Biometric Template Protection and Evaluation

Dr. Marta Gomez-Barrero

Hochschule Darmstadt, CRISP, da/sec Security Group

WIFS, Hong Kong, December 2018

Outline

- About da/sec
- Introduction
- Vulnerabilities of Biometric Systems
- Biometrics & Privacy
- Security and Privacy Evaluation
- Cancelable Biometrics Based on Bloom Filters
- BTP Based on Homomorphic Encryption
- Summary

For non-biometric experts

Motivation – why are we doing this?

The main concepts

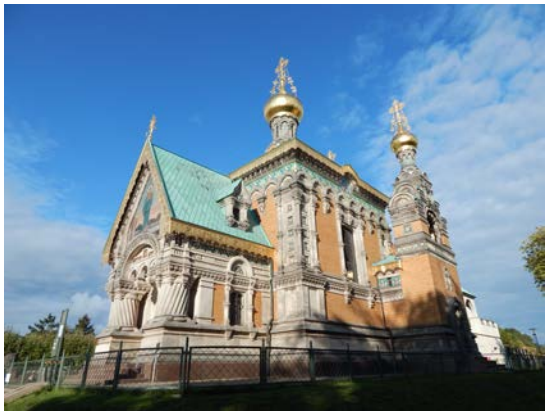
About da/sec

Where is da/sec?

- We are located in Darmstadt, a German city close to Frankfurt

About da/sec

Where is da/sec?



About da/sec?

- We are affiliated with the Hochschule Darmstadt
- And do research on biometrics and internet security – more info on https://dasec.h-da.de/

About da/sec

- Research topics and projects with partners in the US, Germany, Switzerland, France, Norway, etc.:
  - Fingerprint Presentation Attack Detection (PAD) – US IARPA and German BSI
  - Efficient mobile (face, voice, iris) biometric recognition with PAD and Biometric Template Protection (BTP) – German DFG and French ANR
  - Efficient biometric identification - Hessen Agentur (DE)
  - Attack detection (e.g. presentation and morphing attacks) for facial biometric systems – German BSI
  - PAD and BTP for voice based biometrics – Hessen Agentur (DE)
  - ... And more

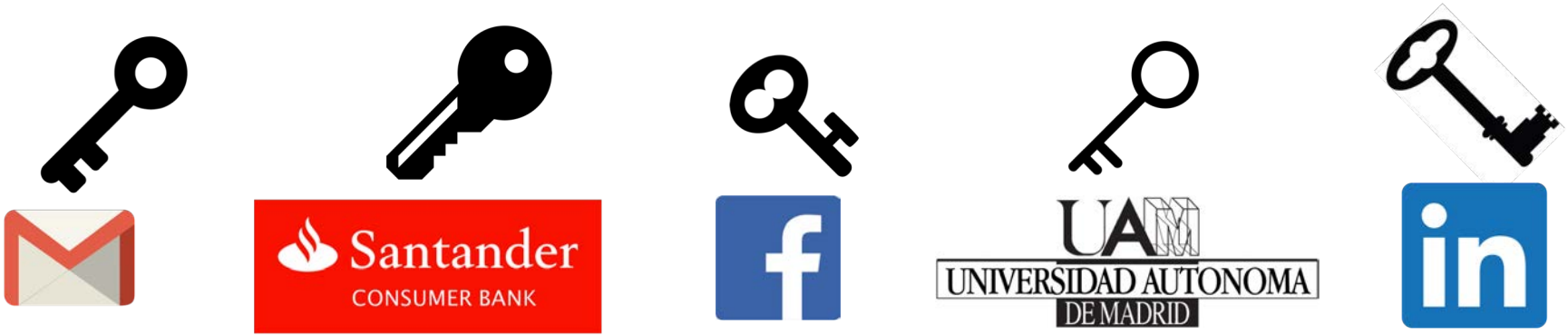
More info on https://dasec.h-da.de/projects/current-projects/

And you can contact any of us for an internship!

Introduction

Why biometric recognition?

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



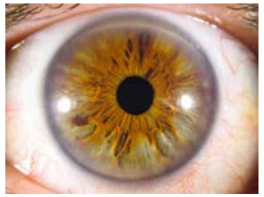
- Losing or forgetting our password / token is easy

Why not use our body features or behavioural patterns?

Biometric characteristics

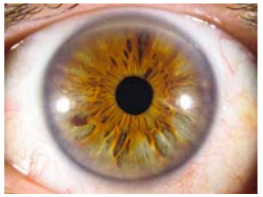
- Classification:
  - Physiological
  - Behavioural

- 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



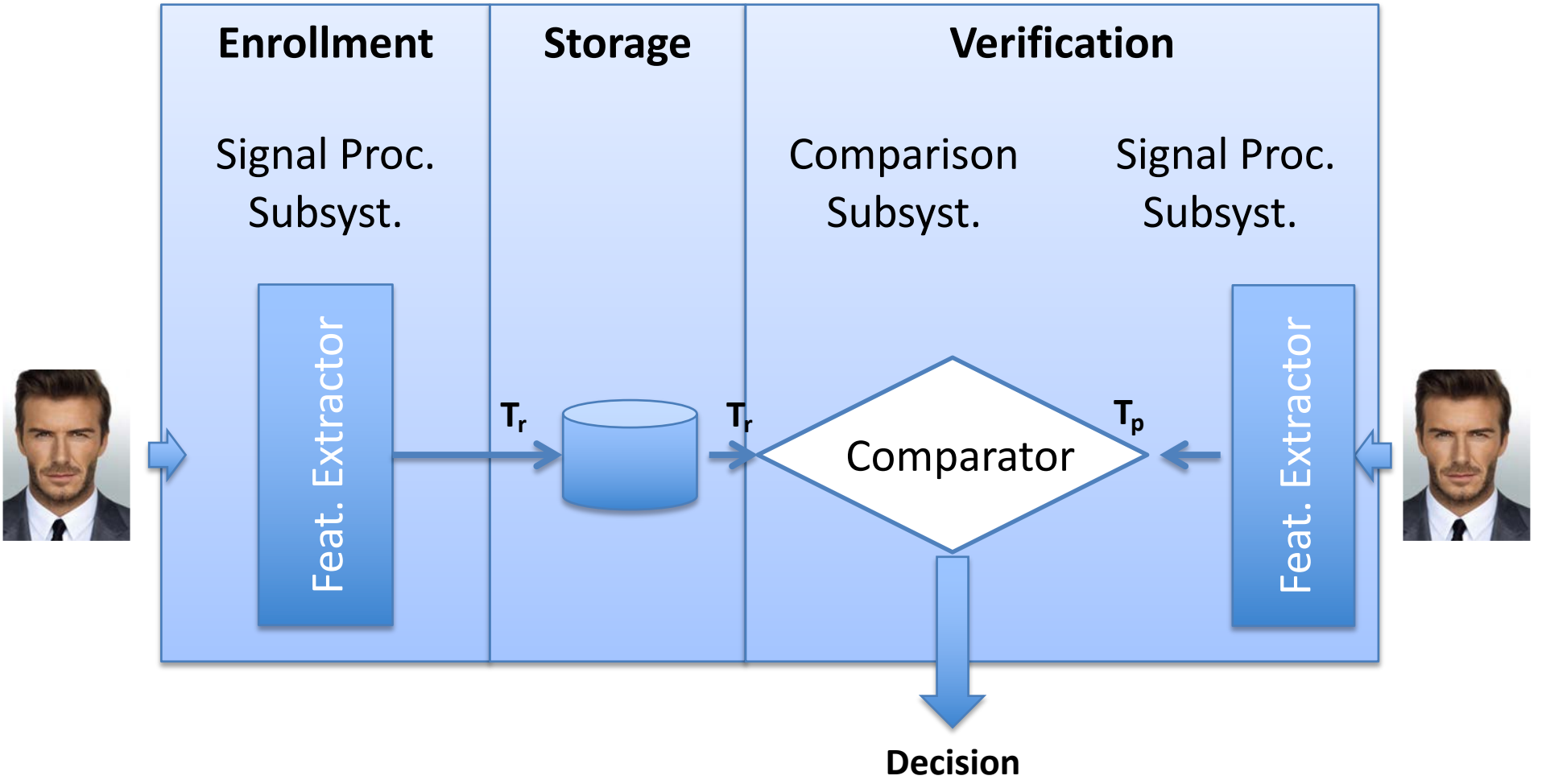
Advantages and disadvantages of biometrics

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

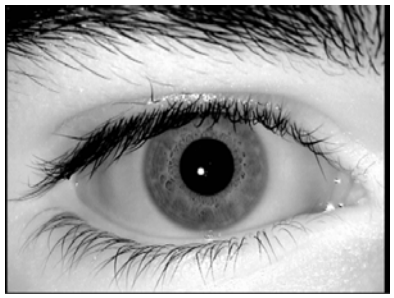


- Presentation Attacks (PA)
- Renewability
- Biometrics are no secrets
- Sensitive information

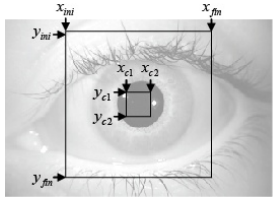
How does it work?

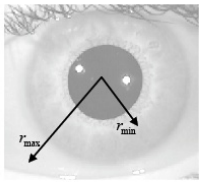


Example I: iris recognition

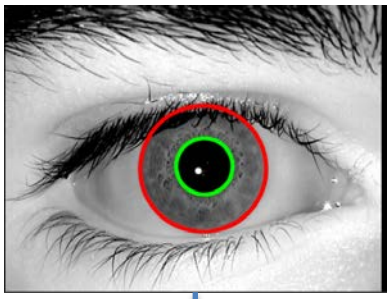


Sample

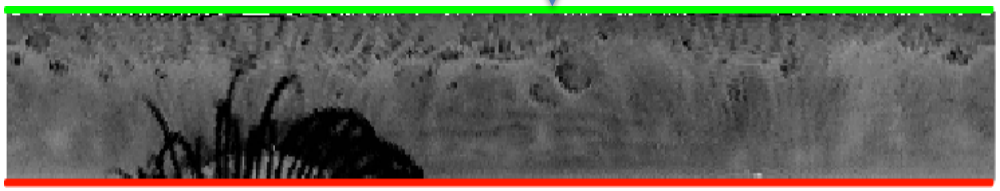


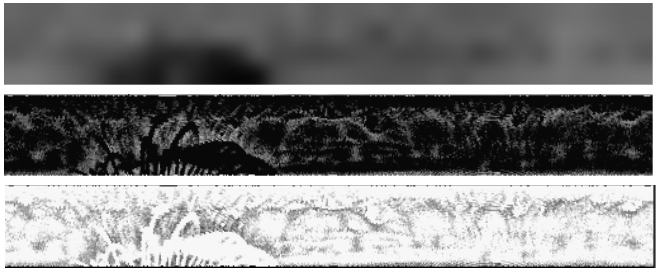


Segmentation



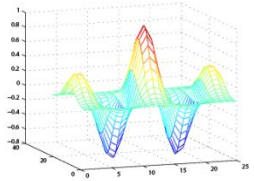
Normalization

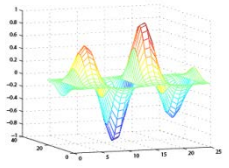




Feature Extraction

Template: T





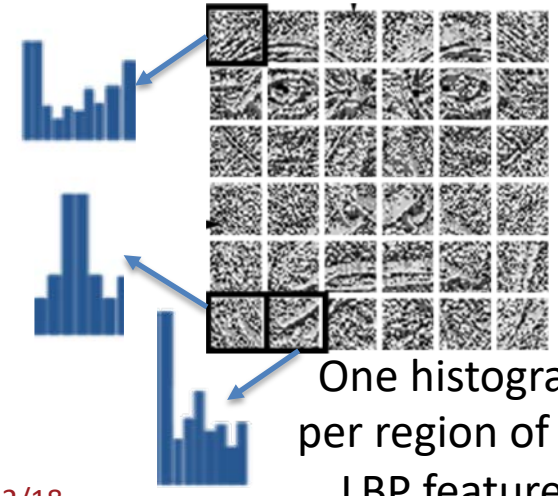
Example II: face recognition

Sample

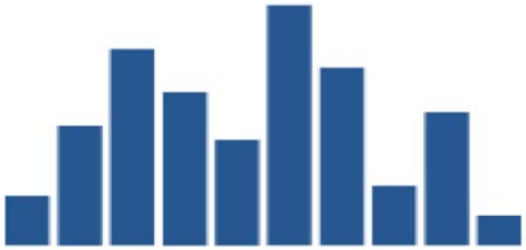
Segmentation

Normalization

Feature Extraction



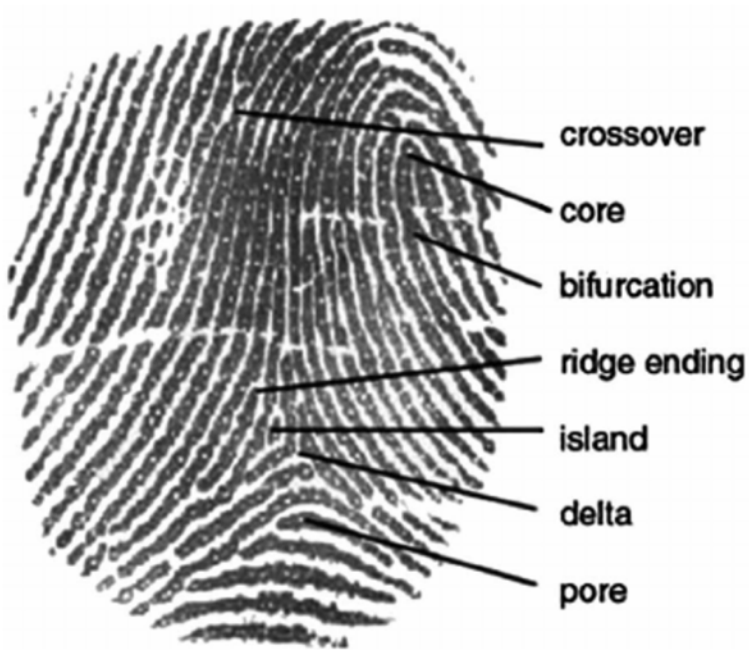
One histogram per region of the LBP features



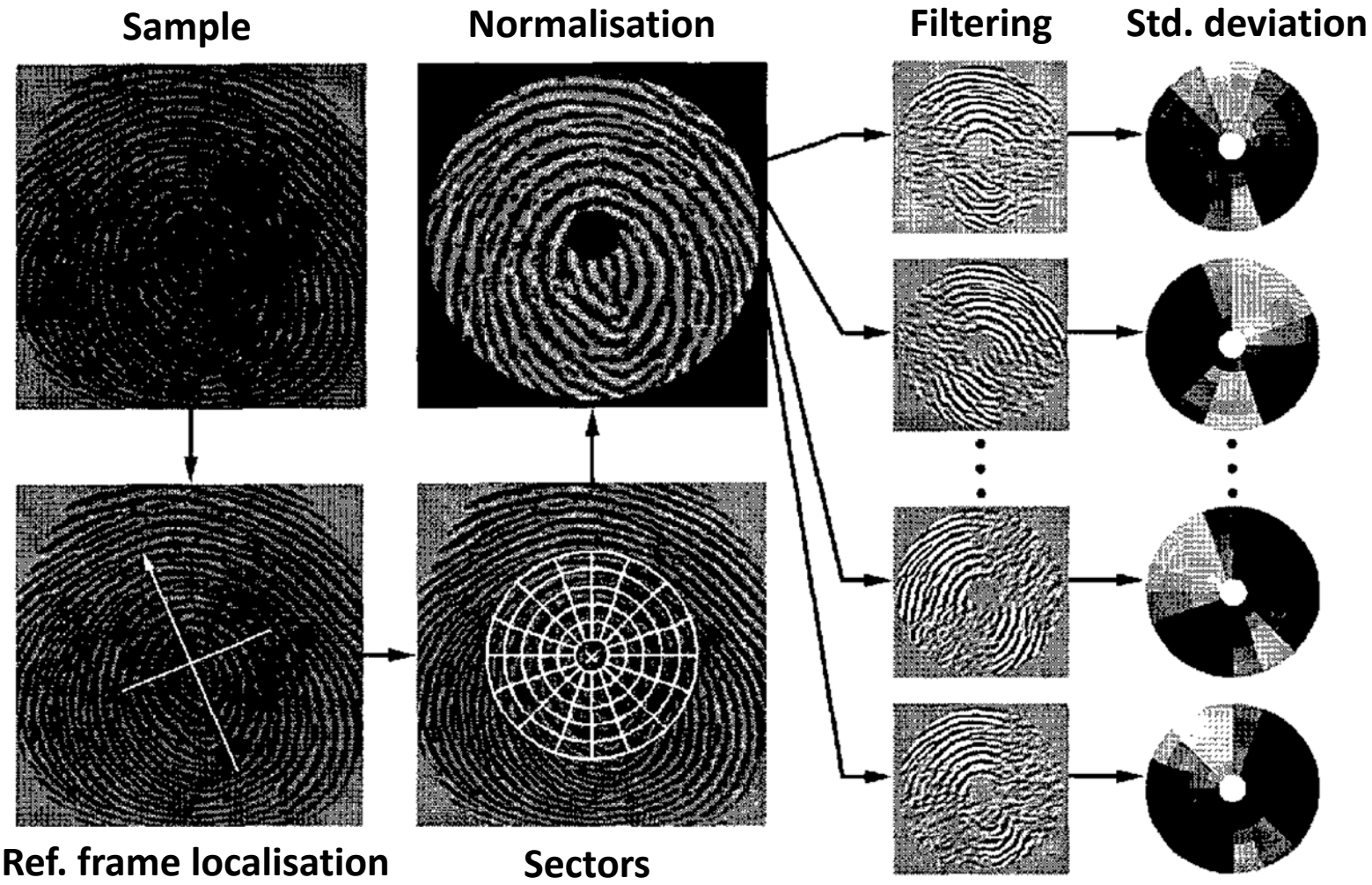
Final Template

Example III: fingerprint recognition (minutiae)

- Most accurate method based on minutiae and Hausdorff distance



Example III: fingerprint recognition (fingercodes)



[A. Jain *et al.*, CVPR'99]

Error rates

[ISO/IEC 2382-37 Harmonized Biometrics Vocabulary (HBV)]

- Two kinds of comparisons:

Mated Comparison

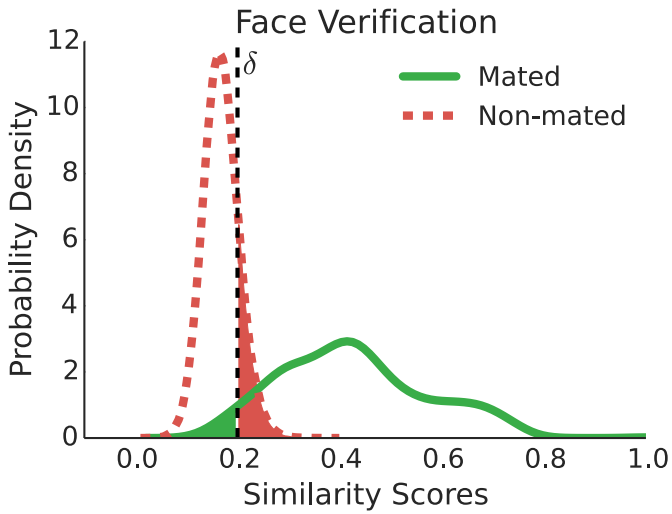
Non-Mated Comparison

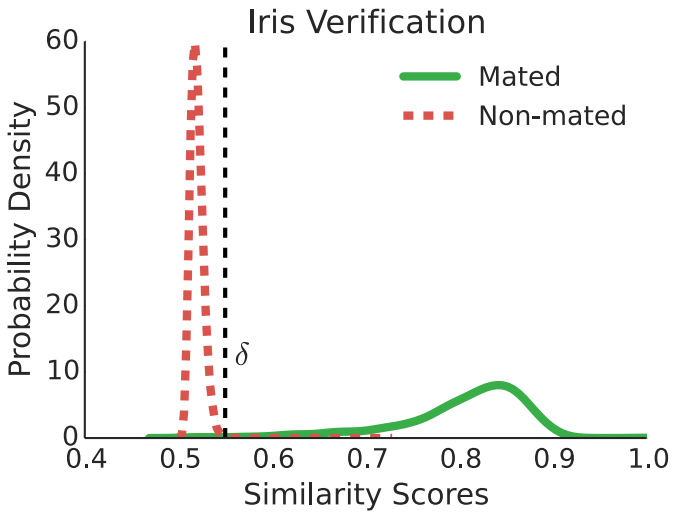
- Two kinds of error rates:
  - **False Match Rate (FMR)** – proportion of falsely accepted non-mated comparison trials
  - **False Non-Match Rate (FNMR)** – proportion of falsely rejected mated comparison trials

Evaluating the accuracy

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

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



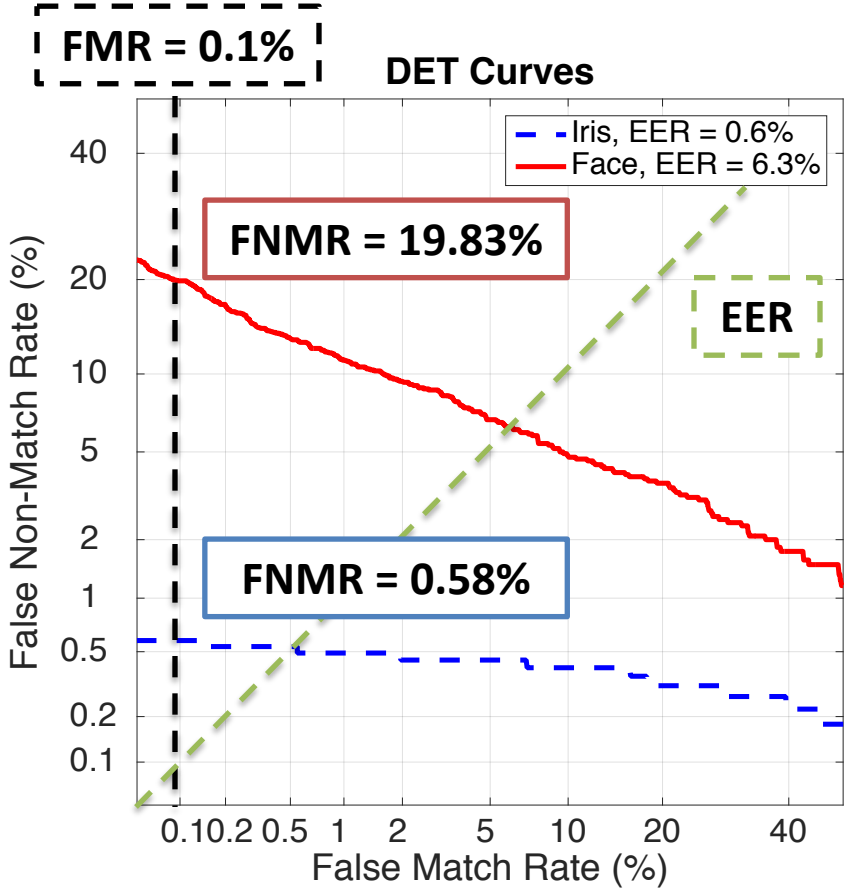


- δ determines the **FMR**
- … and the **FNMR**

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%



Multi-Biometric systems

[ISO/IEC TR 24722 on Multimodal and other multibiometric fusion]

- Advantages
  - Higher accuracy
  - Increased robustness to individual sensor or subsystem failures
  - Decreased number of cases where the system is not able to make a decision
  - Different levels of security
  - …

- Fusion levels:
  - Feature level
  - Score level
  - Decision level

Can be harder to achieve, but it's preferred: reduced storage and higher security

Vulnerabilities of Biometric Systems

Biometric Systems' Attack Points

[ISO/IEC 30137 on Biometric Presentation Attack Detection]

Override or Modify Biometric Database

Data Storage

Modify Biometric Reference

Data Capture

Signal Processing

Comparison

Decision

Presentation Attack at the Sensor

Modify Biometric Sample

Override Signal Processor

Modify Probe

Override Comparator

Modify Score

Override Decision

ISO/IEC 30137 Scope

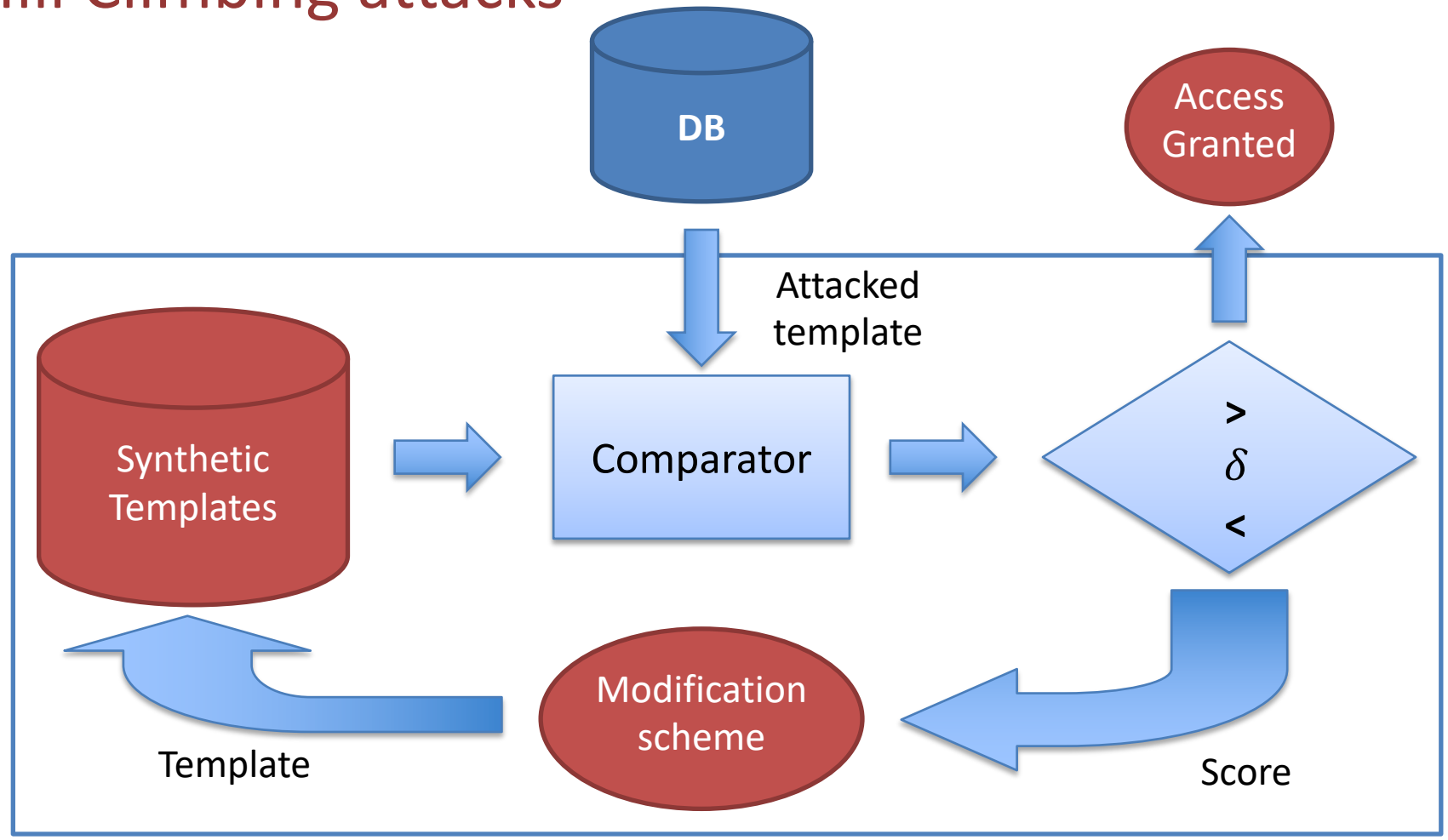
SOFTWARE ATTACKS

Presentation Attacks

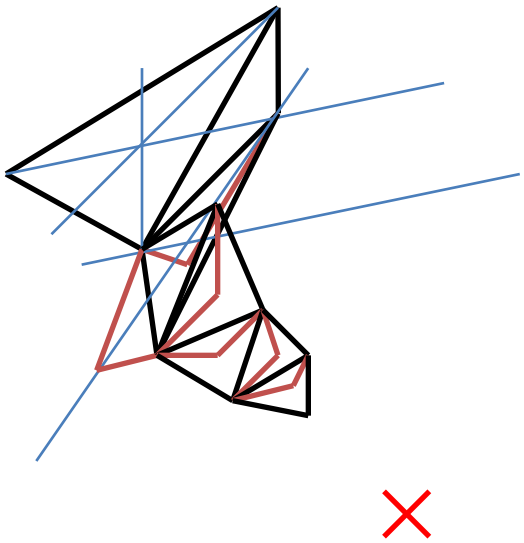
- Definition: presentation to the biometric capture subsystem with the goal of interfering with the operation of the biometric system
  - **Impostor**: the attacker attempts to being matched to someone else's biometric reference
  - **Identity concealer**: the attacker attempts to avoid being matched to their own biometric reference (i.e., to escape from a black-list entry)

Vulnerabilities

Hill Climbing attacks



HC based on the Uphill Simplex algorithm



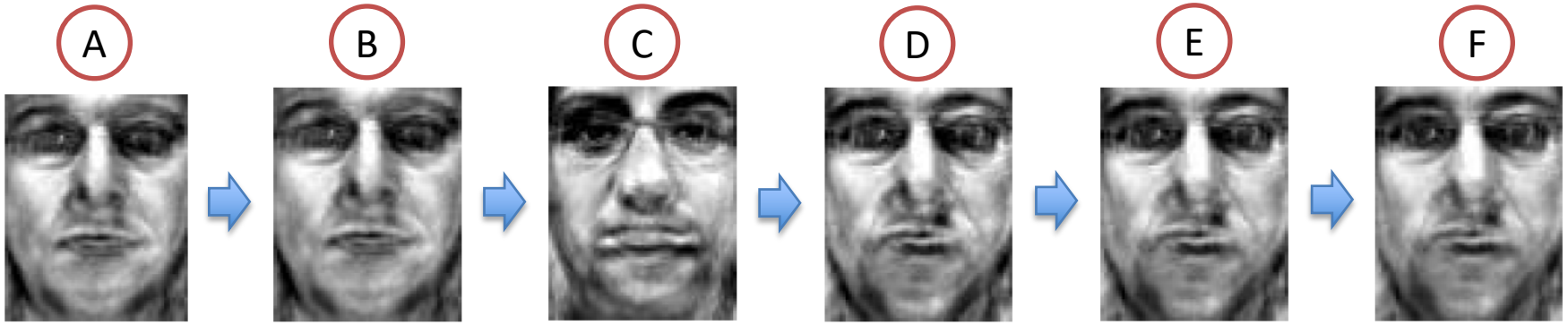
- New point:
  - Compute centroid: $\bar{\mathbf{y}} = \frac{1}{K+1}\sum_i \mathbf{y}_i$
  - Try reflection: $\mathbf{a} = (1+\alpha)\bar{\mathbf{y}} - \alpha\mathbf{y}_l$
  - Try expansion $\mathbf{b} = \gamma\mathbf{a} + (1-\gamma)\bar{\mathbf{y}}$

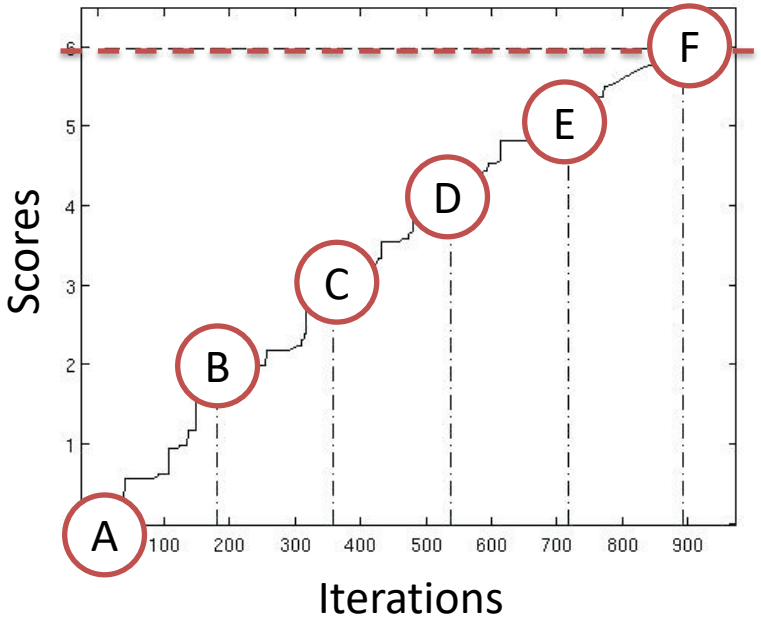
    or contraction: $\mathbf{b} = \beta\mathbf{y}_l + (1-\beta)\bar{\mathbf{y}}$

- Stopping criteria:
  - One of the points of the simplex is close enough => success
  - Maximum number of iterations allowed reached => failure

Vulnerabilities

Example 1: Face





Verification Threshold

The attack was successful, and we only needed access to the scores

Target: Enrolled Sample

Example 2: Face and signature Success Rates (SR)

- We can evaluate how dangerous the attack is in terms of the success rate:

$$\mathrm{SR} = \frac{A_B}{A_T}$$

- At different operation points in terms of FMR

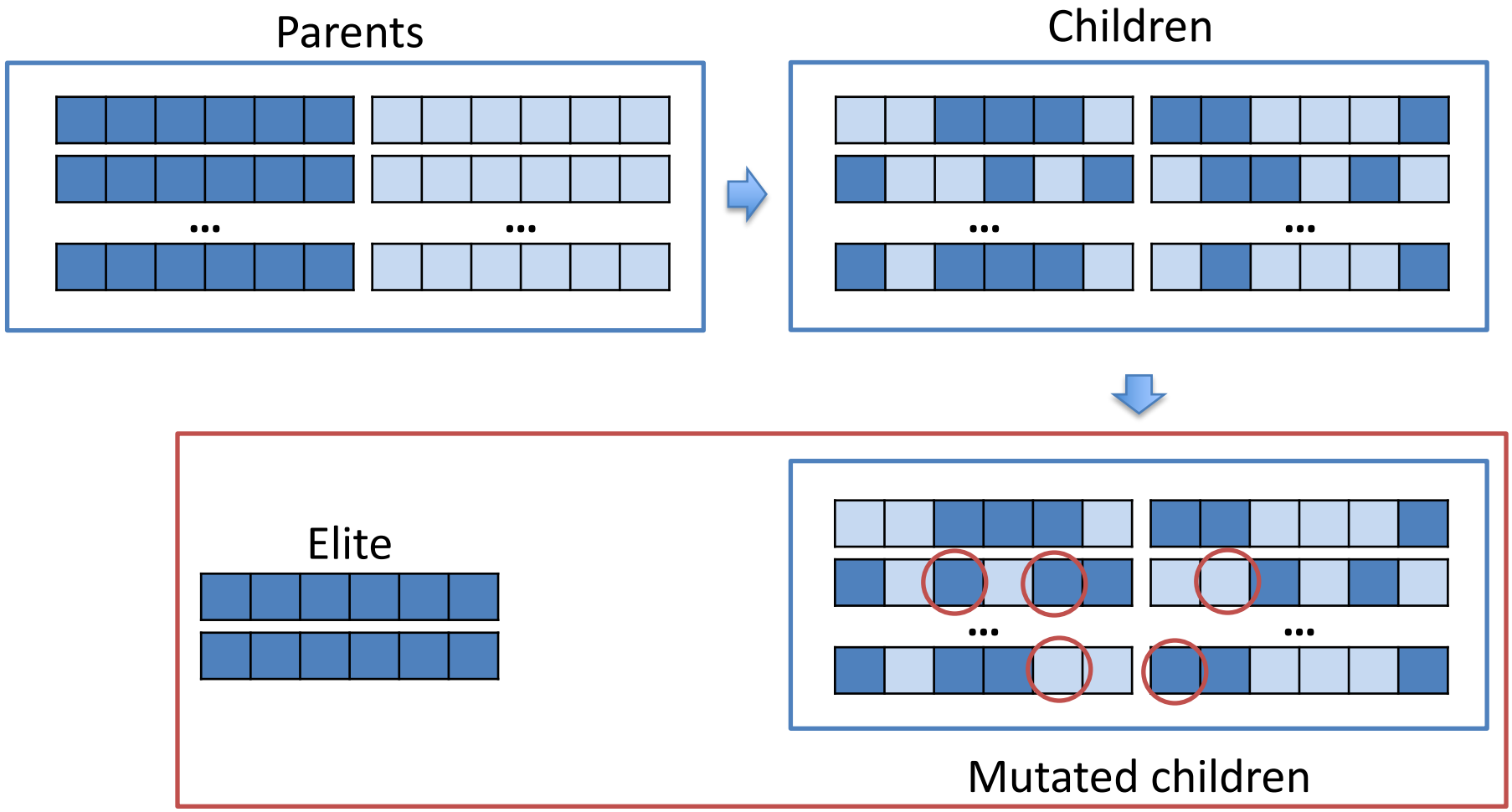
FMR (%)	Face System	Signature System
0.05%	100%	92.69%
0.01%	100%	87.84%

Hill Climbing attacks represent a real challenge to the security offered by biometric systems => Quantized Scores

HC based on genetic algorithms (I)

- We start with a random population of binary individuals
- At each iteration, we generate a new population according to four rules:
  - **Elite**: two individuals
  - **Selection**: stochastic universal sampling
  - **Crossover**: scattered crossover
  - **Mutation**: random changes
- Our fitness function is the similarity score
- Stopping criteria:
  - One of the individuals exceeds the verification threshold => success
  - Score increase in the last generations is very small => failure
  - Maximum number of iterations allowed reached => failure

HC based on genetic algorithms (II)



Vulnerabilities

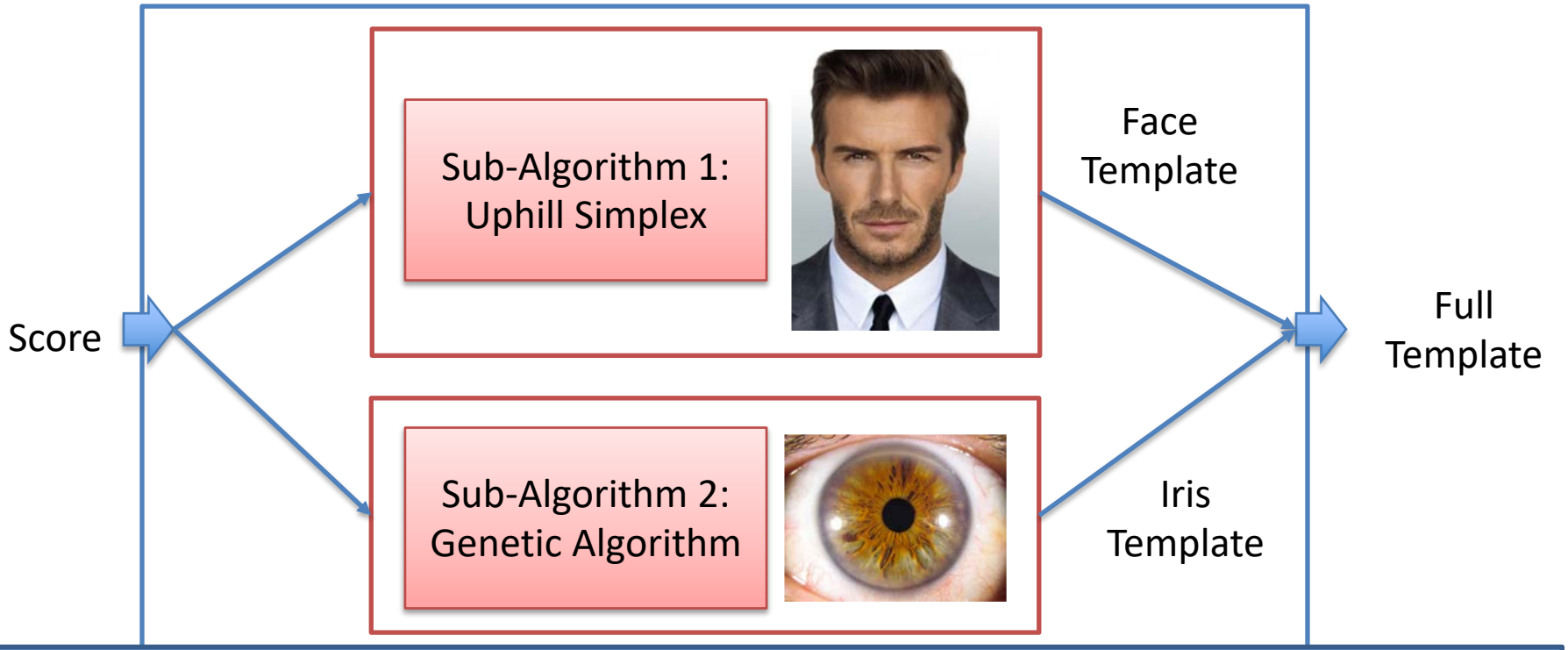
Example: Iris

FMR (%)	Iris System
0.05%	80.89%
0.01%	62.36%

Hill Climbing attacks represent a real challenge to the security offered by biometric systems => Quantized Scores

HC Attacks on multi-biometric systems

- Contrary to the belief that it is more difficult to attack a multi-biometric systems, we can combine these algorithms and succeed in our attack



The multi-biometric system is as vulnerable as the most vulnerable characteristic

Vulnerabilities

Security and privacy protection

- Projects

TABULA RASA

BEAT

BATL
Biometric Authentication with Timeless Learner

ODIN

TURBINE

- Competitions

LivDet
Liveness Detection
Competition Series

- Standards

Common Criteria

ISO

Biometrics & Privacy

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]
  [EU 2016/680 Data Protection Directive]

- And we cannot prevent databases leakage

Biometric symmetry: What is biometric information?

- The term **biometric information** is defined as "the decrease in uncertainty about the identity of a person due to a set of biometric measurements" [A. Adler *et al.*, Proc. CCECE 2006]
- It ultimately depends on the selected feature representation of the biometric data and the comparison algorithm used [Y. Sutcu *et al.*, Proc. HST 2013]
- We can model such decrease using **mutual information** [K. Takahashi and T. Murakami, Image Vision and Computing 2014]:

$$I(X;Y) = H(Y) - H(Y|X)$$

A particular instance

The population

How can we measure biometric information?

- $I(X;Y)$ can be approximated by the Kullback-Leibler divergence of the mated and non-mated score probability distributions

$$I(X;Y) \approx D_{KL}(p_m \| p_{nm}) = \sum_{i=1}^{N} p_m(s^i) \log_2 \left(\frac{p_m(s^i)}{p_{nm}(s^i)} \right)$$

A particular instance: mated

The population: non-mated

- Problem: we don't know those distributions
- Solution: use NN-estimators [Y. Sutcu *et al.*, Proc. ICPR 2010]:

$$I(X;Y) \approx \hat{D}_{KL}(p_m \| p_{nm}) = \frac{1}{N_m} \sum_{1=1}^{N_m} \log \frac{\nu_{nm}(i)}{\rho_m(i)} + \log \frac{N_{nm}}{N_m - 1}$$

where $\rho_m(i) = \min_{j \neq i} \|s_m^i - s_m^j\|$, $\nu_{nm}(i) = \min_j \|s_m^i - s_{nm}^j\|$

Measuring biometric information: multi-biometrics

- Goal: maximise joint entropy

$$\mathrm{H}(X_l, X_r) = H(X_l) + H(X_r) - I(X_l; X_r)$$

- Equivalent to minimising $I(X_l; X_r)$
- As before, we can approximate it as:

$$I(X_l; X_r) \approx \hat{D}_{KL}(p_m \| p_{lr})$$

The population: left *vs* right

- But we need a measure independent of the initial entropy, quantifying only the decrease due to left-right comparisons:

$$\hat{D}_{KL}^{\text{fused}} = 1 - \frac{\hat{D}_{KL}(p_m \| p_{lr})}{\hat{D}_{KL}(p_m \| p_{nm})} \approx 1 - \frac{I(X_l; X_r)}{I(Y; X)}$$

[Gomez-Barrero *et al.*, Proc. EUSIPCO 2017]

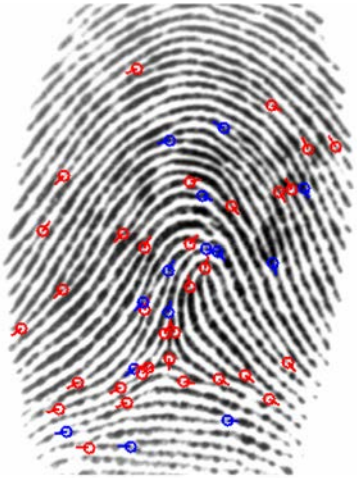
Biometric symmetry: periocular regions

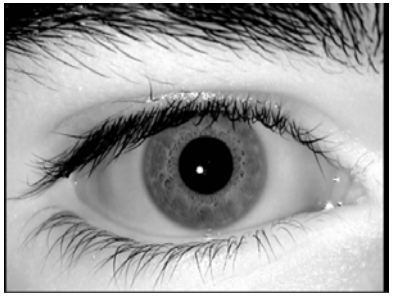
	LBP	BSIF	SIFT	SURF
$\hat{D}_{KL}(p_m \| p_{nm})$	5.45	5.07	7.08	6.07
$\hat{D}_{KL}(p_m \| p_{lr})$	2.07	1.50	3.94	2.51
$\hat{D}_{KL}^{\text{fused}}$	**0.62**	**0.70**	**0.44**	**0.59**

- In all cases, $\hat{D}_{KL}^{\text{fused}} > 0 \Rightarrow$ **there's some correlation**
- **Only 44%** of the information is retained by the SIFT based templates
- **And up to 70%** of the information is retained by the BSIF based templates
- Which means that we always lose at least 30% of the information

Inverse biometrics attacks

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

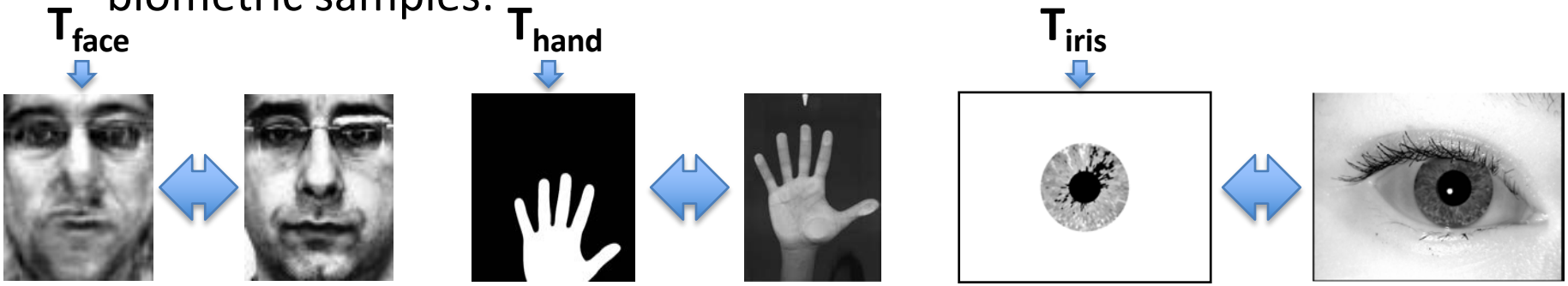


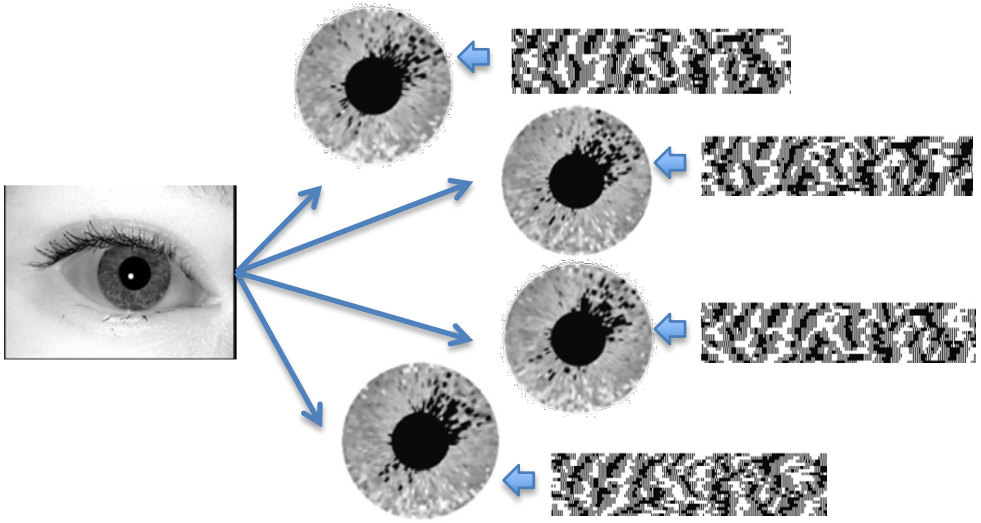


- However, biometric samples can be recovered from the stored unprotected templates

Inverse biometrics attacks: Hill-Climbing

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

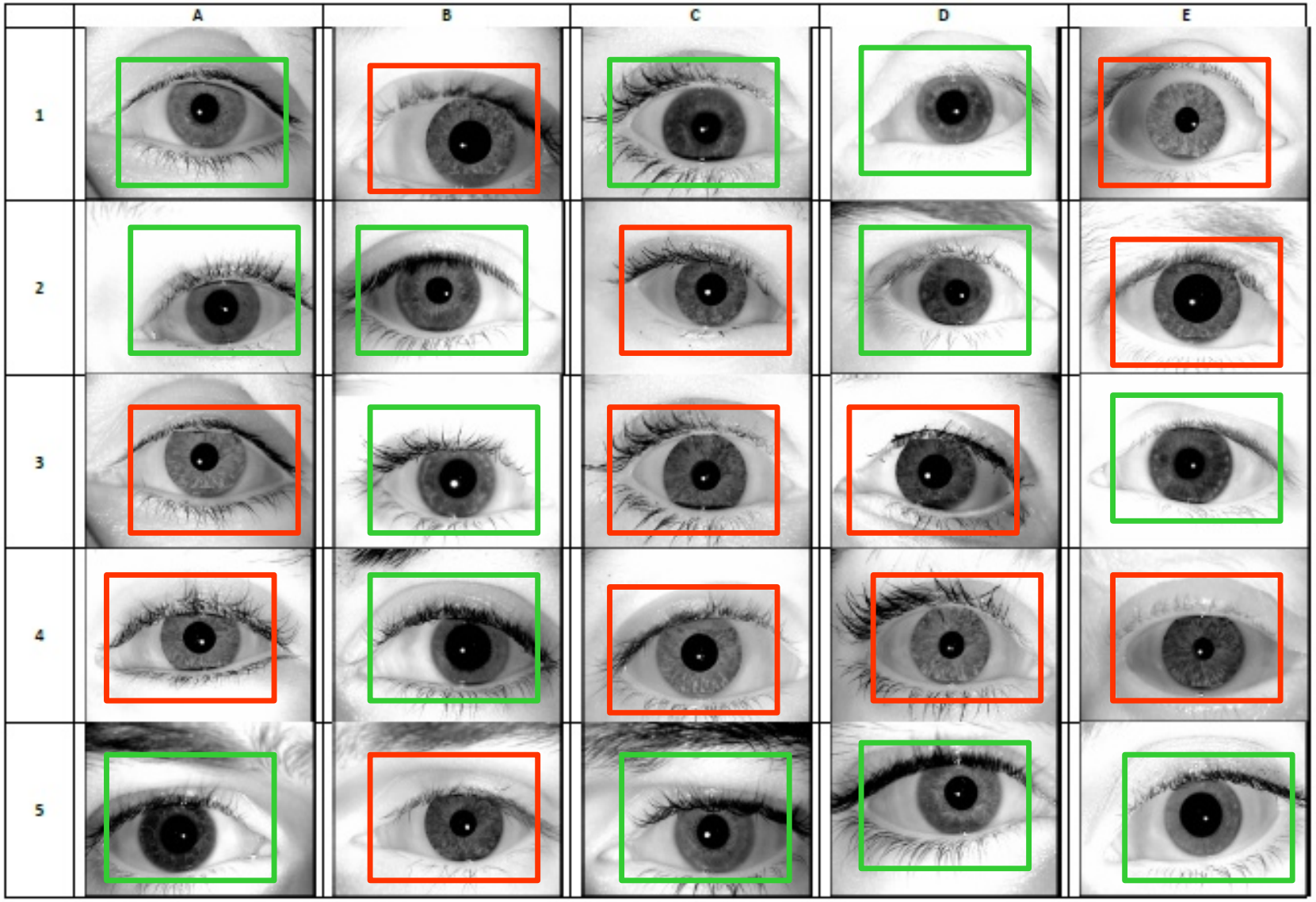




[M. Gomez-Barrero *et al.*, *Int. Conf. on Biometrics*, 2012]
[M. Gomez-Barrero *et al.*, *Information Sciences*, 2014]
[J. Galbally, *et al.*, *Computer Vision & Image Understanding*, 2013]

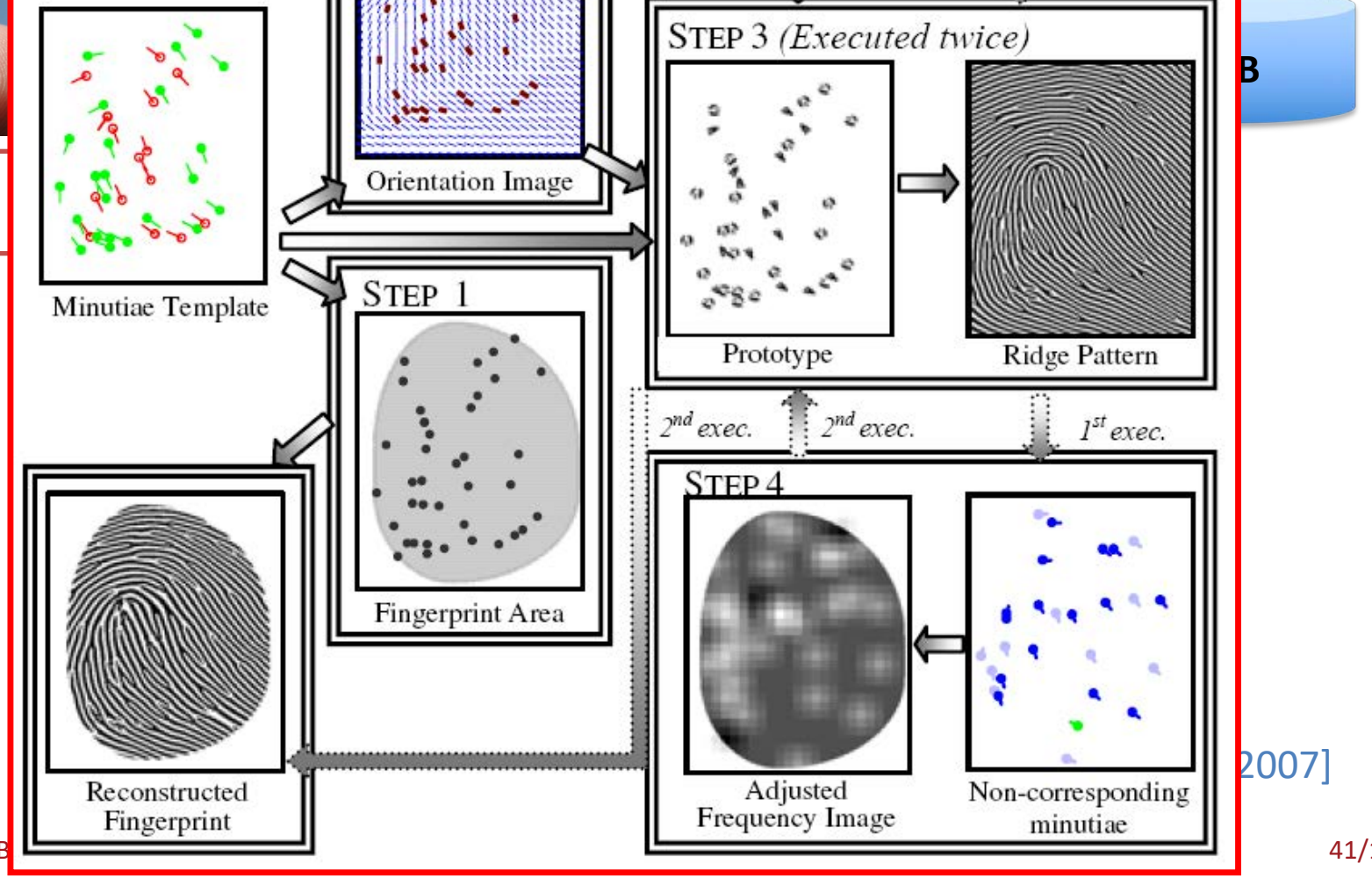
Biometrics & Privacy

Inv



Biometrics & Privacy

Inverse



[Galbally *et* 2009]

2007]

Inverse biometrics attacks: Success Rates

FMR (%)	Iris	Fingerprint (indirect)	Fingerprint (PA)
0.05%	85.1%	98%	78%
0.01%	83.6%	92%	68%

Over 85% of the attacks are successful => Real challenge!

Lower success chances, but more difficult to detect

Templates need to be protected, so that we cannot recover the biometric sample

In addition, Presentation Attacks need to be detected

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

Protecting the subject's privacy

[ISO/IEC IS 24745 on Biometric Information Protection]

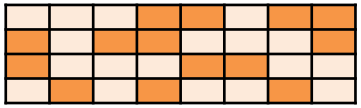
- Requirements of Biometric Template Protection:

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

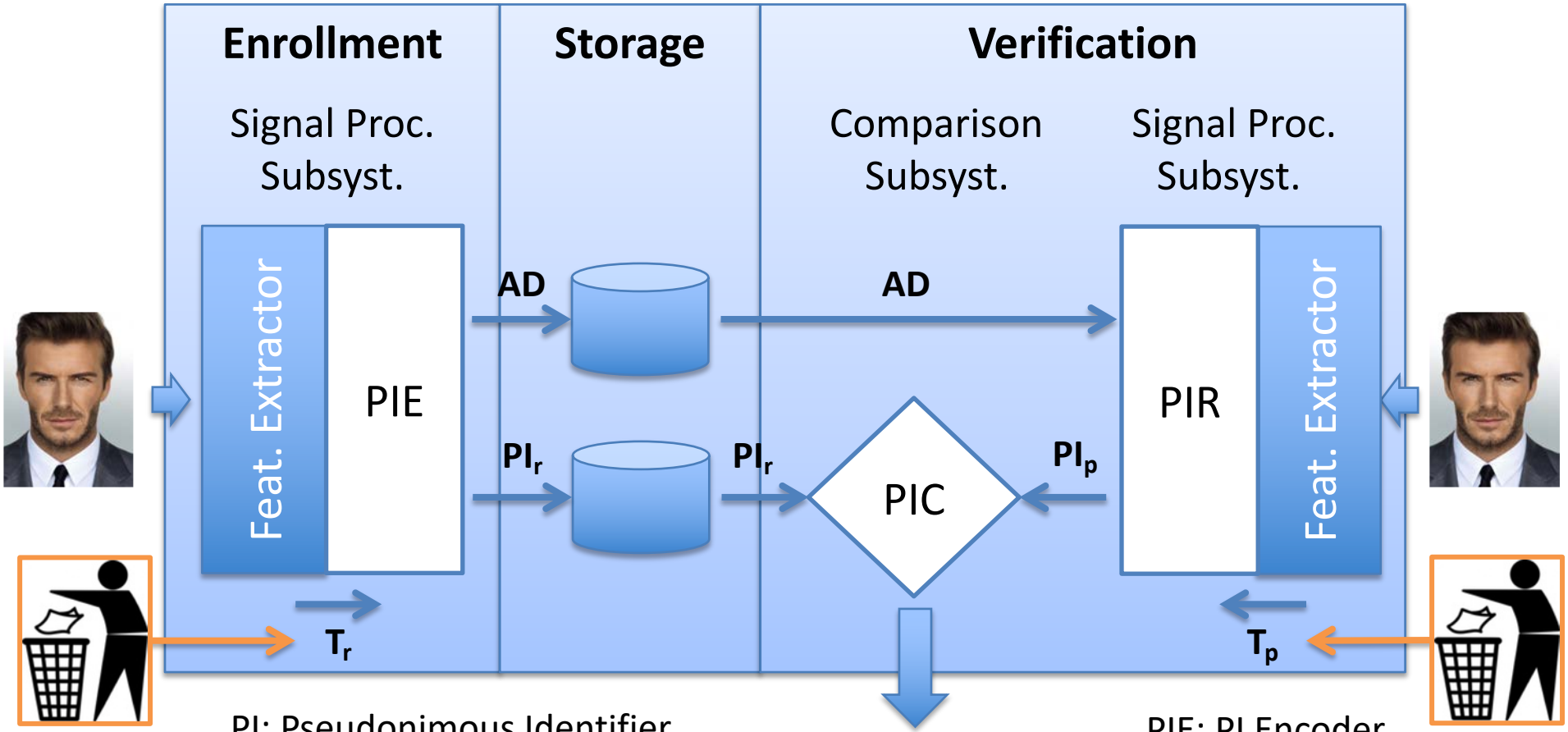
    h(01000101) is not similar, but very different from h(01010101)

Biometrics vs cryptographic protocols

- Conventional cryptography yields two main drawbacks:
  - **Shift of problem**: the encrypted template will be secure only as long as the decryption key is unknown to the attacker.
  - **Decryption at authentication**: the template needs to be decrypted during every authentication attempt, since comparison cannot be directly performed in the encrypted domain (except for homomorphic encryption)

- Potential, but inconvenient solution: store the encrypted template and decryption key in a secure environment within a smart card or a secure chip.

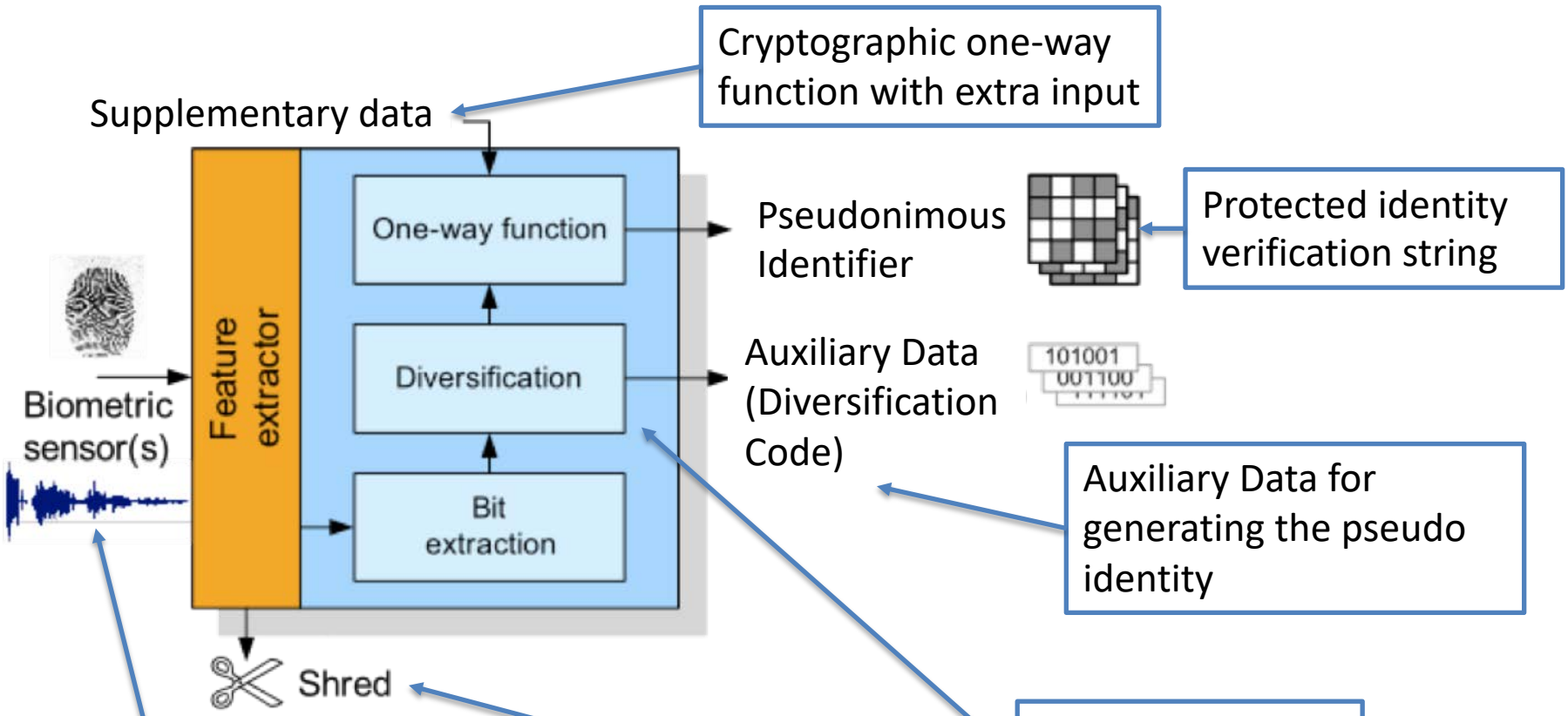
Biometric Template Protection (BTP) architecture



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

Pseudonymous Identifier Encoder (PIE)



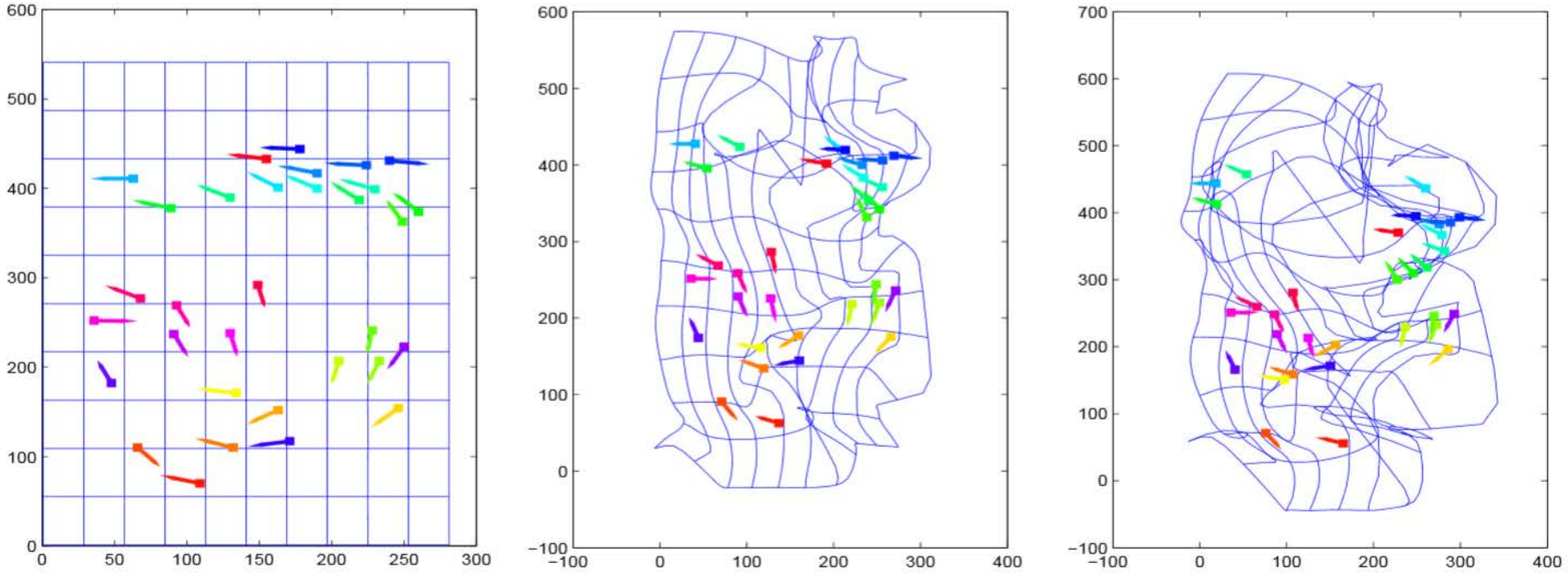
BTP approaches: Cancelable biometrics

- 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

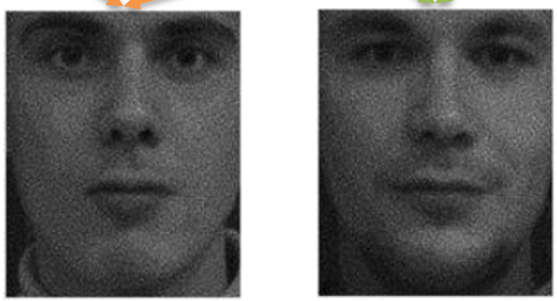
Public "host" images

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

Stored in separate databases!



PI

OR

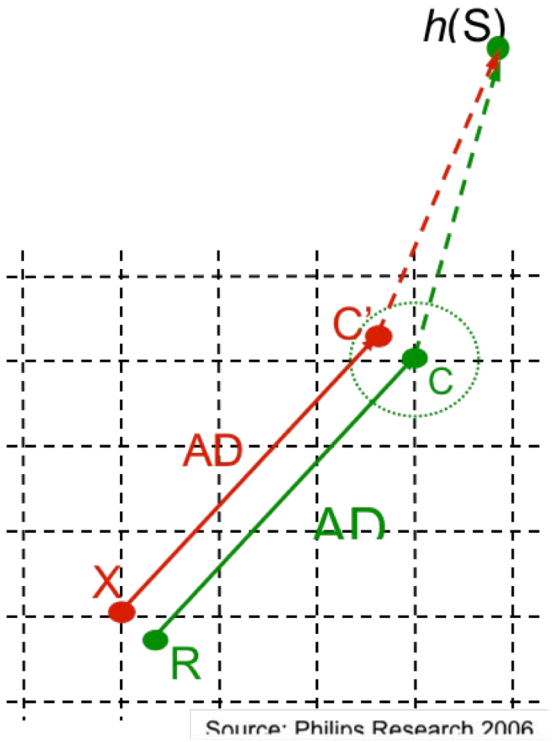
Reconstructed sample

BTP Approaches: Cryptobiometrics

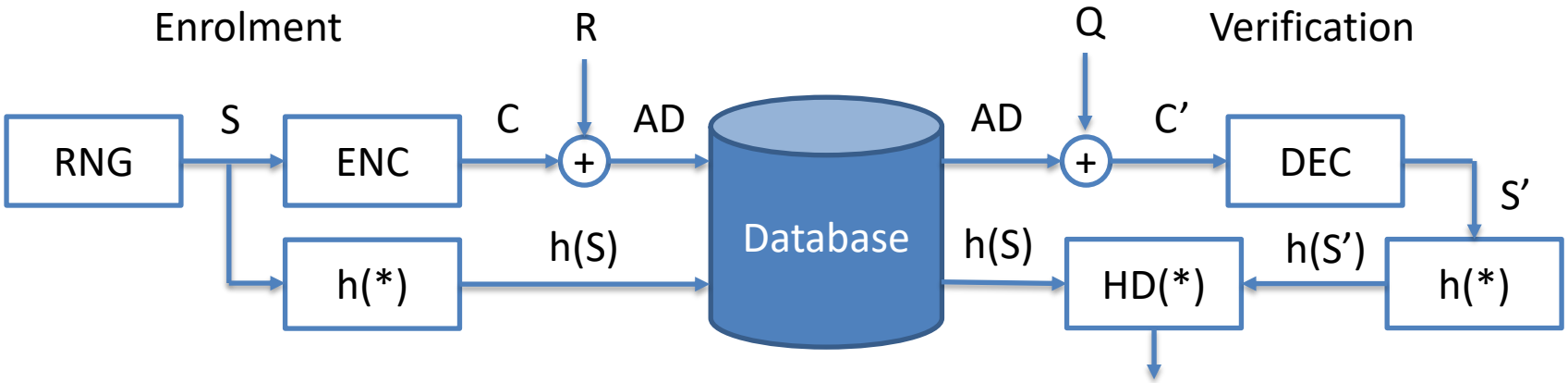
- These methods combine cryptographic keys with transformed versions of the original biometric templates to obtain secure templates.
- In most cases, some public information, known as helper data or auxiliary data, is generated.
- Two types:
  - **Key binding** schemes, where AD are obtained combining 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



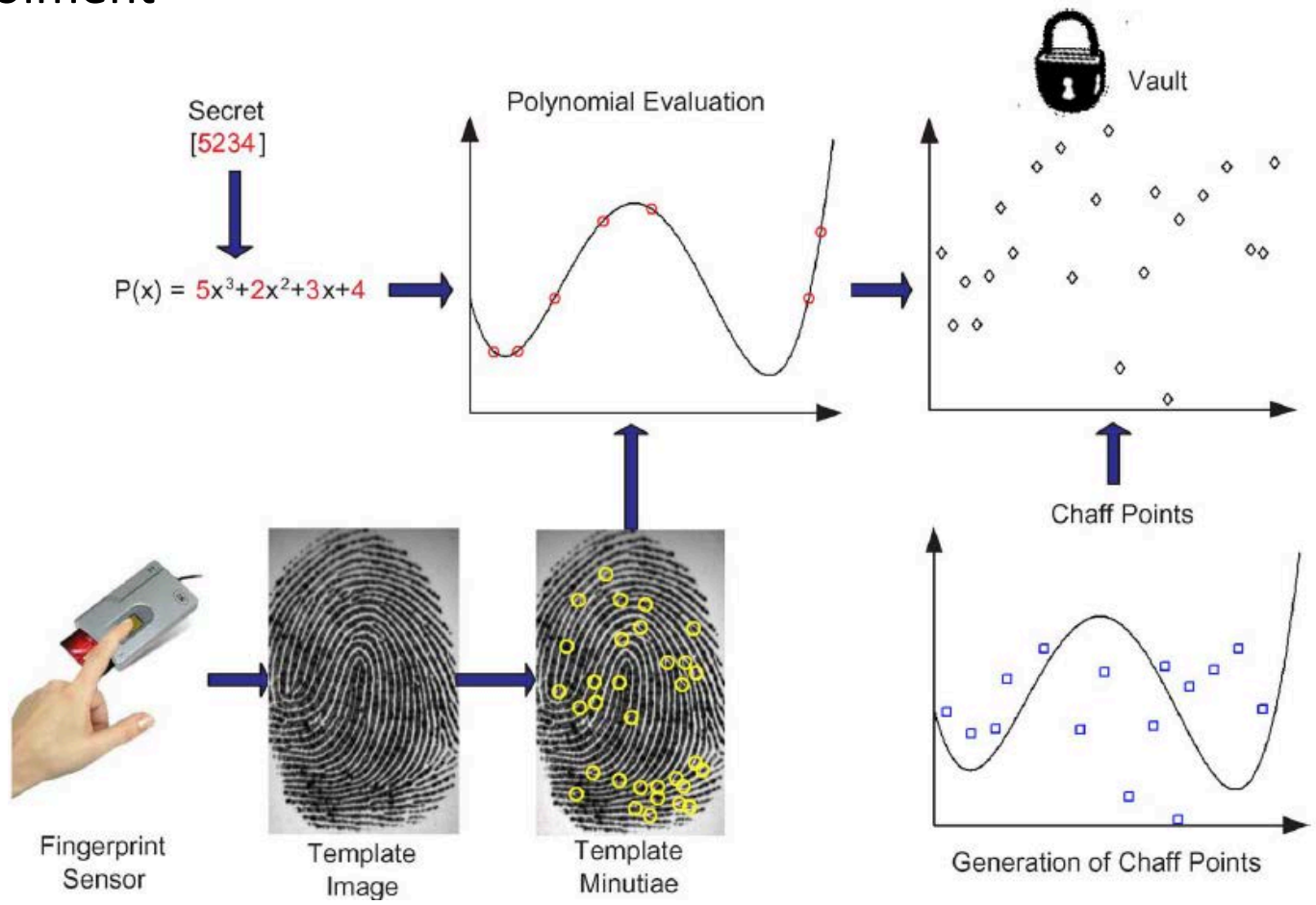
- Enrolment:
  - C is the codeword generated for the random string S
  - R is the binary extract of the reference vector
  - AD = C XOR R is the public AD
  - {h(S), AD} are stored as reference
- Verification:
  - C' = AD XOR Q (query vector)
  - HD(C, C') needs to be smaller than the error correction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

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

Cryptobiometrics: Fuzzy vault

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

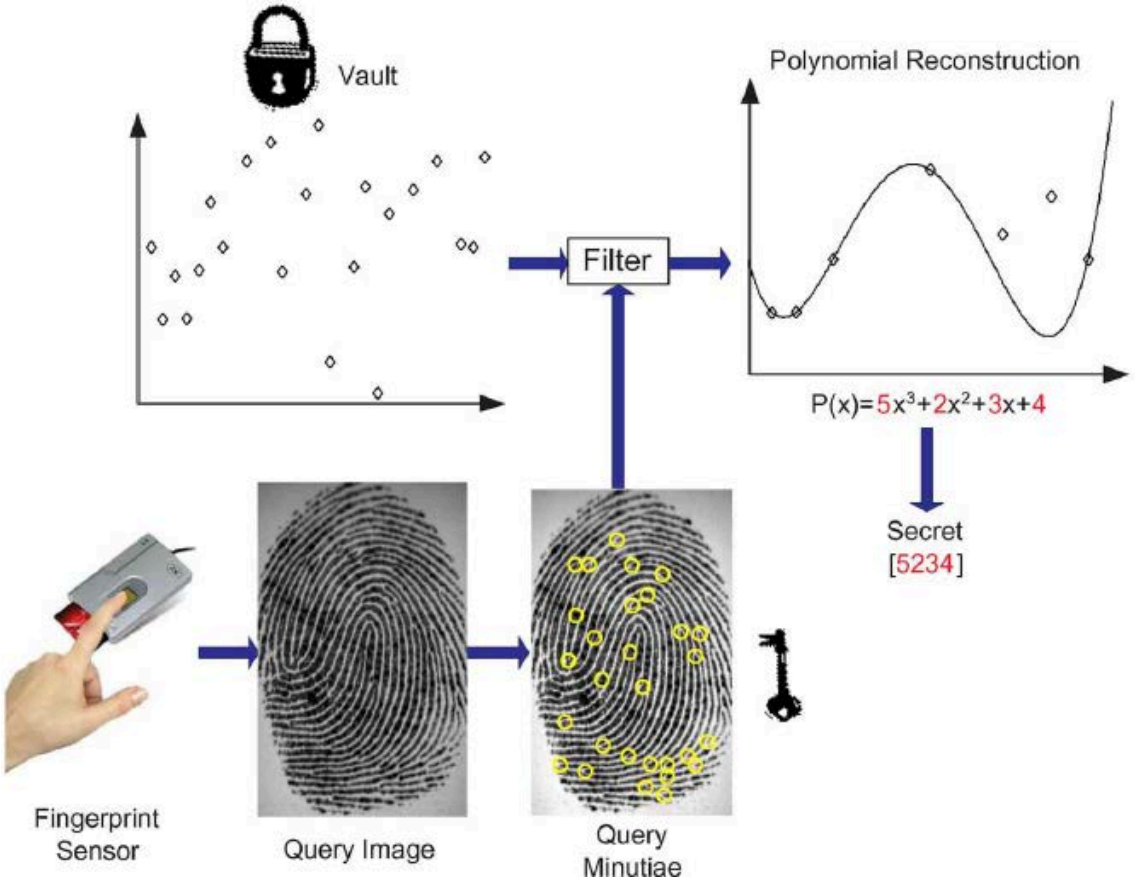
- Enrolment



Cryptobiometrics: Fuzzy vault

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

- Verification



BTP Approaches: Biometrics in the Encrypted Domain

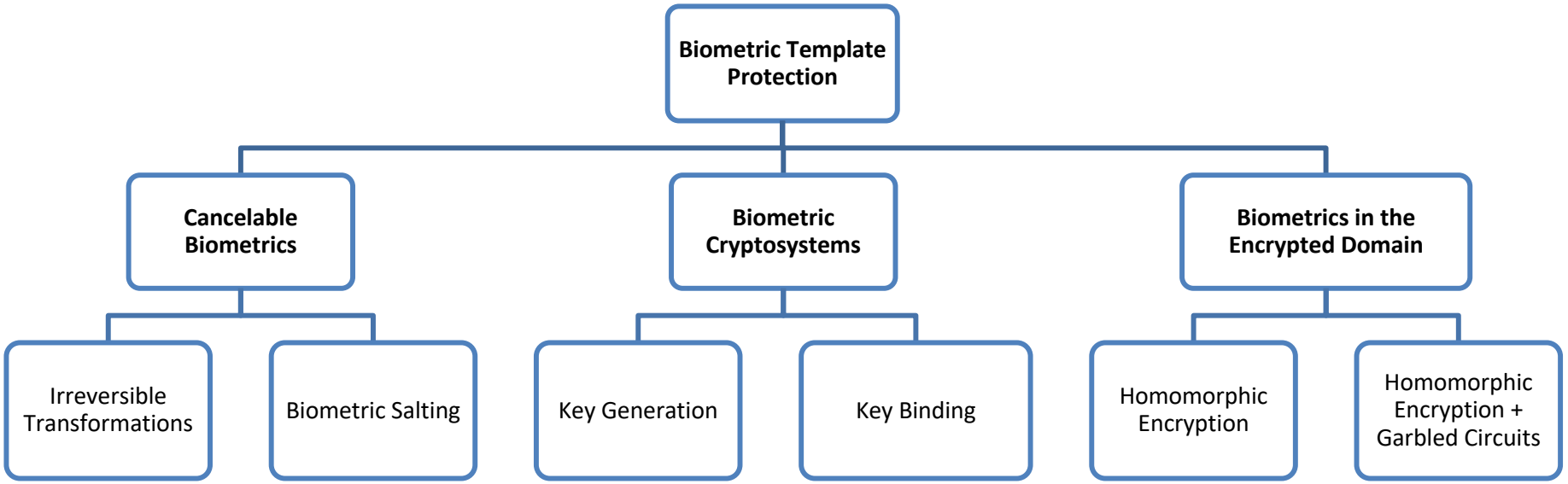
- Homomorphic Encryption (HE) schemes allow for computations to be performed on ciphertexts, with no additional AD, and which generate encrypted results which decrypt to plaintexts that match the result of the operations carried out on the original plaintext

- This solves the issue of decryption before authentication...

- But there is till no free lunch! HE is computationally expensive

- Garbled circuits can also be employed for particular operations

BTP Approaches: Summary



Geometric trans.
Feature mixing
BioTokens
Bloom Filters

Correlation
Random projection
Visual crypto

Quantization

Fuzzy vault
Fuzzy commitment
BioHashing

BTP Approaches: Pros and Cons

Cancelable Biometrics

- **Accuracy** drops
- Permanent **irreversibility**
- **Unlinkability** not analysed
- **Computational Complexity** Preserved

Template Protection based on Bloom filters

Cryptobiometrics

- **Accuracy** drops
- Attacks on AD (**irreversibility** compromised)
- **Unlinkability** not analysed
- **Computational Complexity** Preserved

[Campisi, *Springer* 2013]

Biometrics in the Encrypted Domain

- **Accuracy** preserved
- Permanent **irreversibility**
- **Unlinkability** granted
- **Computational Complexity** increased

Template Protection based on Homomorphic Encryption

Multi-Biometrics and BTP

- Multi-Biometrics:
  - Higher accuracy
  - Different levels of security
  - Three fusion levels: feature, score, decision [ISO/IEC TR 24722]

- Multi-Biometric Template Protection [Rathgeb and Busch, *InTech*, 2012]:
  - Alignment issues
  - Different BTP approaches for different characteristics

Summary

- Do the stored templates reveal any information about the original biometric samples?

IRREVERSIBILITY

- Are my enrolled templates in different recognition systems somehow related to each other?

UNLINKABILITY

- What if someone steals a template extracted from my face? Has it been permanently compromised?

RENEWABILITY

[ISO/IEC IS 24745 on Biometric Information Protection]

Security and Privacy Evaluation

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

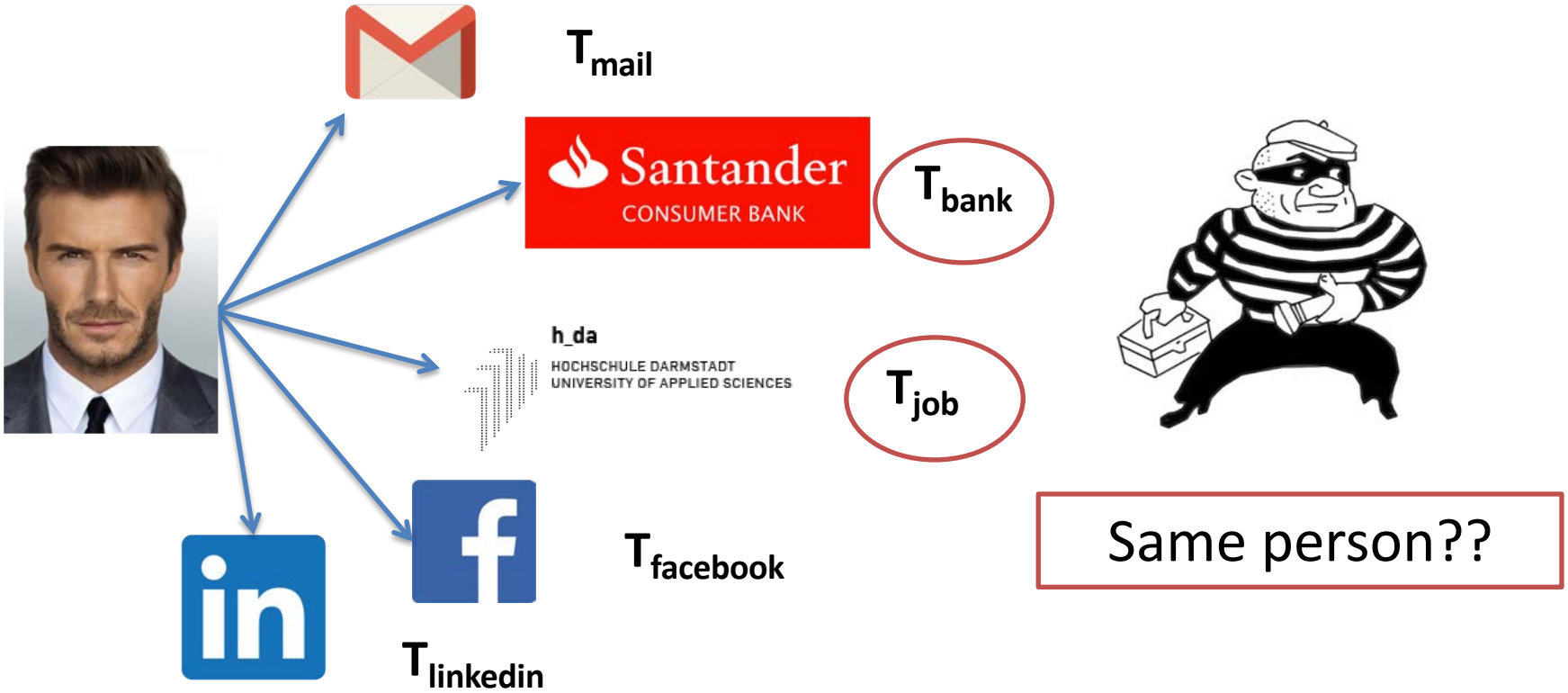
- How can we analyse irreversibility? Following cryptographic paradigms?

- 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

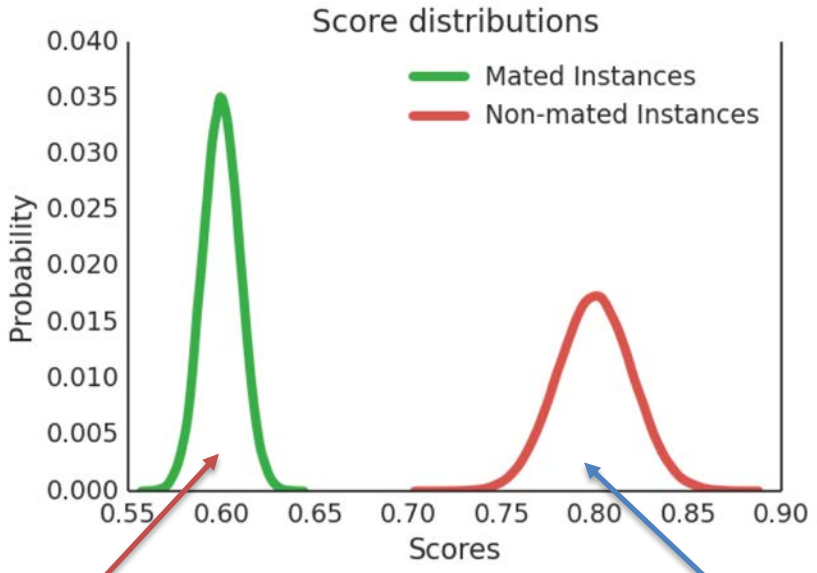
- We can enroll with a single characteristic in different applications



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: Current Status (I)

- Advantage of the attacker over a random guessing in the indistinguishability game
  - Problem 1: assumes uniformity of data – not valid in biometrics
  - Problem 2: only analysed for fuzzy schemes – not straightforward to apply to cancelable biometrics, since calculations rely on ECC properties

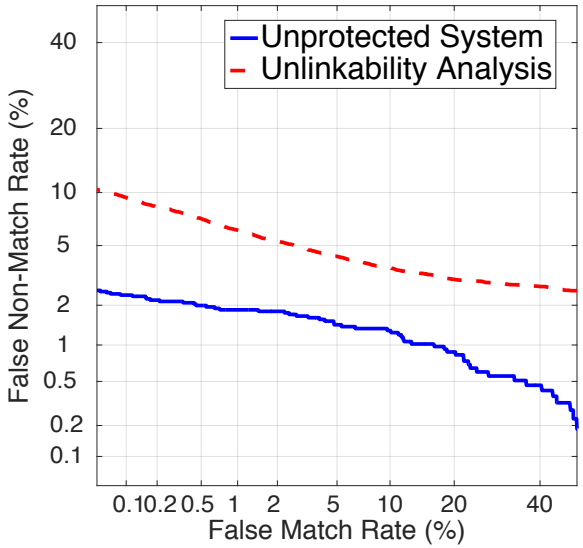
[Simoens09] K. Simoens, P. Tuyls, B. Preneel, "Privacy Weaknesses in Biometric Sketches", *IEEE Symp. On Security and Privacy*, 2009.

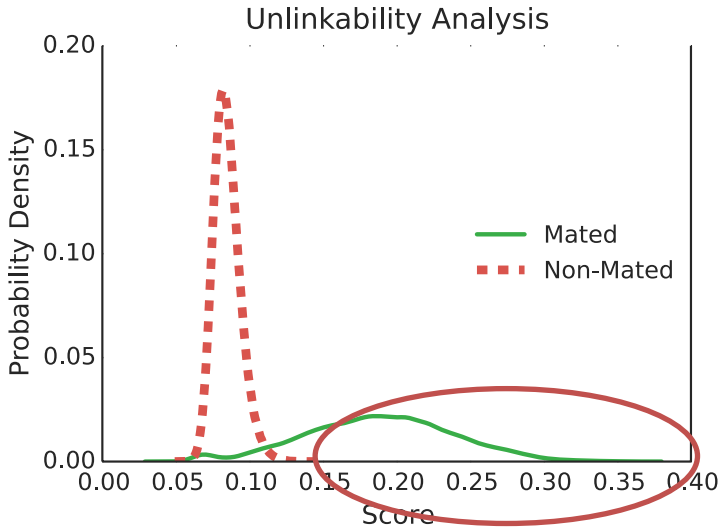
[Buhan09] I. Buhan, J. Breebaart, M. Guajardo *et al.*, "A Quantitative Analysis of indistinguishability for a continuous Domain Biometric Cryptosystem", *Int. Workshop on Data Privacy and Management*, 2009.

[Buhan10] I. Buhan, E. Kelkboom, J. Guajardo, "Efficient Strategies for Playing the Indistinguishability Game for Fuzzy Sketches", *IEEE Workshop on Information Forensics and Security*, 2010.

Unlinkability Analysis: Current Status (II)

- Plot a DET curve of genuine and impostor scores, comparing templates enrolled in different system



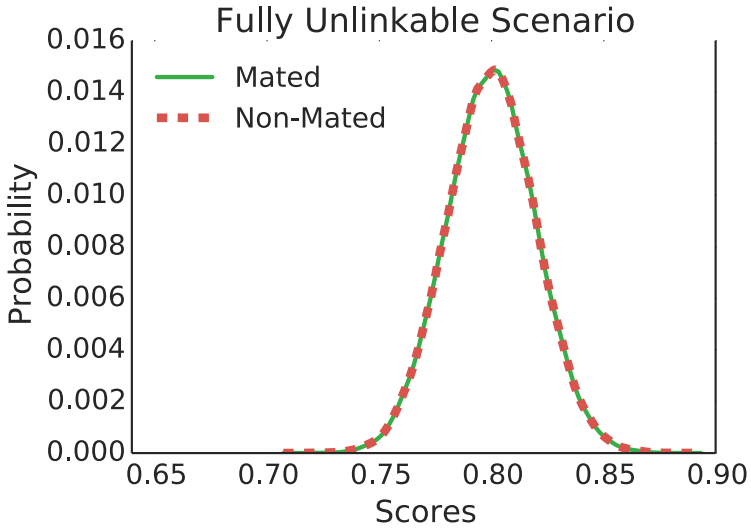


[Nagar10] A. Nagar, K. Nandakumar, A. K. Jain, "Biometric Template Protection Transformation: A Security Analysis", *SPIE, Electronic Imaging, Media Forensics and Security*, 2010.

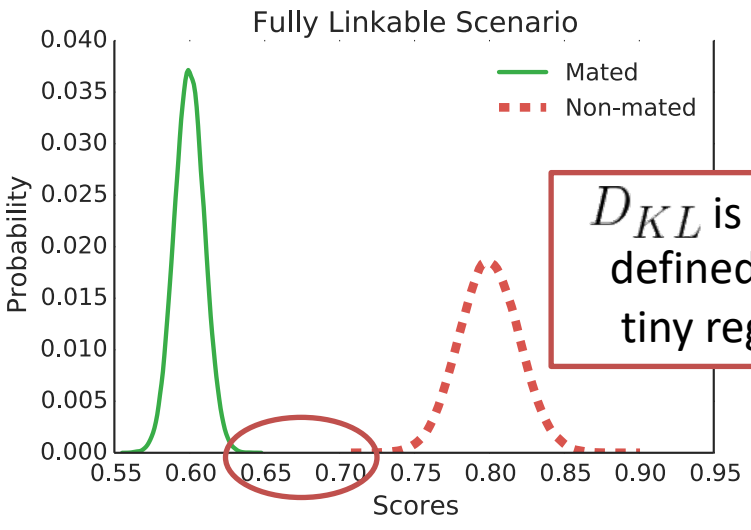
[Kelkboom11] E. Kelkboom, J. Breebart, T. Kevenaar *et al.*, "Preventing the Decodability Attack based Cross-Matching in a Fuzzy Commitment Scheme", *IEEE TIFS*, 2011.

Unlinkability Analysis: Current Status (III)

- Plot *Mated* and *Non-mated samples* distributions, for templates protected with different keys.
- How to analyse those distributions? ⇒ Kullback-Leibler (D_{KL}) divergence



D_{KL} = 0.0



D_{KL} is only defined in a tiny region

D_{KL} = 0.0005

D_{KL} is not bounded: $D_{KL} \in [0, \infty)$ ⇒ difficult to compare systems

Unlinkability Analysis: New Approach

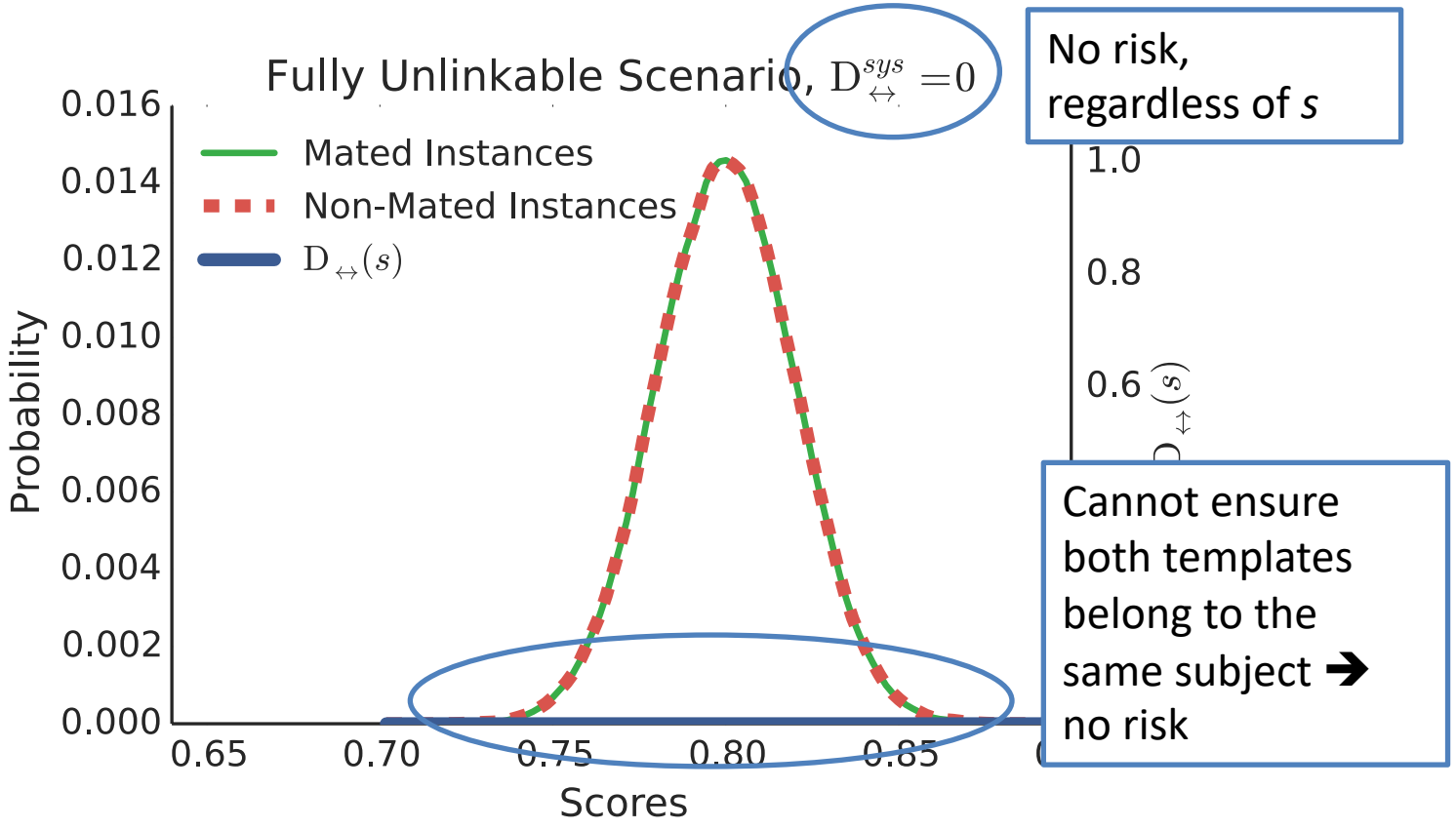
- Two measures:
  - Local measure $D_{\leftrightarrow}(s)$ ➔ for which scores is the system vulnerable?
  - Global measure $\mathrm{D}_{\leftrightarrow}^{sys}$ ➔ how can we compare two systems globally?

- 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]

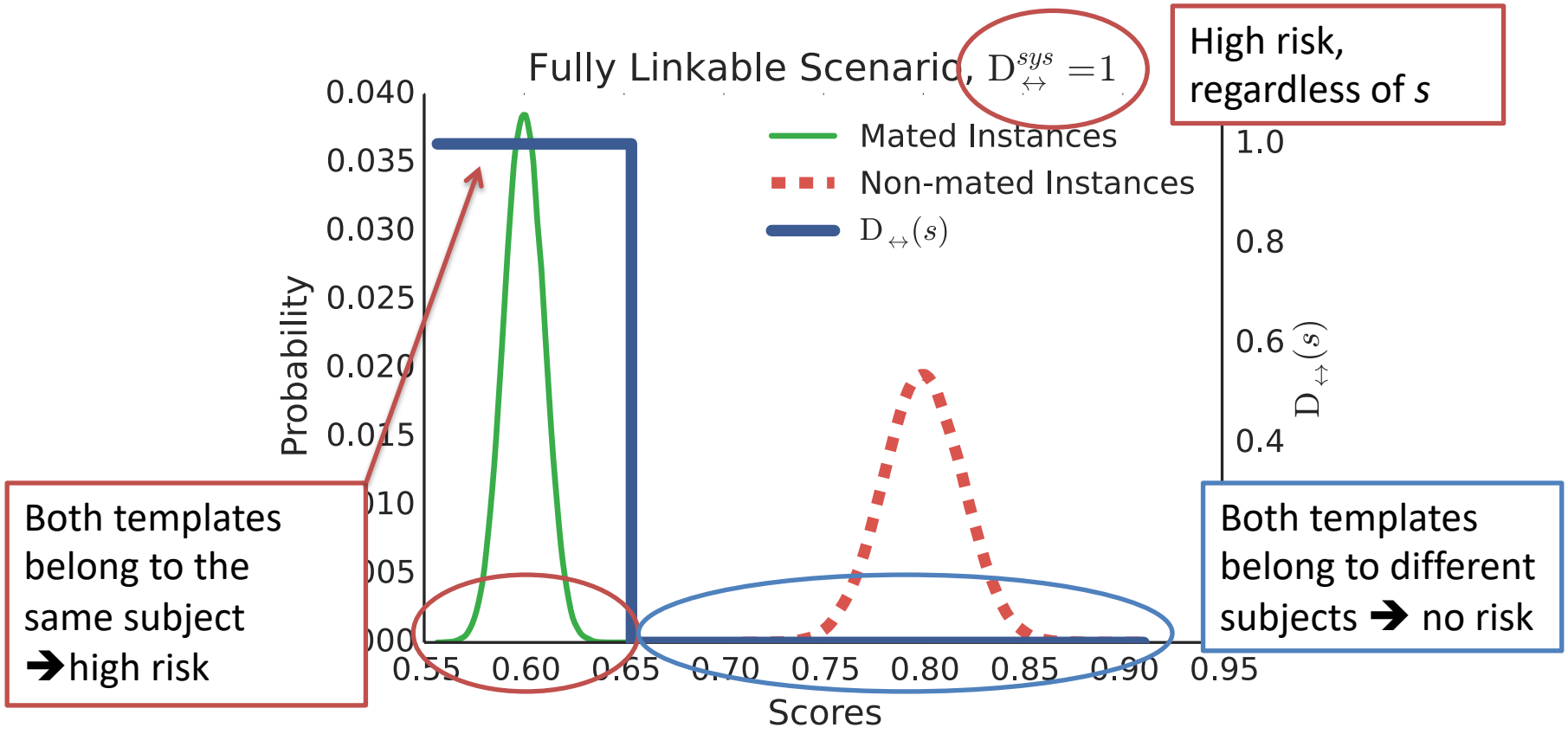
Full Unlinkability



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

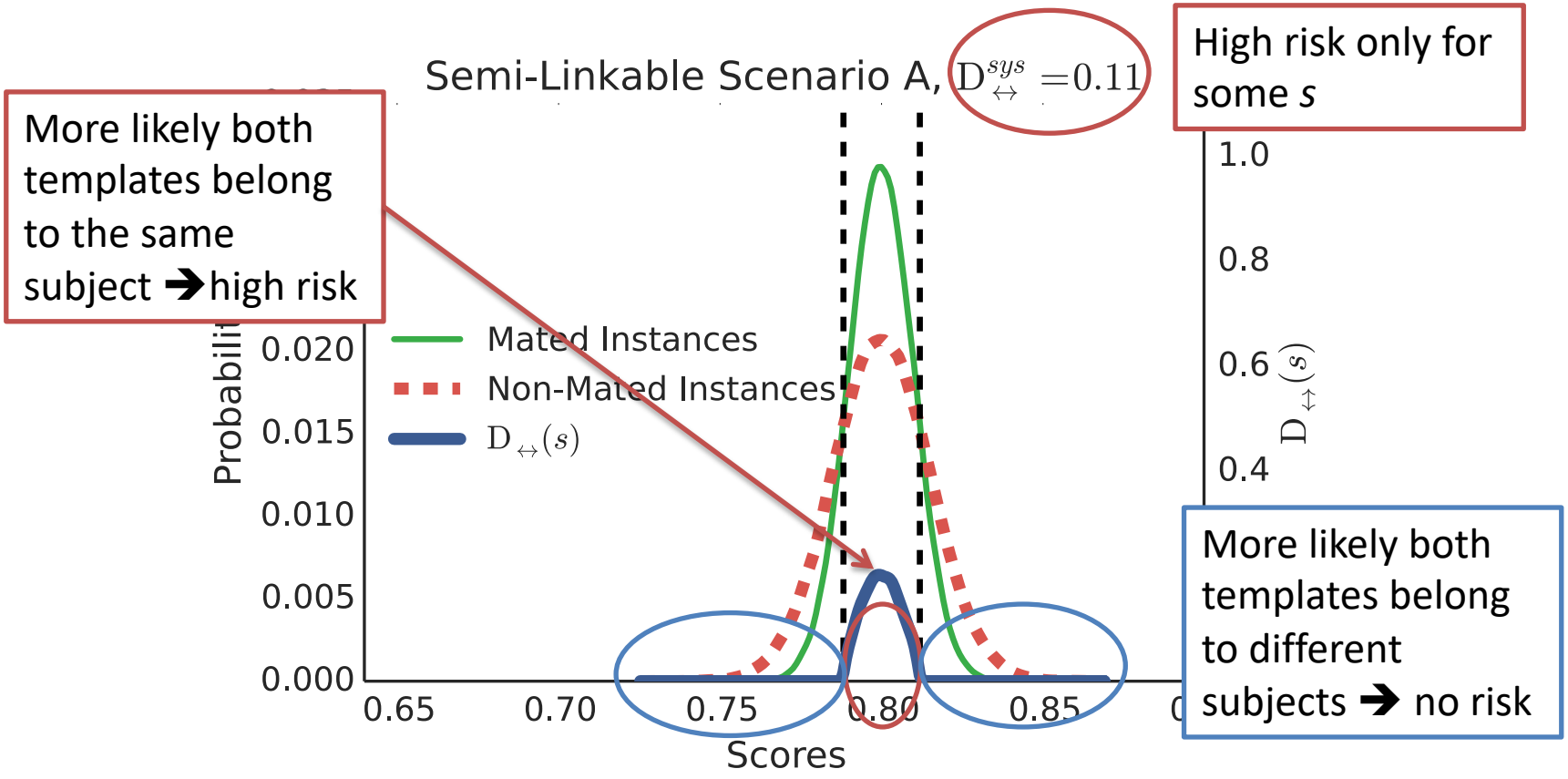
Evaluation

Full Linkability



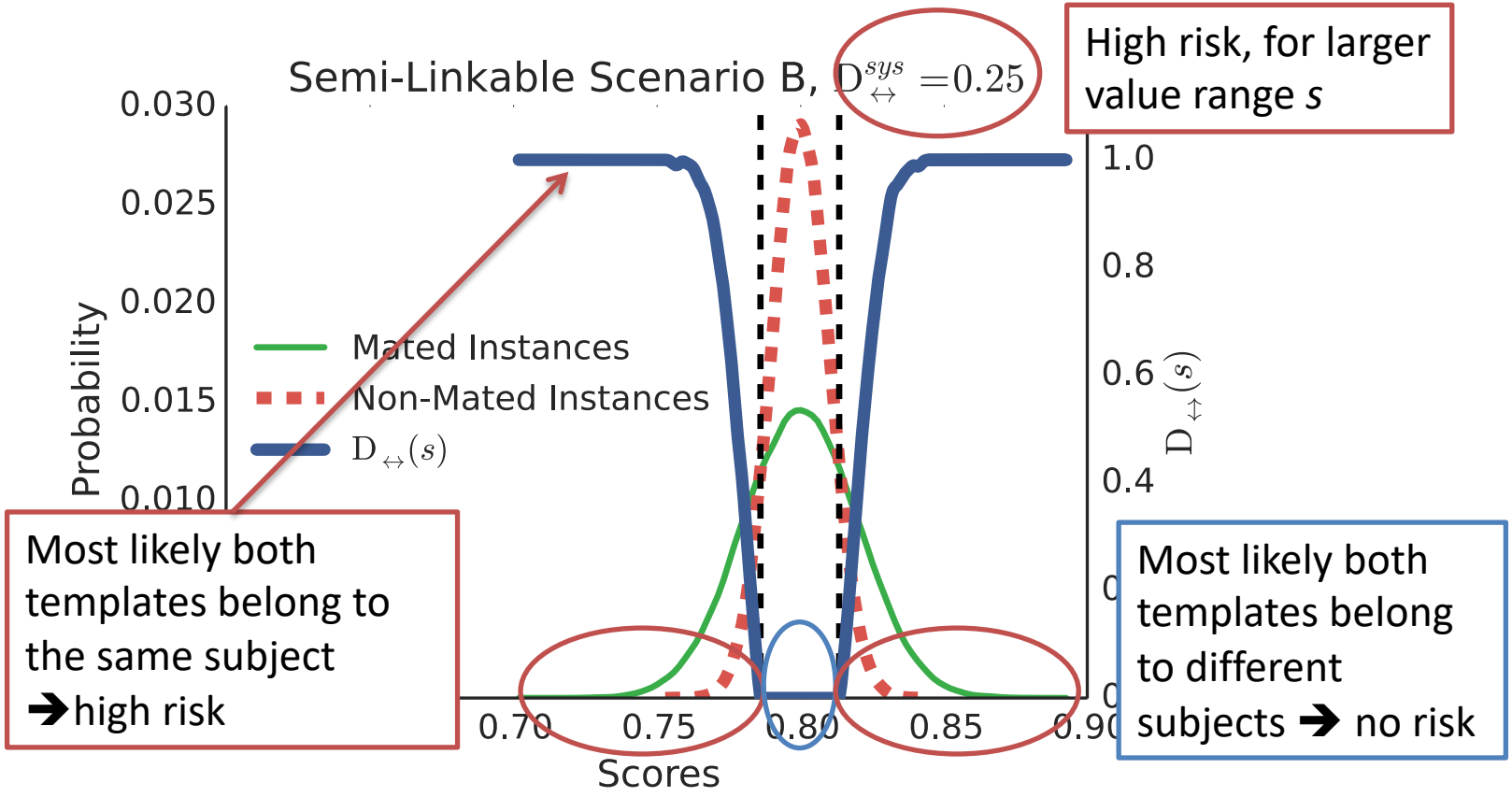
[Gomez-Barrero *et al.*, *IEEE T-IFS*, 2018]

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:

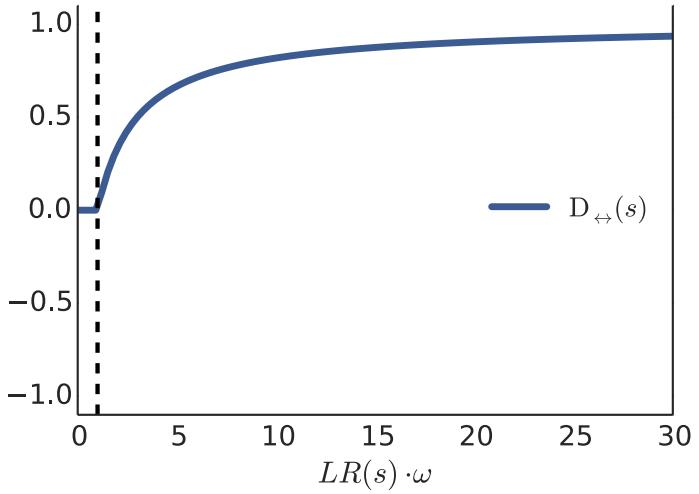
$$p(H_m|s) = \frac{LR(s) \cdot \omega}{1 + LR(s) \cdot \omega} \qquad \omega = p(H_m)/p(H_{nm})$$

[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

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

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

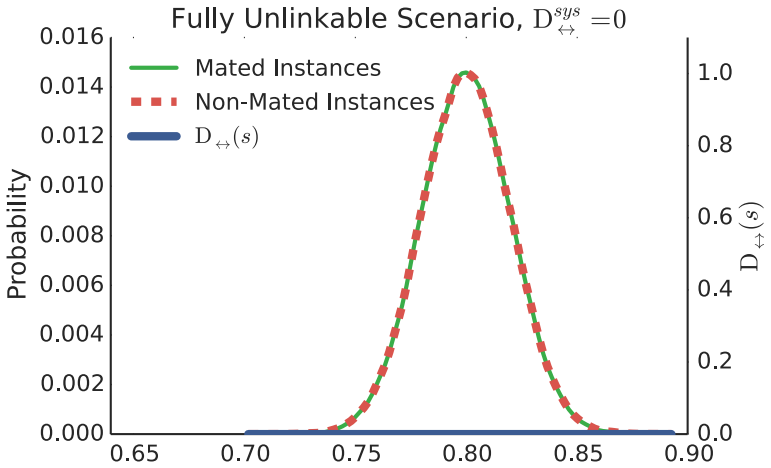
$$p(H_{nm}) \int_{s_{min}}^{s_{max}} p(s|H_{nm}) \cdot (p(H_m|s) - p(H_{nm}|s))\,\mathrm{d}s$$

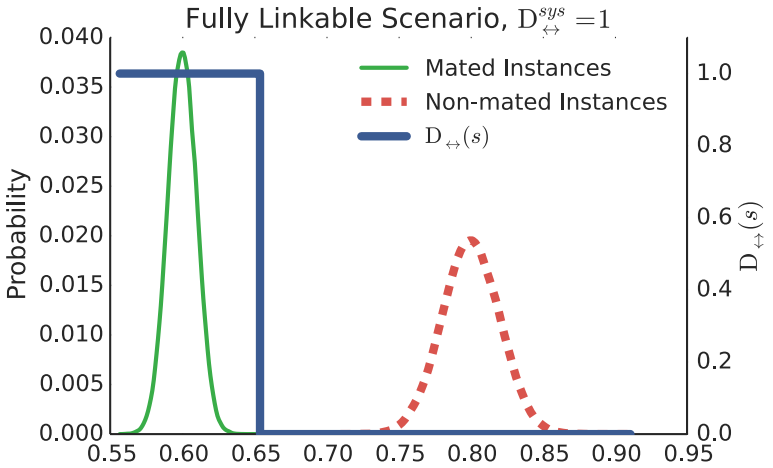
$$p(H_m|s) > p(H_{nm}|s)$$

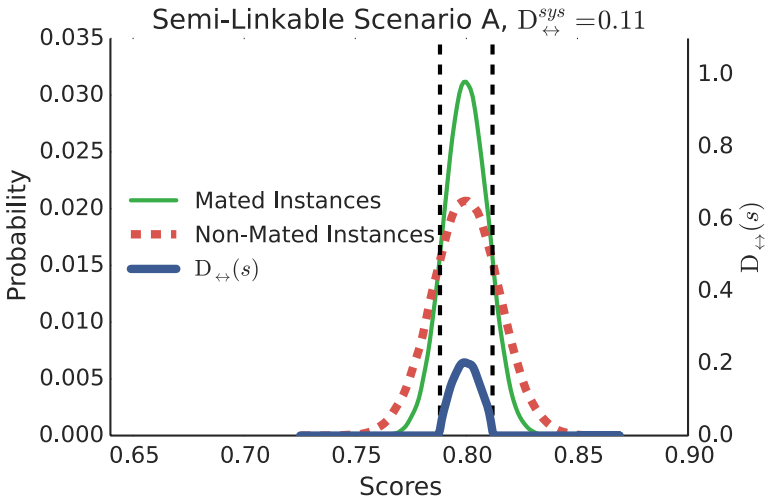
$$\mathrm{D}_{\leftrightarrow}^{sys} = \int_{s_{min}}^{s_{max}} p(s|H_m) \cdot \mathrm{D}_{\leftrightarrow}(s)\,\mathrm{d}s$$

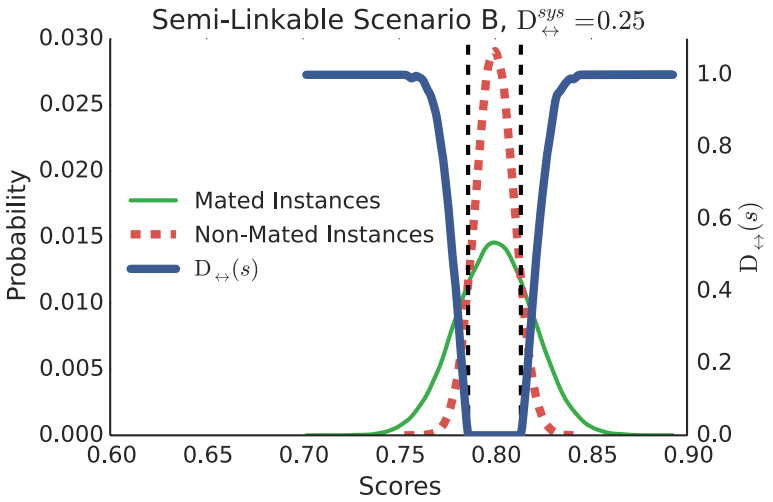
[Gomez-Barrero *et al.*, *IEEE T-IFS*, 2018]

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.

Cancelable Biometrics Based on Bloom Filters

Bloom Filters

Why Bloom filters?

[Bloom, *Comm. of the ACM* 1970]
[Broder and Mitzenmacher, *Internet Mathematics* 2004]

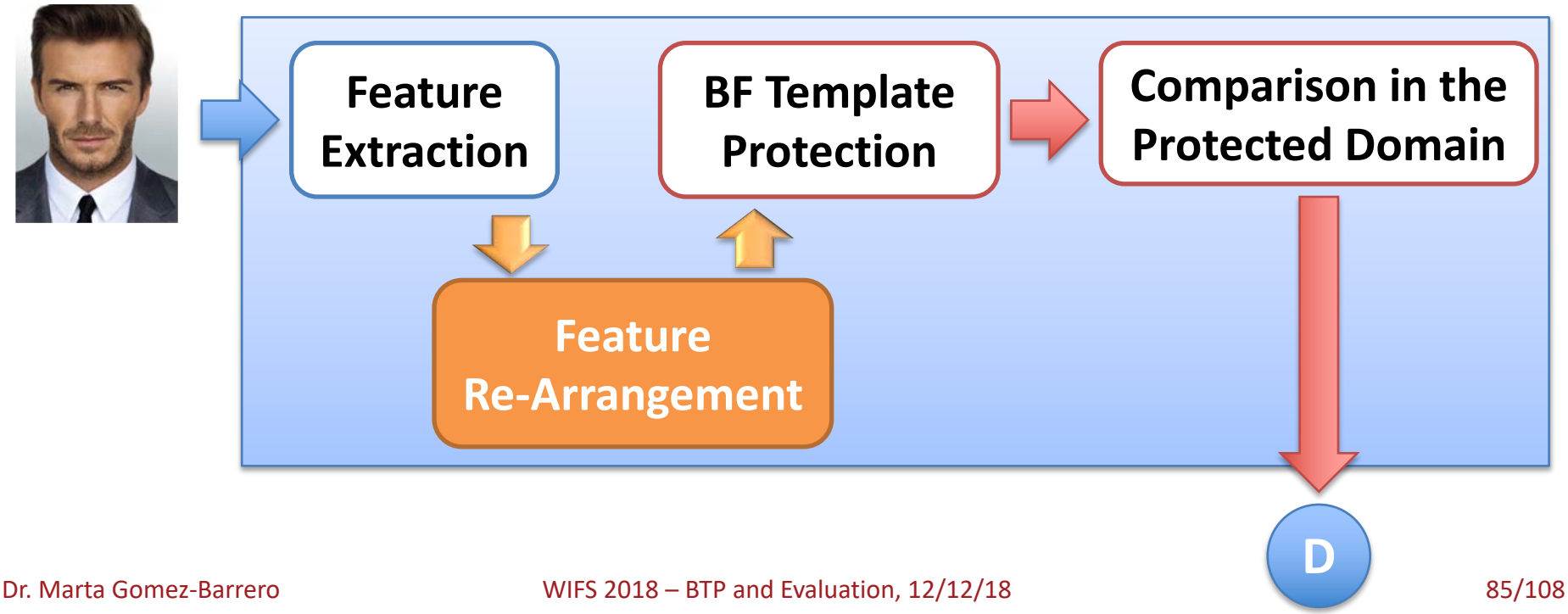
- Biometric Template Protection based on Bloom filters:
  - **General**: successfully applied to iris, face, fingerprint, fingervein
  - **Multimodal**: feature level fusion
  - **Irreversibility** achieved
  - **Accuracy**, depending on the configuration, preserved
  - **Template size**: similar or compressed
  - **Verification speed** similar

- But we need to add **unlinkability**
- And find a way to fuse templates of different sized (**Multi-Biometrics**)

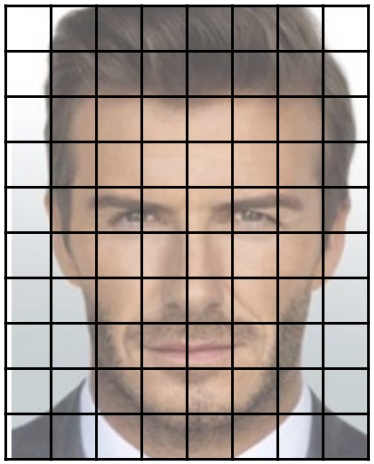
General architecture

- Adding unlinkability:
  - Small complexity
  - Small impact on accuracy

Random shuffling of bits ⇒ ↑EER > 40%



Bloom filters



Re-Arranged Block

0	0	1	0	0	1	1	0	0	1
1	1	1	0	1	0	0	1	1	0
1	0	0	0	1	1	0	0	1	0
0	0	0	1	0	0	1	0	0	1

nBits

nWords

6 4 6 9

0	1	0	1	0	0	1	1	0	1	0	0	0	

2^{nBits}

Bloom Filter

How can we select this parameters?

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

Protected Template

1 BF per block, of 2^{nBits}

Bloom filters

$|\mathbf{b}| = 2.4$

$|\mathbf{b}^{fused}| = 3.2$

If bit is activated here…

… it is also activated here

1 - α

pos

OR

MK-XOR

α

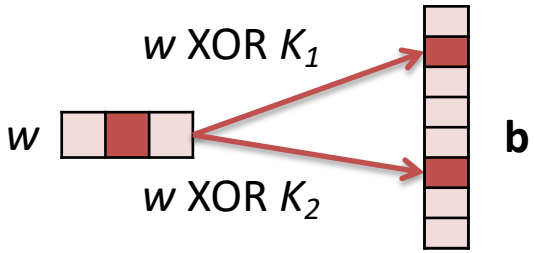
Same size

$|\mathbf{b'}| = 1.6$

$|\mathbf{b'}^{, fused}| = 3.2$

To achieve a a fusion weight α:

Different number of keys => different α



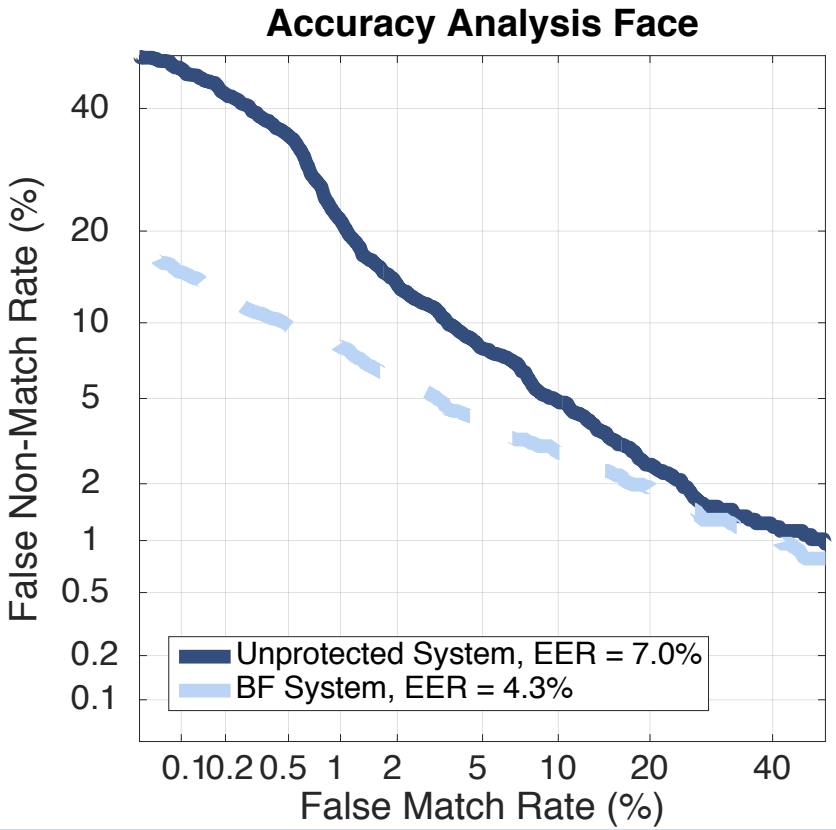
Set number of keys in terms of:
$|\mathbf{b}^{fused}| / |\mathbf{b'}|$

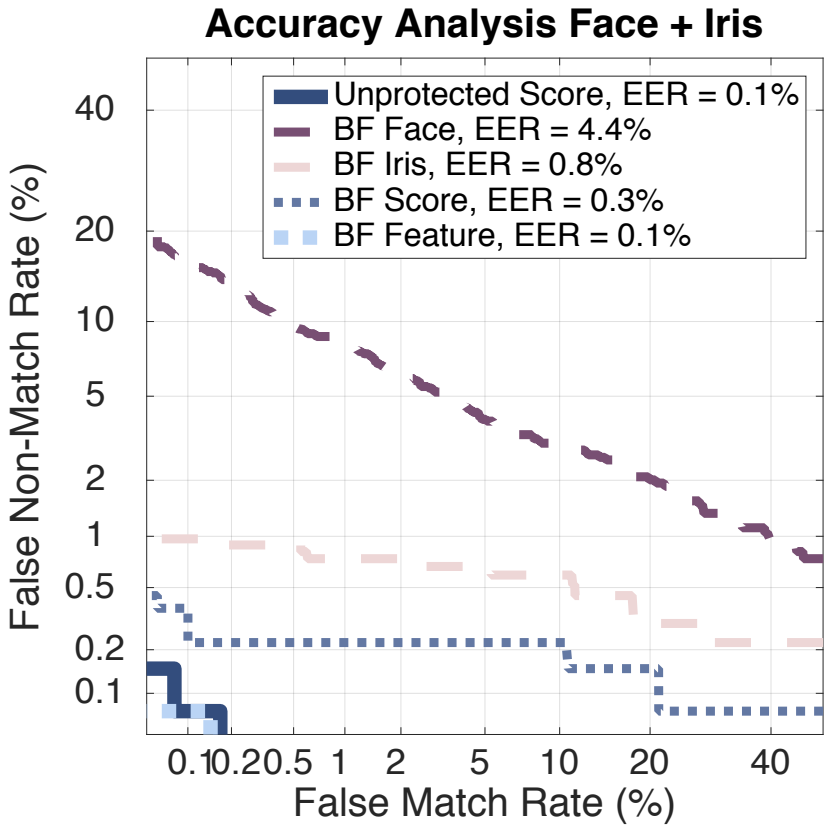
Sequential fusion

- A similar approach can be followed for a sequential fusion, in order to minimise the interaction with the subject.
- Once the decision threshold is reached, access is granted.
- The *i*-th similarity score S_i is obtained comparing the *i*-th fused probe template $\mathbf{C}_q^i$ with the reference template, $\mathbf{C}_r^i$, comprising the information of all the characteristics.
- The templates can be iteratively computed as follows:

$$\mathbf{C}_q^i = \mathbf{C}_q^{i-1} \ \texttt{OR} \ \mathbf{C}$$

Accuracy Analysis



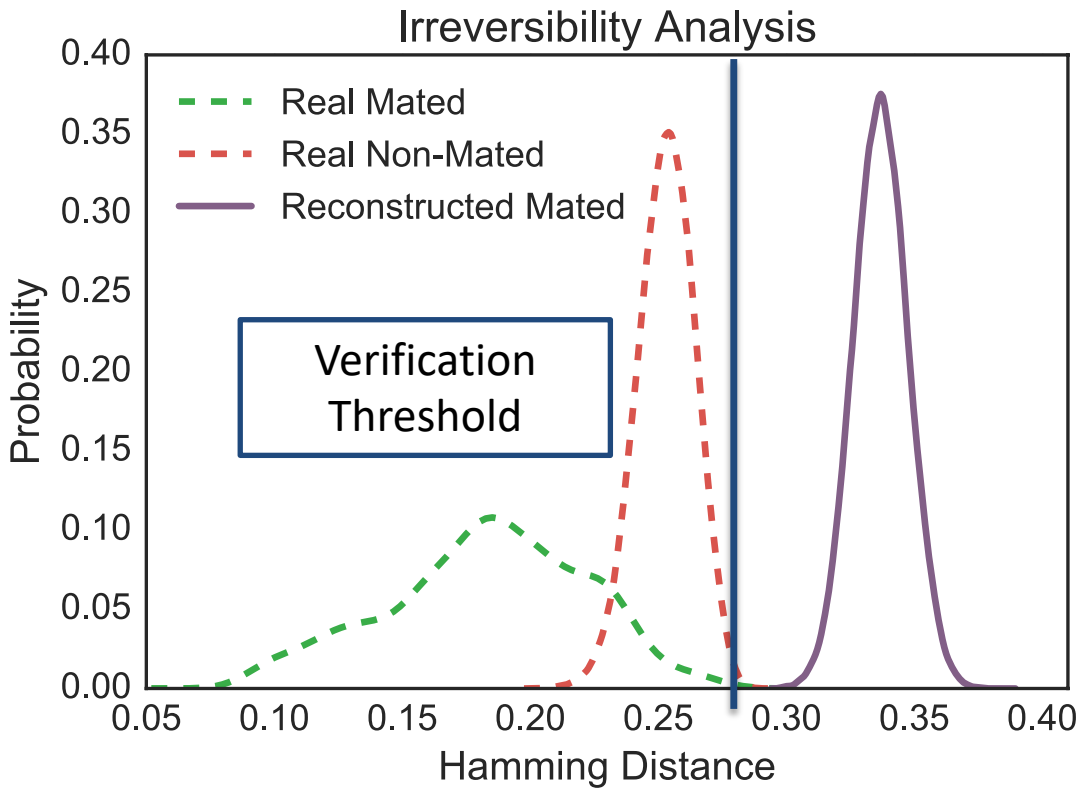


Accuracy is preserved at all operating points

For the fusion, best accuracy for protected feature level

Irreversibility analysis

- Are the reconstructed unprotected templates similar to the original ones?



[Bringer *et al.*, *ICB* 2015]

Irreversible: HD bigger than impostor comparisons

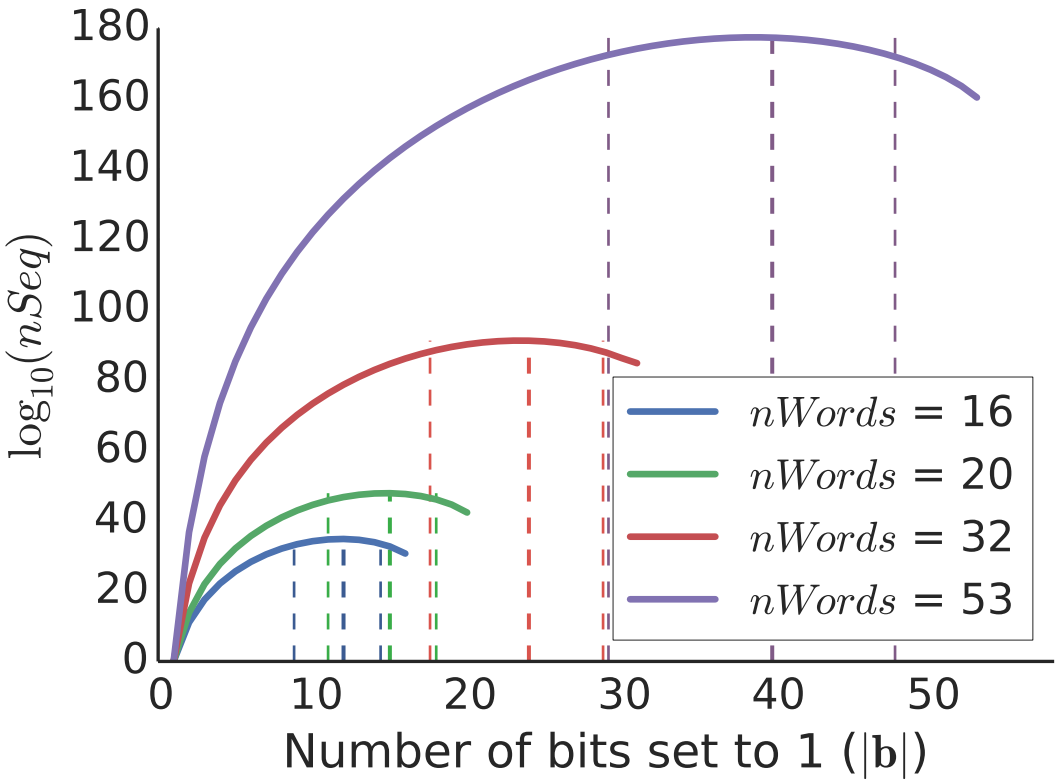
Irreversibility

- Question: How many original sequences lead to a single protected template?
- Bloom filter indexes are visible to an attacker ⇒ the reconstruction of the corresponding binary block involves an arrangement of $|\mathbf{b}| < nWords$ ($|\mathbf{b}|$ = # activated bits) different words to a binary block of length *nWords*.
- By the inclusionexclusion principle, the total number of possible sequences *nSeq* is:

$$nSeq = \sum_{i=1}^{|\mathbf{b}|} (-1)^{|\mathbf{b}|-i} \binom{|\mathbf{b}|}{i} i^{nWords}.$$

- And then, we have to undo the permutation.

Irreversibility



- We estimate |**b**| over a particular database (e.g., Biosecure Multimodal DB):

  |**b**| = 56.3

- With that value, we have

  $nSeq = 2^{40}$

- For a full disclosure model, the probability of a reconstruction is

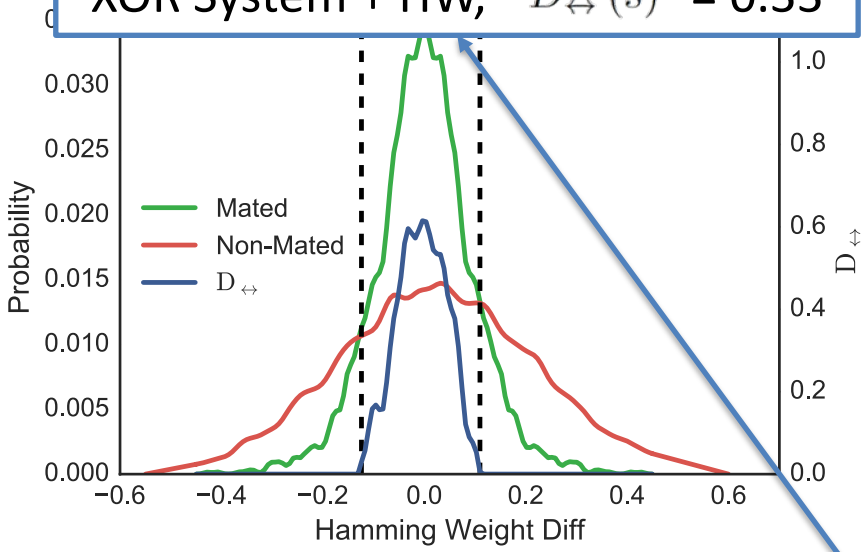
  $2^{-40,960}$

[Gomez-Barrero *et al.*, *Information Sciences* 2016]

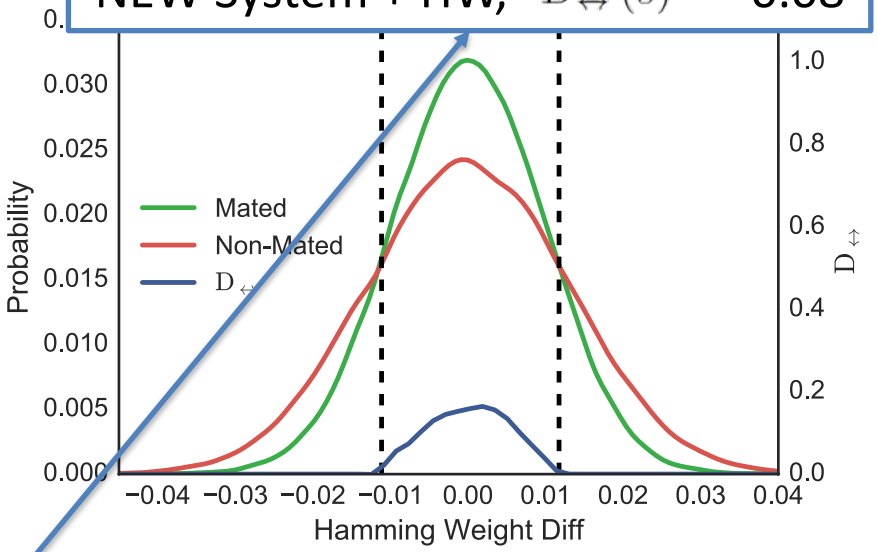
Bloom filters

Unlinkability analysis (I)

XOR System + HW, $D_{\leftrightarrow}(s)$ = 0.33



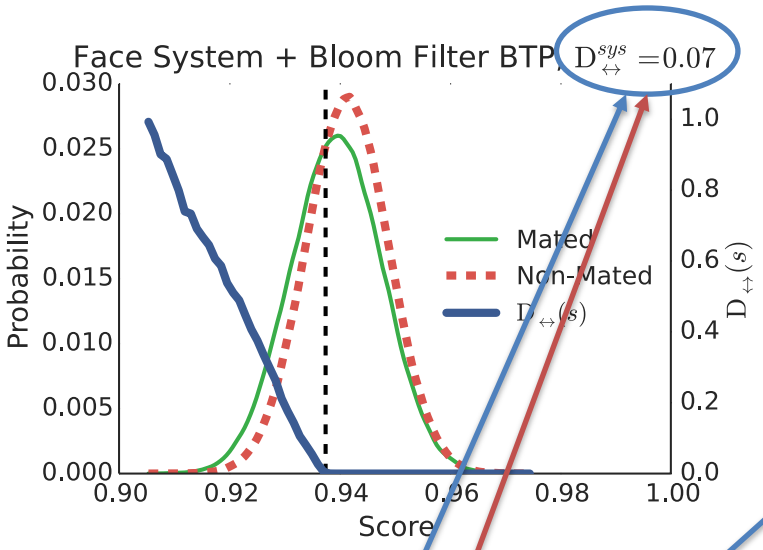
NEW System + HW, $D_{\leftrightarrow}(s)$ = 0.08

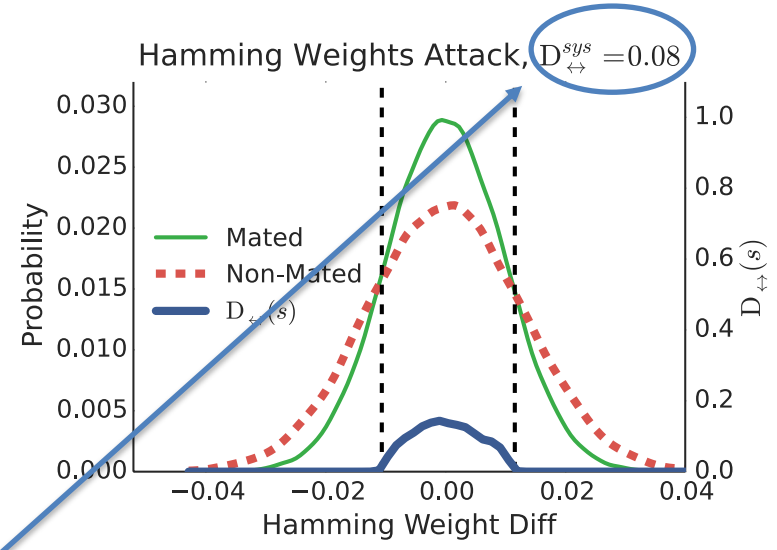


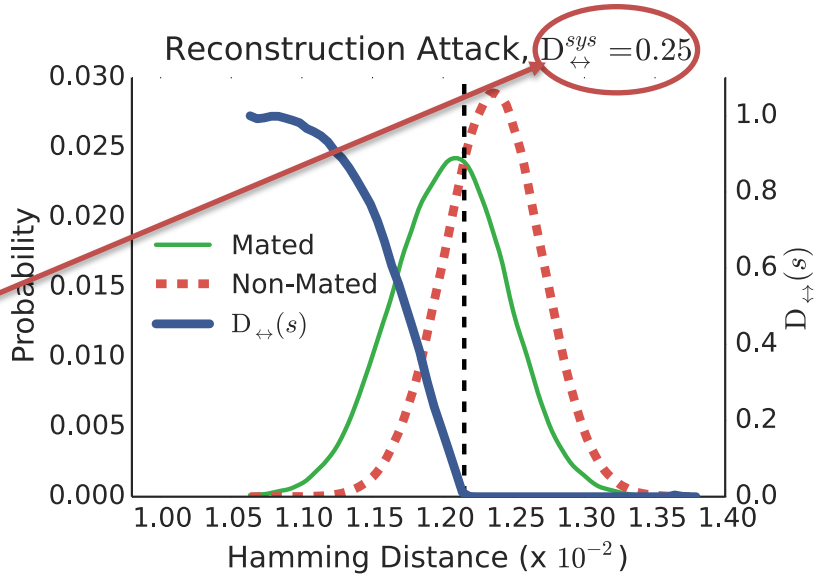
Linkability has decreased! ☺

Bloom filters

Unlinkability analysis (II)







Linkability has barely increased ☺

Still room for improvement

BTP Based on Homomorphic Encryption

Why Homomorphic Encryption?

- BTP based on Homomorphic Encryption:
  - **General**
  - **Accuracy fully preserved**
  - **Permanent protection:** all computations in the encrypted domain
  - **Irreversibility** and **unlinkability** achieved
  - **Renewability** with no re-acquisition

- Limitation on the number of operations in the encrypted domain
- Secret key + protected template = unprotected template compromised

[Fontaine *et al.*, *EURASIP J. Inf. Sec.* 2007]
[Lagendijk *et al.*, *IEEE SP Mag.* 2013]

Homomorphic Encryption

- Practical implementation: Paillier Cryptosystem [P. Paillier, EUROCRYPT, 1999]
- HE- Paillier: based on the DECISIONAL COMPOSITE RESIDUOSITY ASSUMPTION

DCRA: given a composite n and and integer z, it is (very) hard to decide whether there exists y such that:

$$z = y^n \pmod{n^2}$$

Additive 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

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

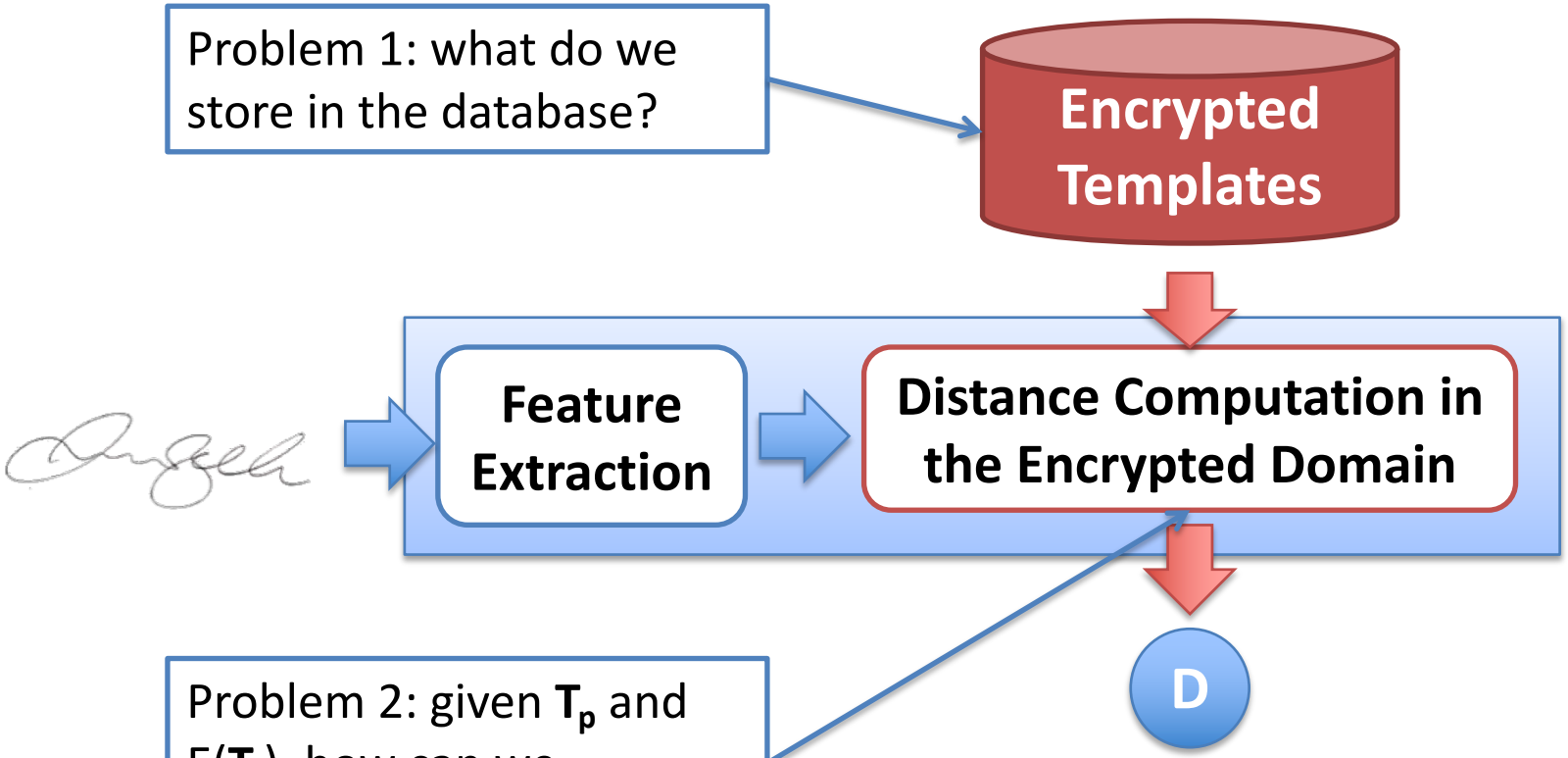
Exponentiation of ciphertext and plain text

Product of plain texts

General architecture

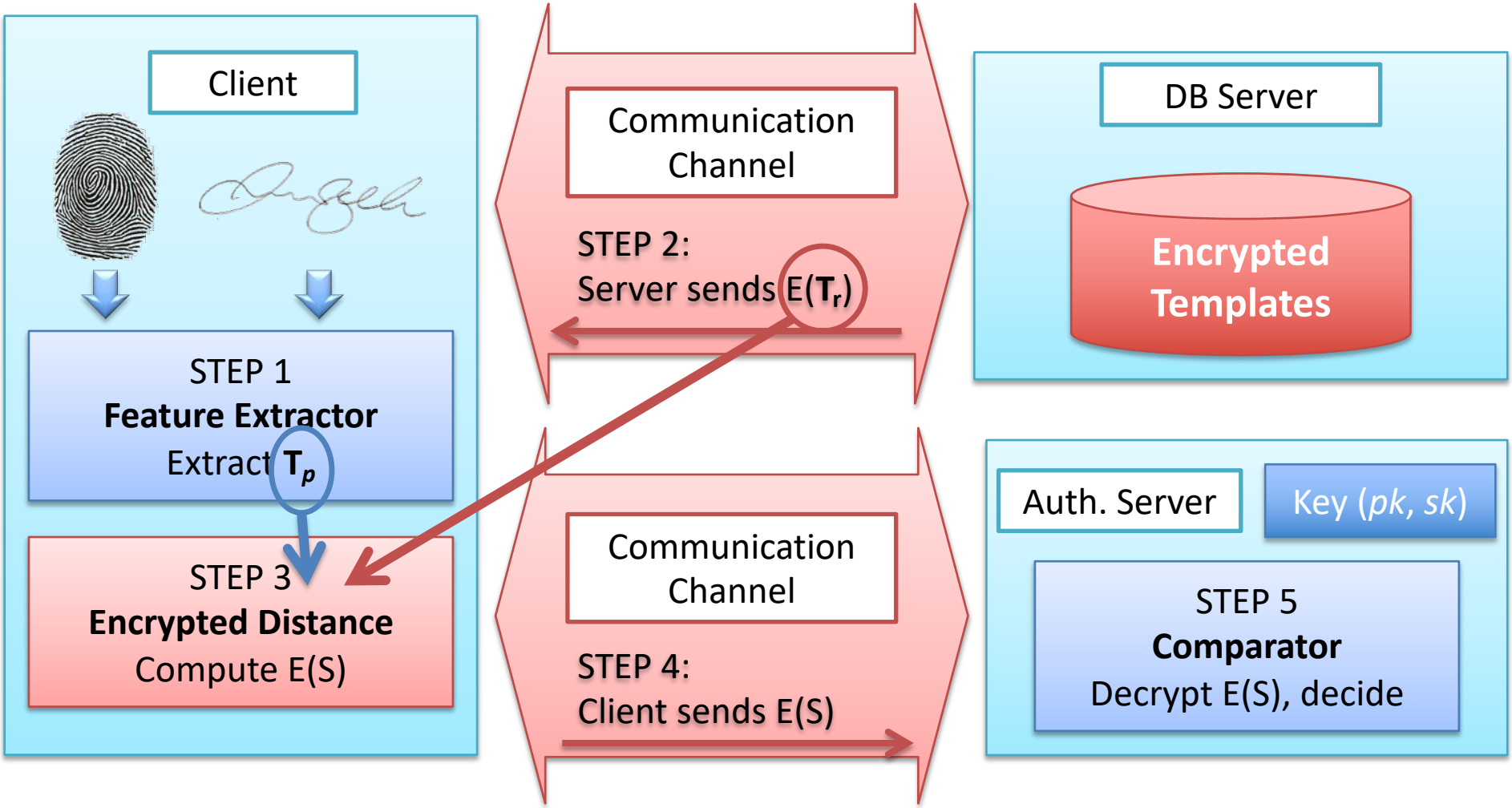
Additionally, only integer values can be handled

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-Biometrics



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

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

Encrypted reference template stored in DB

Probe template

BTP & HE

Cosine similarity: Given two vectors $\mathbf{T}_p$ and $\mathbf{T}_r$, of length F

$$d_{cos}(\mathbf{T}_p, \mathbf{T}_r) = \frac{\mathbf{T}_p \cdot \mathbf{T}_r}{\|\mathbf{T}_p\| \cdot \|\mathbf{T}_r\|} = \sum_{f=1}^{F} \frac{p_f \cdot r_f}{\|\mathbf{T}_p\| \cdot \|\mathbf{T}_r\|}$$

$$d_{cos}(\mathbf{T}_p, \mathbf{T}_r) \in [0, 1] \Rightarrow S_{cos} = 10^{12} d_{cos}(\mathbf{T}_p, \mathbf{T}_r)$$

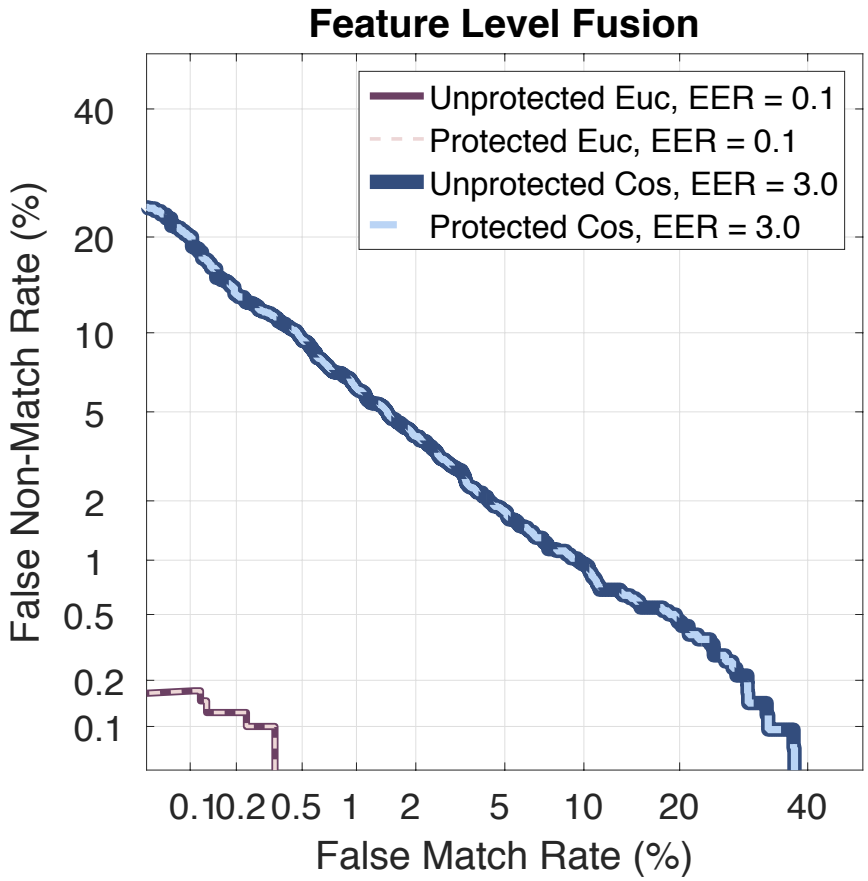
Encrypted Cosine similarity: Given two vectors $\mathbf{T}_p$ and $E(\mathbf{T}_r)$, of length F

$$E(S_{cos}) = \prod_{f=1}^{F} E\left(\frac{10^6 r_f}{\|\mathbf{T}_r\|}\right)^{10^6 p_f / \|\mathbf{T}_p\|}$$

Encrypted reference template stored in DB

Probe template

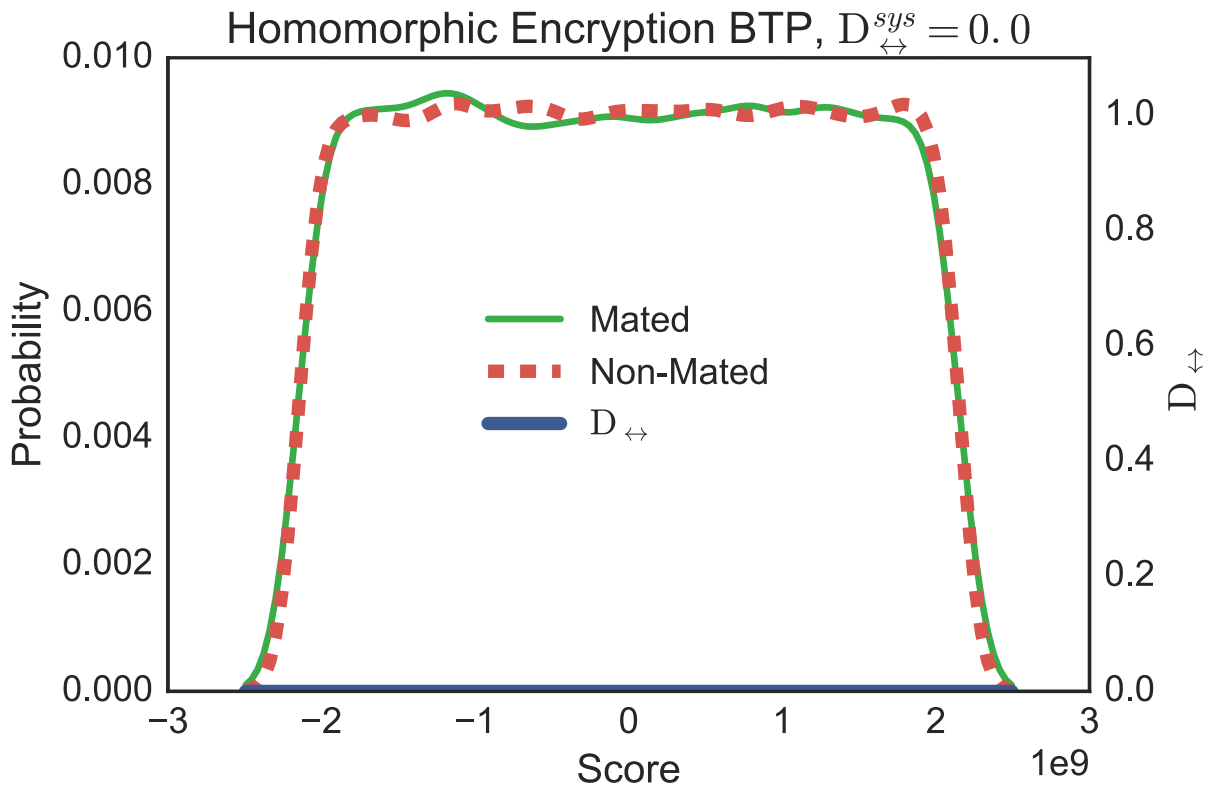
Accuracy Evaluation



BioSecurID DB [Fierrez *et al.*, *PAA* 2009]
Global Features Sign. [Martinez-Diaz *et al.*, *IETBio* 2014]
Fingercodes [Jain *et al.*, *CVPR* 1999]
4,200 mated + 17,500 non-mated scores

Accuracy is fully preserved at all operating points

Unlinkability Analysis



Full unlinkability, as long as the secret key is not compromised

Computational Overhead

- 1 real value (16 bits) ➔ 2,048 bits encrypted ➔ x 128 increase factor
- Depending on distance, more values need to be stored

Unprotected template:
F real values ➔ 0.27 KB

Euclidean distance template:
2*F* + 1 encrypted values ➔ 70.25 KB

Cosine distance template:
F encrypted values ➔ 35 KB

Storage requirements and communication bandwidth multiplied by 128 - 256

However, templates are still small enough for real time apps

Summary

Summary

- Biometric data is sensitive data, which needs to be protected, providing **irreversibility, unlinkability, renewability and accuracy preservation**.
- Unprotected templates can be reconstructed using inverse biometrics methods, where only access to similarity scores is required.
- Current BTP schemes can be classified as cancelable biometrics, cryptobiometric systems, or biometrics in the encrypted domain.
- We need to follow a standardised methodology for a standardised security and privacy evaluation of BTP schemes.
- Case studies:
  - BTP schemes based on Bloom filters or Homomorphic Encryption comply with ISO/IEC IS 24745.
  - MBTP schemes can achieve higher accuracy and privacy protection

Summary

- Bloom filters advantages:
  - **Compressed** templates
  - **Irreversibility** even if **key is compromised**
  - **Low** computational **load**

- HE advantages:
  - **Full accuracy** preservation
  - Revocability with **no re-acquisition**
  - **Higher** degree of **unlinkability**

- Bloom filters limitations:
  - **Some accuracy degradation** depending on feature extractors
  - **Some accuracy degradation** at low FMRs

- HE limitations:
  - **Key compromised → reversible**
  - **Storage** requirements **x 128**

Dr. Marta Gomez-Barrero
(marta.gomez-barrero@h-da.de)

da/sec
BIOMETRICS AND INTERNET-SECURITY
RESEARCH GROUP

CRISP
Center for Research
in Security and Privacy

References

- **ISO/IEC 24745 on Biometric information protection**
- **ISO/IEC 30136 on Performance testing of biometric template protection schemes**
- M. Gomez-Barrero, C. Rathgeb, G. Li, R. Raghavendra, J. Galbally, C. Busch, "Multi-Biometric Template Protection Based on Bloom Filters", *Information Fusion*, vol. 42, pp. 37-50, 2018
- **M. Gomez-Barrero, J. Galbally, C. Rathgeb, C. Busch, "General Framework to Evaluate Unlinkability in Biometric Template Protection Systems", *IEEE Trans. on Information Forensics and Security*, vol. 3, no. 6, pp. 1406-1420, 2018**
- M. Gomez-Barrero, J. Galbally, A. Morales, J. Fierrez, "Privacy-Preserving Comparison of Variable-Length Data with Application to Biometric Template Protection", *IEEE Access*, vol. 5 (1), pp. 8606-8619, 2017
- M. Gomez-Barrero, E. Maiorana, J. Galbally, P. Campisi, J. Fierrez, "Multi-Biometric Template Protection Based on Homomorphic Encryption", *Pattern Recognition*, vol. 67, pp. 149-163, 2017
- E. Martiri, M. Gomez-Barrero, B. Yang, C. Busch, "Biometric Template Protection Based on Bloom Filters and Honey Templates", *IET Biometrics*, Vol. 6 (1), pp. 19-26, 2017
- M. Gomez-Barrero, C. Rathgeb, K. Raja, R. Raghavendra, C. Busch, "Biometric Symmetry: Implications on Template Protection", in *Proc. European Signal Processing Conference* (EUSIPCO), 2017
- M. Gomez-Barrero, C. Rathgeb, J. Galbally, C. Busch, J. Fierrez, "Unlinkable and irreversible biometric template protection based on Bloom filters", *Information Sciences*, vol. 370-371, pp. 18-32, 2016
- C. Rathgeb, M. Gomez-Barrero, C. Busch, J. Galbally and J. Fierrez, "Towards Cancelable Multi-Biometrics based on Adaptive Bloom Filters: A Case Study on Feature Level Fusion of Face and Iris", *Proc. Int. Workshop on Biometrics and Forensics, IWBF*, 2015

References

- **V. M. Patel, N. K. Ratha, R. Chellappa, "Cancelable biometrics: A review", *IEEE Signal Processing Magazine*, vol. 32, no. 5, pp. 54-65, 2015**
- **S. Rane, "Standardization of biometric template protection", *IEEE MultiMedia*, vol. 21, no. 4, pp. 94-99, 2014**
- M. Gomez-Barrero, C. Rathgeb, J. Galbally, J. Fierrez and C. Busch, "Protected Facial Biometric Templates Based on Local Gabor Patterns and Adaptive Bloom Filters", *Proc. IAPR/IEEE Int. Conf. on Pattern Recognition, ICPR*, pp. 4483-4488, Stockholm, Sweden, Aug. 2014
- **C. Aguilar-Melchor, S. Fau, C. Fontaine, G. Gogniat, R. Sirdey, "Recent advances in homomorphic encryption: A possible future for signal processing in the encrypted domain", *IEEE Signal Processing Magazine*, vol. 30, no. 2, pp. 108-117, 2013**
- A. Ross, A. Othman. "Visual Cryptography for Biometric Privacy" *IEEE Trans. on Information Forensics and Security,* vol. 6, no. 1 pp. 70-81, 2011
- **C. Rathgeb, A. Uhl, "A Survey on Biometric Cryptosystems and Cancelable Biometrics", *EURASIP Journal on Information Security*, 2011**
- M. Barni, *et al.,* "A privacy-compliant fingerprint recognition system based on homomorphic encryption and fingercode templates", *Proc. Biometrics: theory applications and systems (BTAS), 2010*.
- A. Juels, M. Sudan, "A fuzzy vault scheme", *Designs, Codes and Cryptography*, vol. 38, no. 2, pp. 237-257, 2006.
- N. Ratha, *et al.* ,"Cancelable biometrics: A case study in fingerprints", *Proc. Int. Conf. on Pattern Recognition (ICPR),* 2006.
- A. Juels, M. Wattenberg, "A fuzzy commitment scheme", *Proc. ACM Conf. on Computer and Communications Security*, 1999.