

Multimedia Security and Privacy

S. Voloshynovskiy

Copyright notice

This course uses in part some materials from

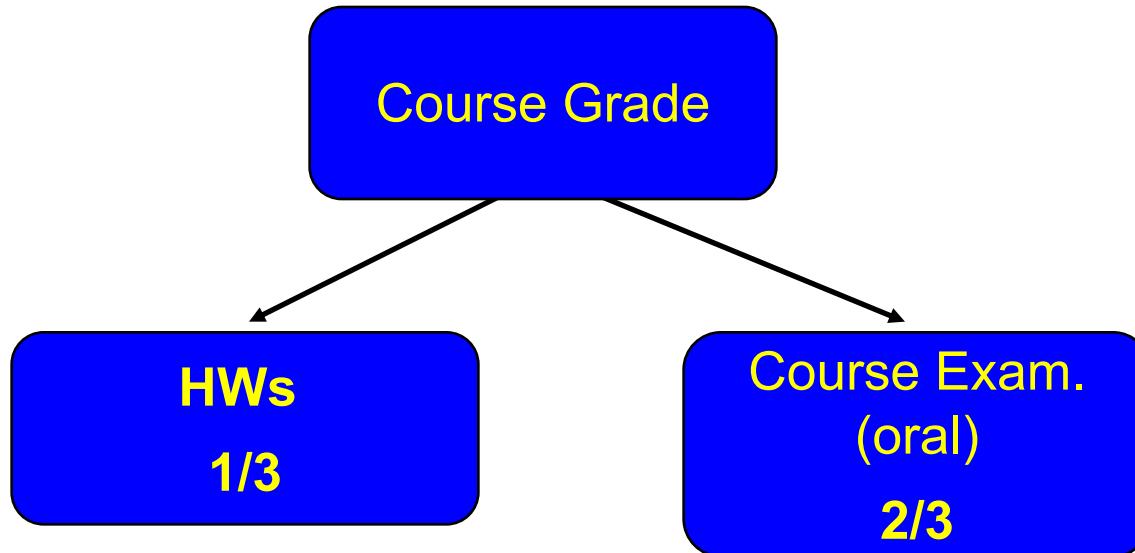
- Book: I. Cox, M. Miller and J. Bloom, “Digital Watermarking: Principles and Practice”, *Morgan Kaufmann Publisher Inc.*, San Francisco, 2001;
- Book: M. Barni and F. Bartolini, “Watermarking Systems Engineering: Enabling Digital Assets Security and Other Applications”, *Marcel Dekker*, 2004;
- Book: M. Barni and F. Bartolini, “Watermarking Systems Engineering: Enabling Digital Assets Security and Other Applications”, *Marcel Dekker*, 2004;
- Book: J. Fridrich, Steganography in Digital Media: Principles, Algorithms, and Applications, 2009

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

The usage of the slides in commercial or educational purposes is prohibited without authorization of the course leaders and permission of the above document authors.

Grading Policy

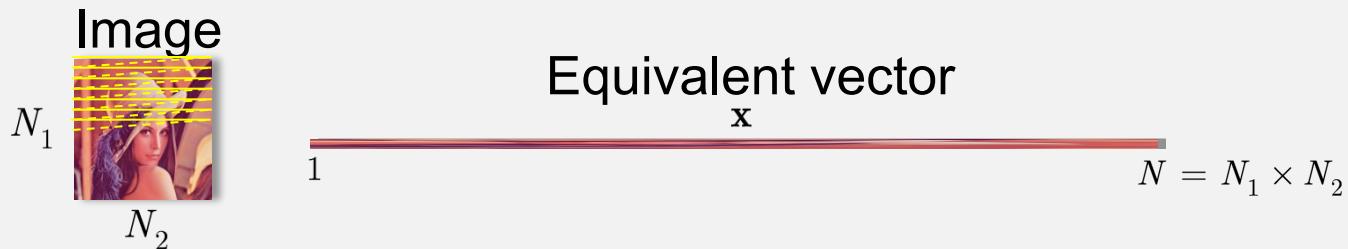
The final grade will be determined based on laboratories and a final examination.



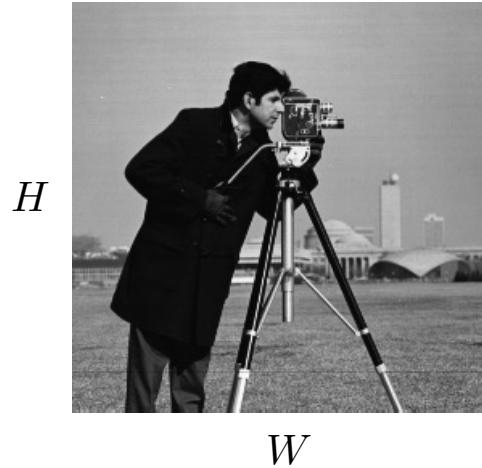
Notations

Notations

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



Notations

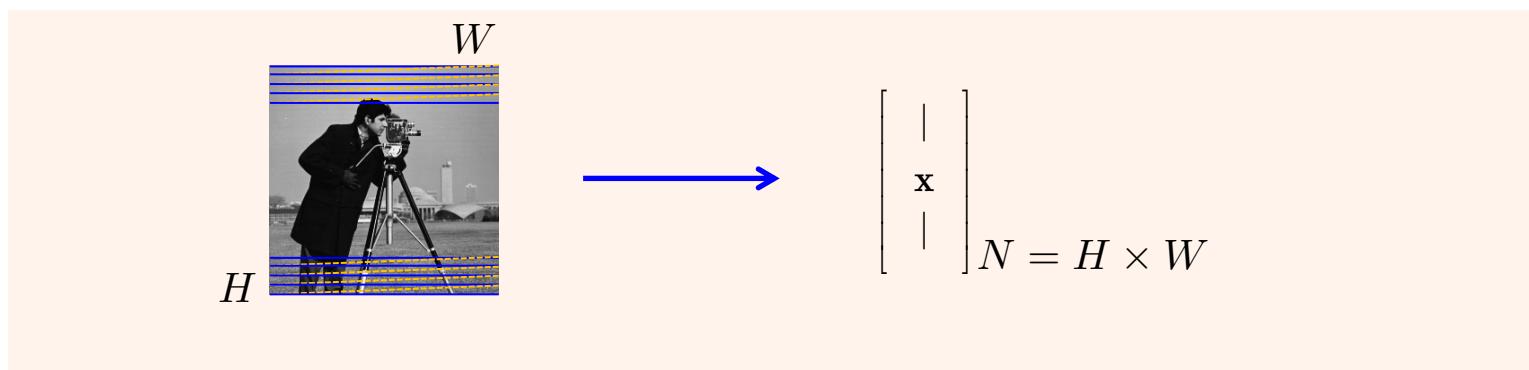


$$H \times W$$

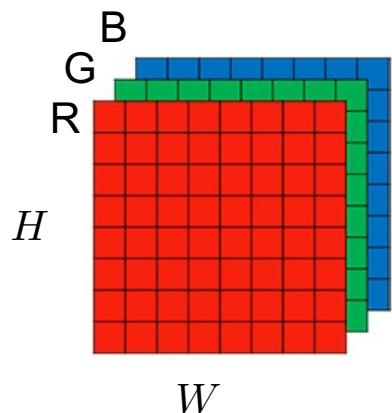
$$x_i \in \mathcal{X} = \{0, \dots, 255\}$$

$$\text{8bits : } 2^8 = 256$$

$$\text{Total size: } (H \times W) \times 256$$



Notations



$$x_i^{R,G,B} \in \mathcal{X} = \{0, \dots, 255\}$$

Total size: $(H \times W) \times 3 \times 256$

Notations



108 Mpx



12 Mpx

Motivation

What is a modern image? Why is it interesting to consider it?

Value: copyright, NFT

Interface between the physical and digital worlds: a link between physical assets and digital tokens

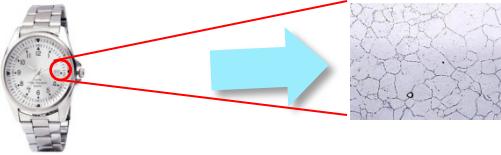
- Any physical object
- Human (biometrics)

As a container of information in forms of :

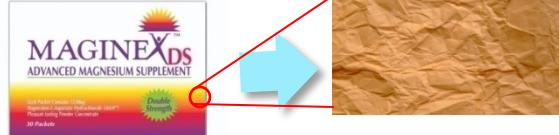
- Digital invisible watermark – copyright protection
- Steganography – hidden/secret communications in multimdeia data
- Tamper proofing – detection of modifications/alterations in data

Applications and main concerns

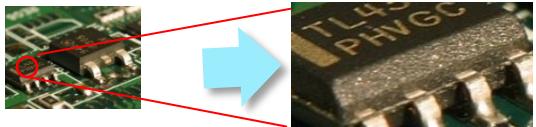
Physical Objects



Luxury products = "Annoying"

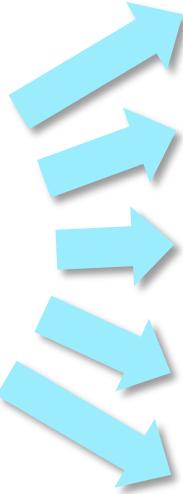
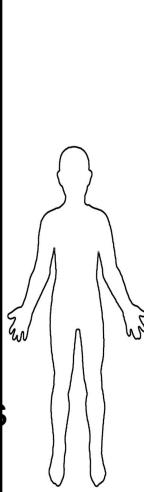


Medicine = "Dangerous"



Electronics = "Security as whole"

Humans



Digital Content

Images

Videos

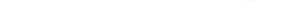
Audos

Text docs



In April 1998, three American students at Harvard University in Cambridge, Massachusetts, created the first peer-to-peer file sharing system, Napster. The three students, Shawn Fanning, Sean Parker, and Nick Frakes, had been working on a system to share their MP3 music collections. They developed a program that would allow them to search for other users who were sharing their music files and download them. The name Napster again refers to both the name of the company and the name of the software they created. In October 1999, the company was sold to America Online (AOL) for \$55 million. AOL used the software to develop its own file sharing service, AOL Instant Messenger. AOL Instant Messenger quickly became one of the most popular ways for people to share files online. After several years of legal battles between Napster and record companies, Napster filed for bankruptcy protection in 2001. While Napster faced the music, Gnutella took over. Gnutella was a decentralized peer-to-peer file sharing system that allowed users to search for files without having to go through a central server. It was very popular in Europe and Asia, and became the reference under download of files. In 2003, Gnutella was taken over by Limewire. In 2005, Limewire was taken over by Kazaa. In 2006, Kazaa was taken over by BitTorrent. BitTorrent has become the most popular peer-to-peer file sharing system in the world. It is estimated that over 100 million people use BitTorrent every day. BitTorrent is used for everything from movies to software to music. It is also used for sharing files between friends and family members. BitTorrent has revolutionized the way we share files online.

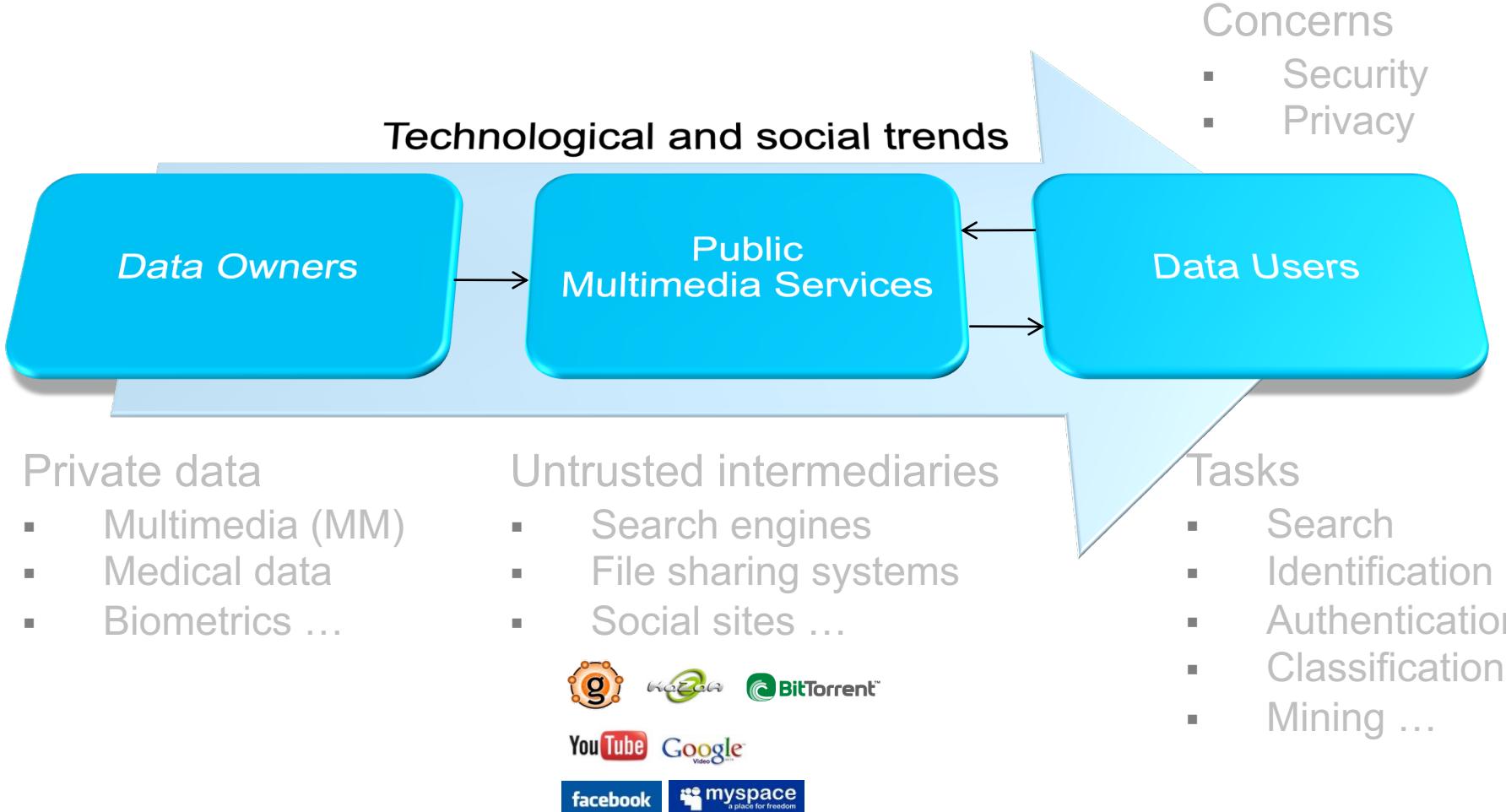
Online sharing services



Main concerns

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

Modern networked multimedia applications

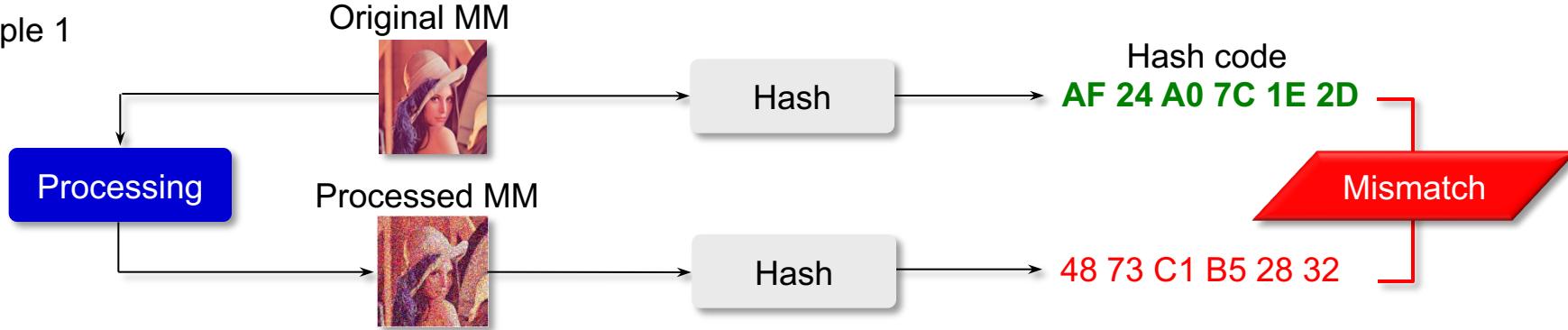


Multimedia security

Why are traditional security tools not suitable for multimedia?

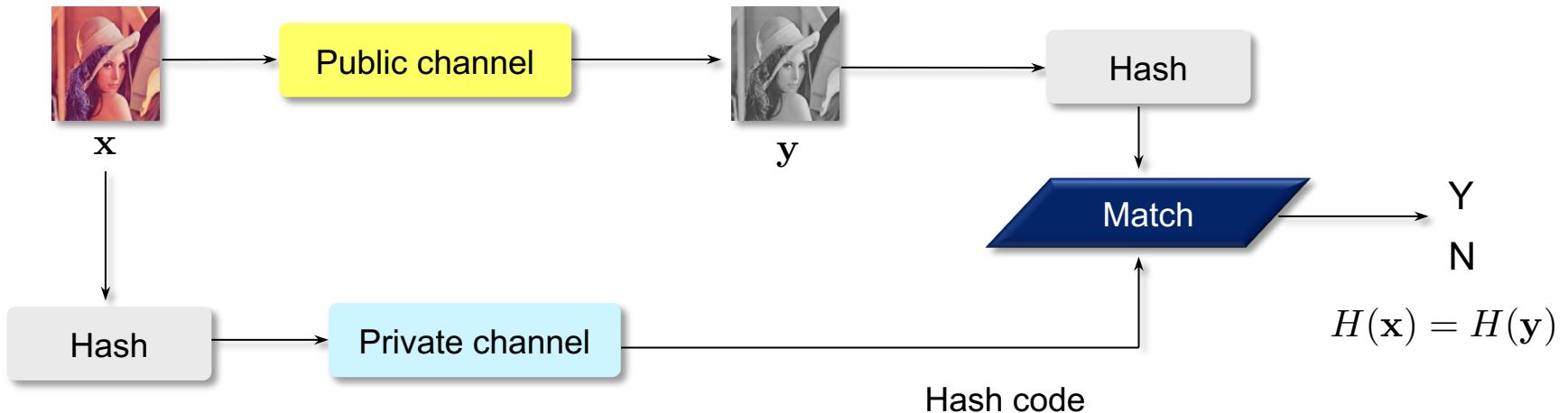
- Traditional *security*:
 - Cryptographic encryption (confidentiality of information)
 - Cryptographic hashes (authentication, trust, access control)
- Main concerns of classical crypto-based algorithms:
 - Sensitivity to noise and unintentional distortions in input data
 - Data handling in the encrypted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

Example 1



Multimedia security

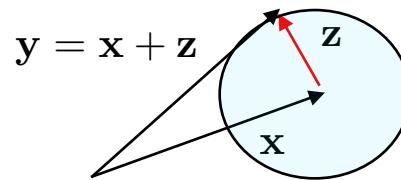
Practical implementation



Multimedia security



$$\mathbf{y} = \mathbf{x} + \mathbf{z}$$



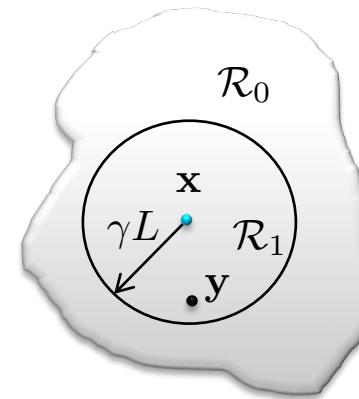
Crypto-hashes

$$\mathbf{x} = \mathbf{y}$$

 \odot

$$H(\mathbf{x}) = H(\mathbf{y})$$

Fingerprtining



Decision region

$$\|\mathbf{y} - \mathbf{x}\|_2^2 = 0$$

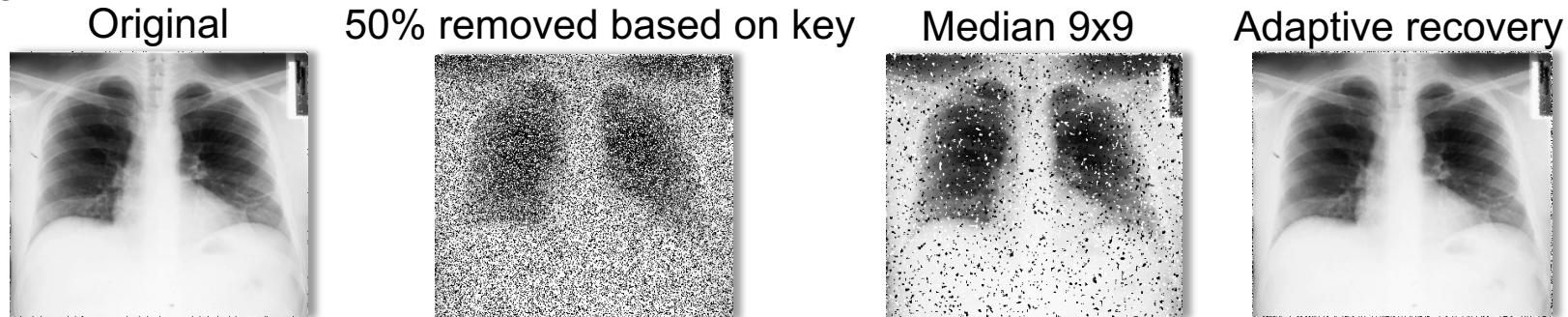
$$\|\mathbf{y} - \mathbf{x}\|_2^2 \leq \gamma L$$

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:



New tools for multimedia privacy and security

Who/When/Where?

Relationship?

Ownership?...



Private
Metadata



Data
Multimedia



Digital Data Hiding (DH)

Self-contained



Encode

Embed



Encrypt



Digital Fingerprinting



Index

Encrypt

Fingerprinting

010101011



Store

Public Domain Multimedia Management

Search

Identification

Authentication

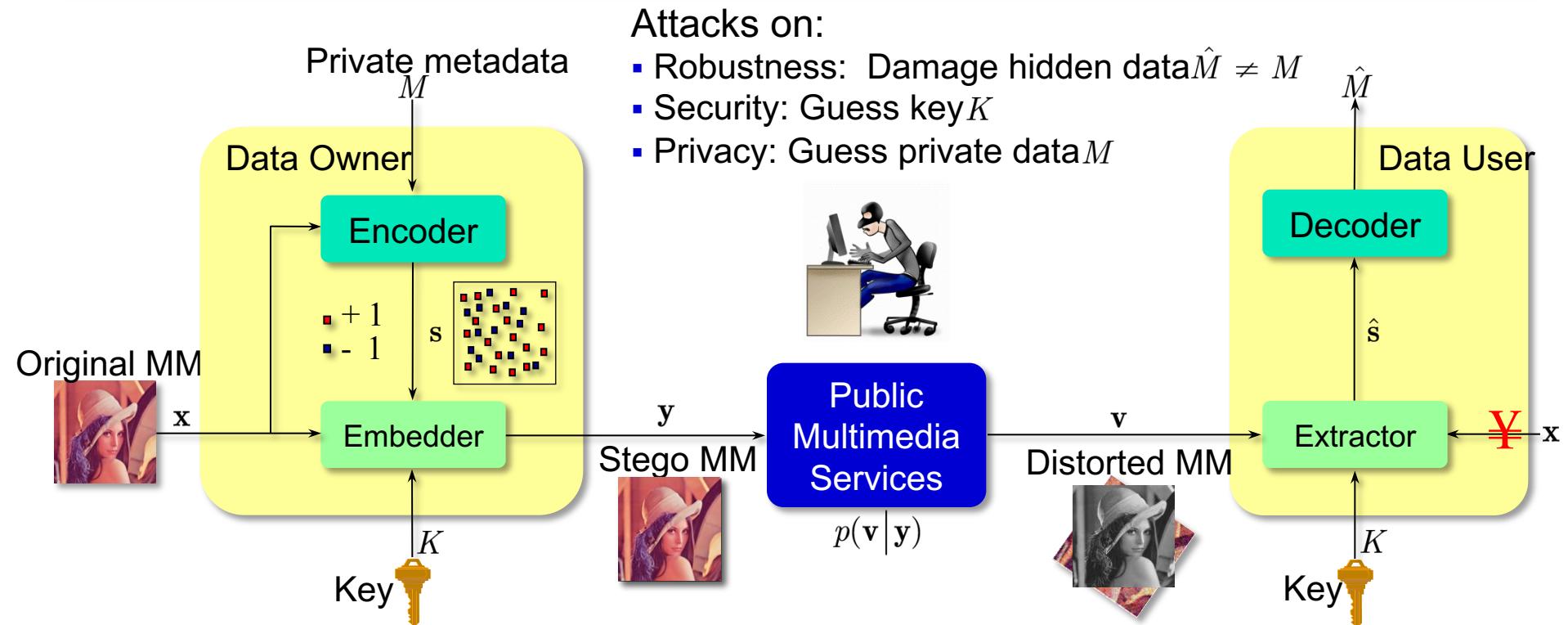
Classification

Mining¹⁵

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.



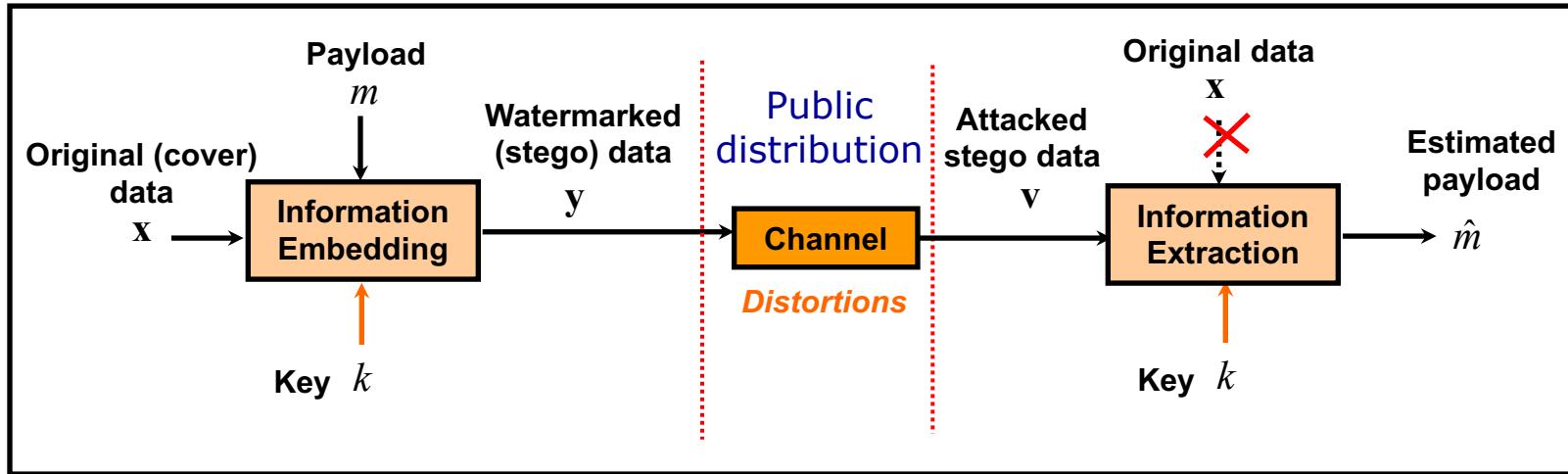
Digital data hiding: definition and concept

Types of digital watermarking

- Robust watermarking
- Steganography
- Tamper proofing

Digital data hiding: definition and concept

Robust watermarking

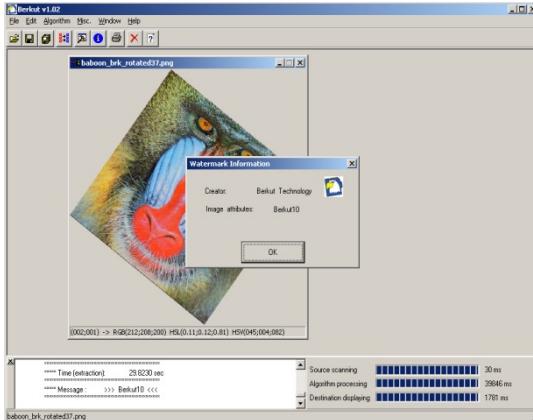


- Copyright protection: who is an owner of content
- Robustness: to ensure reliable extraction of secret message
- Secure: to ensure inability to extract a secret message or to estimate a secret key

Typical message length: 64-128 bits/entire image in all 3 channels

Digital data hiding: definition and concept

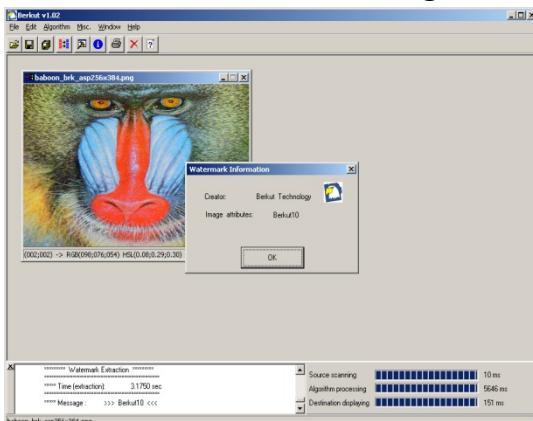
Robust watermarking



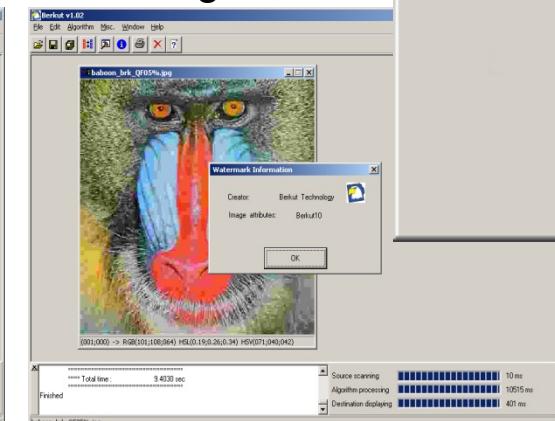
Rotation/shearing



Drawing/erasures



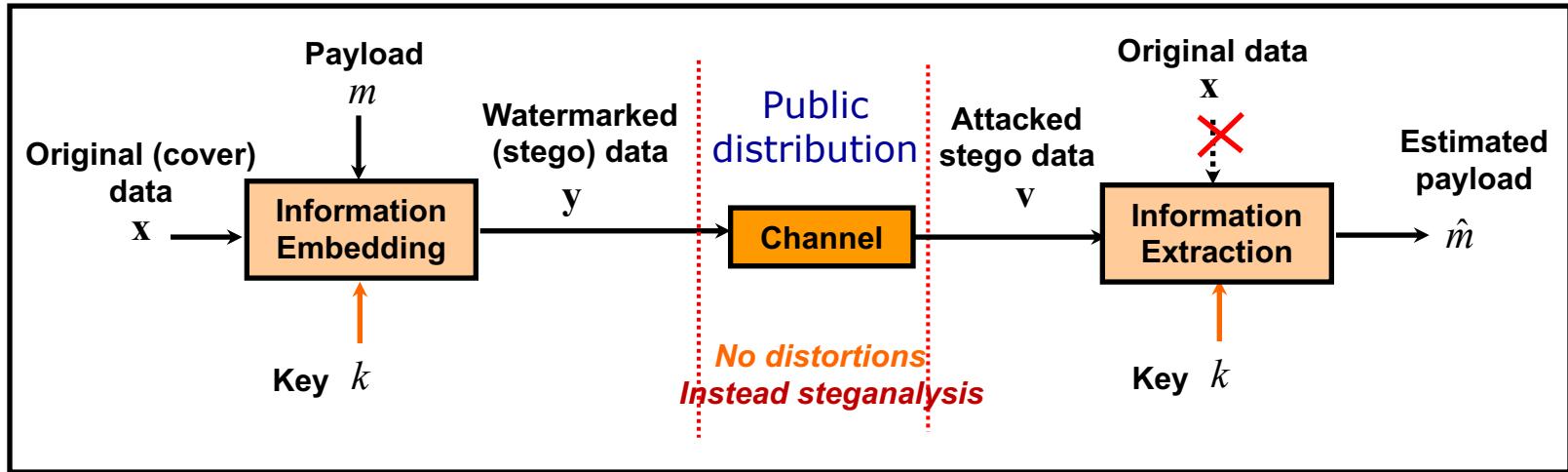
Scaling/aspect ratio



Lossy compression JPEG/JPEG2K

Digital data hiding: definition and concept

Steganography

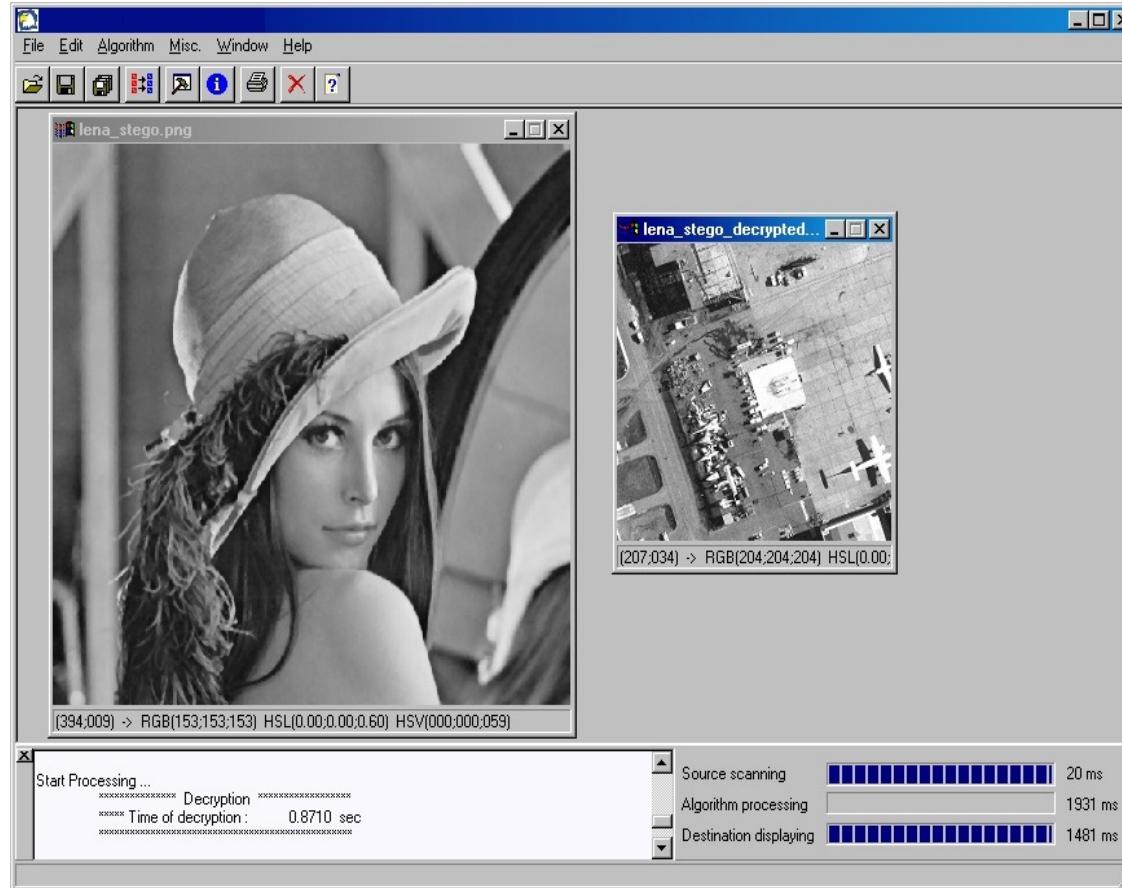


- Secret communications: to hide a fact of message communication
- Robustness: not necessary needed. Instead a fact of secret communication should stay secret as such
- Secure: to ensure inability to extract a secret message or to estimate a secret key

Typical message length: Kbytes per entire image in all 3 channels;
can also be distributed over many images

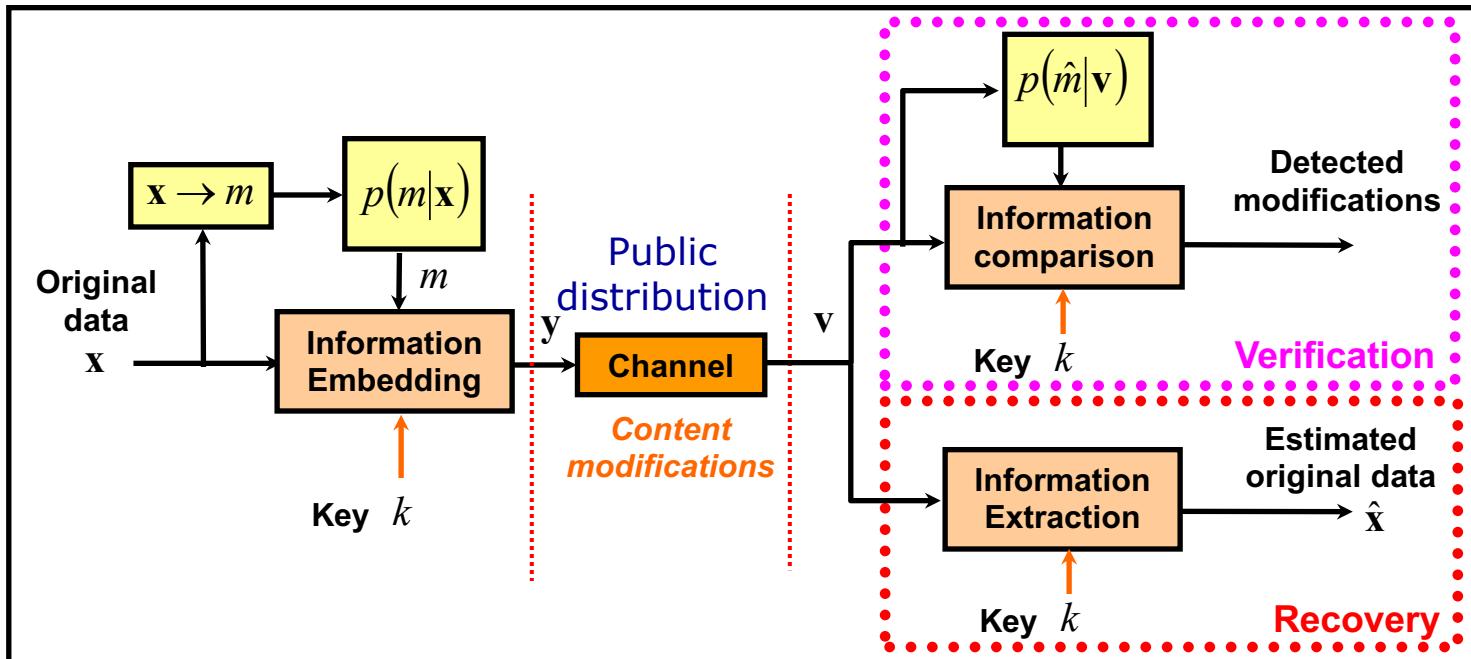
Digital data hiding: definition and concept

Steganography



Digital data hiding: definition and concept

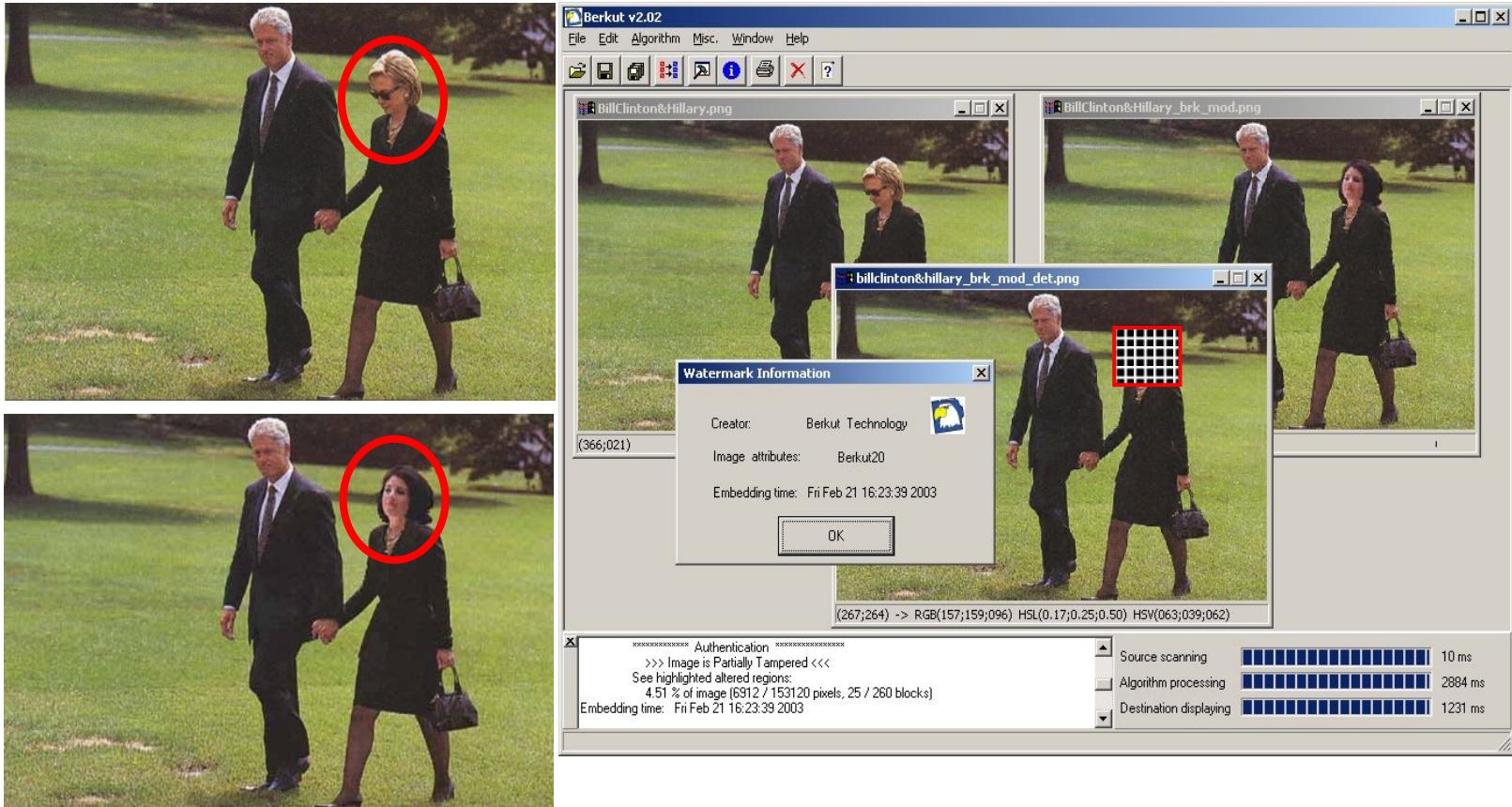
Tamper proofing



- Detect modification in data or localize them
- Robustness: to certain extend only (legitimate distortions)
- Secure: to ensure inability to extract a secret message or to estimate a secret key

Digital data hiding: definition and concept

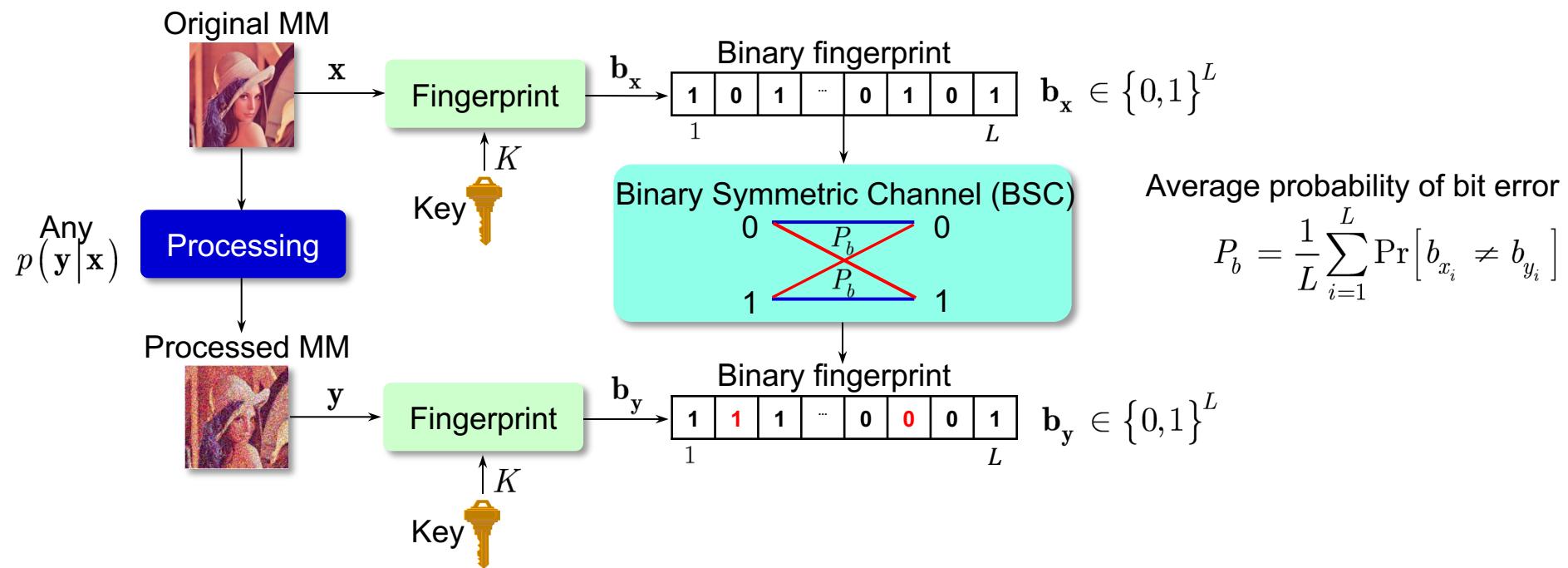
Tamper proofing



Digital fingerprinting: definition and concept

Definition (Digital fingerprinting)

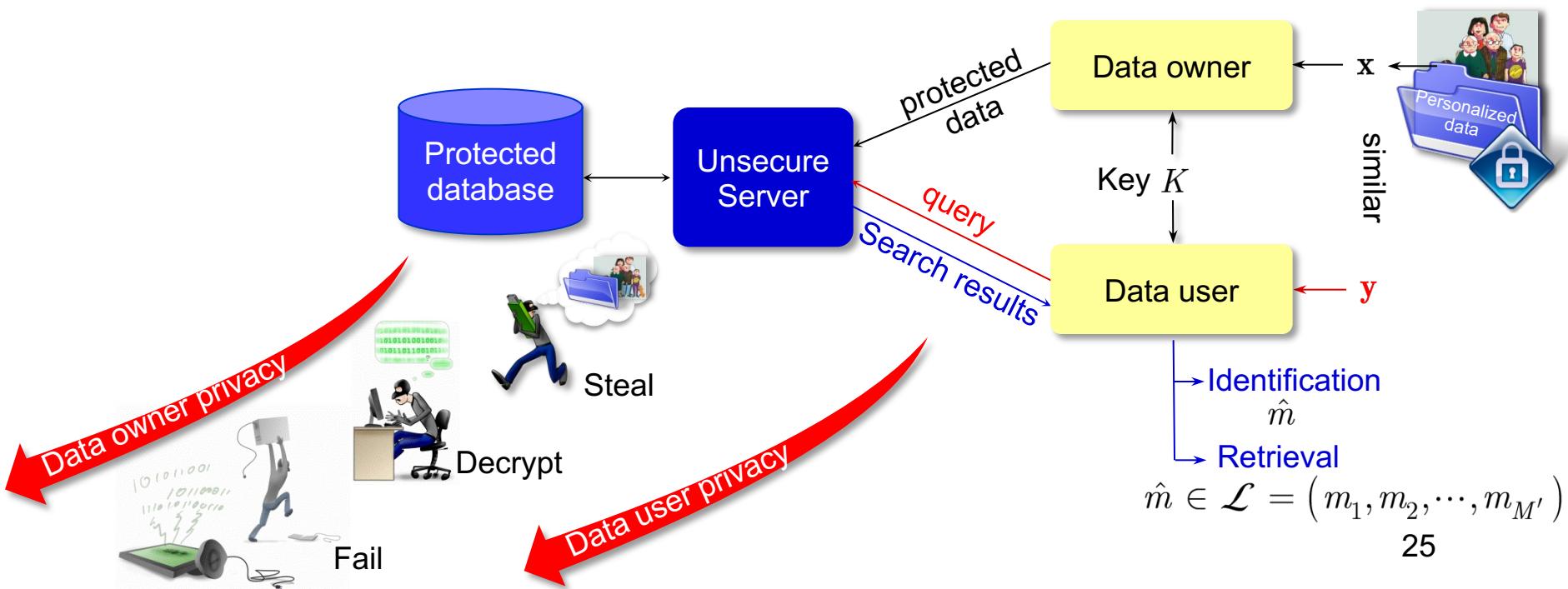
Digital fingerprinting (*a.k.a. robust perceptual hashing*) is a technique for computing a compact robust, secure and private binary representation of multimedia content.



Privacy-preserving search: concept

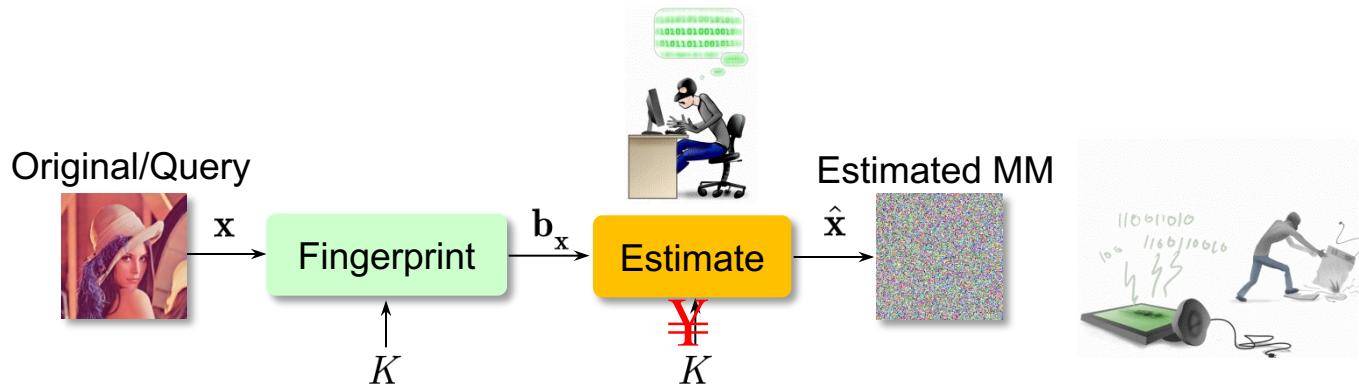
Main concerns of privacy-preserving search in large-scale systems

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



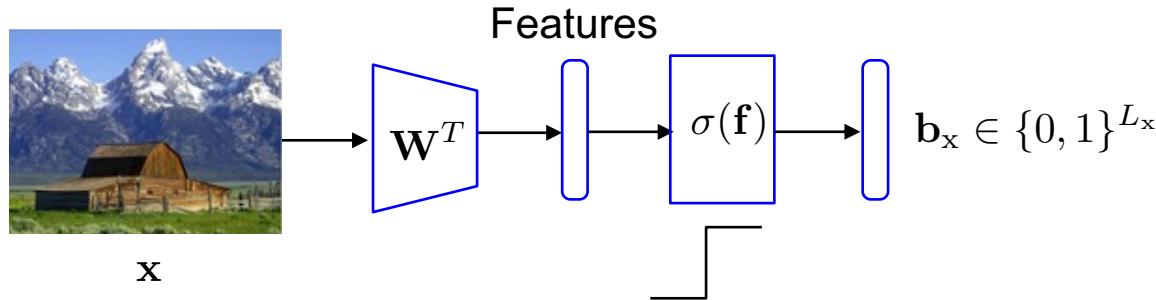
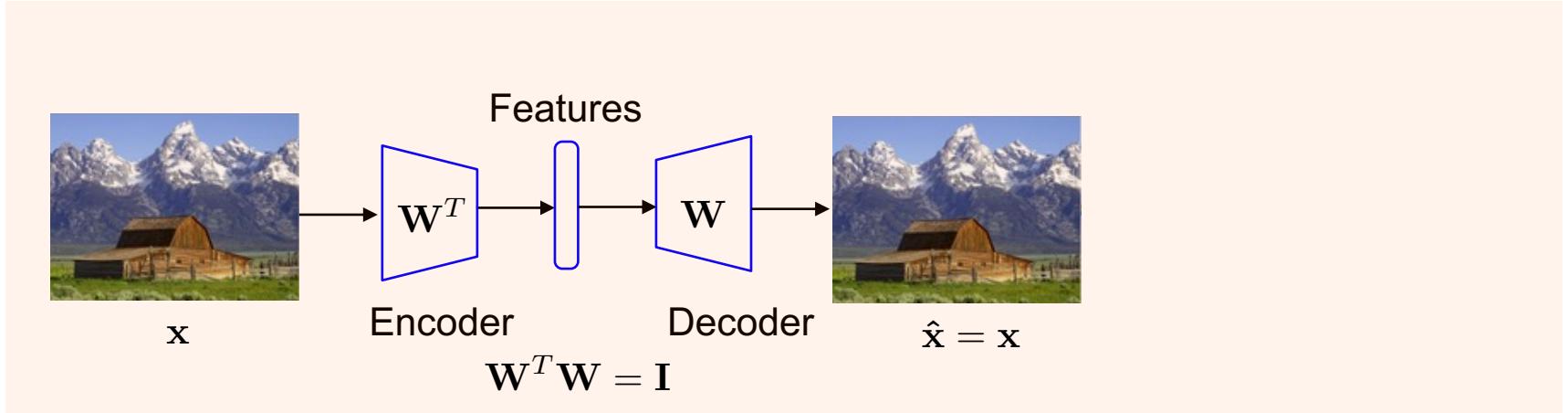
Privacy-preserving search: concept

Privacy

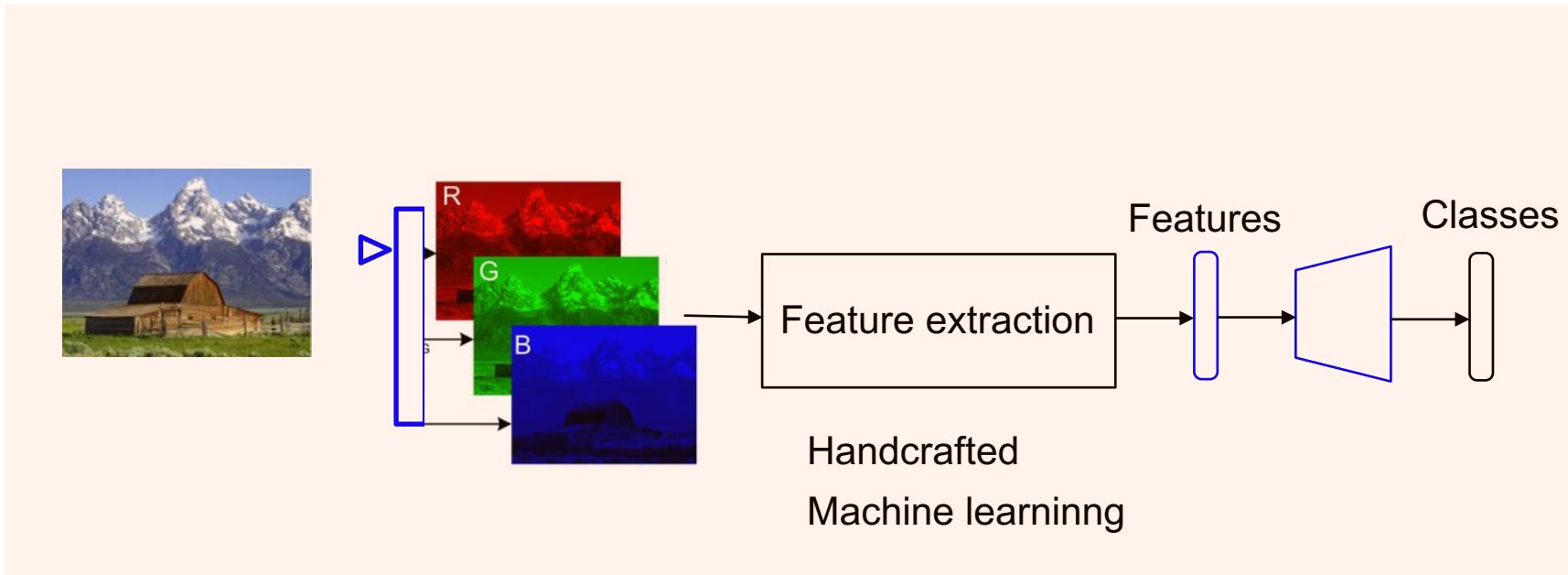


Multimedia security

Hand-crafted based fingerprtining



Multimedia security

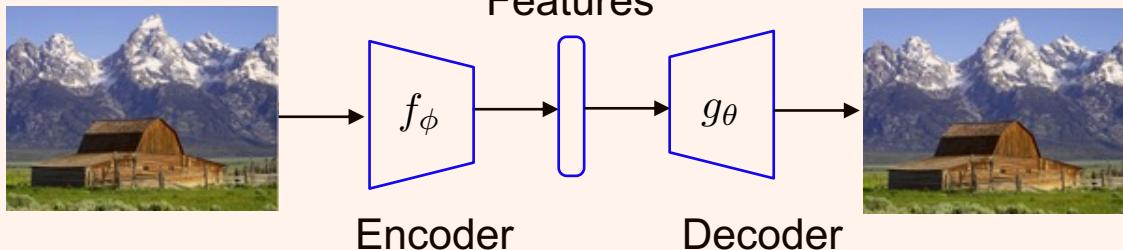


Multimedia security

ML based fingerptining

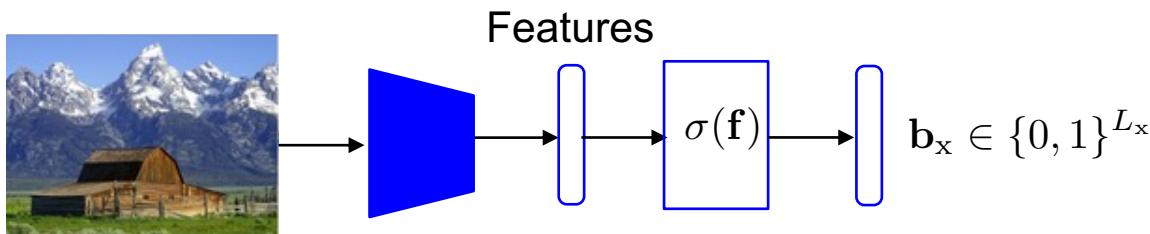
Features taken from auto-encoding

Training



$$f_\phi(\mathbf{x}) = \sigma_K(\mathbf{W}_K \dots (\sigma_1(\mathbf{W}_1 \mathbf{x} + b_1) + b_K)) \quad \phi = \{\mathbf{W}_K, \dots, \mathbf{W}_1; b_1, \dots, b_K\}$$

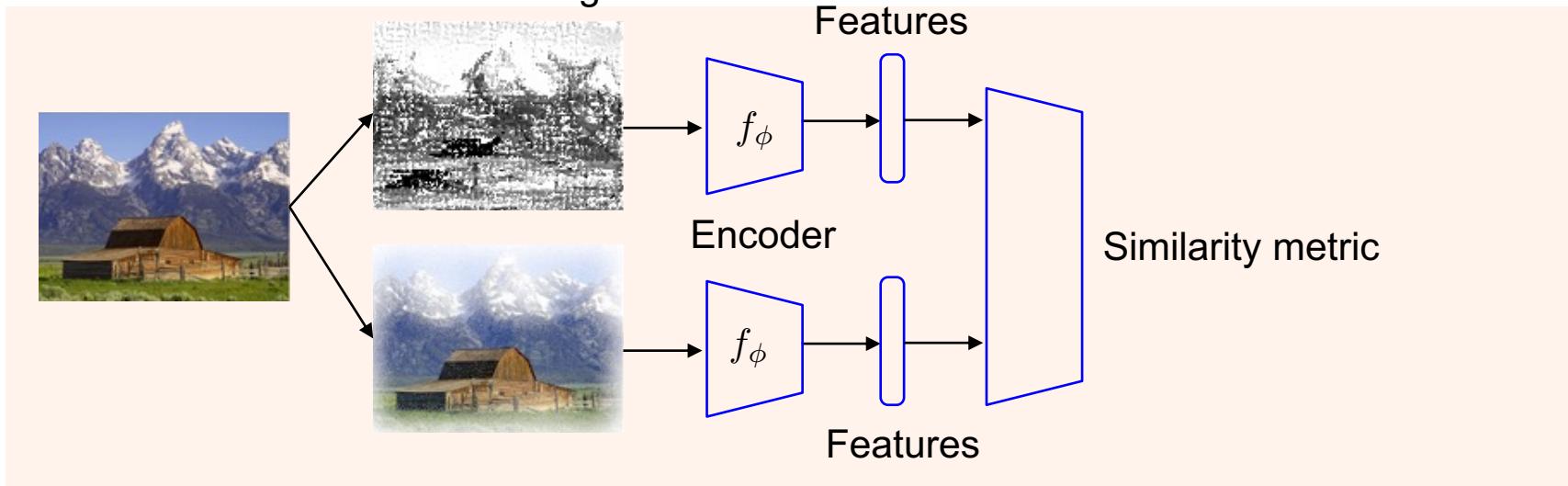
Embedding or feature extraction



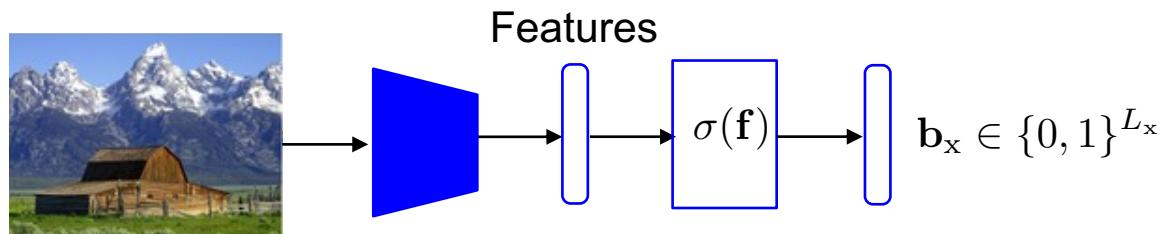
Multimedia security

ML based fingerptining

Features taken from self-learning



Embedding or feature extraction



Course outline

- Digital watermarking
- Digital content fingerprinting
- Privacy protection

Extensions

- Physical world security
- Security based on smart phones

Applications and main concerns

Protection of physical objects

ID docs



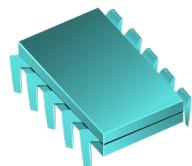
Certificates



Banknotes



Electronics



Luxury objects



Art objects



Packaging



Risks of counterfeiting

- Danger for life
- Market loss
- Damage of brand reputation

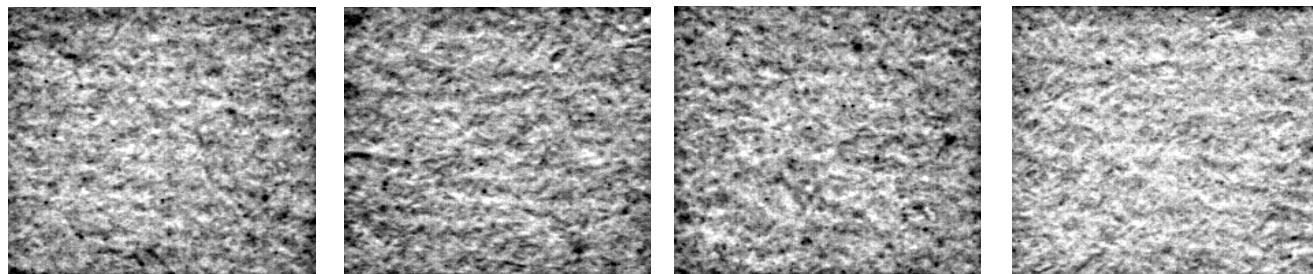
Restraints

- Inefficient authentication technologies
- High cost of track and trace infrastructure
- Lack of awareness for product originality

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)



Item 1, unique surface features on every item



Item 2, unique surface features on every item

u-nica[®]

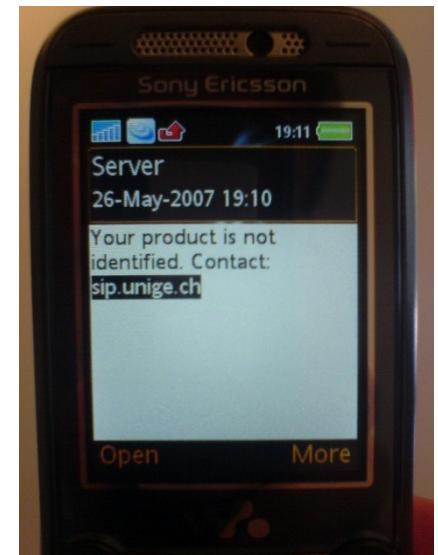
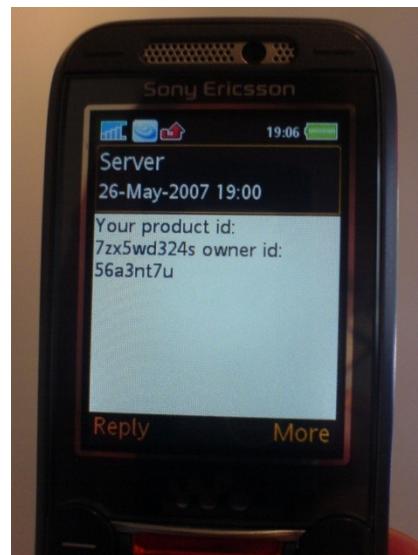
Global
Security
Solutions



© U-NICA Security AG 2011

Verification on mobile phones

Watch identification



Verification on mobile phones



Applications and main concerns

Object recognition

Goal

Accurately recognize each object on mobile phone



Object class

Challenges

- Billions of items
- Very similar



Applications and main concerns

Object recognition

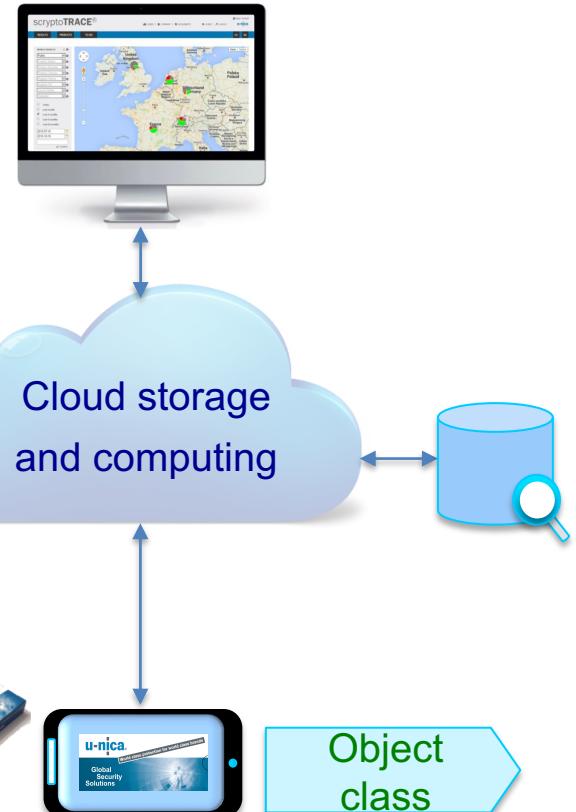
Enrollment from physical object



Enrolled image



Feature extraction



Enrollment from digital template

Applications and main concerns

Object authentication



source: originalideas.info

Observation: if we know the original design, we can easily verify its authenticity.

Question:

- Can we perform the design verification automatically?
- And how accurately (say with the precision about 10-15 microns)?

Applications and main concerns

Object authentication

Integral verification

Text
Graphics
Images
Microstructures
Halftoning



High quality fake



Not authentic

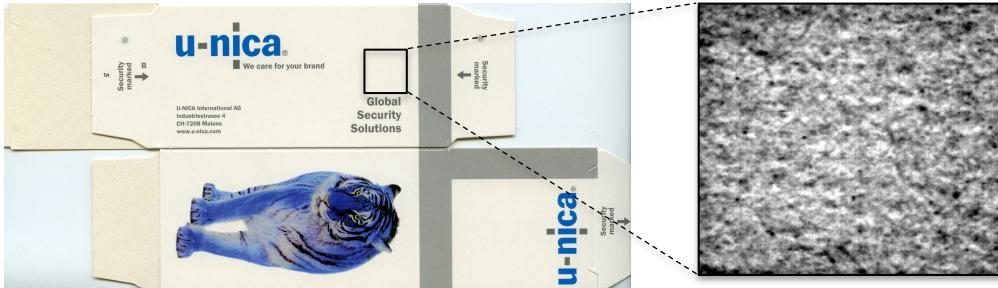


15 micrometers

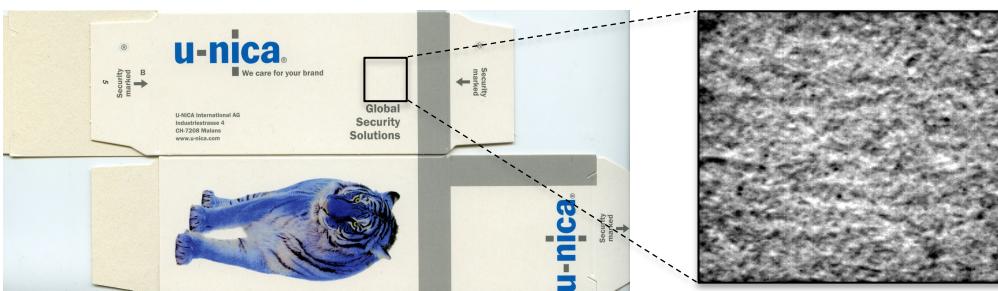
Applications and main concerns

Object identification

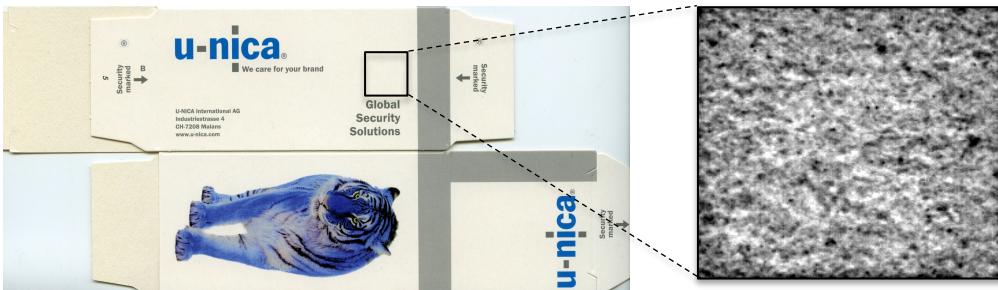
Package 1



Package 2



Package M



Paper microstructures = PUFs

Individually unique PUFs

= unique identifier for
Track&Trace

Visibly packages look identical

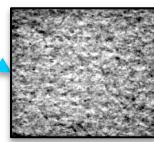
Applications and main concerns

Object identification

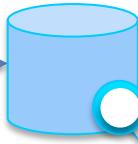
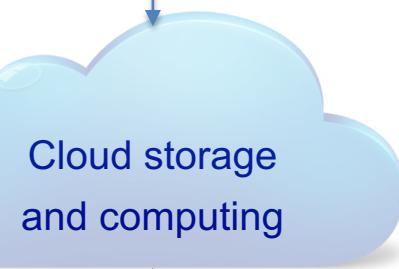
Enrollment from physical object



Enrolled image



Feature extraction



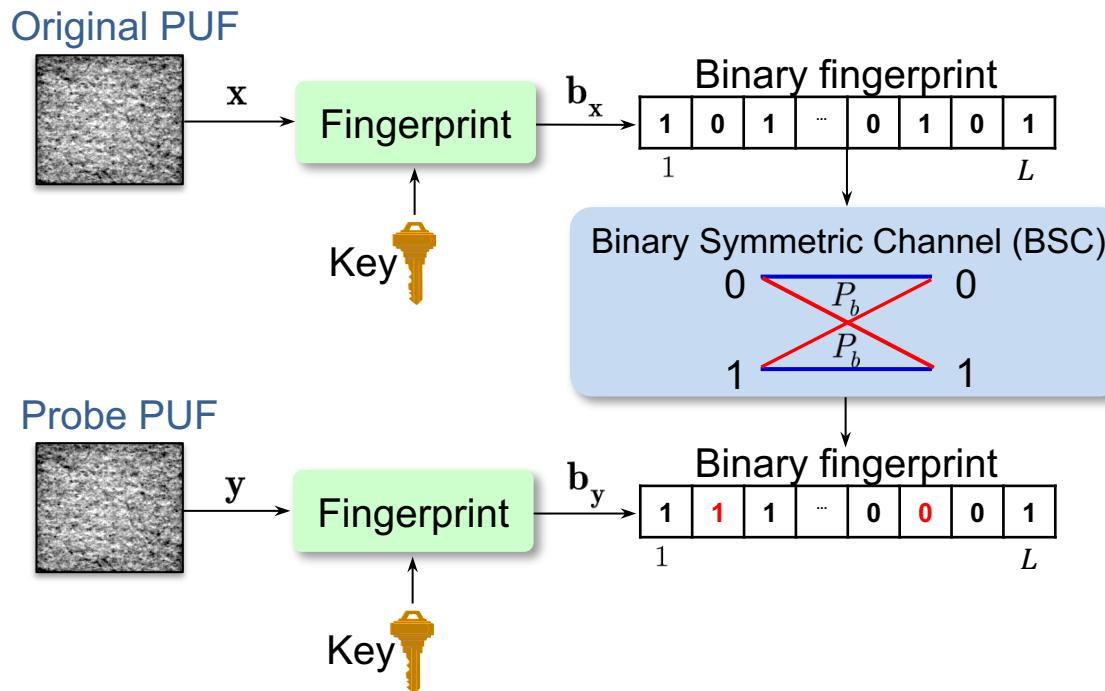
PUF ID

Applications and main concerns

Object identification

Open issue:

Big Data (millions of objects with high-dimensional features)

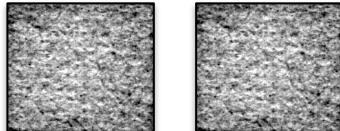


Applications and main concerns

Object identification

Properties of PUFs: Close PUFs = close fingerprints

Correct acceptance

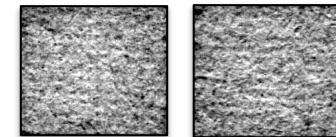


The same object

Two binary vectors representing the state of the same physical object. The top vector has values [1, 0, 1, ..., 0, 1, 0, 1]. The bottom vector has values [1, 0, 0, ..., 0, 1, 1, 1]. Two blue arrows point from the images above to these vectors.

1	0	1	...	0	1	0	1
1	0	0	...	0	1	1	1

Correct rejection

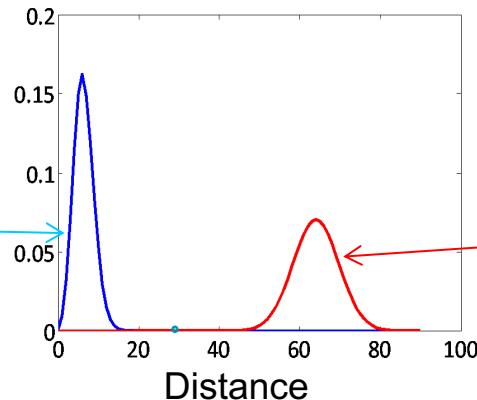


Two different objects

Two binary vectors representing different physical objects. The top vector has values [1, 0, 1, ..., 0, 1, 0, 1]. The bottom vector has values [0, 1, 0, ..., 0, 1, 1, 1]. Two red arrows point from the images above to these vectors.

1	0	1	...	0	1	0	1
0	1	0	...	0	1	1	1

Hypothesis testing



Applications and main concerns

Benefits



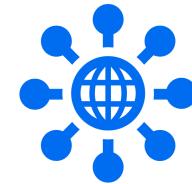
Recognize physical objects

Direct interaction with physical objects



Detect fake objects

Prevent end user consuming fake objects



Track market activity

Tracking goods, market trends and activity



Smartphone app

Public app not requiring any special training



Real-time reporting

Dynamic reporting and visual analytics



Mobile marketing

Consumer engagement and product promotion