**	100 (34)	95 (59)	95 (59)
	Minority	81 (50)	59 (38)	**112 (75)**	89 (56)	89 (56)
	All1	83 (69)	88 (73)	88 (62)	**114 (68)**	84 (68)
	All0	83 (69)	88 (73)	88 (62)	84 (68)	**114 (68)**
c=4	Uniform	**98 (71)**	106 (80)	105 (44)	99 (62)	99 (62)
	Majority	96 (67)	**110 (84)**	105 (17)	97 (50)	97 (50)
	Minority	61 (34)	25 (15)	**128 (91)**	88 (53)	88 (53)
	All1	79 (65)	83 (63)	88 (72)	**121 (67)**	87 (67)
	All0	79 (65)	83 (63)	88 (72)	87 (67)	**121 (67)**
c=5	Uniform	**98 (71)**	110 (83)	110 (33)	100 (58)	100 (58)
	Majority	94 (63)	**120 (93)**	113 (-22)	98 (35)	98 (35)
	Minority	37 (19)	-20 (-17)	**155 (121)**	82 (52)	82 (52)
	All1	77 (59)	83 (47)	90 (90)	**128 (69)**	90 (69)
	All0	77 (59)	83 (47)	90 (90)	90 (69)	**128 (69)**

(Furon et al. IH 08) ; remind that in [SKC08] it is 64 whatever the strategy.

Dynamical optimization of Tardos' scores

- ✓ Kullbach-Leibler Distance between $\mathcal{N}_I$ and $\mathcal{N}_C$ is maximized, and the accusation process is run automatically : traceability is more efficient.
- ✓ For a given Kullbach-Leibler Distance (tracing efficiency), this provides a way to use a shorter code !
- ✓ Works better for large c.
- ☹ The efficiency is better for some strategies than for others, and we do not know why.
- ☹ c is often over-estimated. So it is safer to accuse only the highest score and not the c highest ones.
- ☹ Does not work well for small c (we need at least $c = 8$).
- ϟ Extension of the optimization step to q-ary case is really hard.

How to prevent parties from cheating

- ✓ The first asymmetric fingerprinting protocol compliant with Tardos codes.
- ✓ All the steps were considered in detail.
- ↯ Extension to q-ary Tardos codes.
- ↯ Proofs ?
- ↯ Commutative Encryption Schemes as an alternative to traditional Oblivious Transfer protocols.
- ↯ Extension to anonymous and/or private protocols.

How I found my own way

① INRIA-Rocquencourt, team CODES
cryptography, error correcting codes, information theory

① Univ. Cergy Pontoise

② Univ. Paris XI – LRI, team ALGO
error correcting codes, information theory, algorithmics

③ USTL/CNRS – LIFL, team RD2P
operating systems, smart cards, ad-hoc networks

④ CNRS – IRISA, team TEMICS
source coding, inf. theory, image proc., information hiding

⑤ CNRS – Lab-STICC, team SFIIS
& Télécom Bretagne, dpt. ITI
security, signal and image proc., information hiding