Content Fingerprinting and Security

Gwenaël Doërr – Technicolor R&D France

technicolor

Security Laboratories

Agenda

Generalities

- Definitions
- Applications

Attacks against fingerprinting systems

Security fixes

- Obfuscation techniques
- Crytographic primitives

Conclusion

A Confusing Terminology

Content Fingerprinting

Definition: compact binary representation of multimedia content that is robust to an array of signal processing primitives

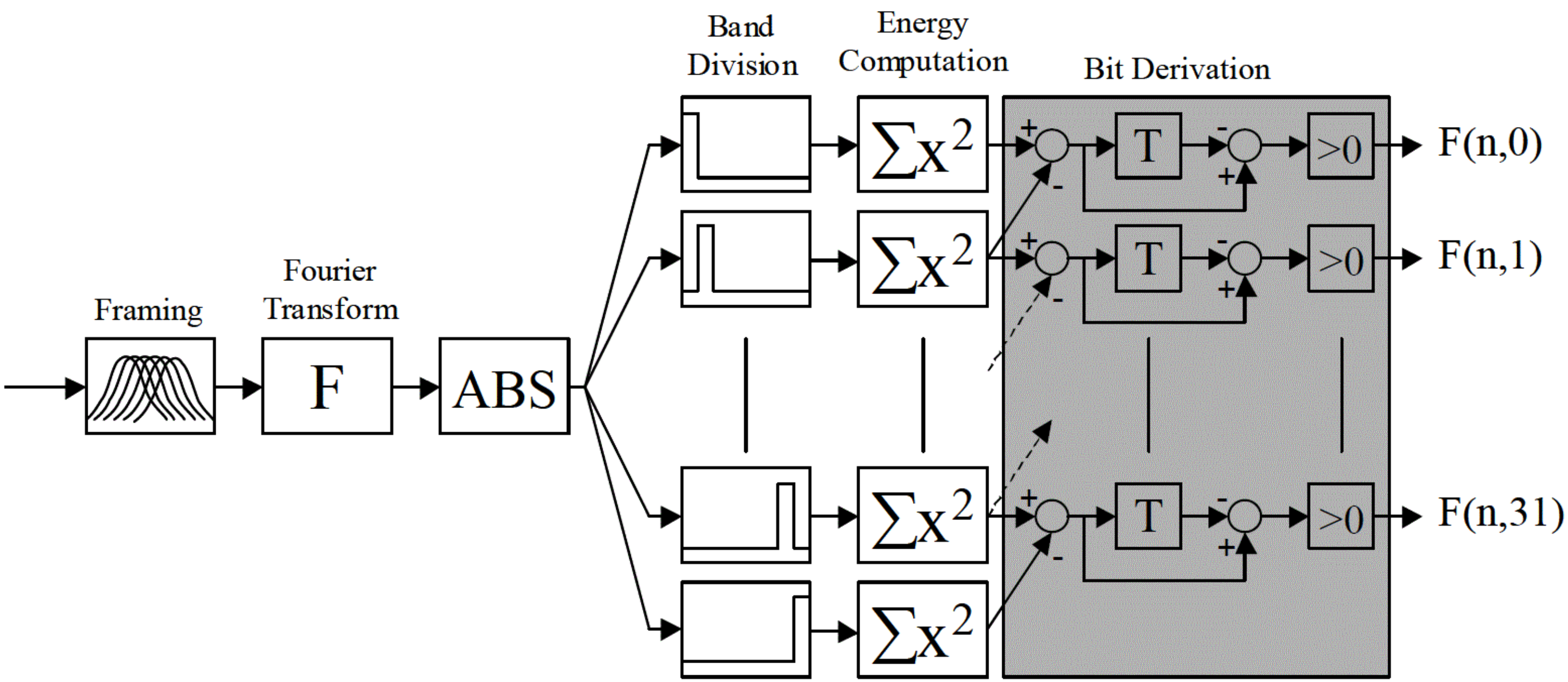
Baseline framework

- Robust representation: filter banks, transforms, features extraction
- Quantization: ad-hoc, K-means, etc
- Binarization

Global fingerprints vs. local fingerprints

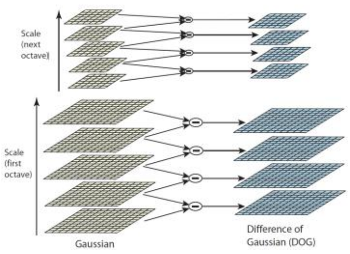
Efficient nearest neighbor search

Example #1: Audio Fingerprinting



J. Haitsma, T. Kalker, and J. Oostveen, “*Robust Audio Hashing for Content Identification*”, CBMI 2001

Example #2: SIFT-based Image Fingerprinting



1. Keypoints detection
   - Scale-space representation e.g. DoG, LoG, etc
   - Local extrema detection ⇒ *location* and *scale*
   - Localization refinement ⇒ <u>*contrast*</u>

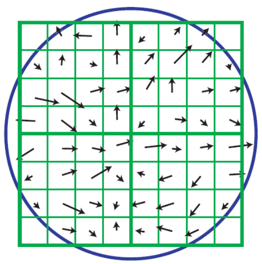
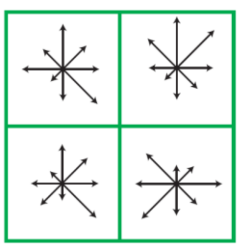


Image gradients

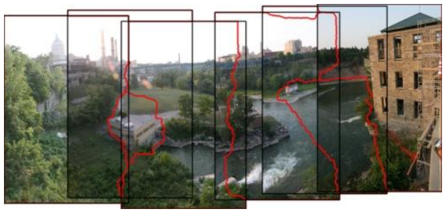
2. Orientation assignment
   - Gradient in a local region around keypoint (orientation and magnitude)
   - Weighted histogram of gradient directions ⇒ <u>*orientation*</u>



3. Keypoint descriptor
   - Array (4×4) of orientation (8) histograms ⇒ *description*

D. G. Lowe, "*Distinctive Image Features from Scale-Invariant Keypoints*", IJCV 2004

Applications



Panorama
2D/3D

Companion screen

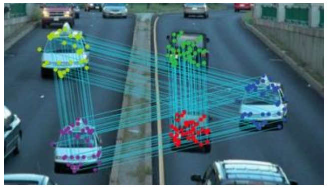
Law enforcement

Semi-blind watermark registration

Content-based
retrieval

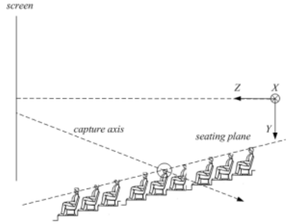
Name that content

Broadcast monitoring
Audience measurement



Copy-move
forgery detection

Near-duplicate
detection



Pirate localization

Content autentication

Adversarial Signal Processing

Potential for money and/or strict laws $\Rightarrow$ pirates and attacks

- Wash out digital watermarks
- Impersonate biometric traits
- Clean-up statistical digital traces
- Etc

Objective of the adversary: learn or infer hidden parameters of the system to modify its expected behavior

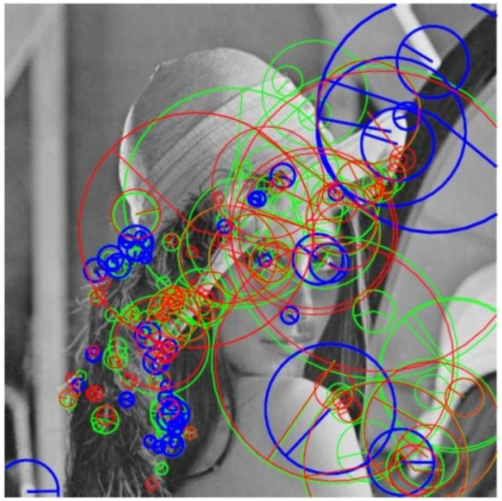
- Sensitivity analysis to learn
- Decision boundaries $\Rightarrow$ switch decisions

Strong links to game theory

- Trade-off robustness $\leftrightarrow$ security

Attacks against SIFT: Remove Key Points

Objective: tamper the local neighborhood of keypoints to make them fall below the detection threshold



- Smoothing attack e.g. local Gaussian blur
- Collage attack
  - Select a patch without keypoint close to the neighborhood of the attacked keypoint
  - Alpha-blending
- Removal with Minimum Distortion (RMD) attack
  - Patch of minimal Euclidean norm that yields a target contrast value (≈ Mexican hat)
- Alternate the type of attack depending on the type of keypoint

S.C. Hsu, C.Y. Lu, and C.S. Pei, "Secure and Robust SIFT", ACM MM 2009

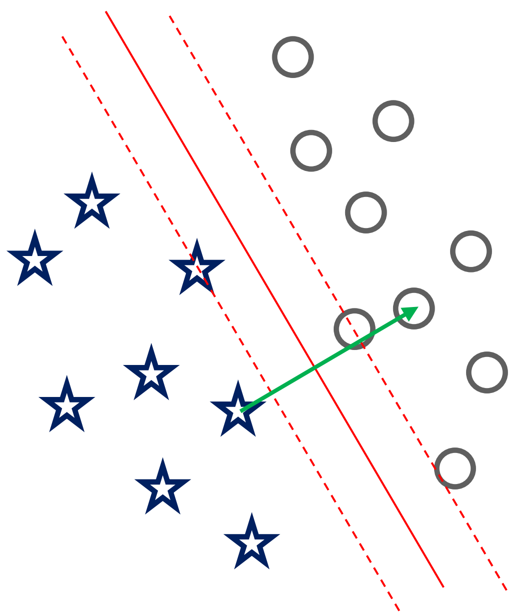
T.T. Do, E. Kijak, T. Furon, and L . Amsaleg, "*Deluding Image Recognition in SIFT-based CBIR systems*", ACM MiFor 2010

I. Amerini, M. Barni, R. Caldelli, and A. Constanzo, "*Counter Forensics of SIFT-based Copy-move Detection by Means of Keypoints Classification*", EURASIP JIVP 2013:18.

Attacks against SIFT: Change Orientation

Objective: change the orientation of the keypoints

- Different from rotating the whole support region
- Changing the orientation by $\pi/2$ is the most damaging

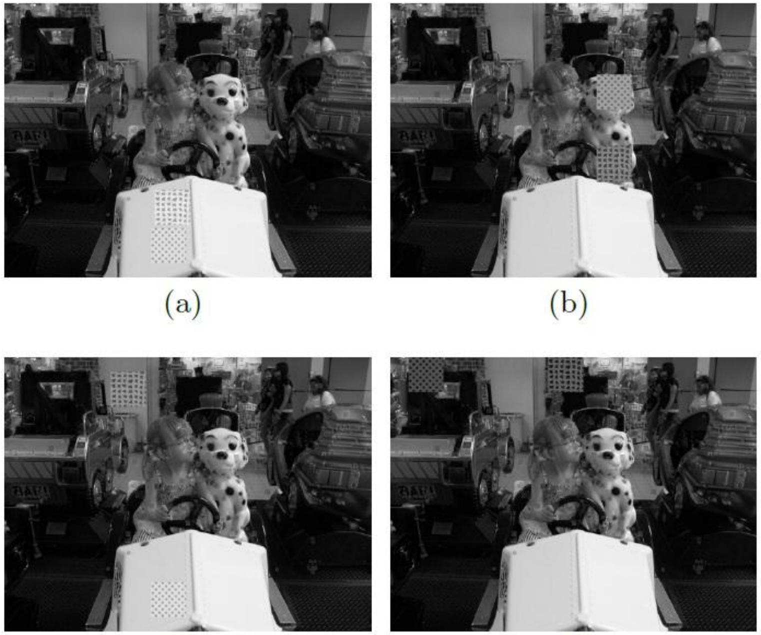


- Collect a large number of SIFT patches
- Train a 2-class SVM for each pair of orientations
  - Hyperplane H_θ separating patch θ and $\theta+\pi/2$
- For each keypoint
  - Identify the associated hyperplane H_θ
  - Add the patch ε that makes the keypoint move in the direction orthogonal to H_θ

T.T. Do, E. Kijak, T. Furon, and L . Amsaleg, "*Enlarging Hacker's Toolbox: Deluding Image Recognition by Attacking Keypoint Orientations*", IEEE ICASSP 2012

Attacks against SIFT: Introduce Distractors

Objective: insert a visual patch to artificially bias the retrieval/ recommendation system towards an intended item



(a) (b) (c) (d)

Design rules

- Small
- High density of keypoints

Placement guidelines

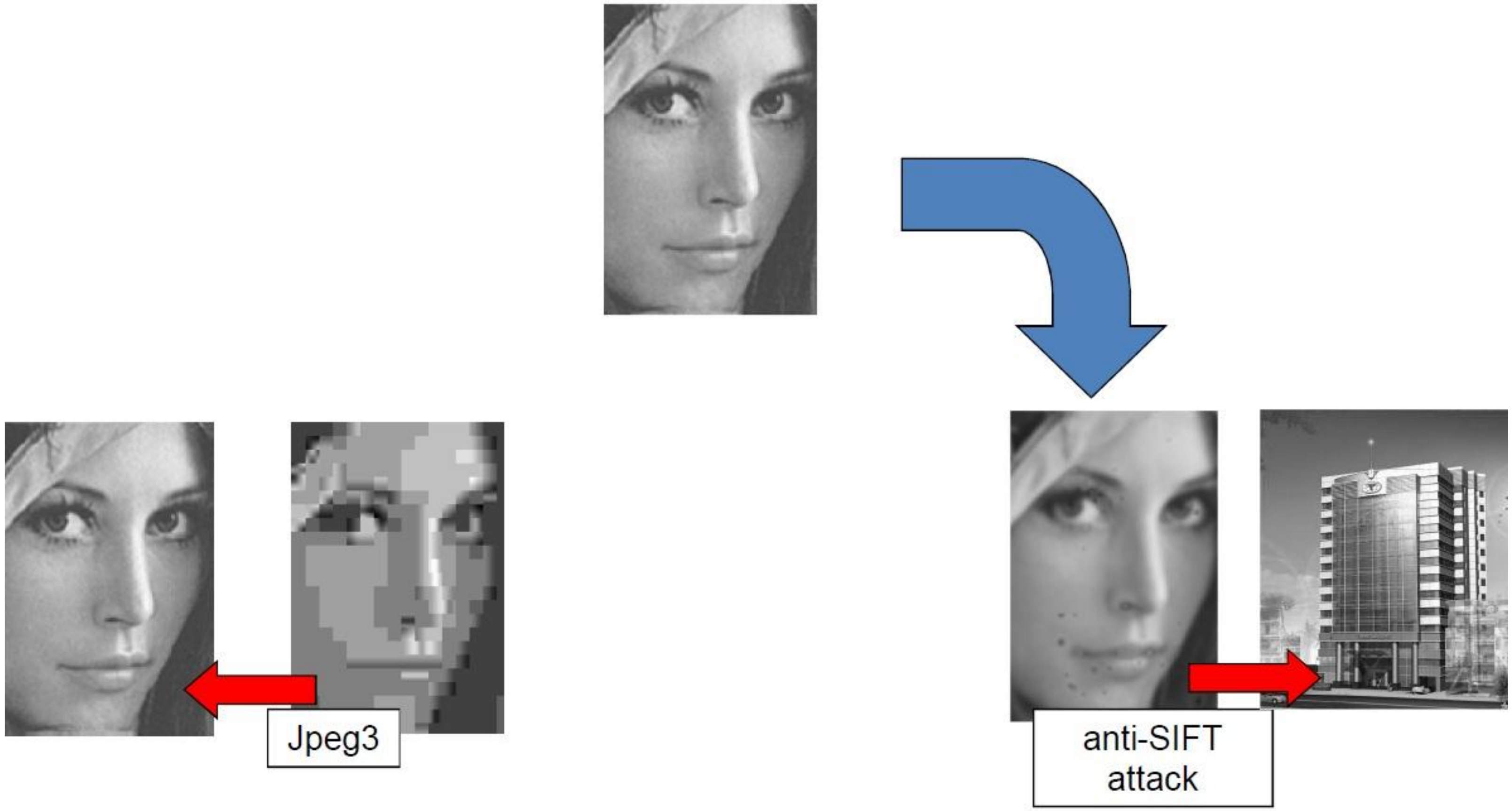
- Low induced distortion
- High original density of keypoints
- Account for visual attention

Obfuscation recommendation

- Blur patch to avoid separation

T.T. Do, L. Amsaleg, E. Kijak, and T. Furon, “*Security-oriented Picture-in-Picture Visual Modifications*”, ACM ICMR 2012

Combining All Attacks Together



ROBUSTNESS ≠ **SECURITY**

Attacks against SIFT: Confidentiality (*Privacy*)

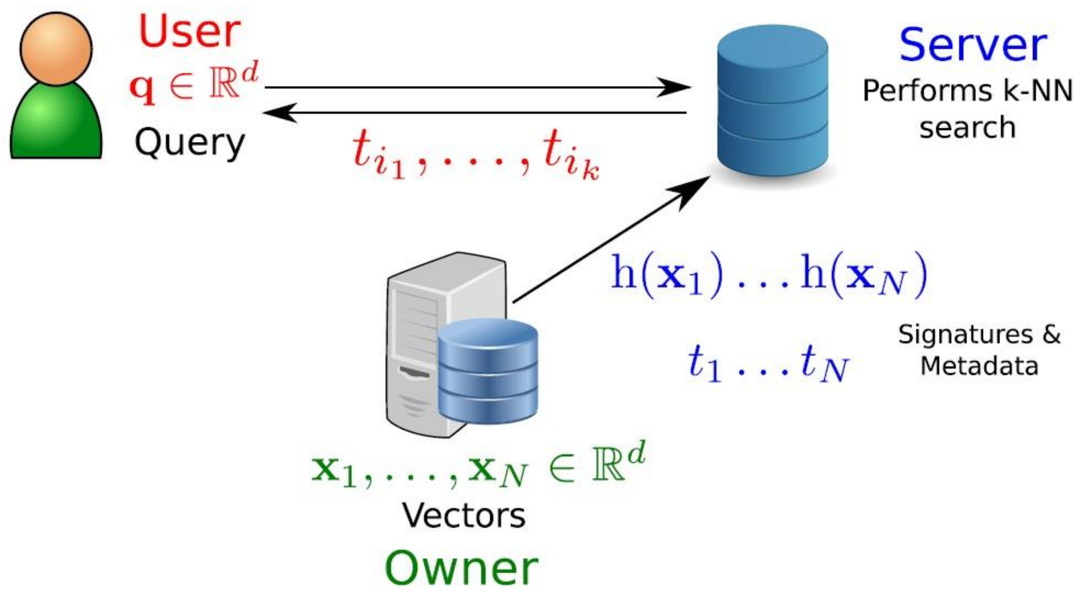
Original image

Reconstruction from SIFT description

+ inpainting

P. Weinzaepfel, Hervé Jégou, and Patrick Pérez, "*Reconstructing an Image from its Local Descriptors*", CVPR 2011

Threat Analysis of a Content-based Retrieval System



Curious but honest Server

- Reconstruct $\mathbf{x}_i$ from h($\mathbf{x}_i$)
- Cluster the database vectors from {h($\mathbf{x}_i$)}
- Reconstruct **q** from h(**q**)
- Detect similar queries (from one or different users)

Defense Mechanims

Obfuscation techniques

- Security by obscurity
- Key-dependent parametrization of the system

Cryptographic techniques

- Hash function
- Homomorphic encryption
- Zero-knowledge protocols
- Etc.

Obfuscation Techniques

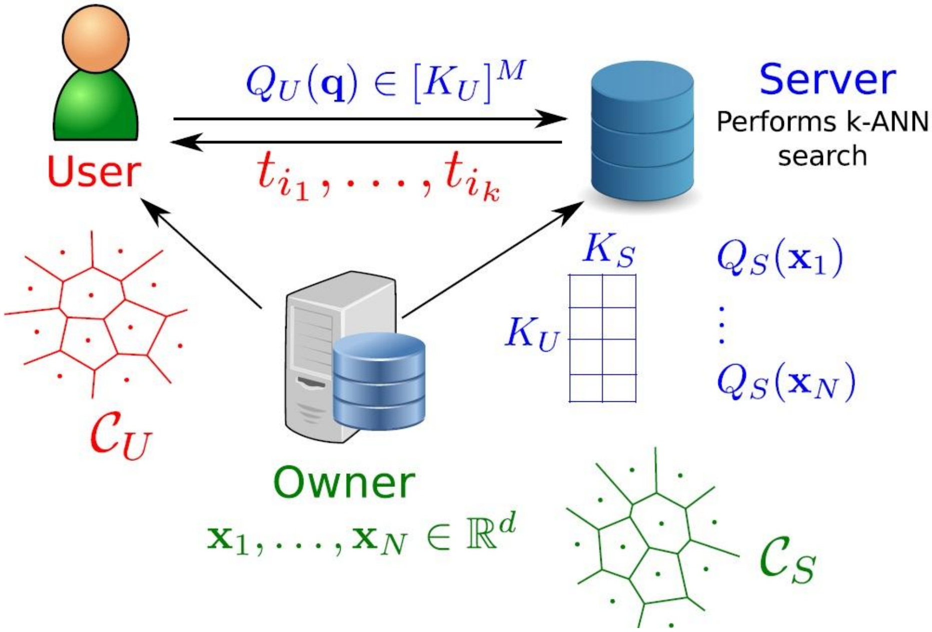
1. Random tiling of the image
2. Compute some statistics for each tile
   e.g. mean, variance, etc
3. Randomized rounding

R. Venkatesan, S.-M. Koon, M. H. Jakubowski, and P. Moulin, "*Robust Image Hashing*", ICIP 2000

1. Generate low-pass pseudo-random patterns
2. Project the content onto those patterns
3. Take the sign of the correlation value
4. Generate the binary digest with a heuristic design

J. Fridrich and M. Goljan, "*Robust Hash Functions for Digital Watermarking*", ICIT 2000

Obfuscation Techniques: Randomizing the Quantizer



Baseline idea: randomize the quantizer & different quantizer for Server and User

Randomized quantizers

- Random training subset
- Random initialization vector
- Stop before convergence

Curious but honest Server

- ~~Reconstruct $\mathbf{x}_i$ from h($\mathbf{x}_i$)~~
- Cluster the database vectors

- ~~Reconstruct **q** from h(**q**)~~
- *Detect similar queries*

B. Mathon, T. Furon, L. Amsaleg, and J. Bringer, "*Secure and Efficient Approximate Nearest Neighbors Search*", ACM IHMMSec 2013

The Issue of Security Assessment

How much security is provided by heuristic obfuscation techniques?

- Uniformly distributed fingerprints?
- Different keys $\Rightarrow$ different fingerprints

Several metrics based on information theory

☺ Mutual information, differential entropy, etc

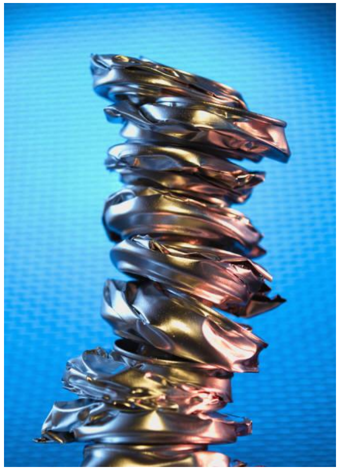
☹ No security proof

What does it mean to be “more secure”?

A. Swaminathan, Y. Mao, and M. Wu, “*Robust and Secure Image Hashing*”, IEEE TIFS 2006

V. Monga and K. Mihcak, “*Robust and Secure Image Hashing via Non-negative Matrix Factorizations*”, IEEE TIFS 2007

The Cryptographic Approach: Hash Function



Cryptographic hash functions
(typically used for authentication)

- High sensitivity: $a \approx b \Rightarrow h(a) \neq h(b)$
- Non invertibility
- Small collision probability

Visual hash: content fingerprint + hash function

- Inherits robustness from the fingerprint and security from the hash
- Does not really work in practice
  - Content fingerprinting is not strictly robust (even with ECC decoder hack)

The Cryptographic Approach: Homomorphic Encryption

$$\mathrm{E}_K(A+B)=\mathrm{E}_K(A)\times\mathrm{E}_K(B)$$

Linear operations directly in the encrypted domain

- Signal processing in the encrypted domain
- Privacy enhancement technologies

☺ Provides all the security features that you could dream of

☺ Recent leap forward with Gentry's fully homomorphic scheme

☹ Many operations not supported e.g. thresholding, trigonometry, …

☹ Overhead: big and slow!

R.L. Lagendijk, Z. Erkin, and M. Barni, "*Encrypted Signal Processing for Privacy Protection*", IEEE SPM, 2013

The Cryptographic Approach: Miscellaneous

Secure Multiparty Computation

Hamming embedding

Attribute-based encryption

… and many more

P. Boufounos and S. Rane, “Secure Binary Embeddings for Privacy Preserving Nearest Neighbors”, IEEE WIFS 20101
S. Rane and W. Sun, “An Attribute-based Framework for Privacy Preserving Image Querying”, IEEE ICIP 2012

Conclusion

Content fingerprinting is now part of the signal processing toolbox

- Depending on the application case, security may be an issue

First attacks on fingerprinting systems

- Rudimentary & focus on disrupting fingerprint matching

No ideal security fix yet

- Obfuscation techniques are ad-hoc and provide no provable security
- Cryptography-based solutions are not practical

Is this relevant in practice or only an academic mind game?

Questions (/ Answers)

Email: gwenael.doerr@technicolor.com

technicolor
Security Laboratories