Information Security

Prof. Dr. Marta Gomez-Barrero

Hochschule Ansbach

VISUM Summer School, 09/07/2020

Outline

- About me & hs-ansbach
- Introduction to Information Security
- Cryptography
- User Authentication
- Biometrics
- Case Study: Cancelable Biometrics Based on Bloom Filters

About Marta

Who am I and where do I come from?

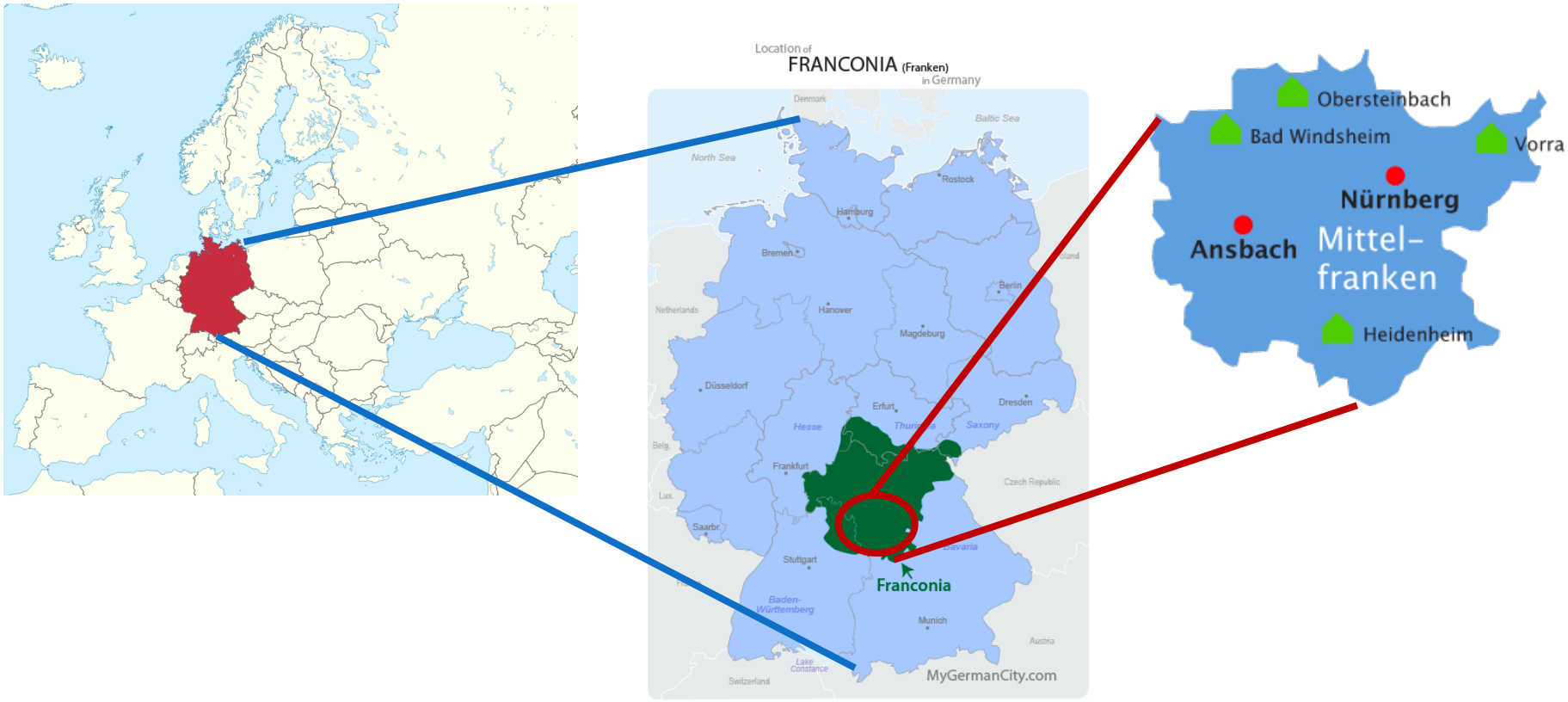
- I come from Madrid, in Spain
- There I studied Mathematics and Computer Science, and finished my PhD on IT-Security (biometrics)
- After that, I moved to Germany in 2016, to Darmstadt, as PostDoc in ATHENE (German National Centre for Applied Cybersecurity)
- And now I am Prof. for IT-Security and technical data privacy at the Hochschule Ansbach

About Marta

About hs-ansbach

Where is hs-ansbach?

- We are located in Ansbach, a German city close to Nurnberg



About hs-ansbach

About Ansba

ki/Bocksbeutel

About hs-ansbach

- The Hochschule Ansbach is a young, modern Bavarian university for applied sciences with Bachelor and Master studies focusing on economics, technology, media and natural sciences

- Research focus on:
  - Technical Innovation (biotechnology, big data, AI, ML, …)
  - Multimedia Innovation (journalism, communication, didactic, …)
  - Operational organization (digital business, business intelligence, …)

Introduction to Information Security

What is information security?

- "Preservation of confidentiality, integrity and availability of information. Note: In addition, other properties, such as authenticity, accountability, non-repudiation and reliability can also be involved." (ISO/IEC 27000:2009)

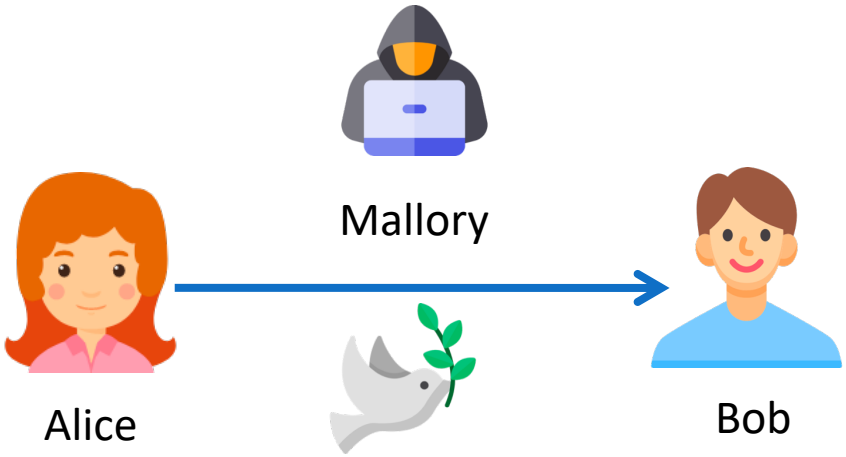
- "Information security is the protection of information and minimizes the risk of exposing information to unauthorized parties." (Venter and Eloff, 2003)

What is information security?

- "Information Security is a multidisciplinary area of study and professional activity which is concerned with the development and implementation of security mechanisms of all available types (technical, organizational, human-oriented and legal) in order to keep information in all its locations (within and outside the organization's perimeter) and, consequently, information systems, where information is created, processed, stored, transmitted and destroyed, free from threats.Threats to information and information systems may be categorized and a corresponding security goal may be defined for each category of threats. A set of security goals, identified as a result of a threat analysis, should be revised periodically to ensure its adequacy and conformance with the evolving environment. The currently relevant set of security goals may include: *confidentiality, integrity, availability, privacy, authenticity & trustworthiness, non-repudiation, accountability and auditability.*" (Cherdantseva and Hilton, 2013)

Our scenario

- Alice and Bob want to share information (i.e., send messages m)
- The communication channel is not secure
- Nonetheless, they don't want a stranger (Mallory) overhearing or even manipulating their messages



Security goals

- **Confidentiality**: Mallory cannot read the messages sent between Alice and Bob
- **Integrity**: the messages between Alice and Bob cannot be altered, or Alice and Bob can detect if the messages were altered
- **Authenticity**:
  - **Data**: Bob can establish that the message was sent by Alice
  - **Entity**: Bob can establish that Alice is who she said she is
- **Non-repudiation**: Bob can prove to a third party that Alice sent the message

Data Privacy Security Goals (GDPR)

- **Availability**: an IT-System maintains availability, when the authorised subjects cannot be prevented from exercising their right by unauthorised subjects
  - Backups, substitution rules or continuous electrical supply can be used as countermeasures

- **Anonymity**: Personal data are changed in such a way, that these cannot be assigned to a person, or only with disproportionate effort

- **Pseudonymity**: Personal data are stored in such a way, that they can only be assigned to a person with knowledge of the assignment rule

Practical security

- **Perfect secrecy** is only possible, when the encryption key is as long as the message (e.g., One time pad) - Claude Shannon, 1948
  - Not valid for 10GB of videos from satellites!

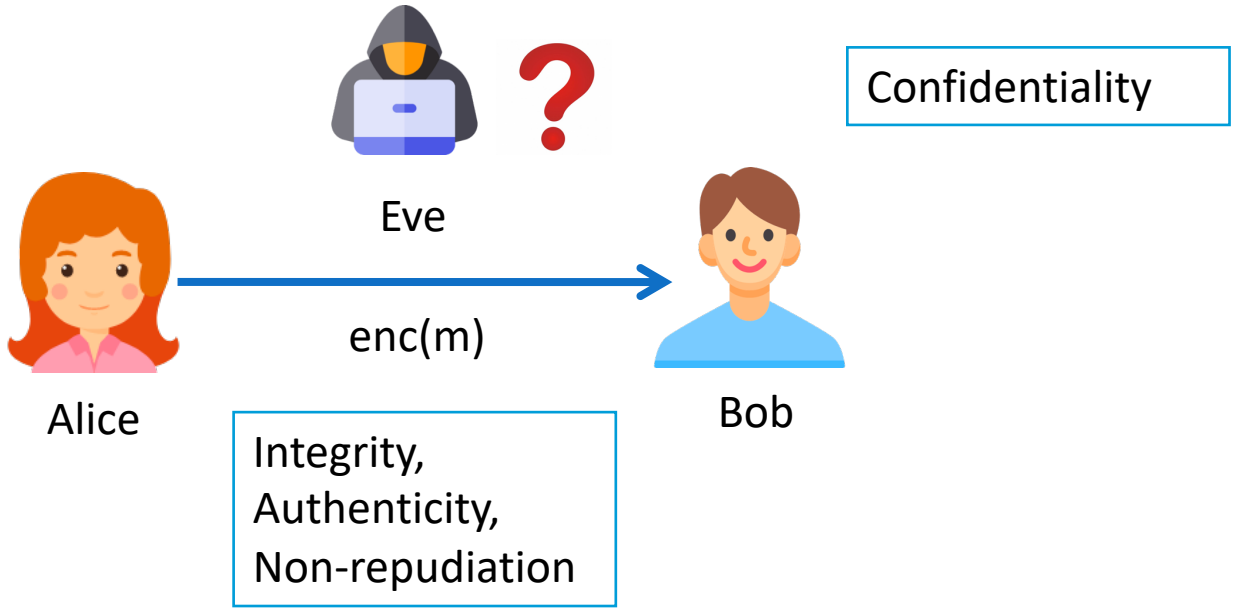
- **Practical security**: the attacker has limited resources in terms of computing capacity and time
  - No need for 10GB long keys, but e.g., 512b

- **Security level**: a cryptographic algorithm provides a security level of *n* bits if an attacker needs 2^n attempts to break the algorithm (e.g., to obtain the plaintext)
  - $n \geq 100$ bit is considered practically secure

Cryptography

What is cryptography?

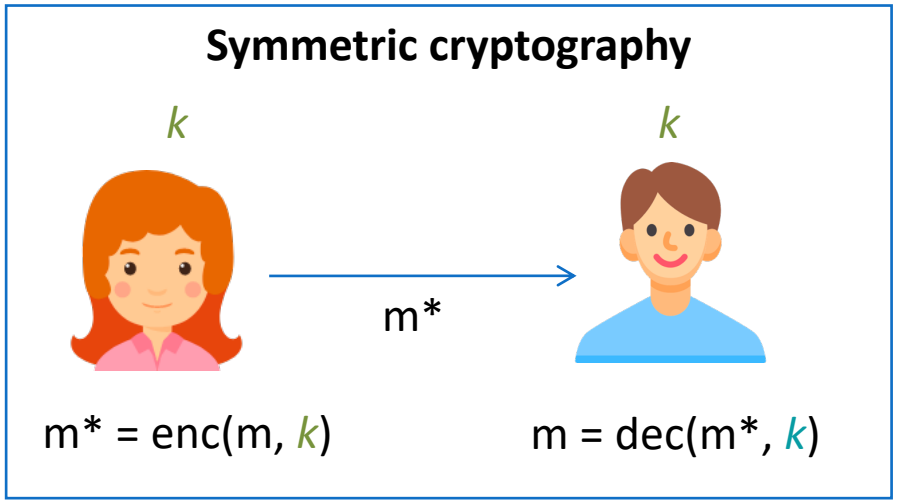
- Cryptography = κρνπτοσ (secret) + γραφειν (write) denoted originally the art of encrypting messages.

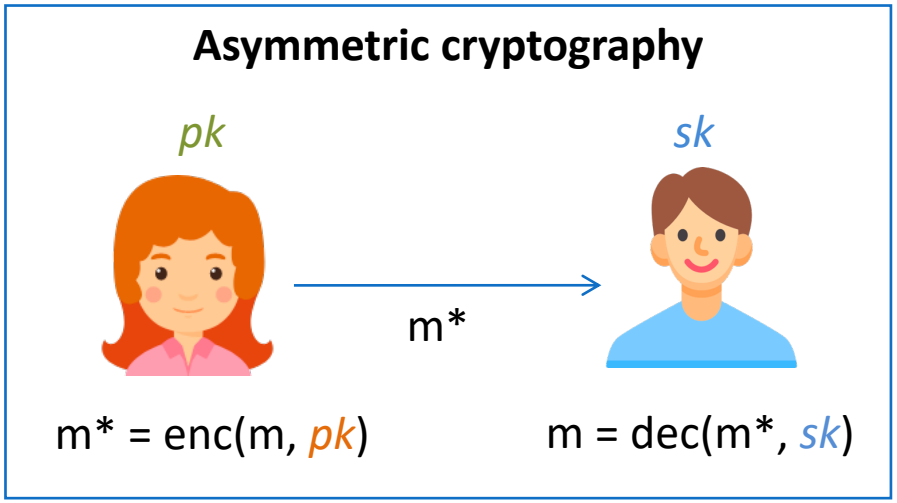


- Modern cryptography (since about 1975) takes care of secure communication in general (i.e., more security issues).

Symmetric *vs.* asymmetric cryptography

- There are two main types of encryption schemes:
  - **Symmetric schemes**: a single key *k* is used to encrypt and decrypt
  - **Asymmetric or public key schemes**: a public key *pk* is used to encrypt, and a secret key *sk* to decrypt





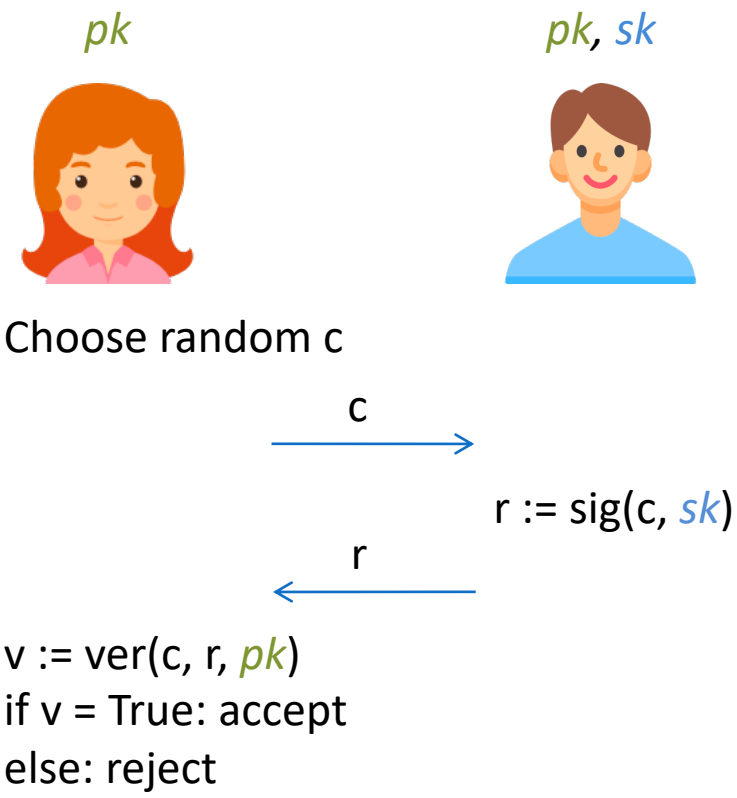
How to?

- The main idea is to find a one-way function *f* with a trapdoor
  - *f* is easy to compute
  - f^{-1} is practically unfeasible to compute
  - Trapdoor: With some additional knowledge (e.g., the secret key *sk*) we can compute f^{-1} efficiently

- To find *f*, we rely on hard mathematical problems:
  - Integer factorization: Rabin cryptosystem, RSA
  - Discrete logarithms: ElGamal cryptosystem
  - Decoding general linear codes (known to be NP-hard): McEliece cryptosystem
  - Elliptic curve cryptography

Challenge-response authentication

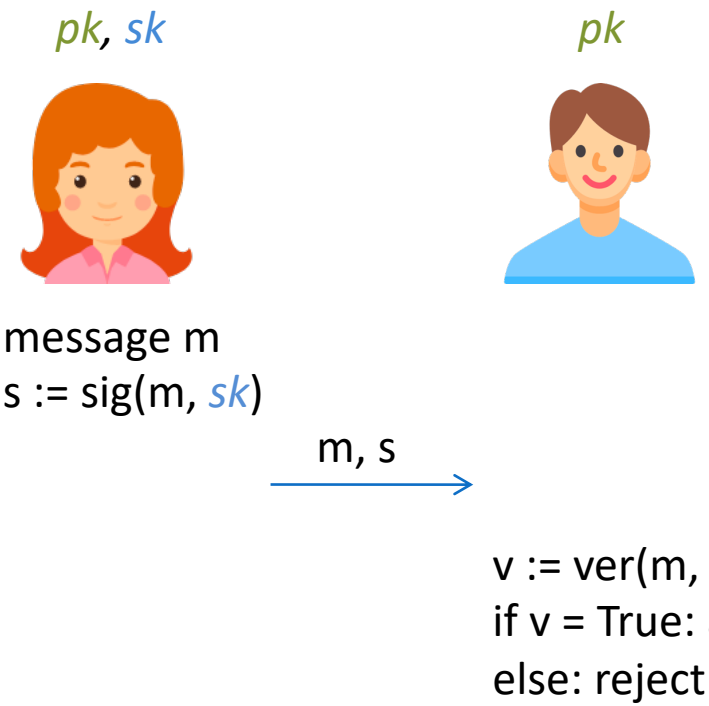
- Goal: user authentication, without revealing a secret (password)



- As long as only Bob knows *sk*, Alice can establish that Bob is who he claims to be
- Examples:
  - ssh's challenge-response system based on RSA
  - Secure Remote Password (SRP)
  - Challenge-Handshake Authentication Protocol (CHAP)

Digital signatures

- Goal: non-repudiation, data authenticity, integrity



- As long as only Alice knows *sk*, Bob can establish that the message m came from Alice
- Examples:
  - RSA-Signatures
  - Digital Signature Algorithm (DSA)
  - Elliptic Curve Digital Signature Algorithm (ECDSA)

Secure communication

- Goal: confidentiality, non-repudiation, data authenticity, integrity

sk_A, pk_A

sk_B, pk_B

message m
s := sig(m, sk_A)
y := enc(m||s, pk_B)

y →

m||s := dec(y, sk_B)
v = ver(m, s, pk_A)
if v = True: accept
else: reject

- Now we have two pairs of keys:
  - {sk_A, pk_A} for Alice
  - {sk_B, pk_B} for Bob
- As long as only Bob knows sk_B, only he can read the message m
- As long as only Alice knows sk_A, Bob can establish that m came from Alice

Practical example: RSA

- Published in 1977, named after Rivest, Shamir and Adleman
  - One of the first asymmetric cryptosystems
- Integer factorization problem
  - Given a natural number $n = pq$, where p and q are prime numbers, find the prime factors p and q
- Establishing the keys:
  - For a security level t, choose two prime numbers $p, q \geq 2^t$
  - $n := pq$, $\Phi(n) = (p - 1)(q - 1)$
  - Choose $0 < e < \Phi(n)$ with $GCD(e, \Phi(n)) = 1$ (e, $\Phi(n)$: coprime numbers)
  - Compute $0 < d < \Phi(n)$ with $ed = 1 \bmod \Phi(n)$ (d: multiplicative inverse of e)
  - Private key: d, public key: $\{n, e\}$

Using RSA

- Private key, *sk*: *d*, public key, *pk*: {*n*, *e*}
- Message encryption and decryption:
  - Encryption: $c = m^e \bmod n$
  - Decryption: $m = c^d \bmod n$
- Digital signatures
  - In addition to selecting *n*, *d*, *e*, we need a hash function H
  - Reminder: Hash functions take care of the integrity of a message, by generating a fixed-length short digest from a given input message. They are non-invertible functions with collision resistance

Using RSA: Digital Signatures

$\{n, e\}, d$

message m
h := H(m)
$s := h^d \bmod n$

m, s

$\{n, e\}$

h = H(m)
$h^* := s^e \bmod n$
If h = h*: accept
else: reject

Here RSA is only used for the digital signature, the message m is sent on the clear!

Hybrid Cryptography

What is it?

- Public-key cryptosystems do not require the sender and receiver to share a common secret in order to communicate securely (among other useful properties) – convenient!
  - However, they often rely on complicated mathematical computations and are thus generally much more inefficient than comparable symmetric-key cryptosystems.
- Hybrid cryptosystems combine the convenience of a public-key cryptosystem with the efficiency of a symmetric-key cryptosystem
- We need two separate cryptosystems:
  - A public-key cryptosystem for **key encapsulation**
  - A symmetric-key cryptosystem for **data encapsulation**

Encryption

sk_B, pk_B

Message m
Random key k
z := $enc_a(k, pk_B)$
y := $enc_s(m, k)$

Asymmetric cryptography used only to exchange the symmetric key k

y, z

k := $dec_a(z, sk_B)$
m := $dec_s(y, k)$

Symmetric cryptography used to encrypt (long) messages

Hybrid Cryptography

Secure communication

Example: Pretty Good Privacy (PGP) for E-mail communication
https://www.openpgp.org/

sk_A, pk_A

sk_B, pk_B

Message m
Random key k
$z := enc_a(k, pk_B)$
$s := sig(m, sk_A)$
$y := enc_s(m||s, k)$

Asymmetric cryptography used only to exchange the symmetric key k and sign the messages

Symmetric cryptography used to encrypt (long) messages

y, z

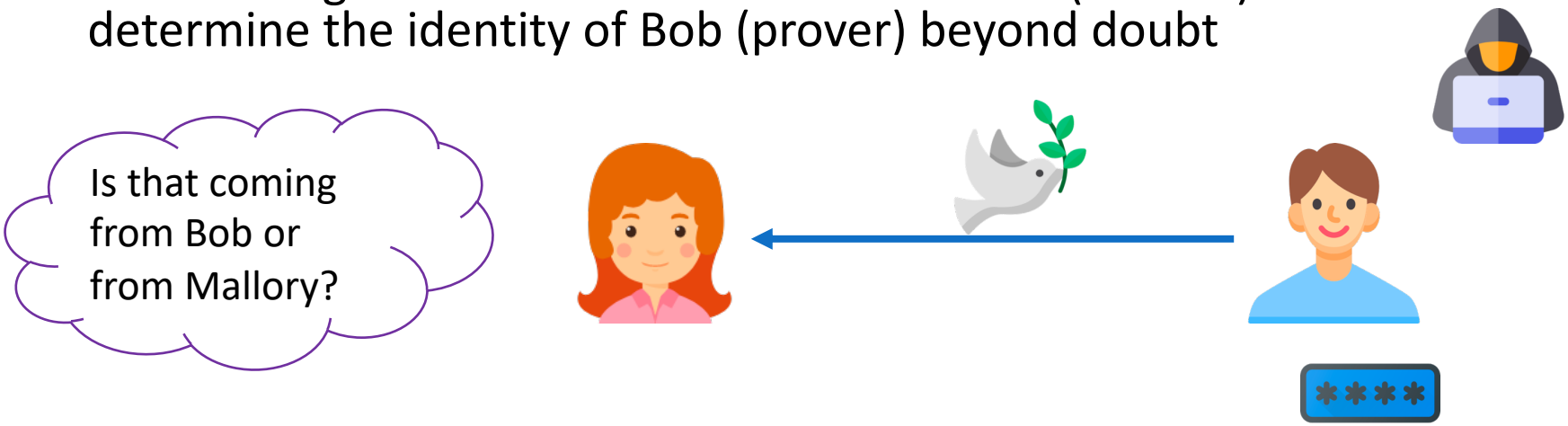
$k := dec_a(z, sk_B)$
$m||s := dec_s(y, k)$
$v := ver(m, s, pk_A)$
if v = True: accept
else: reject

User Authentication

User Authentication

Auhenticity

- Protection goal instance authentication: Alice (verifier) can determine the identity of Bob (prover) beyond doubt



- Examples of instance authentication are:
  - What you know (e.g. user name / password)
  - What you are (e.g. biometric characteristics such as fingerprints)
  - What you know, without having to disclose the secret (challenge-response procedure)
  - What you have (e.g. a bank card)
  - n-factor authentication (e.g. cash machines: bank card + PIN)

Passwords

- Typical example: username / password
  - Login of a client
  - Login for web-services (E-Mail, Forum, Online-Banking)

- They are usually not stored in the clear, but protected with a hash: $p \rightarrow h(p)$

- Disadvantages: susceptible to
  - Dictionary attacks
  - Espionage (e.g. phishing, keylogging, overhearing the channel)
  - Replay-Attacks: overhearing and re-using the message
  - Man-in-the-Middle Attacks

One-Time Passwords

- Enhancement: One-Time Passwords
  - Each password is only used once
  - Replay-Attacks can be prevented
- Problem: both sides have to know the passwords
- Two possibilities:
  - Lists of passwords
  - Password generators

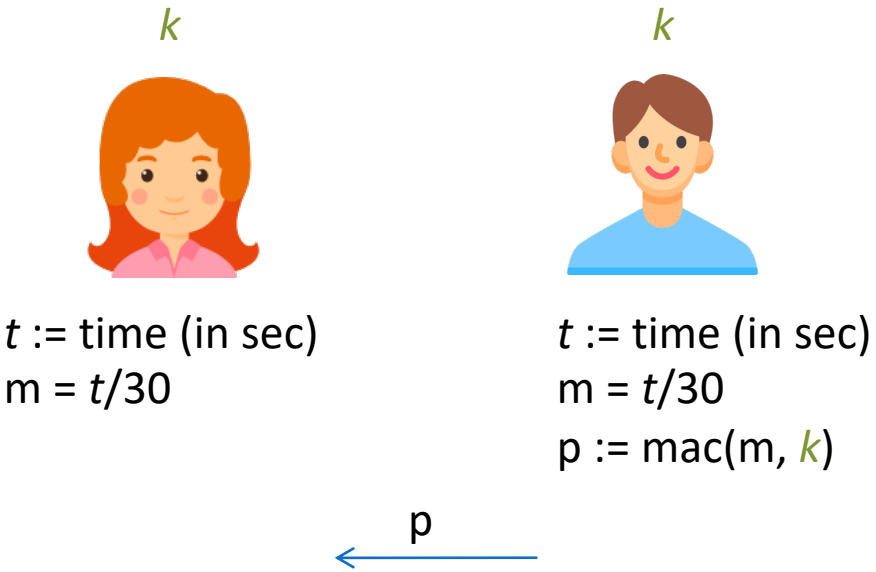
One-Time Passwords: Lists of Passwords

- Typical case study: transaction authentication number (TAN) in Online-Banking, e.g. to confirm a wired transfer
- Both communication partners get a list of passwords
- The passwords are then used as follows;
  - Sequential selection: top to bottom
  - Indexed selection: the verifier selects which password will be used next (e.g. ranking in the list)

One-Time Passwords: Password Generators

- Passwords are derived from a previously exchanged secret *k*
- We can differentiate:
  - Time-controlled generators
  - Event-controlled generators
  - Challenge-Response generators

Time-controlled Generators



- The password is generated from
  - *k* (previously exchanged secret) und
  - *t* (authentication time). Problem: the time by the prover and the verifier is not exactly the time → tolerance (e.g. 30 sec)

- Example:
  - Google Authenticator

Event-controlled Generators: Lamport-Hash

- Generation of the one-time password from
  - k (previously exchanged secret)
  - Random number r (does not need to be kept secret)
  - Seed $S = h(r||k)$ (h is a hash function)

- Generation of the one-time password :
  - First password: $p_1 = h^N(S) \rightarrow$ N applications of h
  - Second password: $p_2 = h^{N-1}(S) \rightarrow$ $N-1$ applications of h
  - t–th password: $p_t = h^{N-(t-1)}(S)$

Event-controlled Generators : Lamport-Hash

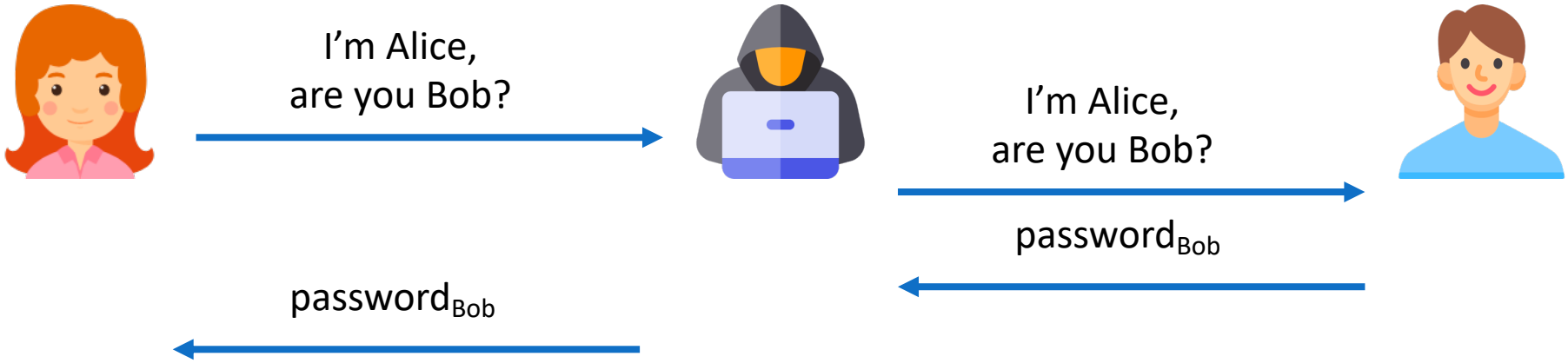
- We cannot compute $p_{t+1} = h^{N-(t-2)}(S)$ from $p_t = h^{N-(t-1)}(S)$
  - $h^{N-(t-1)}(S) \rightarrow h^{N-(t-2)}(S)$ is the inversion of h to $h^{N-(t-1)}(S)$
  - A good hash function does not allow the computation of h^{-1}!!

- Problem: at some point, we will have produced N passwords
  - Reinitialise: choose a new random value r'
  - Create a new seed $S = H(r'||k)$

One-Time Passowrds: Summary

- Advantages:
  - Secure against passive attacks (espionage): a new password each time
  - Replay attacks will be thus prevented

- Still possible: Man-in-the-Middle Attack (active attack)



- How to prevent the attacks? Double authentication

2-Factor Authentication v1

- 2-factor authentication based on
  - Property (security element) and
  - Knowledge (PIN)
- Implementation: store the cryptographic key and the PIN in a non-readable area of the chip
- Authentication:
  - With a Challenge-Response protocol
  - The cryptographic key will be released by using the PIN

2-Factor Authentication v2

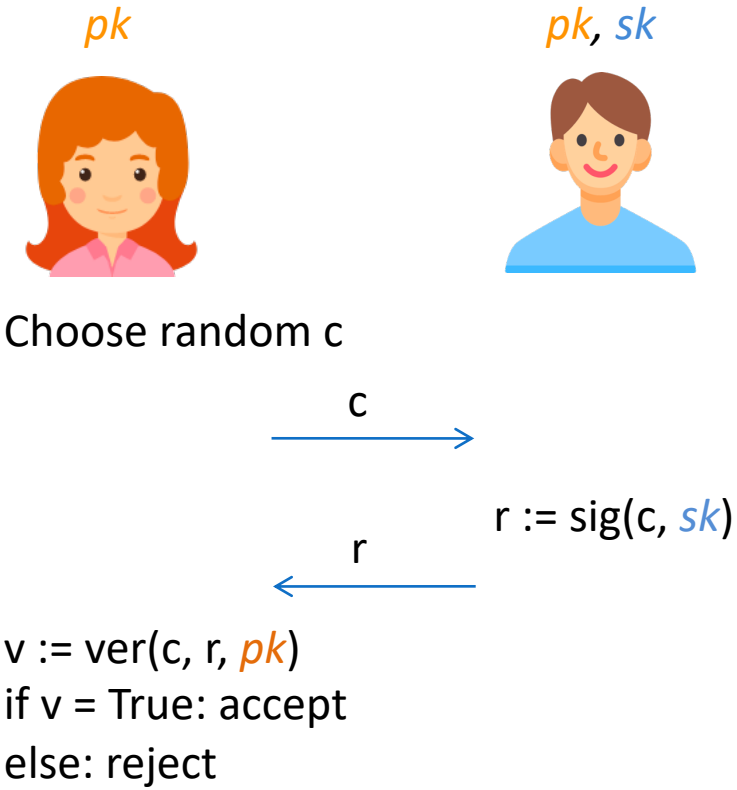
- 2-factor authentication based on
  - Characteristic (a biometric characteristic) and
  - Property (pre-registered smartphone)

- Authentication:
  - Biometric authentication of the subject
  - Approval of a wired transfer through a PIN-Code sent via SMS or with a Secure-TAN App in the smartphone

Zero-Knowledge-Protocols

- Even if a secure one-way hash function is used for password encryption, user authentication is not yet secure
  - The password must be entered by the user in plain text!
- With Zero-Knowledge-Protocols, the prover knows a secret, and he can convince any verifier that he knows the secret, and thus identify himself
- No one learns anything about the secret

Challenge-Response Protocol with SIgnatures



- As long as only Bob knows *sk*, Alice can be sure that Bob is who he says he is
- Examples:
  - ssh's Challenge-Response-System with RSA
  - Secure Remote-Passwort (SRP)
  - Challenge Handshake Authentication Protocol(CHAP)

Fiat-Shamir Identification Protocol: Keys

- Zero-Knowledge-Protocol, similar to RSA

- The prover (Bob) chooses two prime numbers p, q and computes $n = pq$
- He then chooses $s \in \mathbb{Z}_n$ and computes $v := s^2 \, mod \, n$
- Bob's public jey is $\{v, n\}$, and his secret key is s, a square root of $v \, mod \, n$.

- Bob wants to prove to Alice that he knows a square root s of v
  - Finding the secret, when n is big, is impossible nowadays!

Fiat-Shamir Identification Protocol

pk = (v, n)

pk = (v, n), sk = s

Choose $r \in \mathbb{Z}_n$
Compute $x := r^2 \bmod n$

x ←

Choose $b \in \{0,1\}$

b →

Compute $y = rs^e \bmod n$

y ←

$y' = xv^e \bmod n$
If $y^2 = y'$, accept
Else: reject

$y^2 = (rs^e)^2 \bmod n$
$= r^2(s^2)^e \bmod n$
$= xv^e \bmod n$

Digital Signatures

- A good signature has the following characteristics:
  - It is **authentic**, i.e., it shows that the signatory has signed willingly
  - It is **unforgeable**. It proves that the signatory and no other person signed the document
  - It is **not reusable** The signature cannot be copied to another document - no replay attacks!
  - The signed **document is unchangeable**. It cannot be changed after signing
  - The signature is **binding**. The signatory cannot later claim that she has not signed the documen
- We have already seen an example: RSA-Signatures
- Others include the use of the discrete logarithmus problem: DSA

Blind Signatures

- A signature is called blind if the signer cannot see the document when signing it
  - Important for electronic currency!

- Exampel with RSA-signatures:
  - Classic RSA-Signature: $s := (hash(m))^d \ mod \ n$
  - The blind version uses a randon number r, such that r co-prime to n (i.s. $\mathrm{GCD}(r, n) = 1$).

Blind Signatures with RSA

Choose random r with gcd(r, n) = 1

t := m r^e mod n

t ←

s' := t^d mod n

s' →

s := s' r^{-1} mod n

s' r^{-1} mod n = t^d r^{-1} mod n = (m r^e)^d r^{-1} mod n = m^d r r^{-1} mod n = m^d mod n

Biometrics

Why biometric recognition?

- We need to identify ourselves in a daily basis
- Impossible to remember 100 different passwords

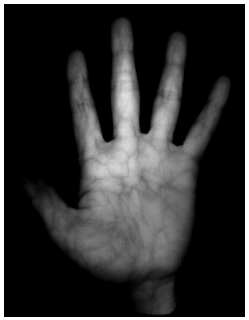
- Losing or forgetting our password / token is easy

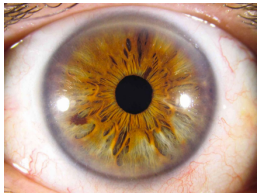
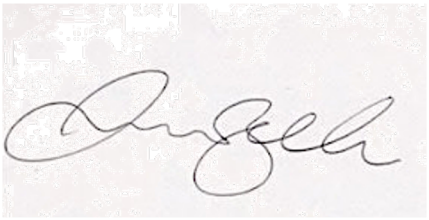
Why not use our body features or behavioural patterns?

Biometric characteristics

- Classification:
  - Physiological
  - Behavioural

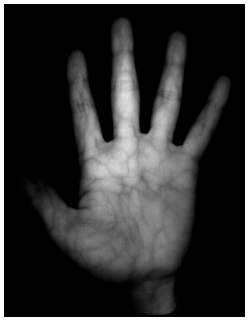
- Properties:
  - **Universality**: everybody should possess it
  - **Distinctiveness**: should have enough intervariability
  - **Permanence**: should not vary through time
  - **Collectability**: should be easy to acquire
  - **Performance**: should have good error rates
  - **Acceptability**: user should not be reluctant to use it
  - **Circumvention**: difficult to bypass

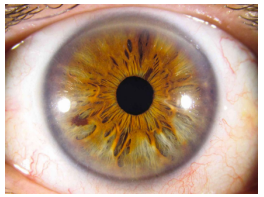



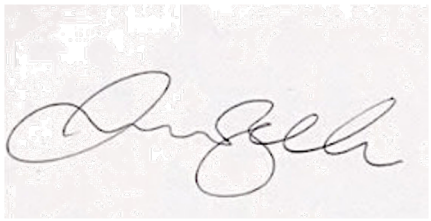



Advantages and disadvantages of biometrics

- No need to remember passwords or carry tokens
- Impersonation can be detected
- A single characteristic can be used in multiple applications, without security decrease



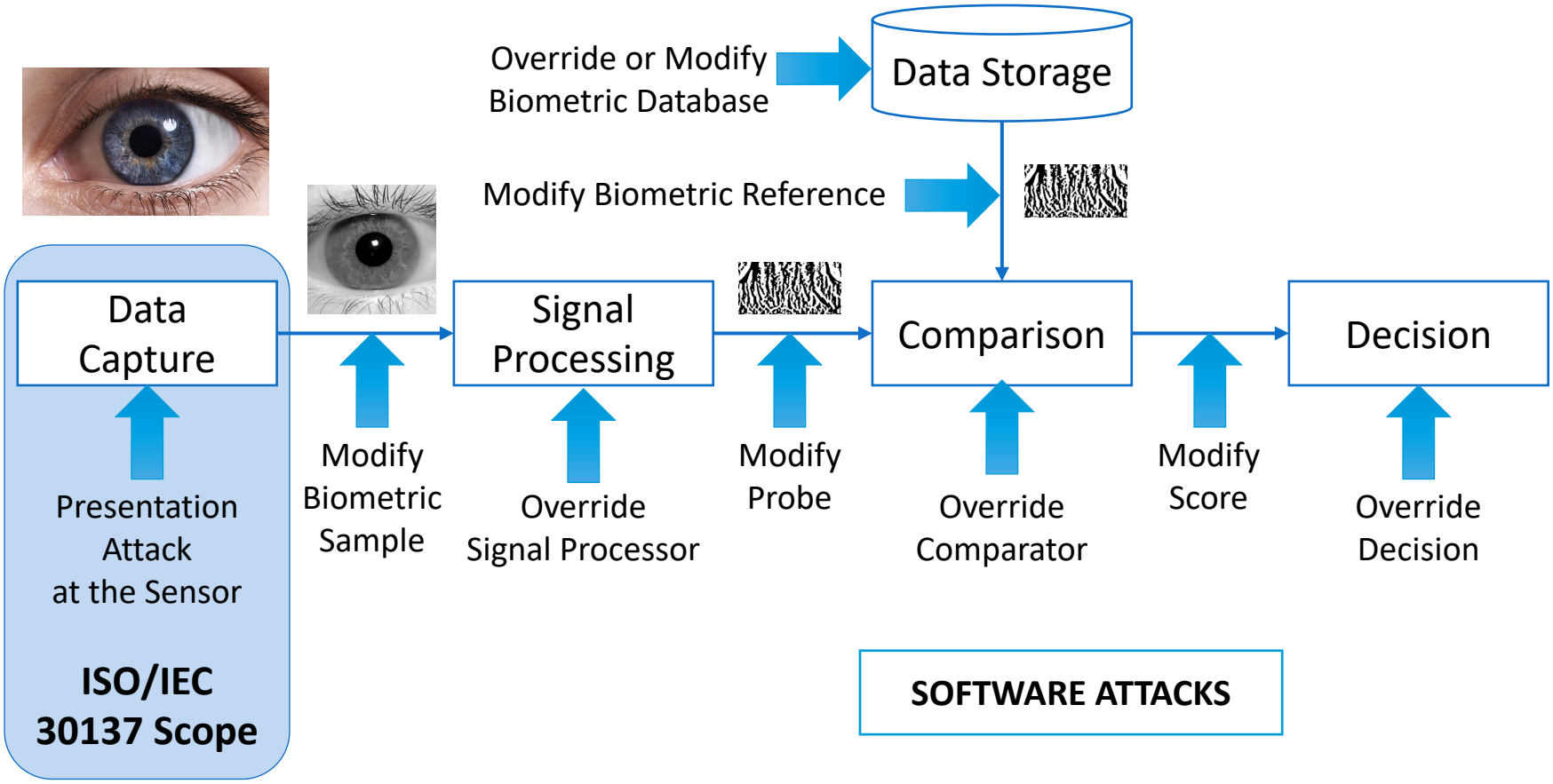




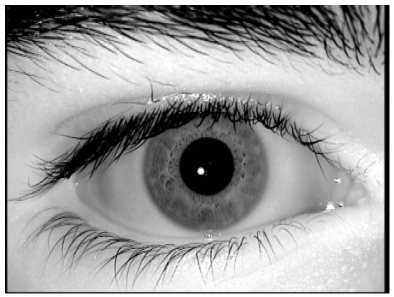
- Presentation Attacks (PA)
- Renewability
- Biometrics are no secrets
- Sensitive information

Biometric Systems & Attacks

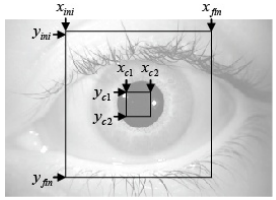
[ISO/IEC 30137 on Biometric Presentation Attack Detection]

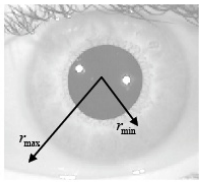


Example I: iris recognition

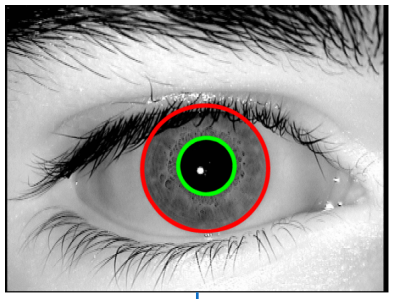


Sample

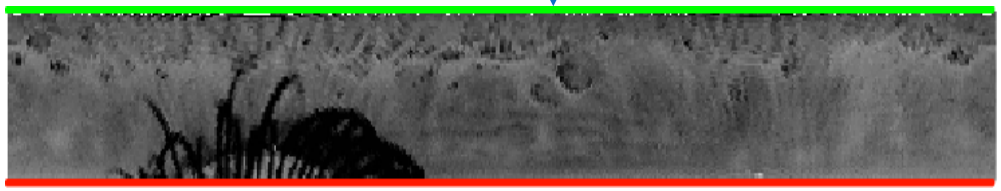


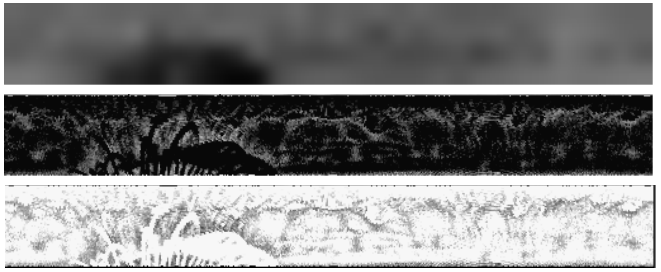


Segmentation



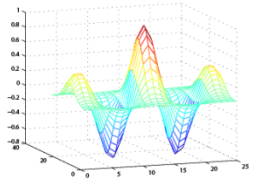
Normalization

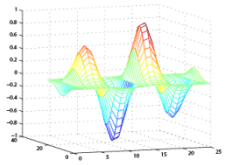




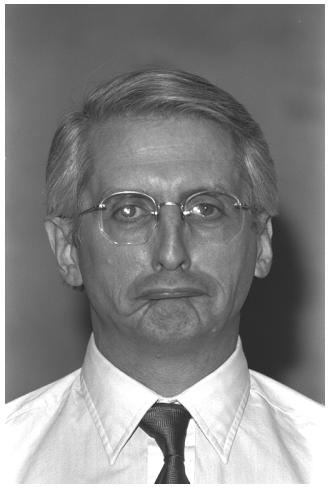
Template: T

Feature Extraction





Example II: face recognition

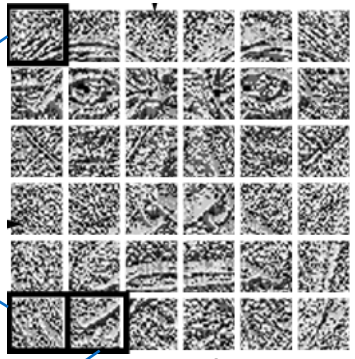


Sample

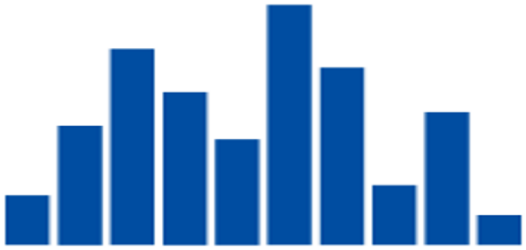
Segmentation

Normalization

Feature Extraction



One histogram per region of the LBP features

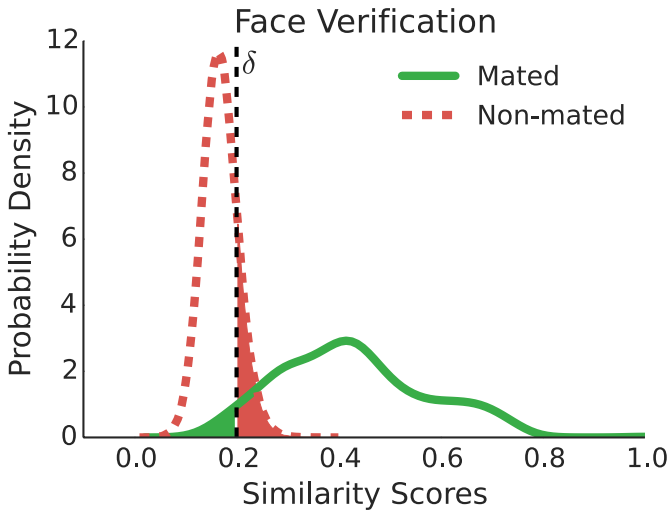


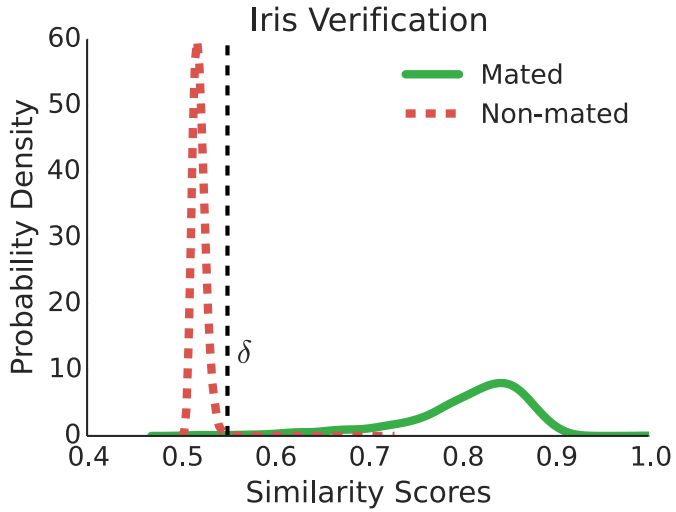
Final Template

Evaluating the accuracy

[ISO/IEC 19795 on Biometric performance testing and reporting]

- Plot mated and non-mated score distributions
- Establish a verification threshold: δ



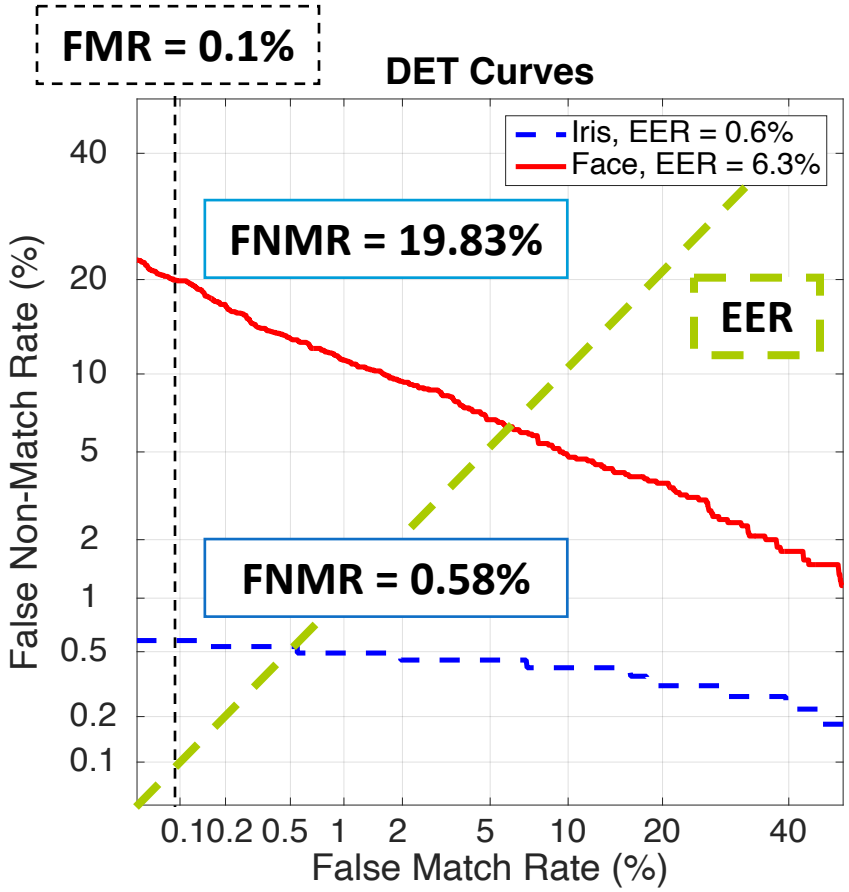


- δ determines the **FMR**
- … and the **FNMR**

Benchmarking systems

[ISO/IEC 19795 on Biometric performance testing and reporting]

- Compare all operating points with a **Detection Error Trade-off (DET)** curve
- The point at which FMR = FNMR is defined as **Equal Error Rate (EER)** - the lower, the better
- Report FNMR at fixed FMR – e.g., FMR = 0.1%



Biometrics: sensitive data

AADHAAR

European Commission

- Wide deployment of biometrics:
  - Large scale national and international projects
  - Banking apps, ATMs
  - Smartphone unlocking

We are family

CTBC BANK 中國信託銀行

Pay

iPhone

android

MasterCard

ING DIRECT

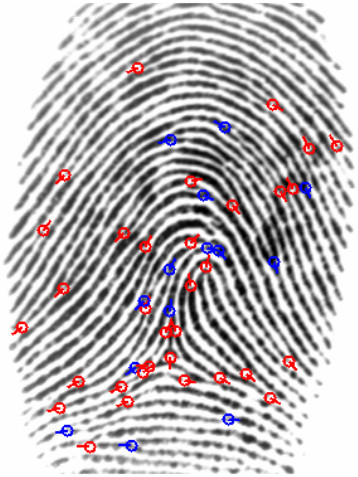
- Biometrics are classified as sensitive data

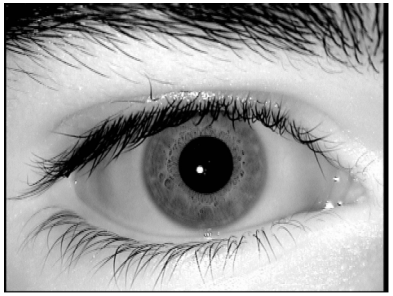
[EU 2016/679 Data Protection Regulation]

- And we cannot prevent databases leakage

Inverse biometrics attacks

- It was a common belief that the stored templates revealed no information about the biometric characteristics:

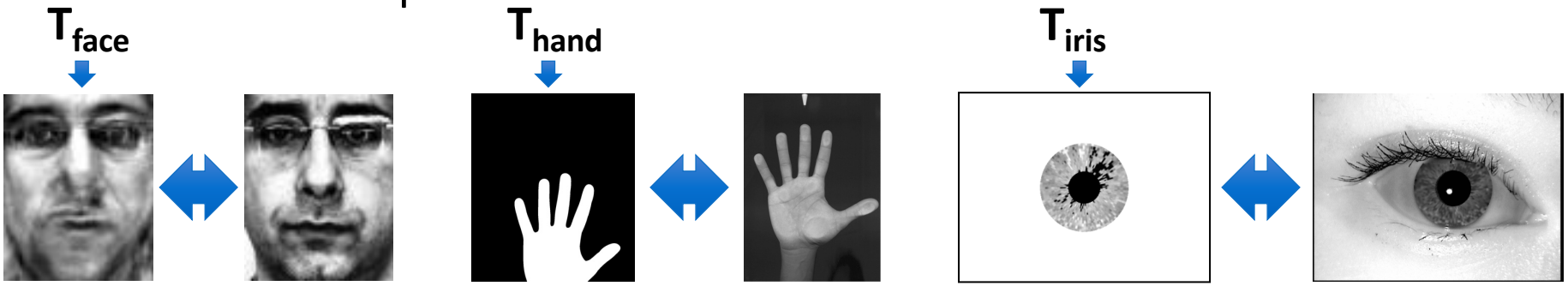


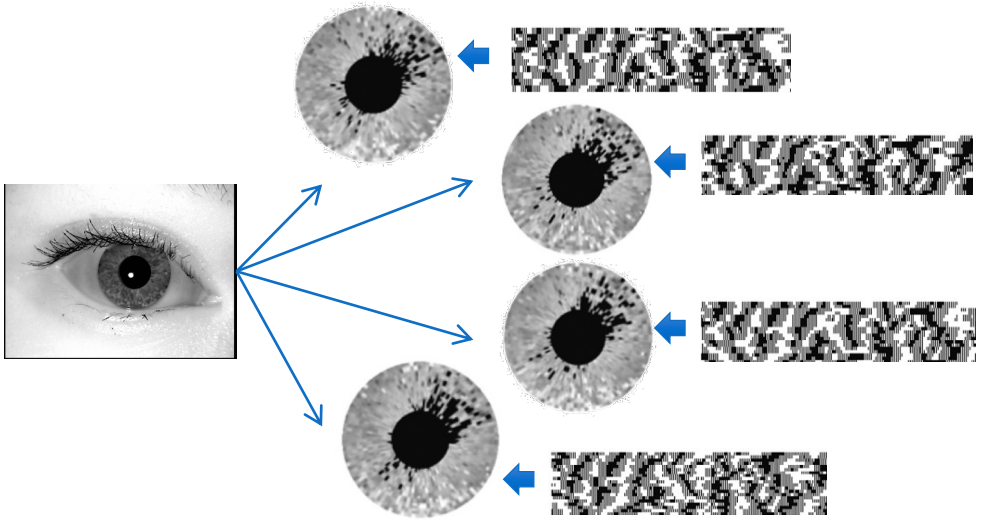


- However, biometric samples can be recovered from the stored unprotected templates

Inverse biometrics attacks: Hill-Climbing

- Based on the HC algorithms presented before, we can reconstruct biometric samples:

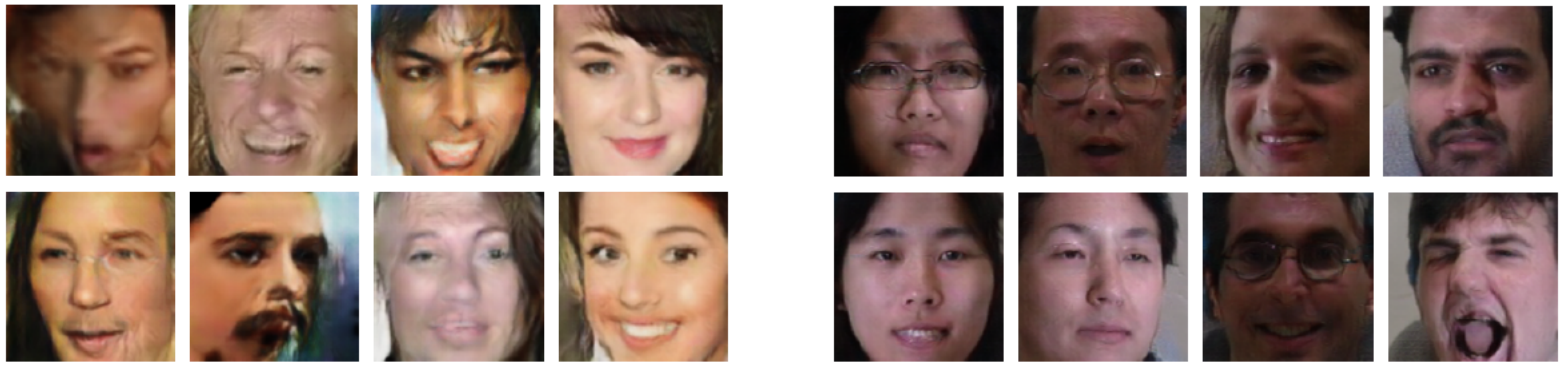




[M. Gomez-Barrero *et al.*, *Int. Conf. on Biometrics*, 2012]
[M. Gomez-Barrero *et al.*, *Information Sciences*, 2014]
[J. Galbally, *et al.*, *Computer Vision & Image Understanding*, 2013]

Inverse biometrics attacks: deep learning

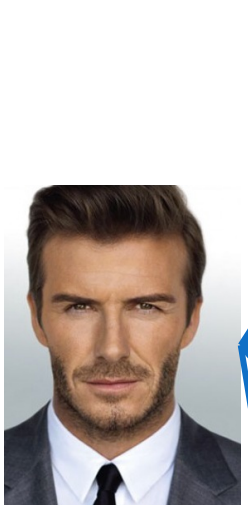
- Also vulnerable to inverse biometrics attacks!
- A neighbourly de-convolutional network (NbNet) can be used to reconstruct facial templates from FaceNet [Schroff *et al., Proc. CVPR*, 2015]
- Same assumptions as before
- Over large open access databases, success rates over 73% and 95% are achieved



[Mai *et al., IEEE T-PAMI*, 2018]

Cross-matching attacks

- We can enroll with a single characteristic in different applications



T_{mail}

Santander CONSUMER BANK

T_{bank}

HOCHSCHULE ANSBACH

T_{job}

$T_{facebook}$

$T_{linkedin}$

Same person??

Templates need to be protected, so that no one can find out on which applications we are enrolled

Protecting the subject's privacy

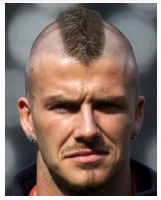
[ISO/IEC IS 24745 on Biometric Information Protection]

Male,
white, 40s…

Irreversibility

=?

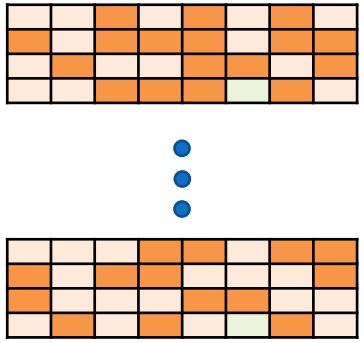
Unlinkability



Renewability

K_1

K_n



At the same time, **accuracy**, **template size** and **verification speed** must be preserved.

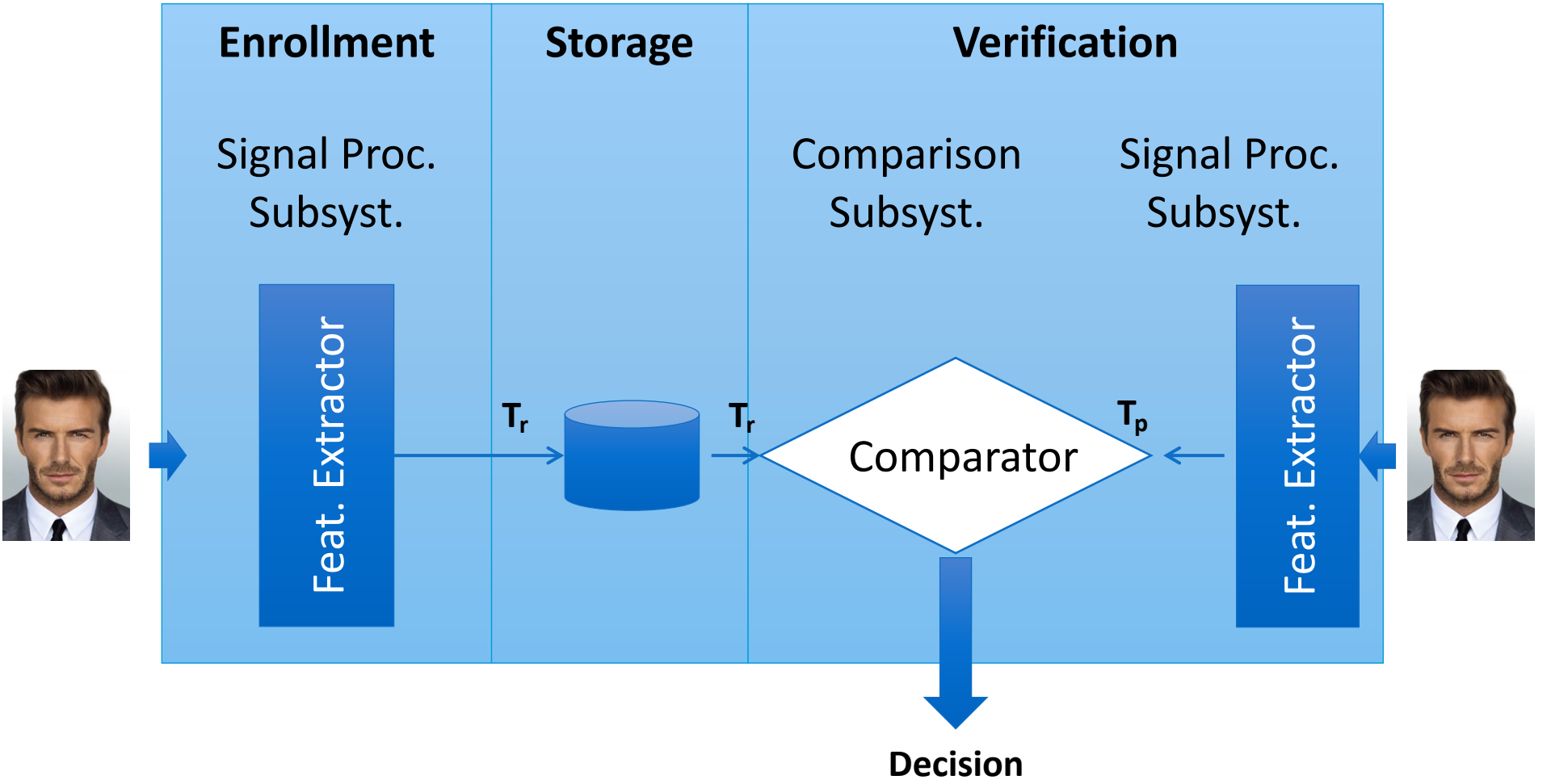
Biometrics vs cryptographic protocols

- How can we solve this issue? Encryption of the reference? Hashing?
- Difference between passwords and biometric samples
  - Biometric measurements are influenced by noise

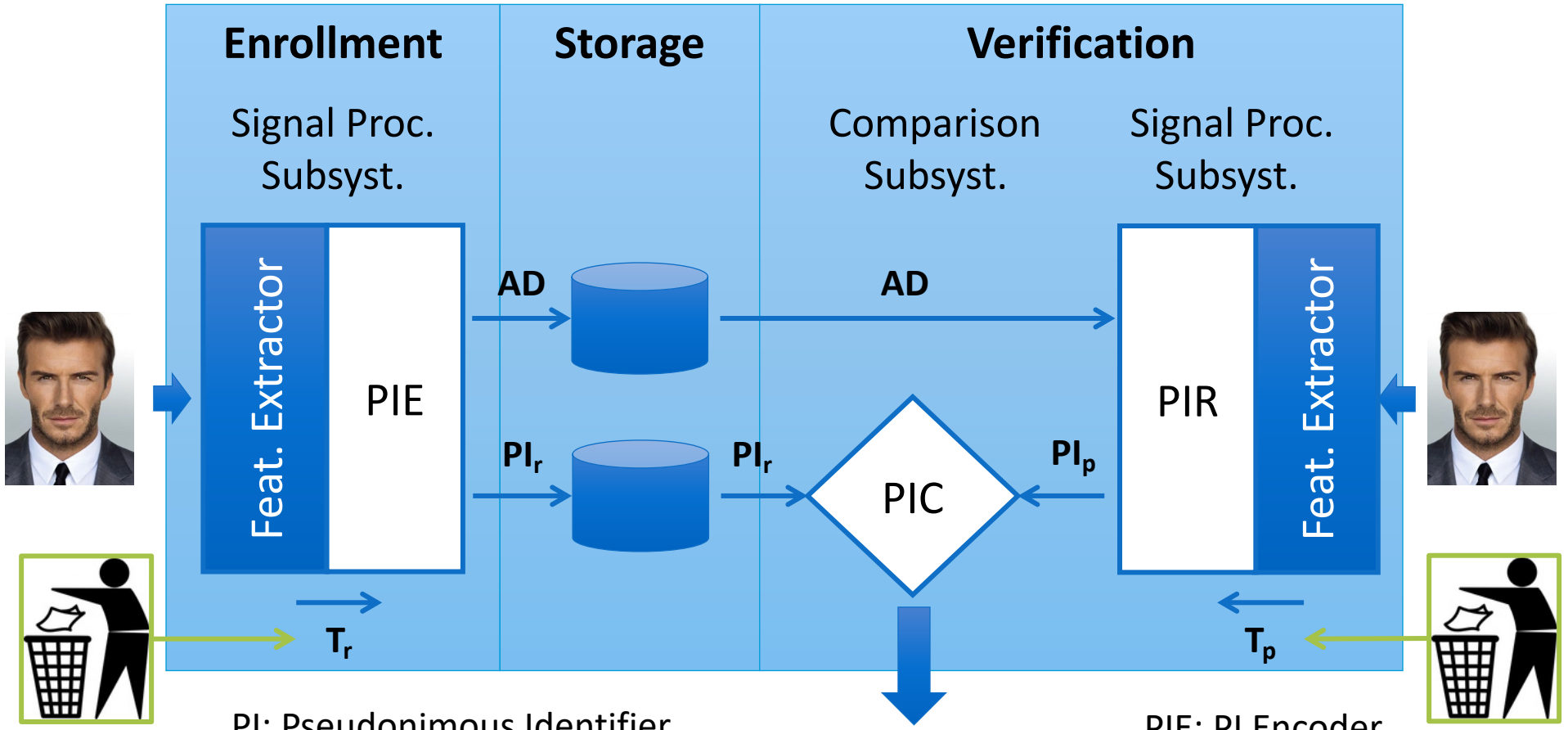
  - Cryptographic one way functions (e.g. hashes) are (by purpose) extremely sensitive to smallest changes in the input data

    h(01000101) is not similar, but very different from h(01010101)

Traditional architecture



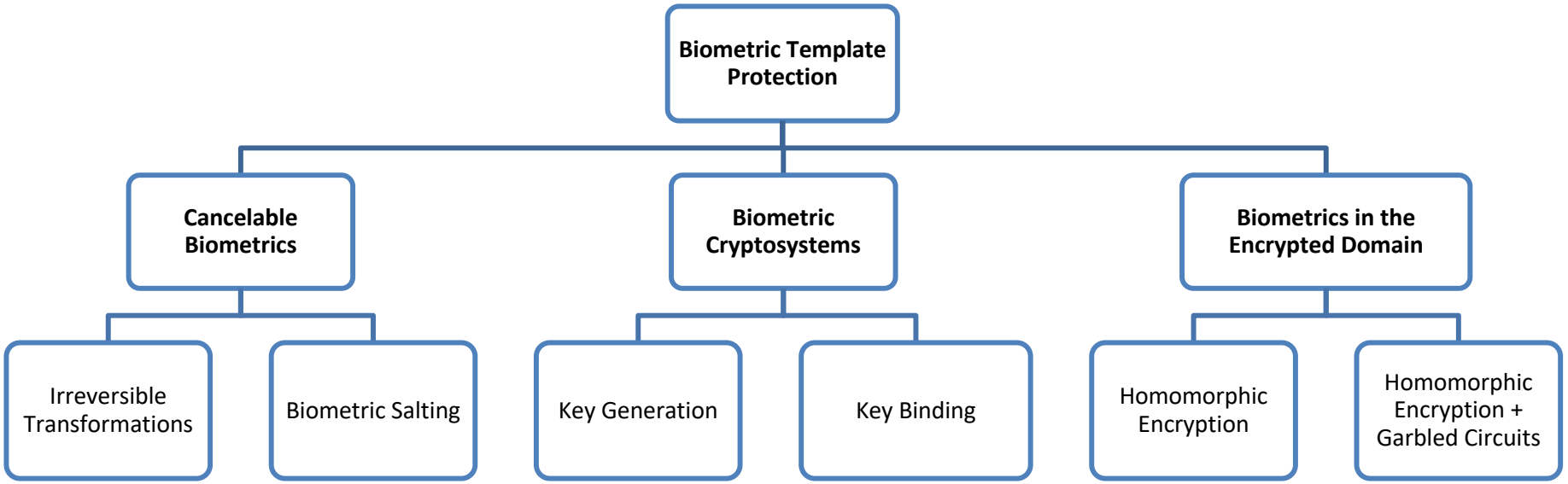
Biometric Template Protection (BTP) architecture



Pseudonymous Identifier (PI) Framework

- Two-stage conversion of captured biometric samples to protected templates.
  - For permanent protection: protected storage, transmission and comparison
- Impossible to retrieve the original biometric sample from the protected template
- A template represents identification data for a specific purpose or application only

BTP Approaches: Summary



Geometric trans.
Feature mixing
BioTokens
Bloom Filters

Correlation
Random projection
Visual crypto

Quantization

Fuzzy vault
Fuzzy commitment
BioHashing

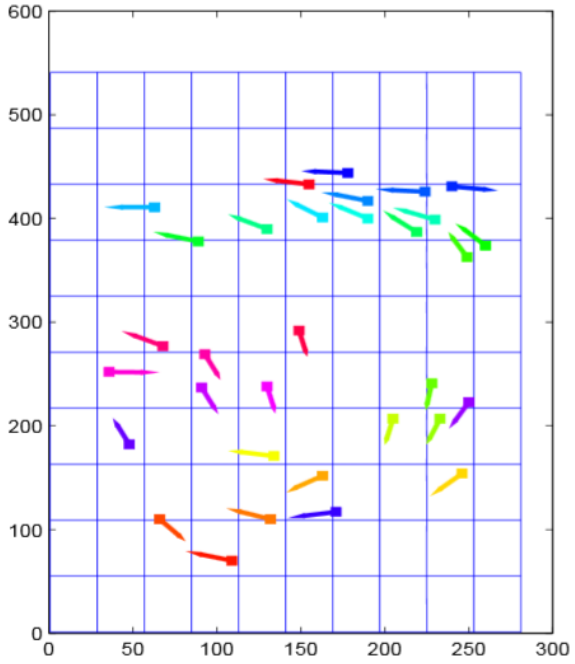
BTP approaches: Cancelable biometrics

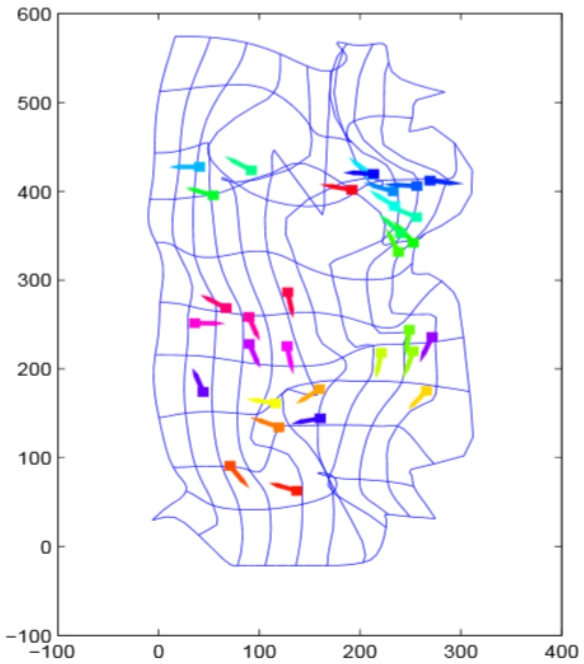
- Cancelable biometrics consist of intentional, repeatable distortions of biometric signals based on transformations which provide a comparison of biometric templates in the protected domain.

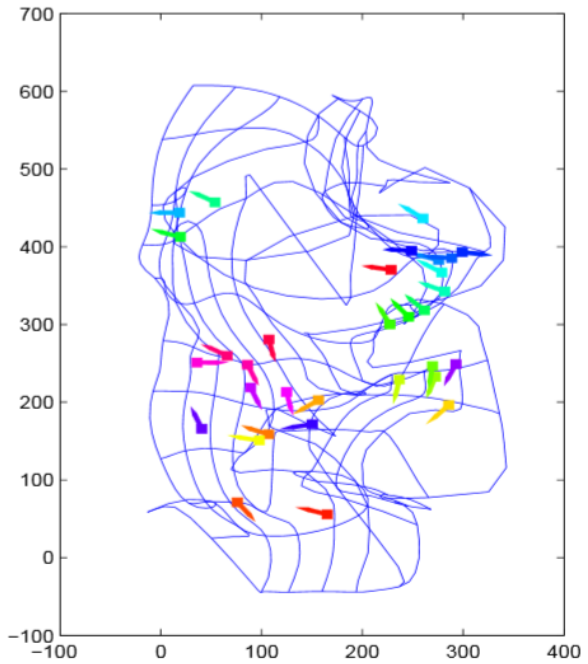
- Two types:
  - **Non-reversible transformations** of the biometric data or unprotected templates.
  - **Biometric salting**, in which Auxiliary Data (AD) is blended with biometric data to derive a distorted version of the biometric template.

Cancelable biometrics: Surface folding

- One of the first approaches is based on surface folding







[Ratha *et al.*, IEEE T-PAMI 2007]

Cancelable biometrics: Visual cryptography

Public "host" images

Reference sample



[Ross and Othman, *IEEE T-IFS,* 2011]
[Naor and Shamir, *Proc. EUROCRYPT,* 2011]

Only with **access to all sheets** can we reconstruct the sample


PI

OR

Reconstructed sample

Stored in separate databases!

BTP Approaches: Cryptobiometrics

- These methods combine cryptographic keys with transformed versions of the original biometric templates to obtain secure templates.

- In most cases, some public information, known as helper data or auxiliary data, is generated.

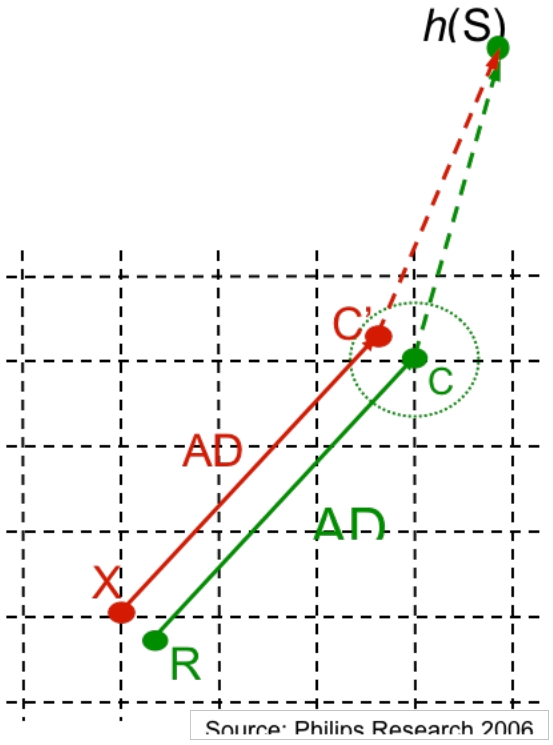
- Two types:
  - **Key binding** schemes, where AD are obtained combining the key with the biometric template. At verication time, applying an appropriate key retrieval algorithm to the probe biometric sample, the key is obtained from the AD.
  - **Key generation** schemes, where both the AD and the key are generated directly from biometric data. Again, at verication time, a key is recovered from the probe sample using the AD.

Cryptobiometrics : Fuzzy extractor

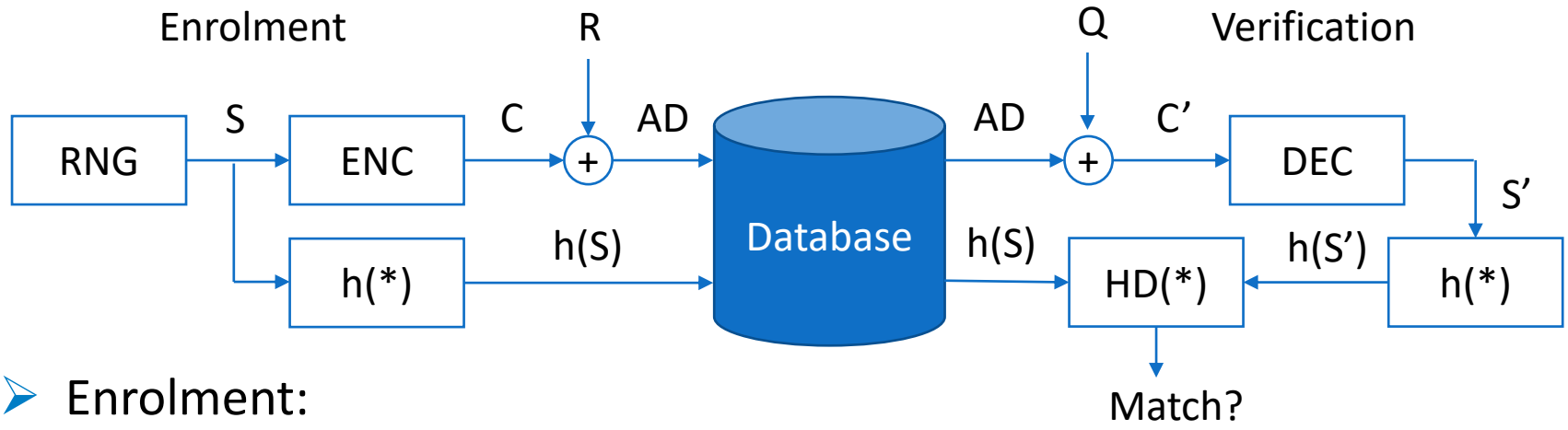
- To address the variability accross samples, Error Correcting Codes (EECs) are used (grid points represent the ECC code words)

- At enrolment:
  - A random codeword C is chosen
  - R is the binary biometric reference template
  - Helper data: AD = C - R
  - Store AD and h(S) = h(DEC(C))

- Verification
  - X is binary probe template
  - X + AD=C'
  - S' = DEC(C')
  - h(S) == h(S')?



Cryptobiometrics: Fuzzy commitment



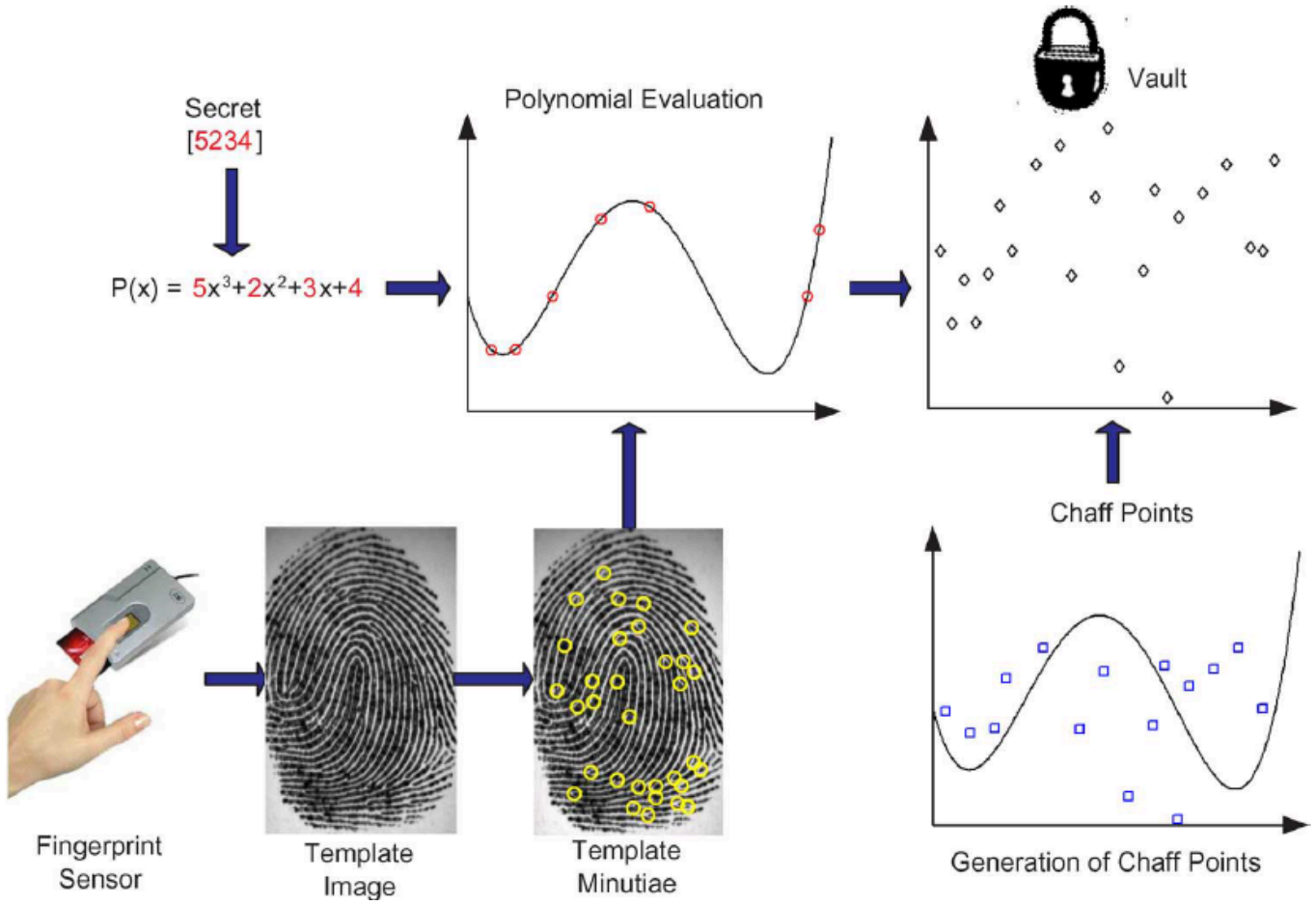
- Enrolment:
  - C is the codeword generated for the random string S
  - R is the binary extract of the reference vector
  - AD = C XOR R is the public AD
  - {h(S), AD} are stored as reference
- Verification:
  - C' = AD XOR Q (query vector)
  - HD(C, C') needs to be smaller than the error correction capabitlities

[Jules and Wattemberg, *Proc. ACM CCCS,* 1999]

Cryptobiometrics: Fuzzy vault

[Juels and Sudan, *Designs, Codes and Cryptography*, 2006]
[Nandakumar *et al.*, *IEEE TIFS*, 2007]

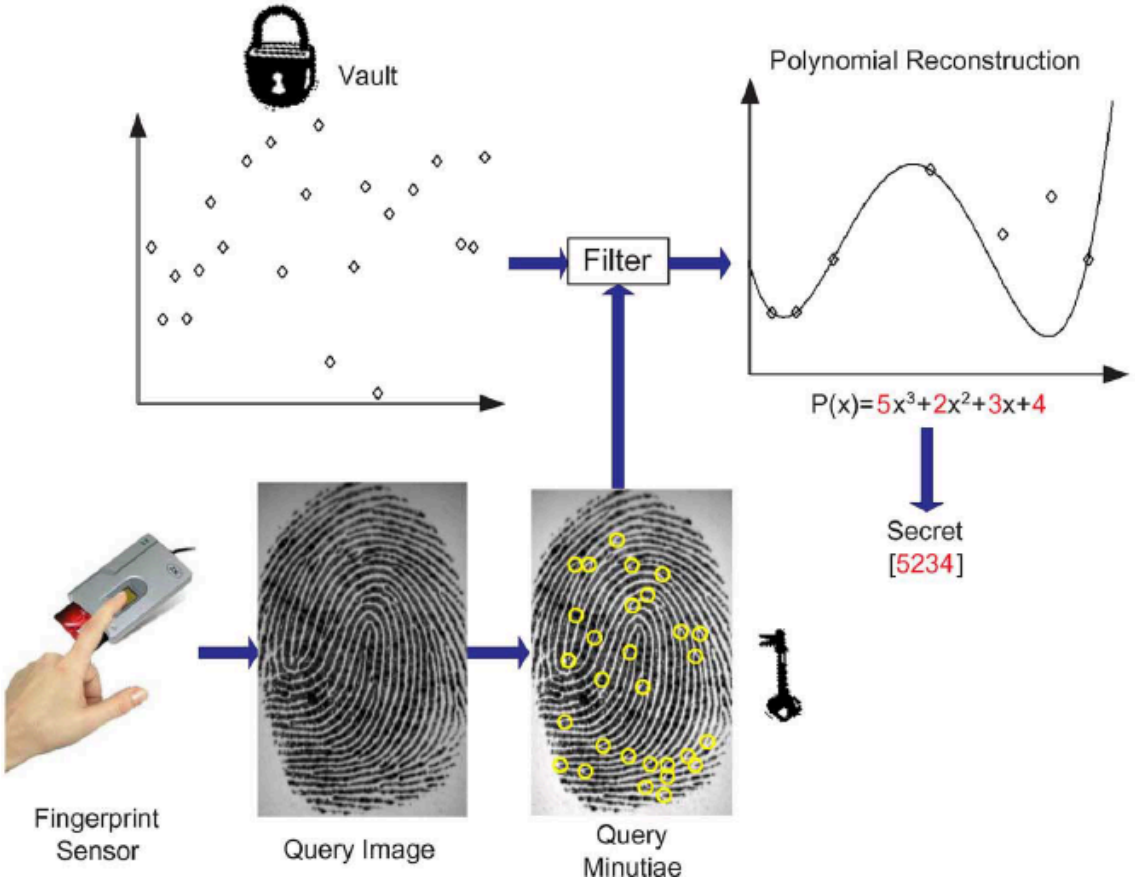
- Enrolment



Cryptobiometrics: Fuzzy vault

[Juels and Sudan, *Designs, Codes and Cryptography*, 2006]
[Nandakumar *et al.*, *IEEE TIFS*, 2007]

- Verification



BTP Approaches: Biometrics in the Encrypted Domain

- Homomorphic Encryption (HE) schemes allow for computations to be performed on ciphertexts, with no additional AD, and which generate encrypted results which decrypt to plaintexts that match the result of the operations carried out on the original plaintext
- This solves the issue of decryption before authentication...
- But there is till no free lunch! HE is computationally expensive
  - Not straightforward to implement for complex comparators
- Garbled circuits can also be employed for particular operations

[Gomez-Barrero *et al.*, *Pattern Recognition* 2017]

[Gomez-Barrero *et al.*, *IEEE Access* 2017]

[Kolberg *et al.*, *Proc. WIFS* 2019]

[Boddeti, *Proc. BTAS* 2018]

Aditive Homomorphic Encryption

$$D_{sk}\left(m_1^* \cdot m_2^* \mod n^2\right) = m_1 + m_2 \mod n$$

Product of ciphertexts

Sum of plain texts

$$D_{sk}\left((m_1^*)^l \mod n^2\right) = m_1 \cdot l \mod n$$

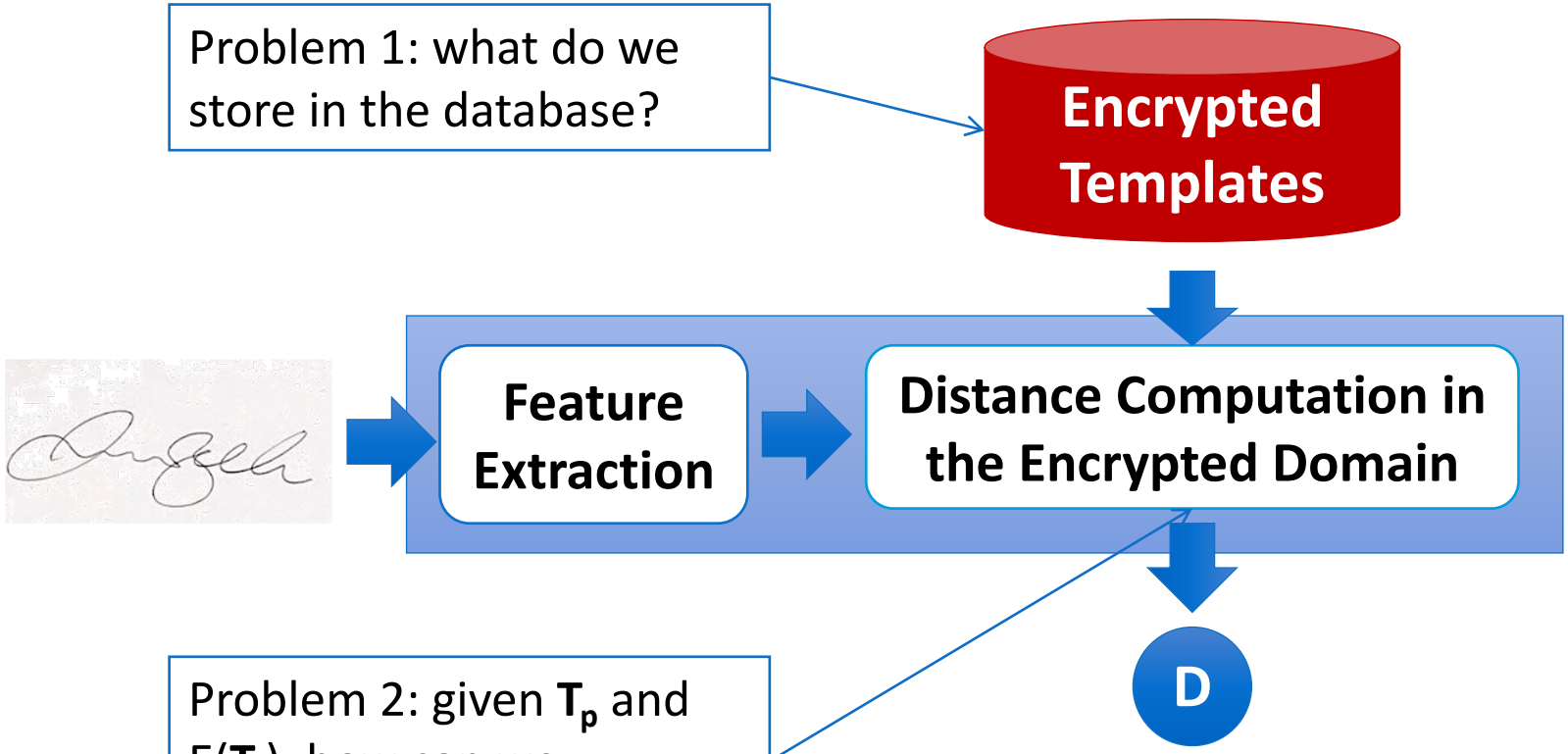
Exponentiation of ciphertext and plain text

Product of plain texts

BTP based on Homomorphic Encryption

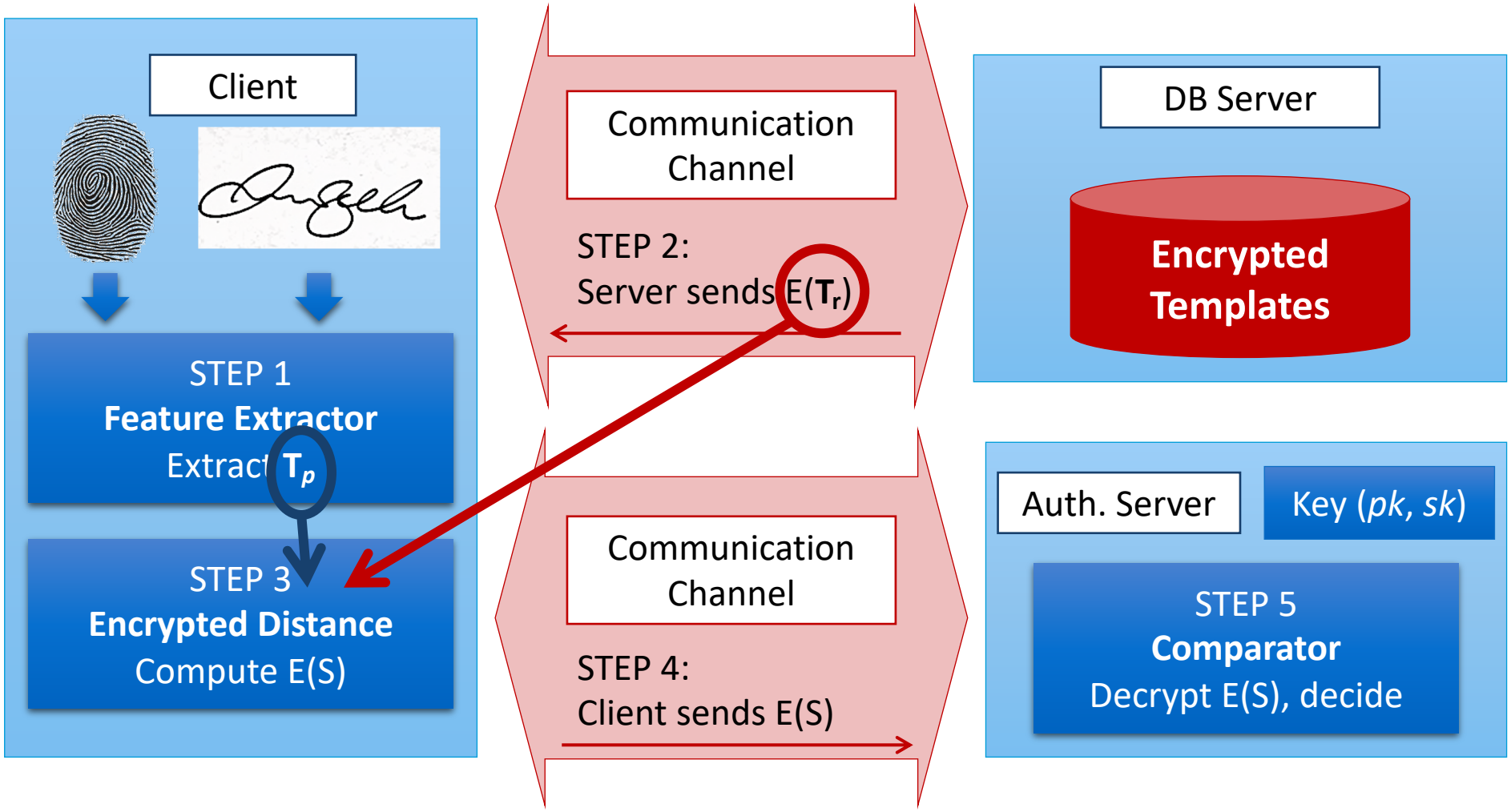
Additionally, only integer values can be handled

Problem 1: what do we store in the database?



Problem 2: given $\mathbf{T_p}$ and $E(\mathbf{T_r})$, how can we compute $E(d(\mathbf{T_p}, \mathbf{T_r}))$?

Multi-BTP based on Homomorphic Encryption



Encrypted distance computation

Euclidean distance: Given two vectors $\mathbf{T}_p$ and $E(\mathbf{T}_r)$, of length F

$$S_{euc} = \sum_{f=1}^{F} p_f^2 + r_f^2 - 2p_f r_f$$

Encrypted Euclidean distance: Given two vectors $\mathbf{T}_p$ and $E(\mathbf{T}_r)$, of length F

$$E\left(S_{euc}\right) = \prod_{f=1}^{F} E\left(1\right)^{p_f^2} \cdot E\left(r_f^2\right) \cdot E\left(r_f\right)^{-2p_f}$$

Encrypted reference template stored in DB

Probe template

Case Study: Cancelable Biometrics Based on Bloom Filters

BTP Evaluation

Reproducible Research

- Public Baseline Systems
- Public DBs
- Knowledge Attacker
- Evaluation Protocol

ISO Requirements Evaluation

- Analysis 1: Accuracy
- Analysis 2: Irreversibility
- Analysis 3: Unlinkability
- Analysis 4a: Robustness to (Cross-Matching) Attacks
- Analysis 4b: Computational Load Increase

Accuracy degradation

- Most BTP schemes transform either the sample (e.g. surface folding) or the template (e.g., fuzzy vault)
- That leads to the addition of noise or information loss, which in turn leads to a decrease in accuracy
- We need to assess such performance loss in accordance with the ISO/IEC 19795:
  - Compute FMR and FNMR for the baseline system AND the BTP scheme
  - Following a common experimental protocol
  - Compare in terms of DET plots
    - The Equal Error Rate (EER), where FMR = FNMR, is not enough!!

Irreversibility analysis

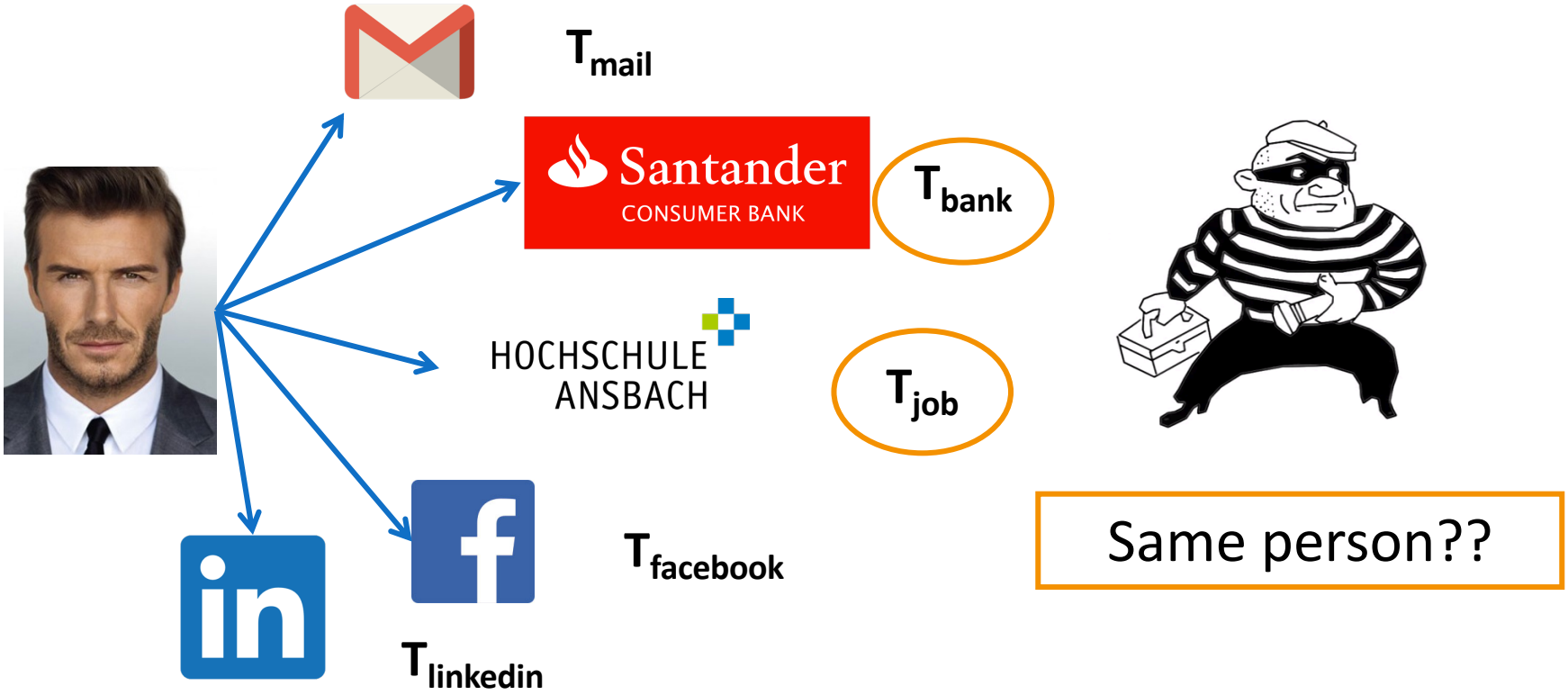
- How can we analyse irreversibility? Following cryptographic paradigms?

- Careful! Some assumptions are not valid:
  - Uniformity of data – neighbouring bits are correlated!!
  - In fact, some biometric templates (e.g., finger vein or fingerprint minutiae spectral representation) are compared in terms of their cross correlation!
  - There are also symmetries

- Therefore, we need to model such correlations and take them into account in the computations

Cross-Matching Attacks

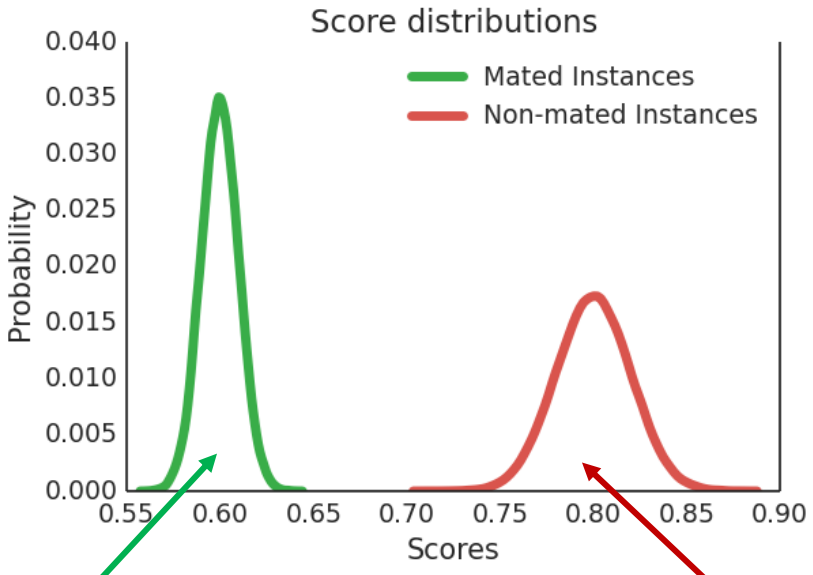
- We can enroll with a single characteristic in different applications



Cross-Matching Attacks: How to?

T_{job} T_{bank}

$s = LS\ (T_{job}, T_{bank})$



s here ➔ success!! ☺

s here ➔ try again!! ☹

s can be the dissimilarity score of the system or any other dissimilarity score, such as values extracted from partial decoding in fuzzy schemes

Unlinkability Analysis: New Approach

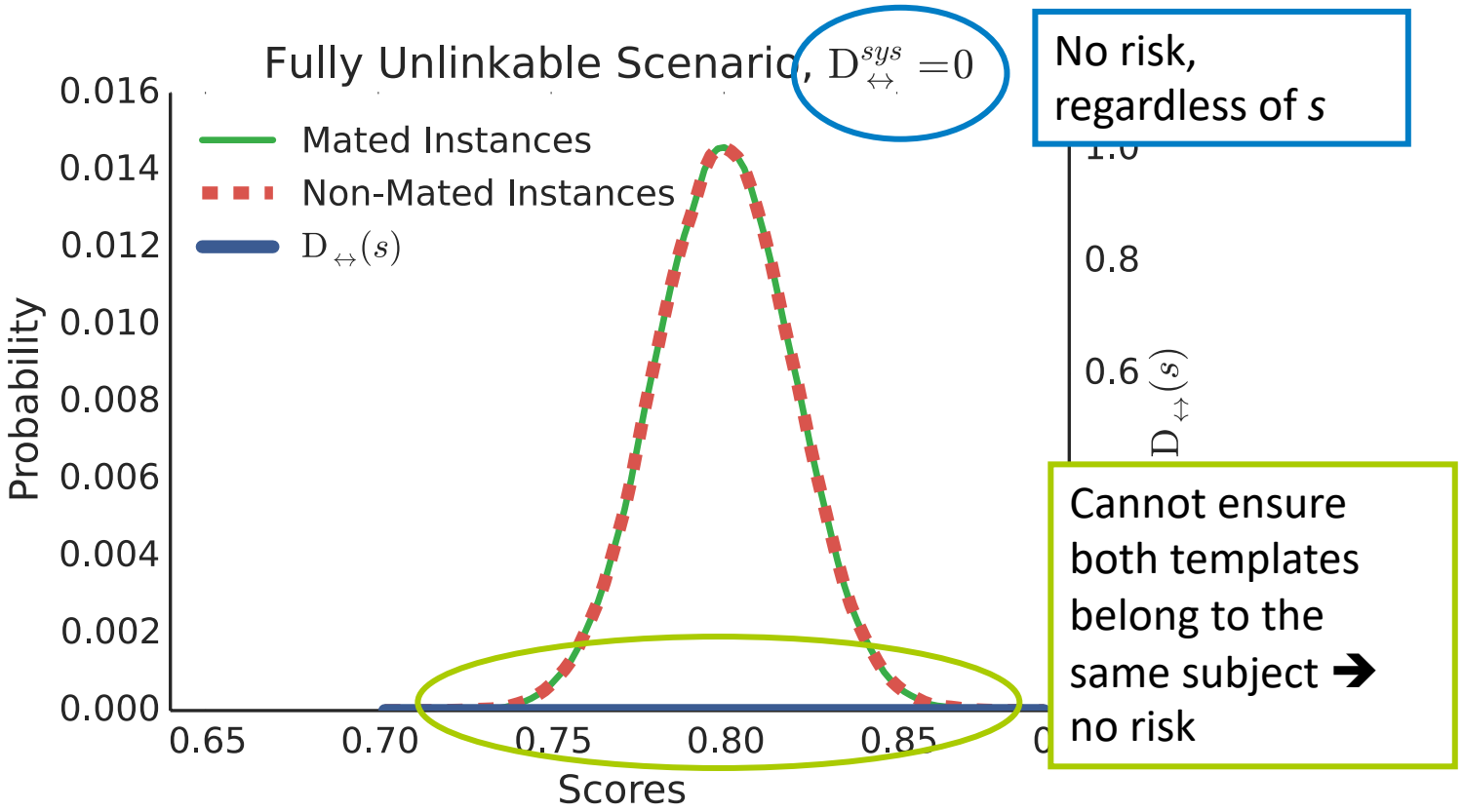
- Two measures:
  - Local measure $D_{\leftrightarrow}(s)$ ➔ for which scores is the system vulnerable?
  - Global measure $\mathrm{D}_{\leftrightarrow}^{sys}$ ➔ how can we compare two systems globally?

- Both bounded in [0,1], and defined for all dissimilarity scores.

- General measures, valid for all BTP schemes

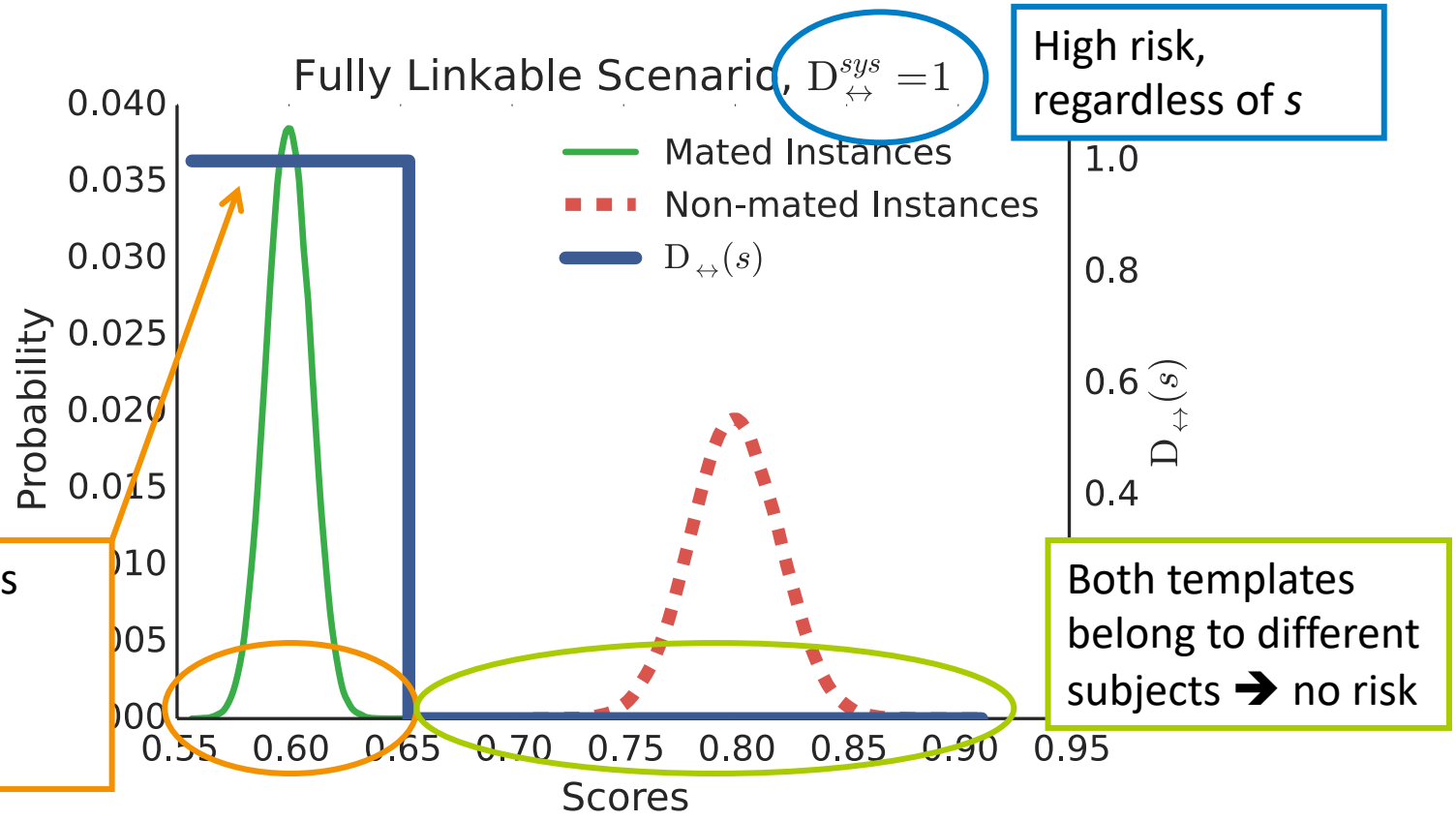
[Gomez-Barrero *et al.*, *IEEE T-IFS*, 2018]

Full Unlinkability



[Gomez-Barrero *et al.*, *IEEE TIFS*, 2018]

Full Linkability



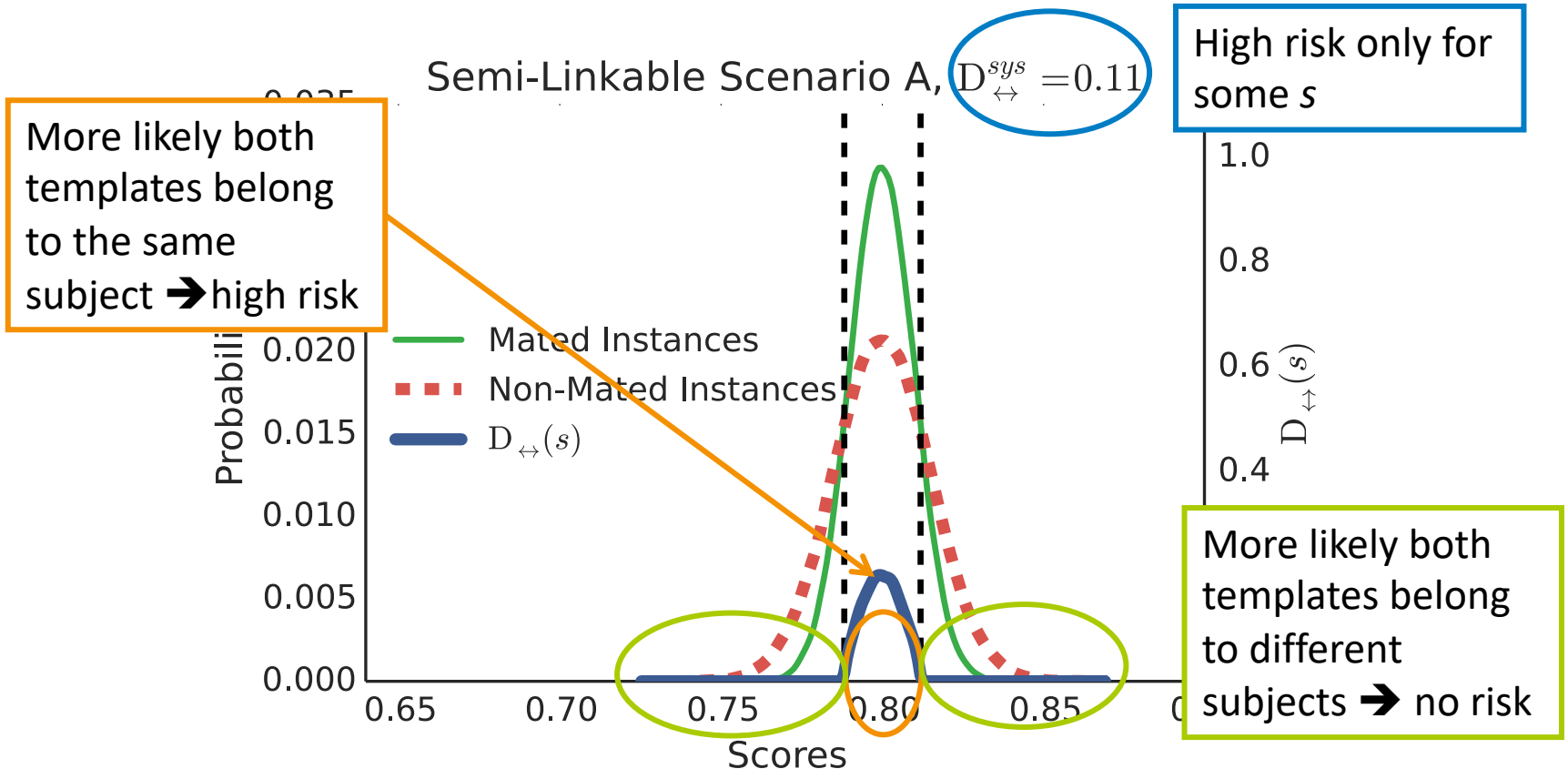
High risk, regardless of *s*

Both templates belong to the same subject ➔high risk

Both templates belong to different subjects ➔ no risk

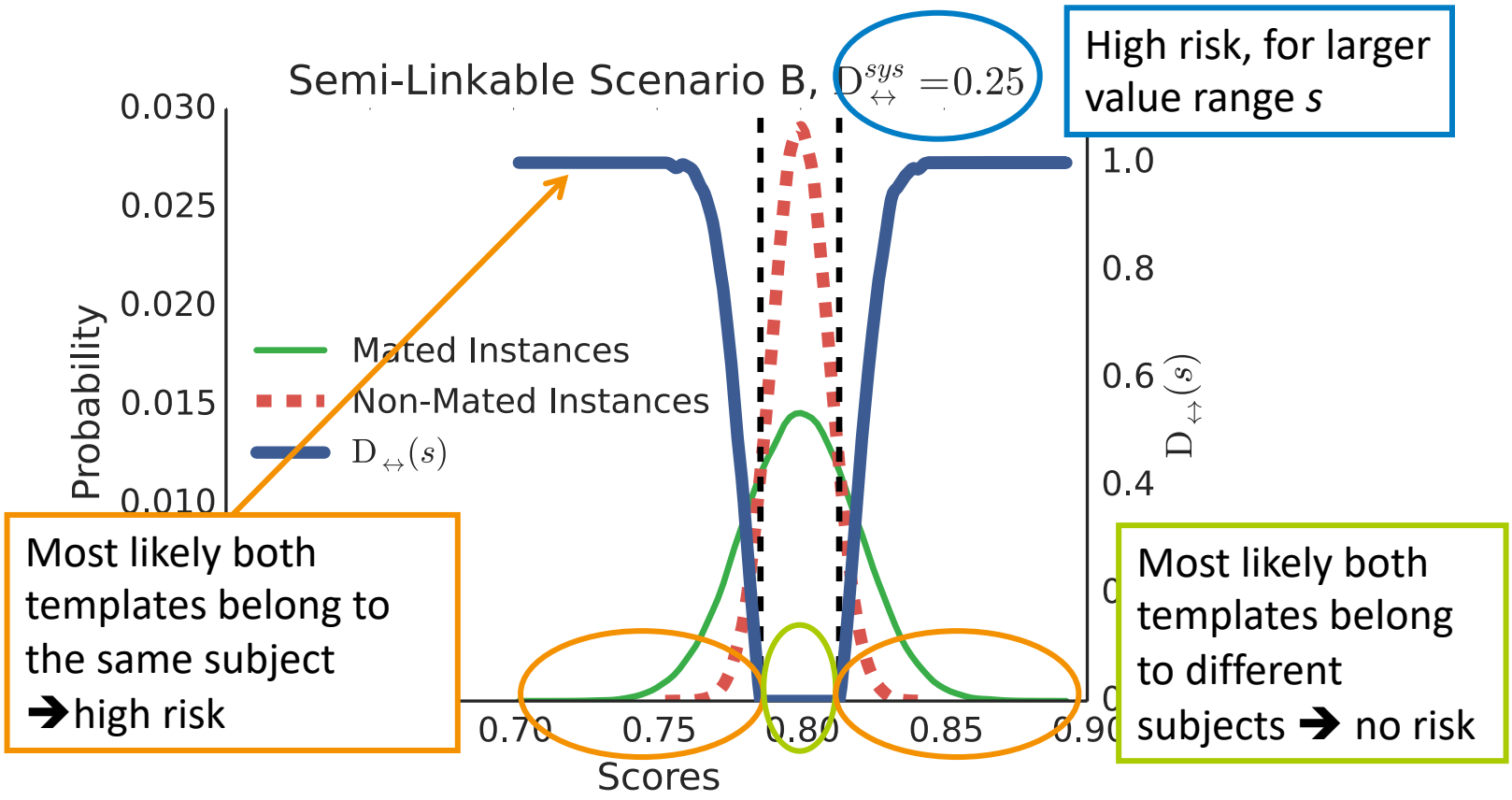
[Gomez-Barrero *et al.*, *IEEE T-IFS*, 2018]

Semi-Linkable Scenario A



[Gomez-Barrero *et al.*, *IEEE T-IFS*, 2018]

Semi-Linkable Scenario B



[Gomez-Barrero *et al.*, *IEEE T-IFS*, 2018]

Local measure: Background

- We are interested in evaluating: $D_{\leftrightarrow}(s) = p(H_m|s) - p(H_{nm}|s)$
- But we don't know $p(H_m|s)$, $p(H_{nm}|s)$
- He can use LRs: $LR(s) = \frac{p(s|H_m)}{p(s|H_{nm})} = \frac{p(H_m|s)}{p(H_{nm}|s)} \cdot \frac{p(H_{nm})}{p(H_m)}$
- Doing some tricks, we get:

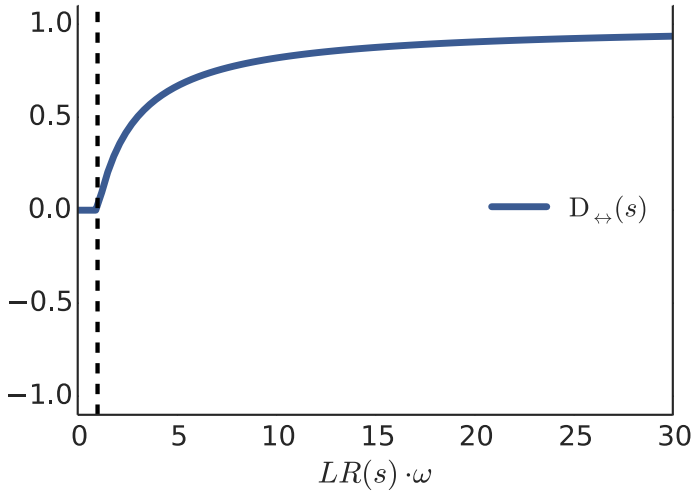
$$p(H_m|s) = \frac{LR(s) \cdot \omega}{1 + LR(s) \cdot \omega} \qquad \omega = p(H_m)/p(H_{nm})$$

[Gomez-Barrero *et al.*, *IEEE T-IFS*, 2018]

Local measure: final definition

$$\mathrm{D}_{\leftrightarrow}(s) = \begin{cases} 0 & \text{if } LR(s)\cdot\omega \leq 1 \\ 2\frac{LR(s)\cdot\omega}{1+LR(s)\cdot\omega} - 1 & \text{if } LR(s)\cdot\omega > 1 \end{cases}$$

- If we know $p(H_m)$, $p(H_{nm})$, use them to set ω
- Otherwise, assume $p(H_m) = p(H_{nm})$ and $\omega = 1$



[Gomez-Barrero *et al.*, *IEEE T-IFS*, 2018]

Global measure

- Global measure

$$\int_{s_{min}}^{s_{max}} p\left(H_m \cap s\right) - p\left(H_{nm} \cap s\right) \mathrm{d}s = \int_{s_{min}}^{s_{max}} p\left(s\right) \cdot \left(p\left(H_m|s\right) - p\left(H_{nm}|s\right)\right) \mathrm{d}s$$

$$= p\left(H_m\right) \int_{s_{min}}^{s_{max}} p\left(s|H_m\right) \cdot \left(p\left(H_m|s\right) - p\left(H_{nm}|s\right)\right) \mathrm{d}s +$$

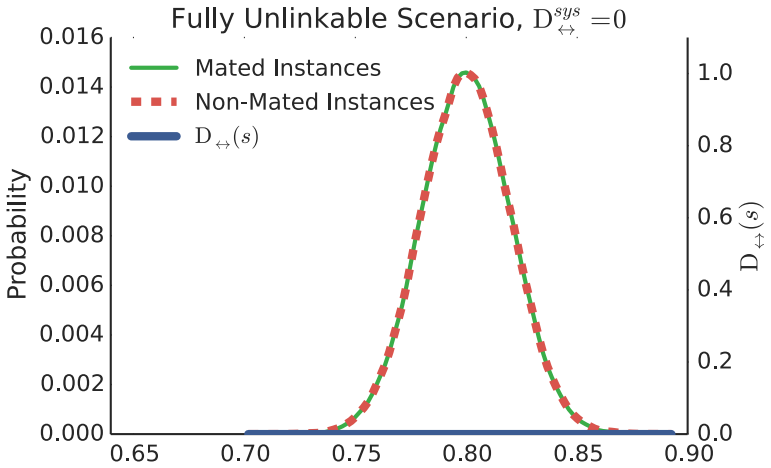
$$p\left(H_{nm}\right) \int_{s_{min}}^{s_{max}} p\left(s|H_{nm}\right) \cdot \left(p\left(H_m|s\right) - p\left(H_{nm}|s\right)\right) \mathrm{d}s$$

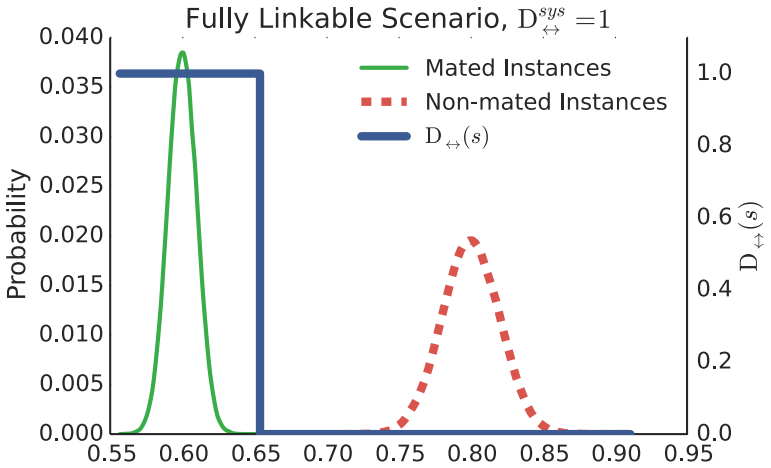
$$p\left(H_m|s\right) > p\left(H_{nm}|s\right)$$

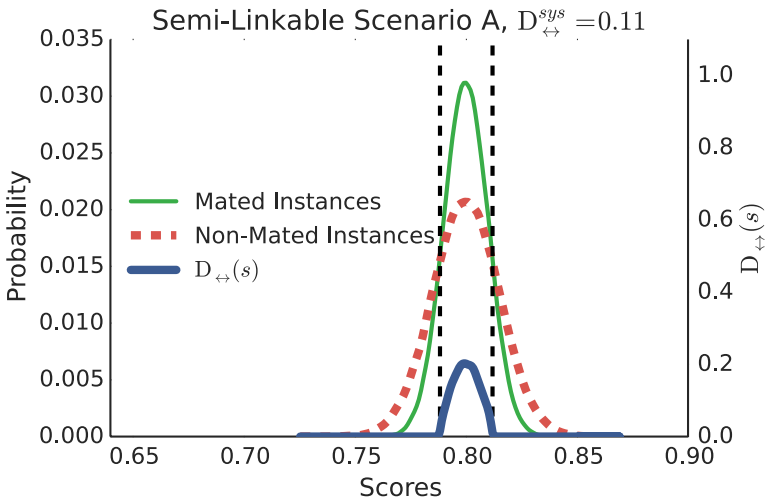
$$\mathrm{D}_{\leftrightarrow}^{sys} = \int_{s_{min}}^{s_{max}} p\left(s|H_m\right) \cdot \mathrm{D}_{\leftrightarrow}\left(s\right) \mathrm{d}s$$

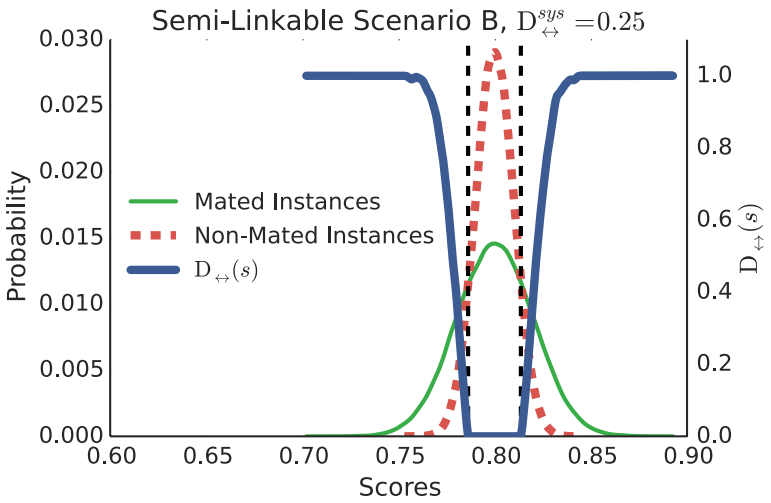
[Gomez-Barrero *et al.*, *IEEE T-IFS*, 2018]

Linkability Scenarios: Summary









Robustness to attacks

- Attackers will always try to exploit weaknesses
- We need to be ahead of them ⇒ security through transparency!
- First, investigate the vulnerabilities
  - C. Rathgeb, A. Uhl, "Statistical attack against fuzzy commitment scheme", *IET biometrics*, *1*(2), 94-104, 2012
  - T. Ignatenko, F. M. Willems, "Information leakage in fuzzy commitment schemes", *IEEE Trans. on Information Forensics and Security*, vol. *5*, no. 2, pp. 337-348, 2010
  - W. J. Scheirer, T. E. Boult, "Cracking fuzzy vaults and biometric encryption", *Proc. Biometrics Symposium*, 2007
- Then, devise countermeasures
  - C. Rathgeb, B. Tams, J. Wagner, C. Busch, "Unlinkable Improved Multi-Biometric Iris Fuzzy Vault", *EURASIP Journal on Information Security*, 2016.

Prof. Dr. Marta Gomez-Barrero
(marta.gomez-barrero@hs-ansbach.de)

HOCHSCHULE
ANSBACH