stochastic information processing

Multimedia Security and Privacy

Slava Voloshynovskiy

Copyright notice

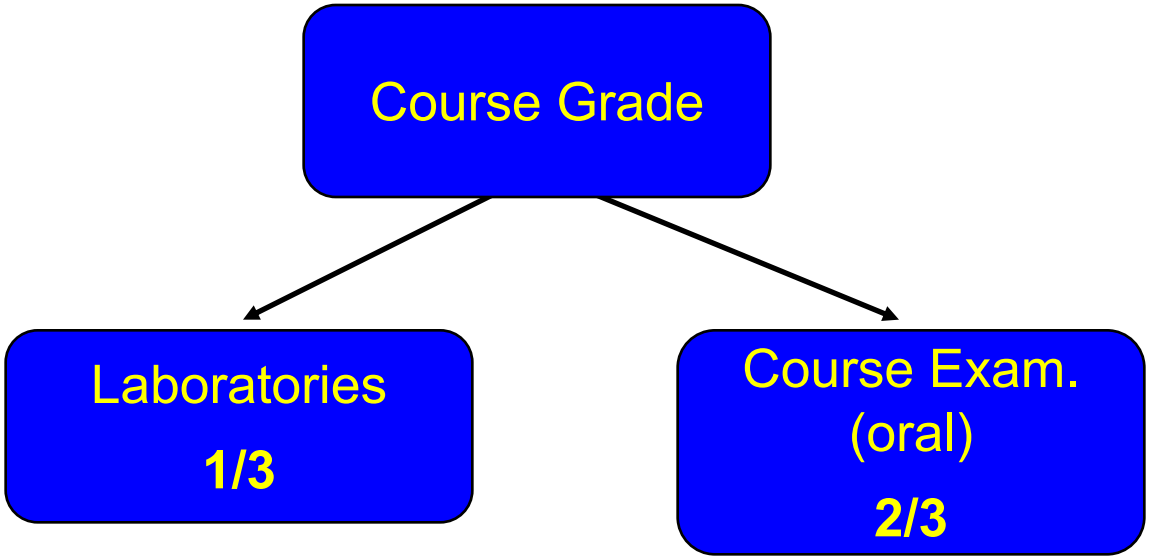
This course uses in part some materials from

- Book: I. Cox, M. Miller and J. Bloom, “Digital Watermarking: Principles and Practice”, *Morgan Kaufmann Publisher Inc.*, San Francisco, 2001;
- Book: M. Barni and F. Bartolini, “Watermarking Systems Engineering: Enabling Digital Assets Security and Other Applications”, *Marcel Dekker*, 2004;

And some papers, books, presentations that will be indicated along the course.

Grading Policy

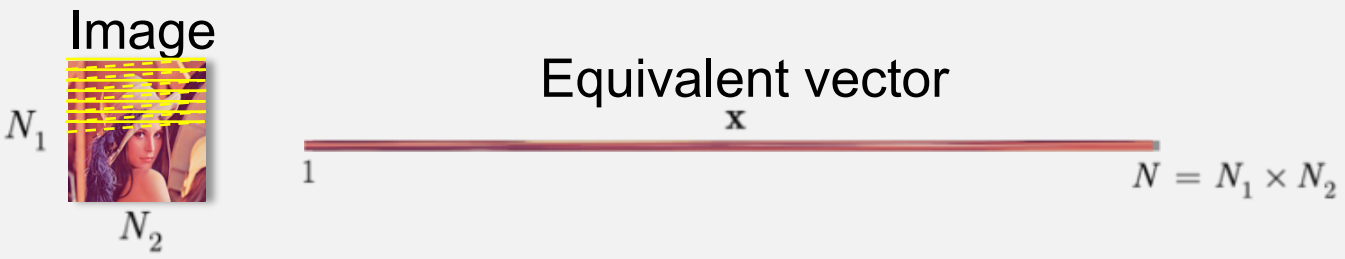
The final grade will be determined based on laboratories and a final examination.



Notations

Notations

- X Random variable with some distribution $p(x)$
- $\mathbf{X}$ Random vector of length N: $\mathbf{X} = \{X_1, X_2, \cdots, X_N\}$
- x Realization of random variable
- $\mathbf{x}$ Realization of random vector $\mathbf{x} = \{x_1, x_2, \cdots, x_N\}$



Notations

	Realization	Random
Scalar	x	X
Vector	$\mathbf{x} = x^N$	$\mathbf{X} = X^N$

$x_i \in \mathcal{X}$ or $x_i \in S_X$

- Discrete alphabets
  - $\mathcal{X} = \mathbb{Z}$ - integer
  - $\mathcal{X} = \{0,1\}$ - binary
  - $\mathcal{X} = \{a, b, \cdots, z\}$
  - $\mathcal{X} = \{January, February, \cdots, Decemeber\}$
- Continuous alphabets
  - $\mathcal{X} = \mathbb{R}$

- Vector $\mathbf{x} \triangleq \begin{bmatrix} | \\ \mathbf{x} \\ | \end{bmatrix} = \begin{bmatrix} x_1 \\ x_2 \\ \vdots \\ x_N \end{bmatrix} =$ 

$$\mathbf{X} \triangleq \begin{bmatrix} | \\ \mathbf{X} \\ | \end{bmatrix} = \begin{bmatrix} X_1 \\ X_2 \\ \vdots \\ X_N \end{bmatrix}$$

Notations

$\Pr\left[X = x \right]$ Probability of event

- Discrete random variables

  $X \sim p_X(x)$ (or $X \sim p(x)$) Probability mass function (**pmf**) of discrete random variable X

  $p_X(x \mid Y = y)$ (or $p_X(x \mid y)$) Conditional pmf of discrete random variable X given Y

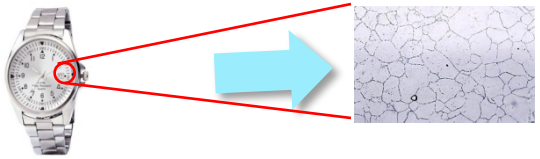
- Continuous random variables

  $X \sim f_X(x)$ (or $X \sim f(x)$) Probability density function (**pdf**) of continuous random variable X

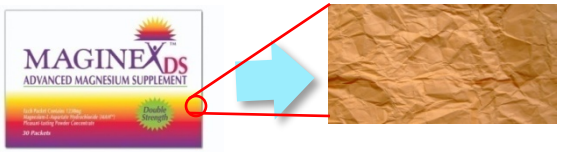
  $f_X(x \mid Y = y)$ (or $f_X(x \mid y)$) Conditional pdf of continuous random variable X given Y

Applications and main concerns

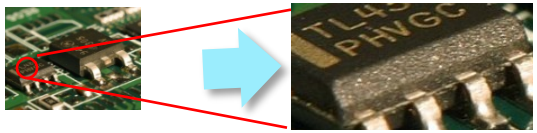
Physical Objects



Luxury products = “Annoying”

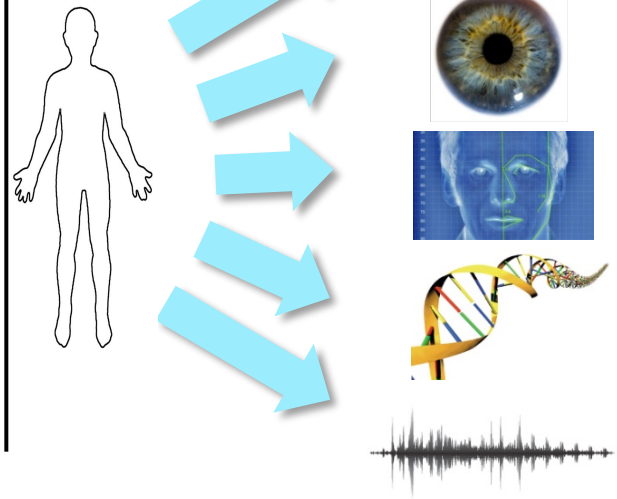


Medicine = “Dangerous”



Electronics = “Security as whole”

Humans



Digital Content

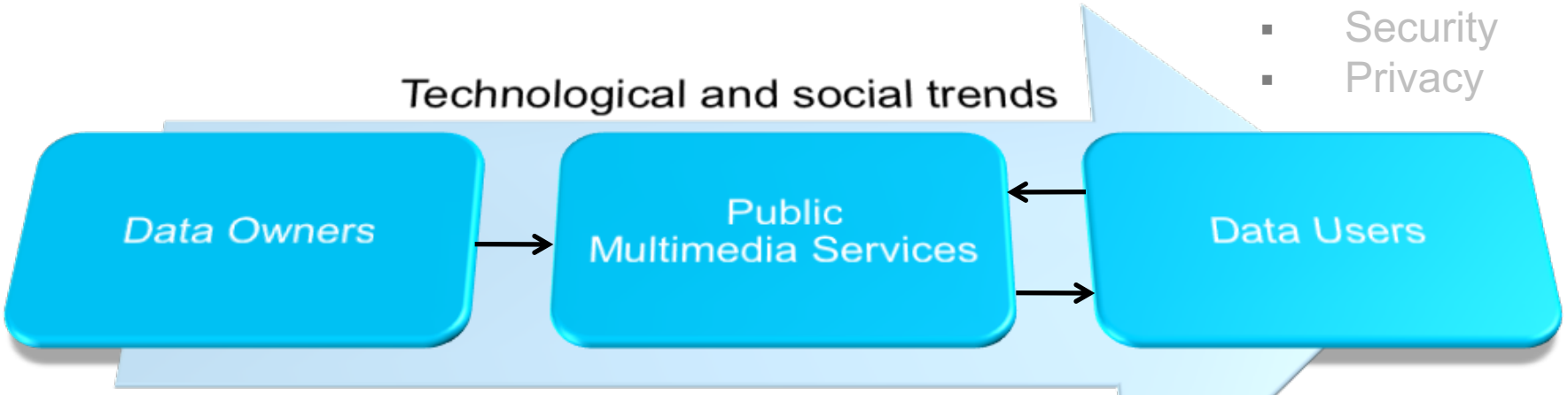
Images Videos Audios Text docs

Online sharing services

Main concerns

- Identification: identity, authenticity, origin, ownership, …
- Tracking and tracing
- Automatic tagging

Modern networked multimedia applications



Concerns
- Security
- Privacy

Private data
- Multimedia (MM)
- Medical data
- Biometrics …

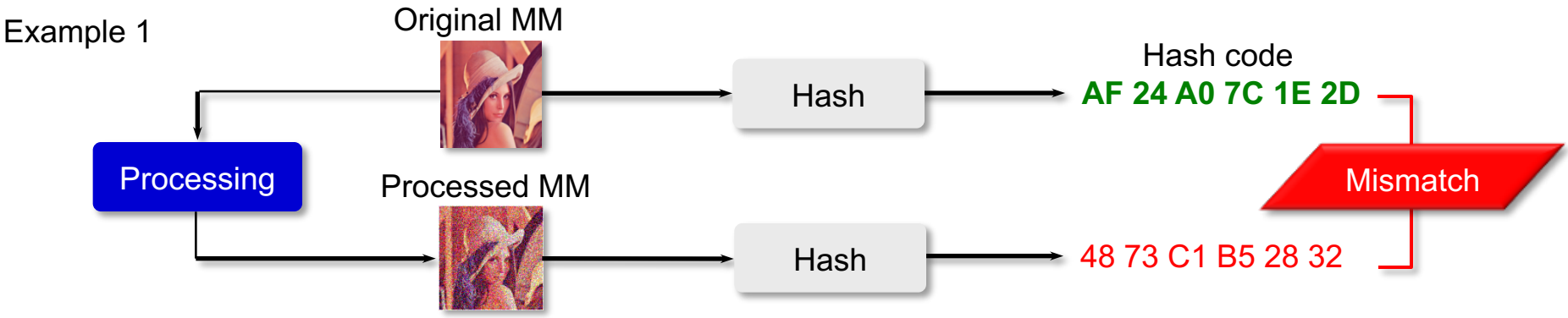
Untrusted intermediaries
- Search engines
- File sharing systems
- Social sites …

Tasks
- Search
- Identification
- Authentication
- Classification
- Mining …

Multimedia security

Why are traditional security tools not suitable for multimedia?

- Traditional *security*:
  - Cryptographic encryption (confidentiality of information)
  - Cryptographic hashes (authentication, trust, access control)
- Main concerns of classical crypto-based algorithms:
  - Sensitivity to noise and unintentional distortions in input data
  - Data handling in the encrypted domain



Multimedia privacy

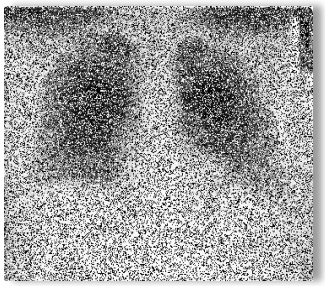
Why are traditional privacy tools not suitable for multimedia?

- Traditional *privacy protection*:
  - Data owner (protect data): based on data degradation and randomization (noise addition, lossy compression, data removal, dimensionality reduction…)
  - Data user (protect requests): based on anonymization, randomized rules
- Main concerns of classical privacy preserving algorithms:
  - Reduction of accuracy (data utility)
  - Not very efficient against experienced attackers (sensitivity analysis)

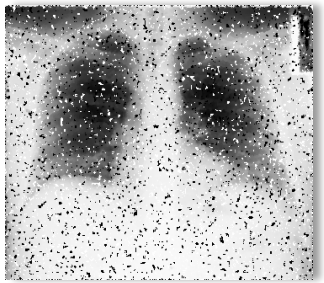
Example 2:

Original

50% removed based on key

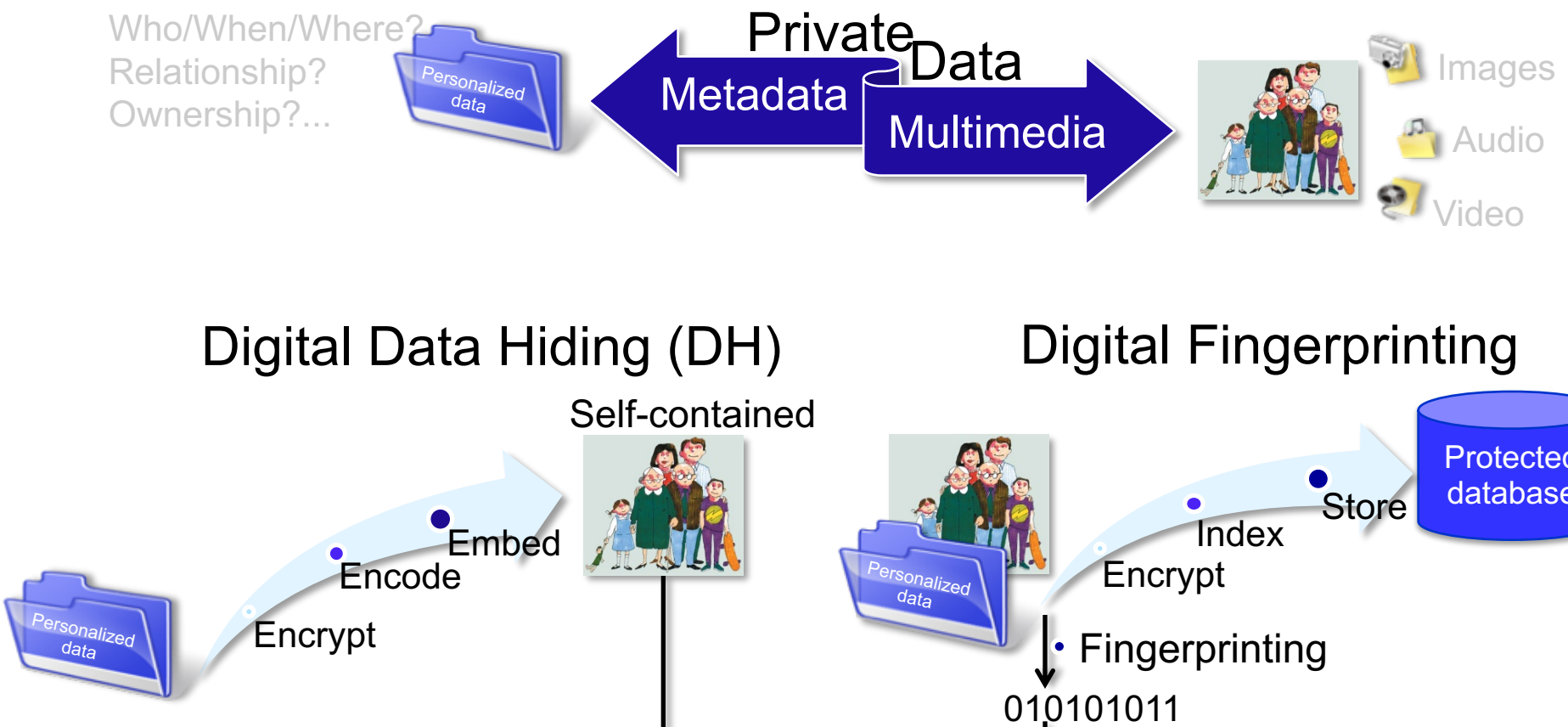


Median 9x9



Adaptive recovery

New tools for multimedia privacy and security

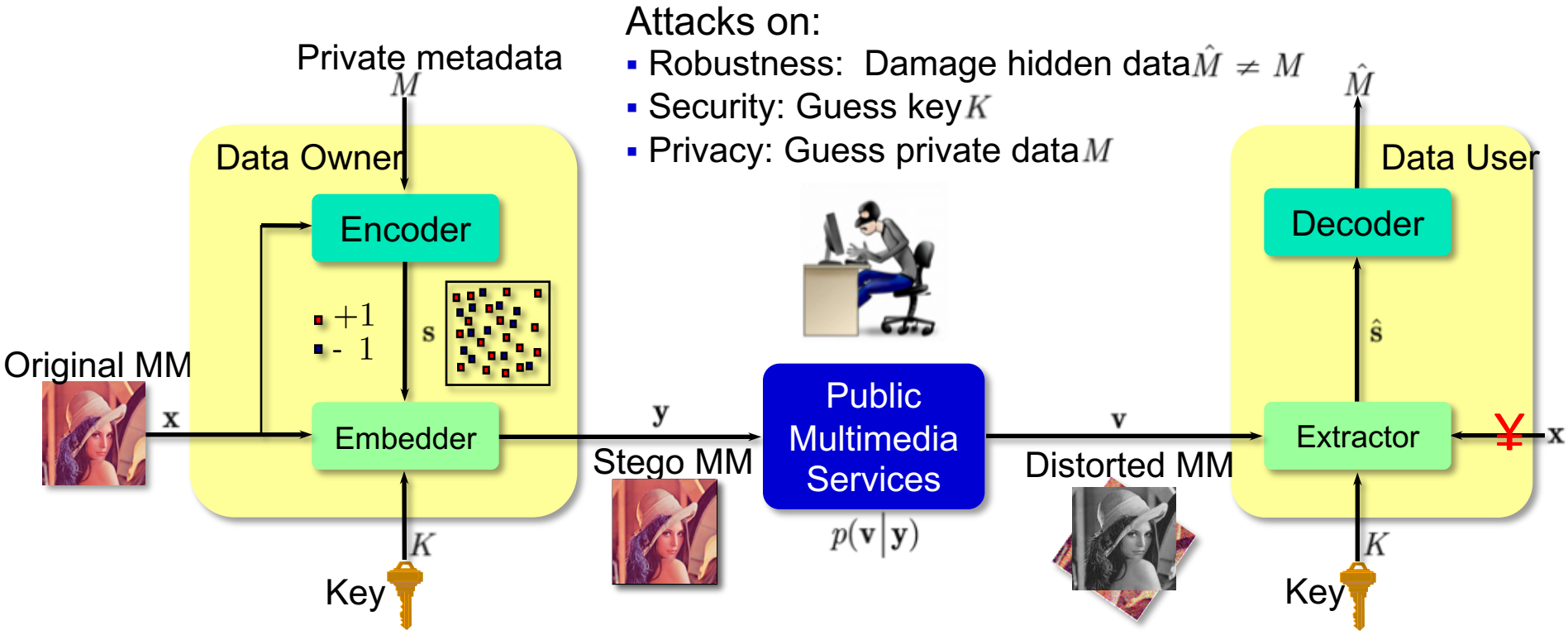


Search Identification Authentication Classification Mining

Digital data hiding: definition and concept

Definition (Digital data hiding)

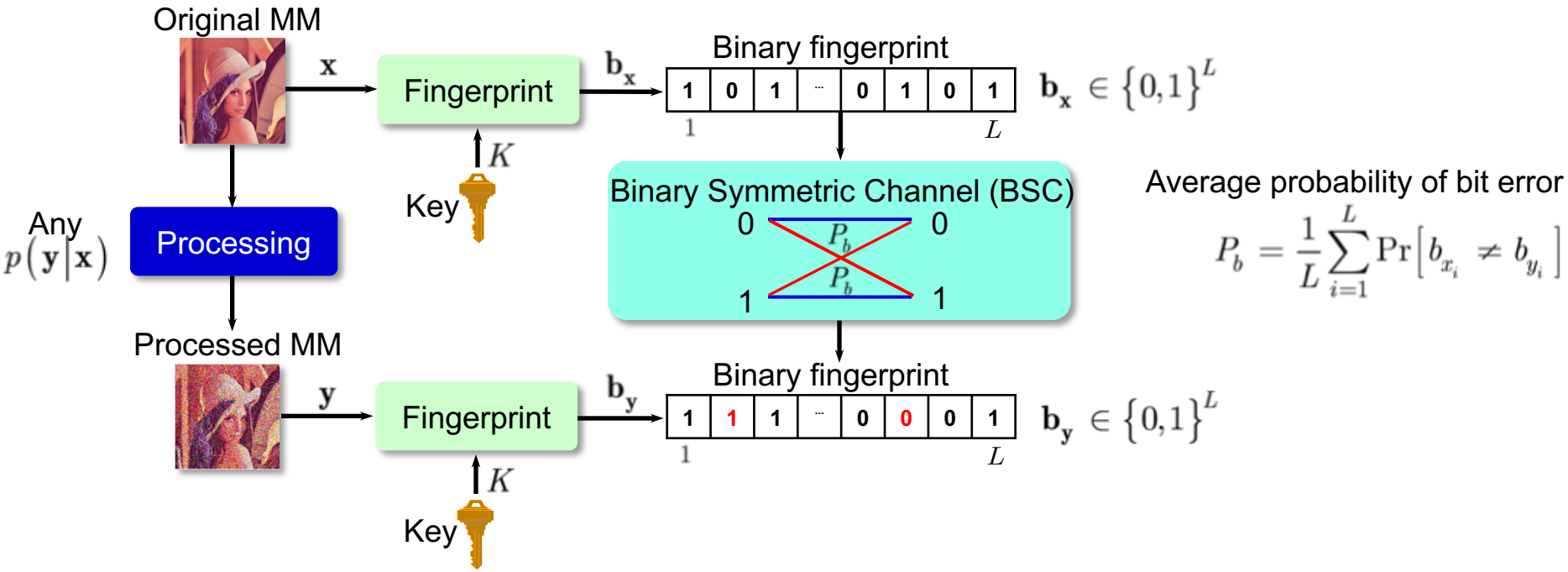
Digital data hiding (*a.k.a. steganography*) is the art of perceptually and statistically undetectable robust information embedding in multimedia content.



Digital fingerprinting: definition and concept

Definition (Digital fingerprinting)

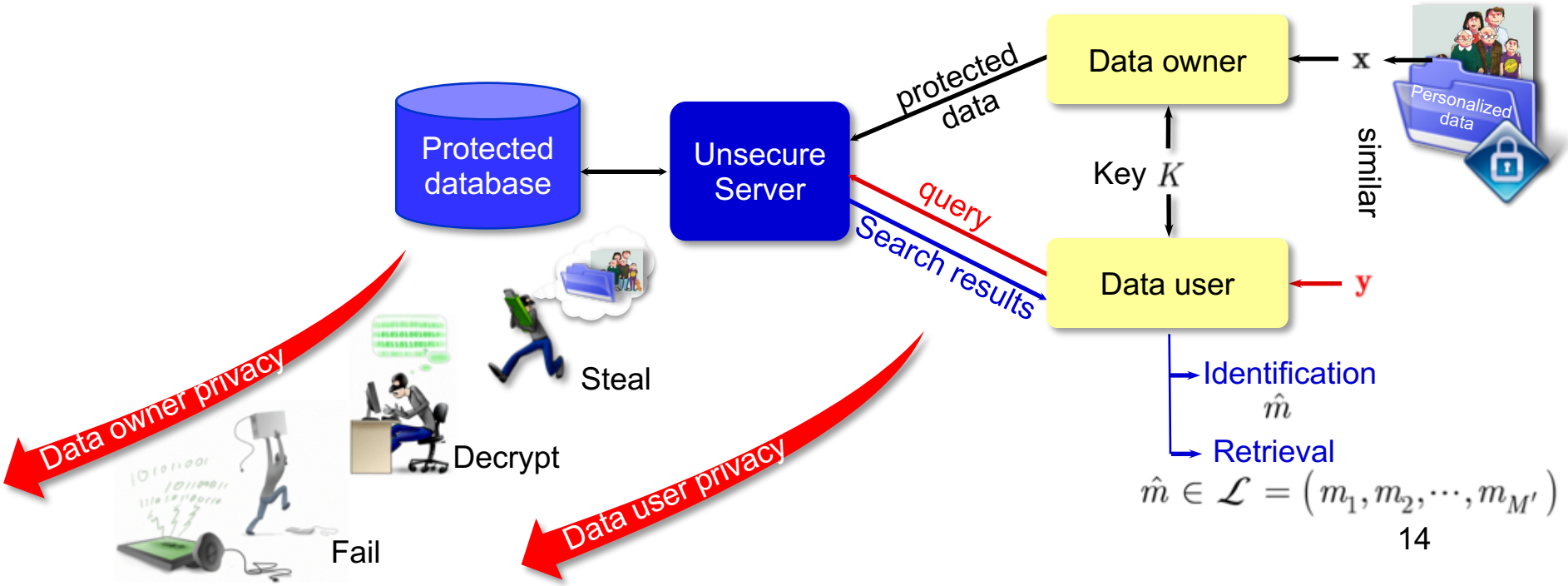
Digital fingerprinting *(a.k.a. robust perceptual hashing)* is a technique for computing a compact robust, secure and private binary representation of multimedia content.



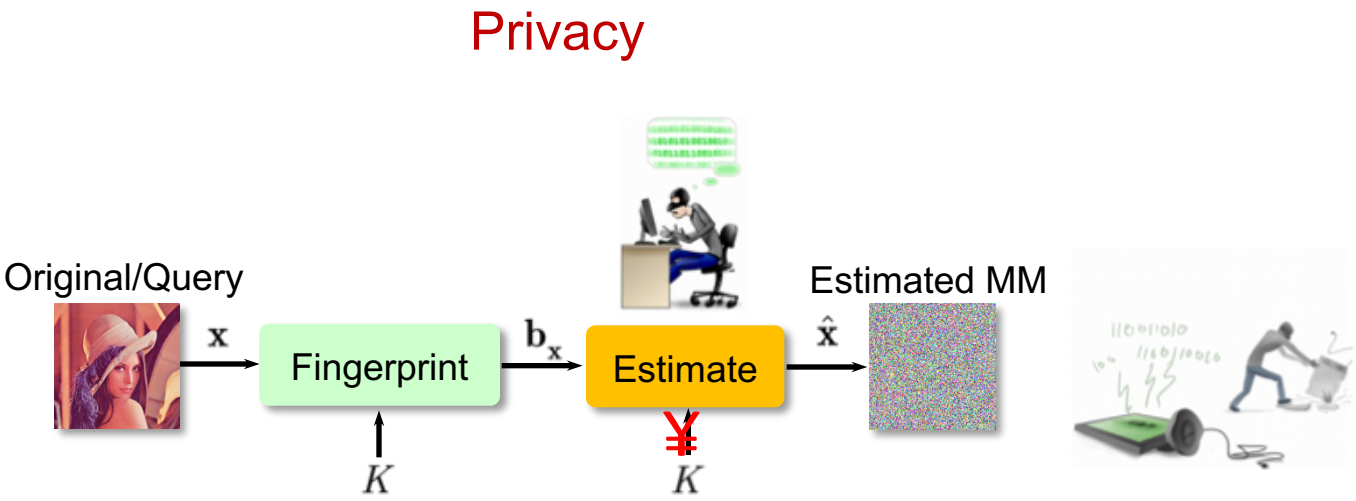
Privacy-preserving search: concept

Main concerns of privacy-preserving search in large-scale systems

- Exact queries and binary distributed hash tables in P2P systems (incl. LSH)
- Complexity:
  - Optimal Maximum Likelihood-based search is a NP-hard problem $\left(\mathcal{O}\left(2^L\right)\right)$
  - Cryptographic homomorphic encryption is computationally expensive



Privacy-preserving search: concept



Course outline

- Digital watermarking
- Digital content fingerprinting
- Privacy protection

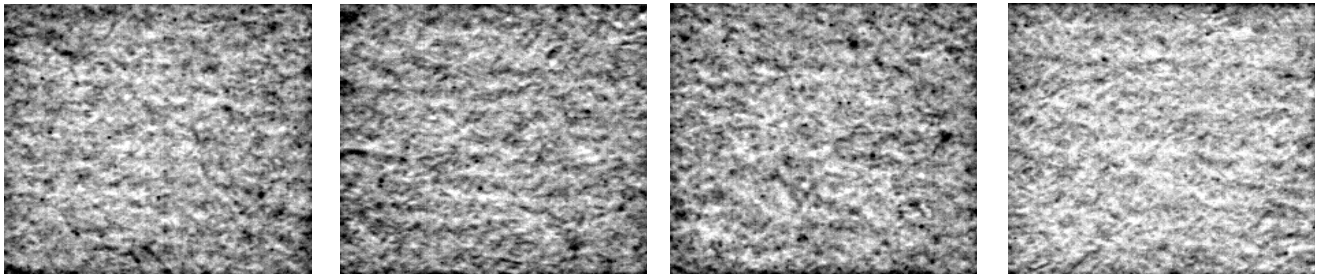
Extensions

- Physical world security
- Security based on smart phones

Protection of physical objects: natural randomness

Definition (Natural randomness)

Natural randomness represents unclonable features created by nature or some uncontrollable randomized process initiated by humans.



Features

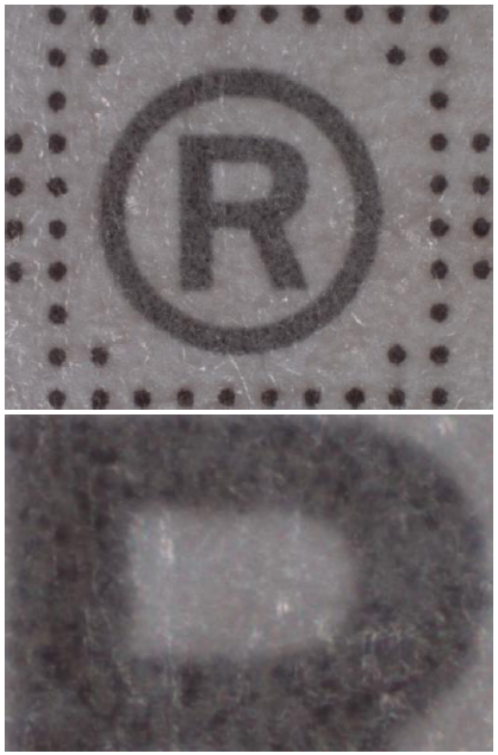
- Easy to evaluate but is hard to characterize
- Structure is unique for each item
- Manufacturer not-reproducible

Protection of physical objects: natural randomness

Sample items with random surface structure (cardboard)

u-nica®

Global Security Solutions

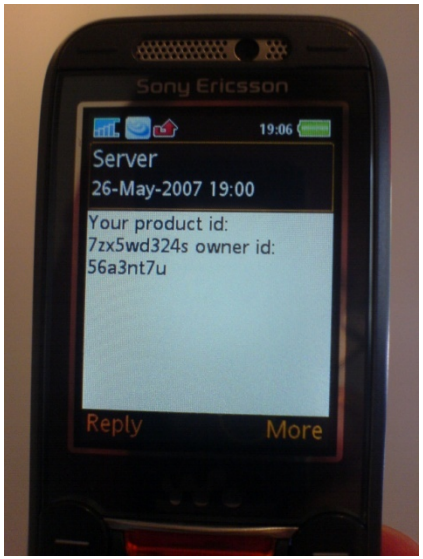


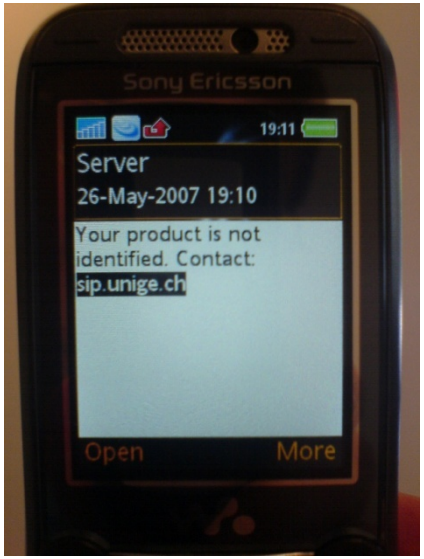
Item 1, unique surface features on every item

Item 2, unique surface features on every item

Verification on mobile phones

Watch identification





Verification on mobile phones