A ZERO-LEAKAGE AND INDISTINGUISHABLE AUXILIARY DATA REPRESENTATION FOR BIOMETRIC CRYPTOSYSTEMS

GABRIEL E. HINE

UNIVERSITÀ DEGLI STUDI ROMA TRE

EUROPEAN BIOMETRICS MAX SNIJDER, RESEARCH, AND INDUSTRY AWARDS 2020

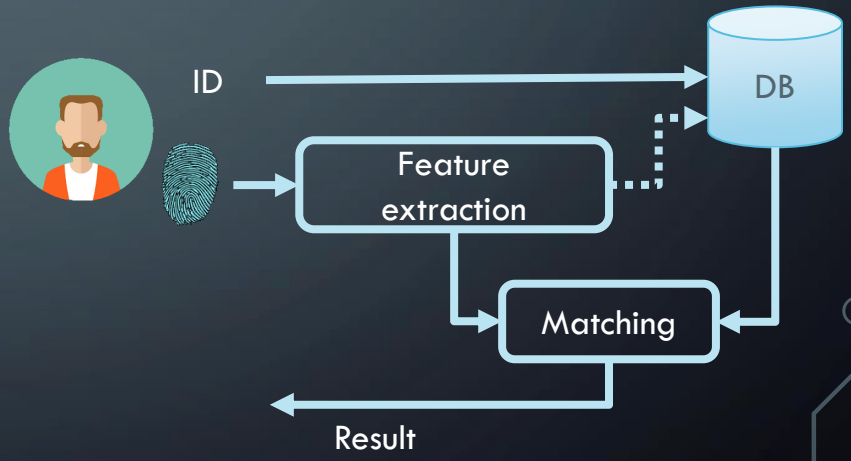
SEPTEMBER 16TH 2020

OUTLINE

- Biometric cryptosystems: requirements
- The proposed zero-leakage cryptosystem
- The enhanced unlinkable system

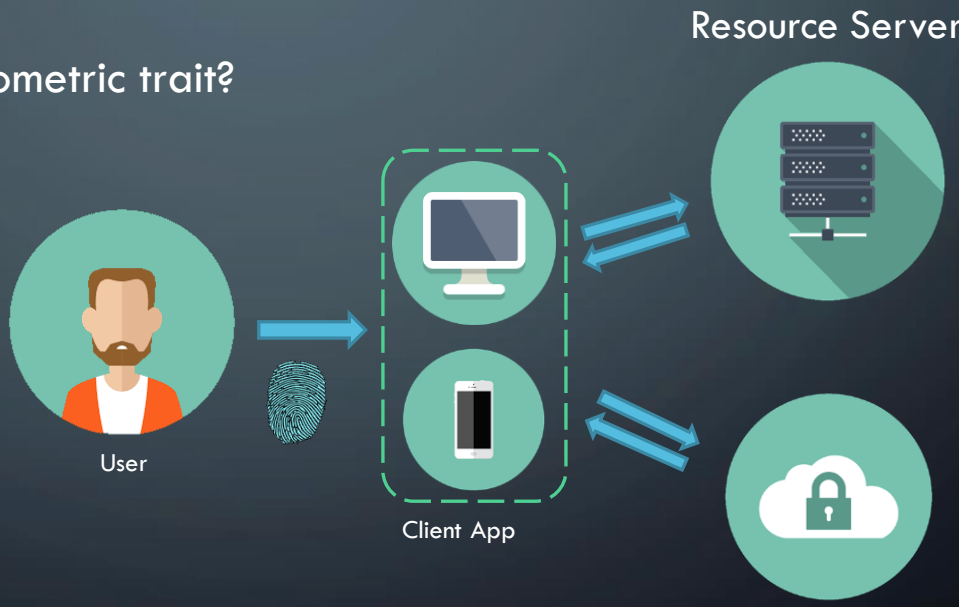
BIOMETRIC VERIFICATION SYSTEM

- Biometric trait:
  - Fingerprint, iris, palm vein, signature…
- Features template
  - Representation of the salient characteristics of the biometric trait
- The system needs to store and process the plaintext template



AUTHENTICATION ARCHITECTURES

- Where to verify the biometric trait?
  - Client side
  - Server side



CLIENT SIDE VERIFICATION

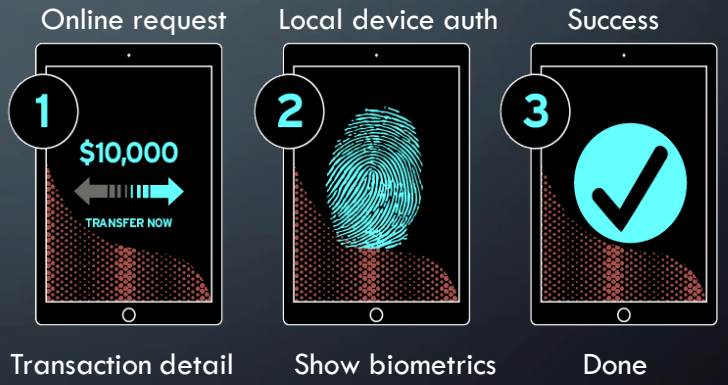
e.g.: UAF standard

- Fido alliance: Google, ING, Visa, MasterCard, Samsung…

- Biometric data never leave the user's device

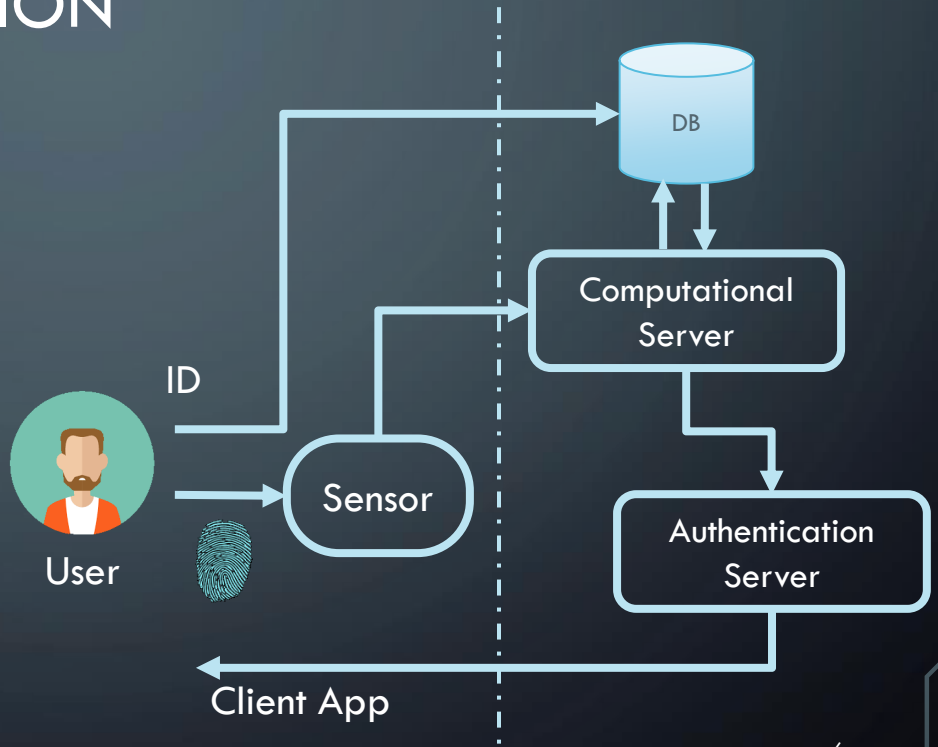
☺ Privacy ok

☹ Security:
The device may be compromised



SERVER-SIDE VERIFICATION

- Centralised Database
  - Attractive for attackers
  - Privacy threats
  - EU GDPR Regulation



EU GENERAL DATA PROTECTION REGULATION

- Biometrics are sensitive & personal data

→ Processing, storage, and transmission have many restrictions

→ Minimisation of data collection: proportionality

→ Mandatory to implement data protection 'by design and by default'

EU GENERAL DATA PROTECTION REGULATION

- Biometrics are sensitive & personal data

→ Processing, storage, and transmission have many restrictions

→ Minimisation of data collection: proportionality

→ Mandatory to implement data protection 'by design and by default'

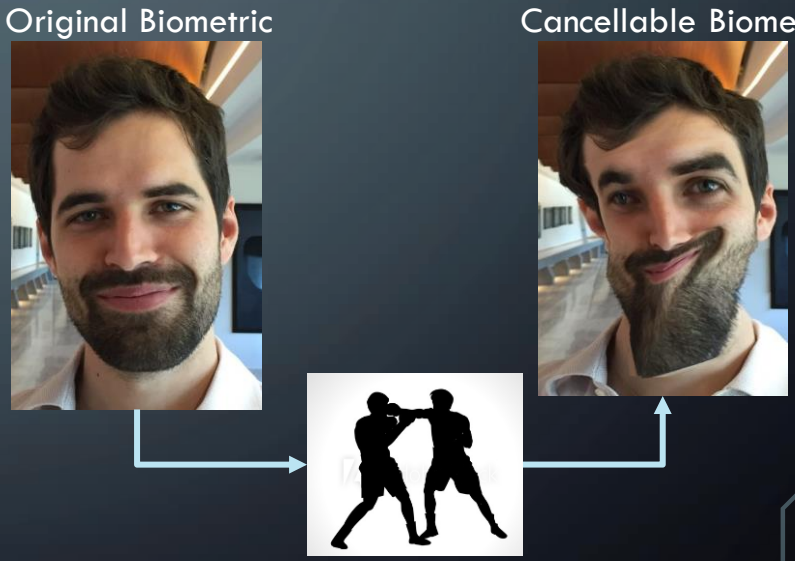
How to deal with these limitations?

PROPOSED SOLUTIONS

- Cancellable biometrics
  - “not invertible” transformation of the signal
- Biometric Cryptosystems
  - Hash of the biometric trait
- Secure computation (Homomorphic Encryption)
  - Comparison in the encrypted domain

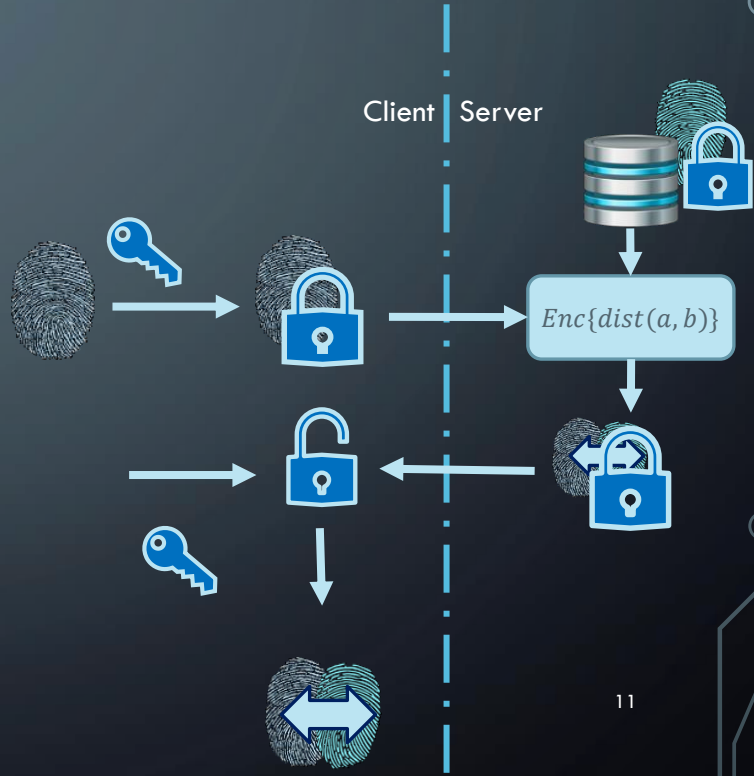
CANCELLABLE BIOMETRICS

- Parametric transformation of the biometric.
- Security analysis is weak ☹
- Bad recognition performance ☹



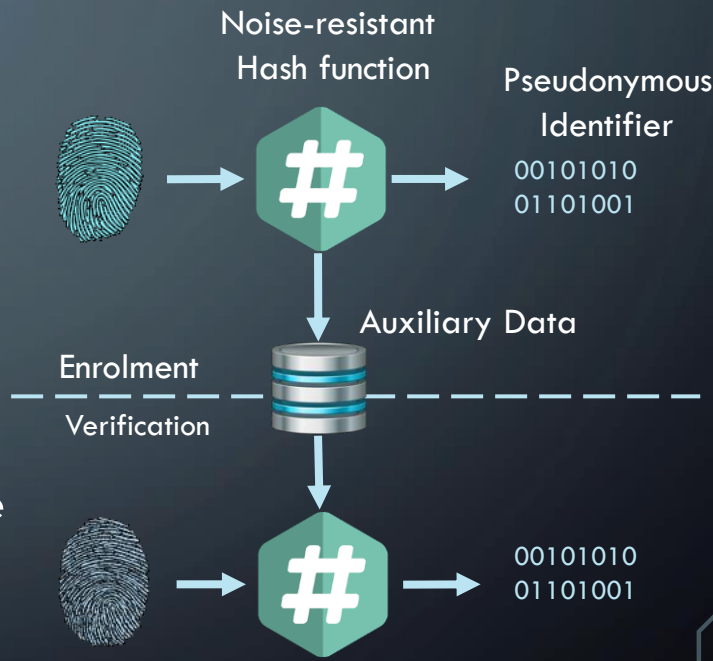
SECURE COMPUTATION

- Biometric comparison in encrypted domain
- Well known cryptographic techniques ☺
- Scalable security: ☺
  - unbounded key length
- Intrinsically double-factor ☹
  - Sometimes it is inconvenient
- Computationally expensive ☹



BIOMETRIC CRYPTOSYSTEMS

- Emulate hash-based password checking
- The biometric is mapped to a noiseless Pseudonymous Identifier
- Auxiliary Data are needed to absorb the intra-class variability



PASSWORD HASHING

USER	PASSWORD
GABRIEL	123456
PATRIZIO	ADMIN
EMANUELE	12345678
EMANUELA	ABC123
RIDVAN	111111
HIMANKA	PASSWORD
EBEN	QWERTY

USER	HASH(PSW)
GABRIEL	N2YRKL7IUFHK
PATRIZIO	0GF65INES9K0
EMANUELE	OHVNP560X0Z9
EMANUELA	BE80DIEZSPBQ
RIDVAN	KKRA32PLB5AU
HIMANKA	1E6LWP6S0FGJ
EBEN	U49V6ZPM3Y4N

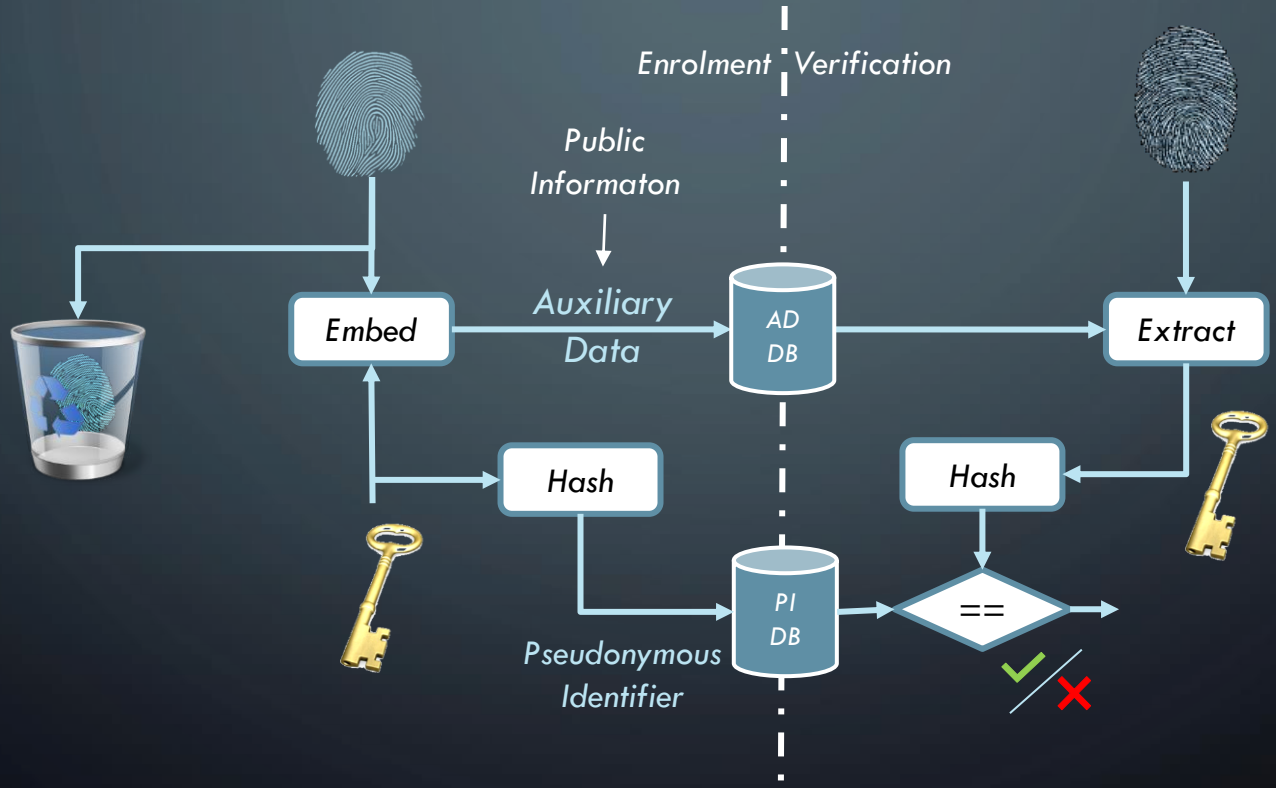
BIOMETRIC HASHING?

- What's the problem? The Noise!

```
gabriel   →  647431b5ca55b04fdf3c2fce31ef1915
gabriele  →  8bc674f8b3278ec1de6112accd643b4f
```

- By construction, the hash loses any distance metric
- Error correcting methods are needed

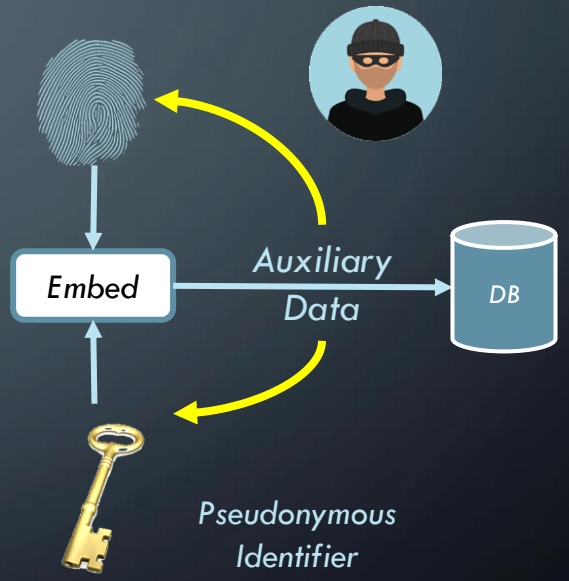
FUZZY EMBEDDERS



FUZZY EMBEDDERS REQUIREMENTS

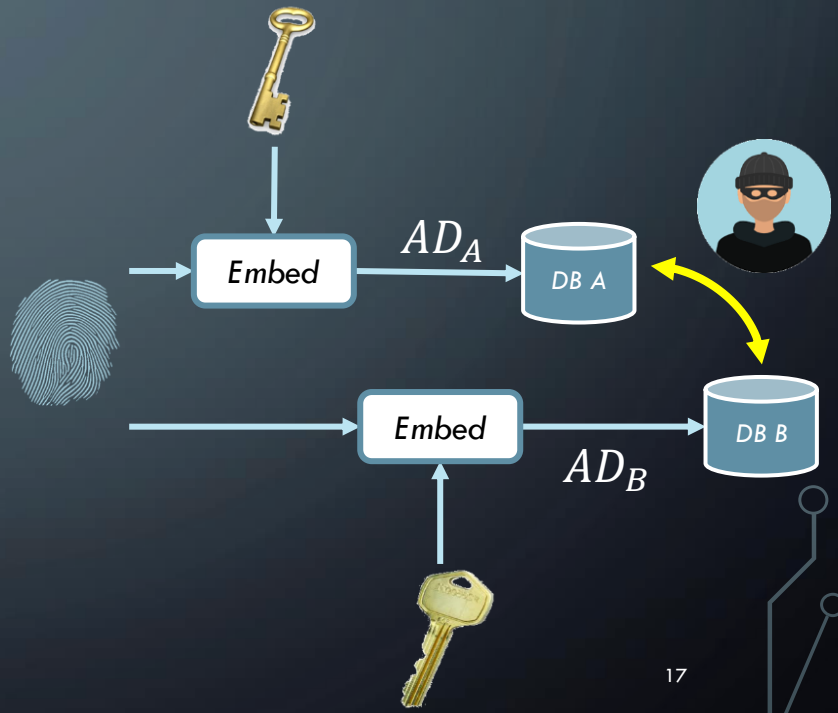
- The stored Auxiliary Data is public
- Security: AD should not leak any information about the Pseudonymous Identifier

  (Zero Leakage Cryptosystem)
- Privacy: AD should leak as little information as possible on the biometric data (some is needed to absorb noise)

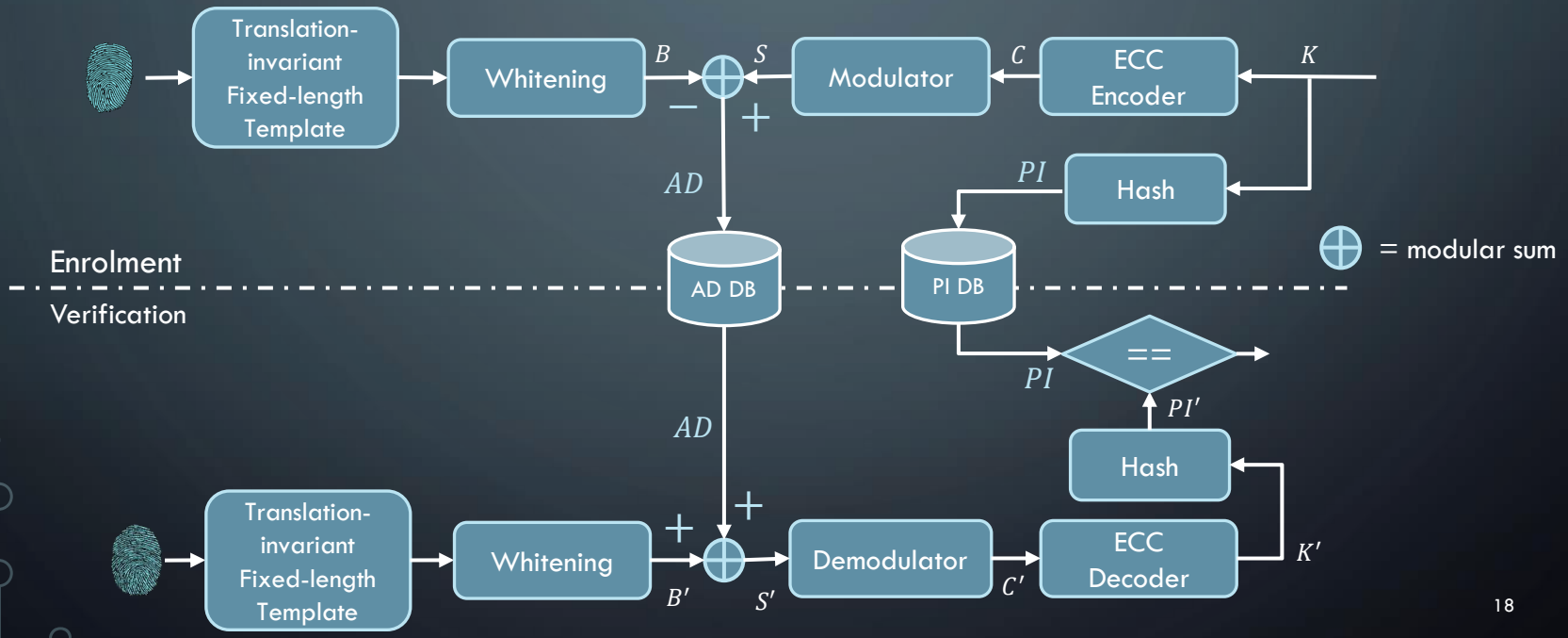


LINKABILITY

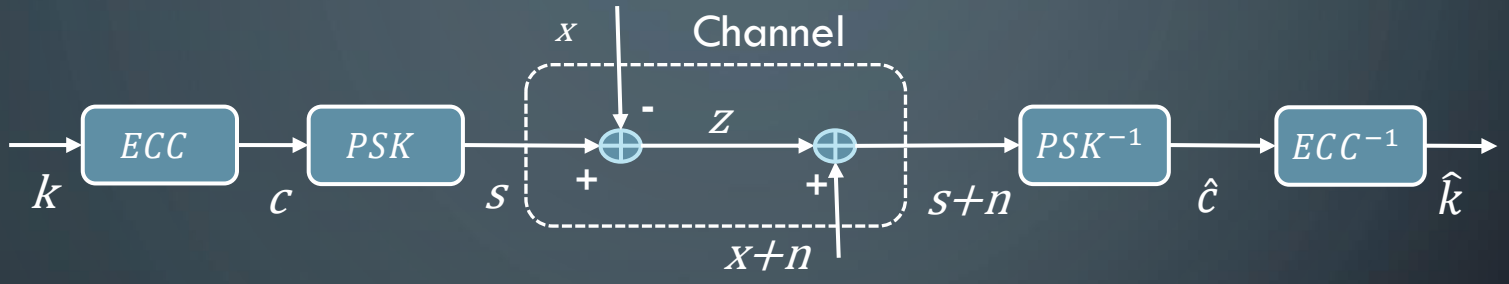
- AD has some trace of the original biometric
- The traces could be used to link ADs of the same user among different servers
- Confusion techniques are needed



CODE-OFFSET FUZZY EMBEDDER SCHEME

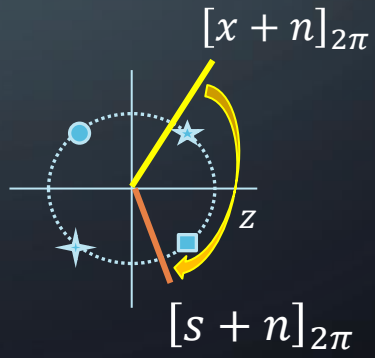


FUZZY EMBEDDER: TX/RX



Tx Symbol =

= 00
= 01
= 11
= 10



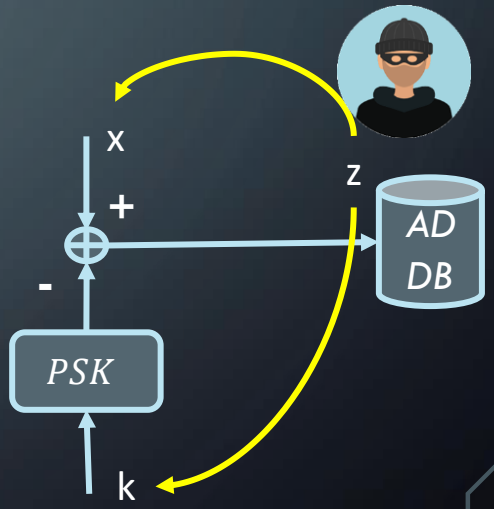
SYSTEM DESIGN REQUIREMENTS

- The Auxiliary Data Z should be considered public
- Security

  $\rightarrow I(K,Z) = 0$ (Zero Leakage Crypto-system)

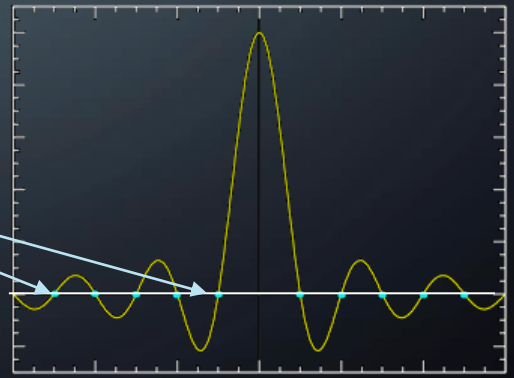
- Privacy
  - $I(X,Z) > 0$ To absorb X intra-class variability
  - Proposed privacy evaluation

$$P = \frac{E_{X,K}\{(\hat{x}(z) - x)^2\}}{E_X\{x^2\}} \rightarrow 1$$



PERFECT SECURITY (ZERO-LEAKAGE) CONDITION

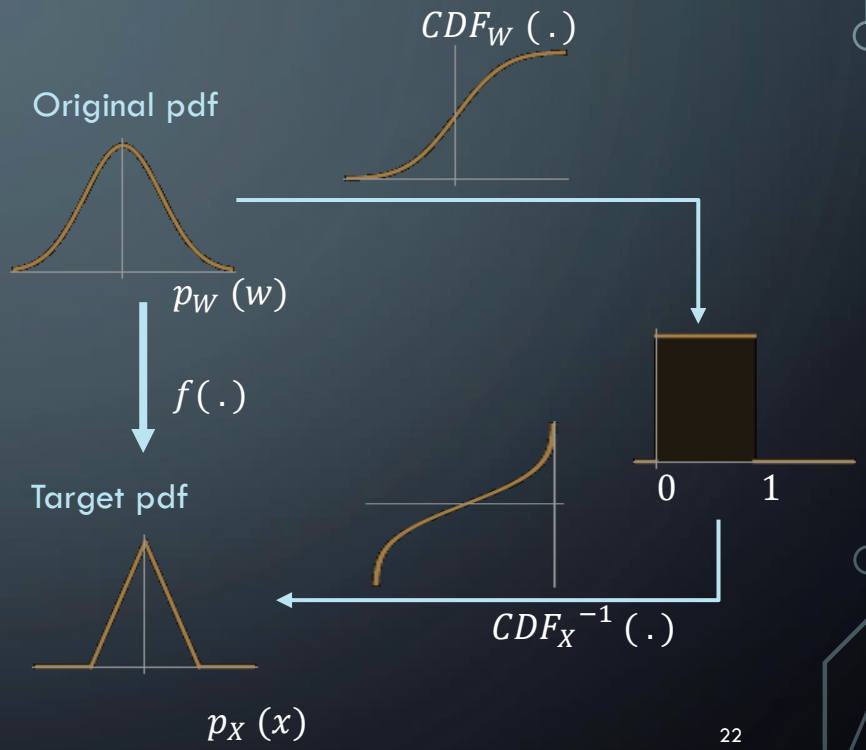
- Def: zero-leakage crypto-system: $I(K, Z) = 0$
- Zero Mutual Information between key and helper data
  - Necessary and sufficient condition: $\phi_X\left(m\frac{2\pi}{\Delta}\right) = 0,\ \mathrm{m} = \pm 1, \pm 2, \pm 3, \ldots$
    - Fourier transform of the probability density function of X should contain sinc-like zeros



PROPOSED METHOD

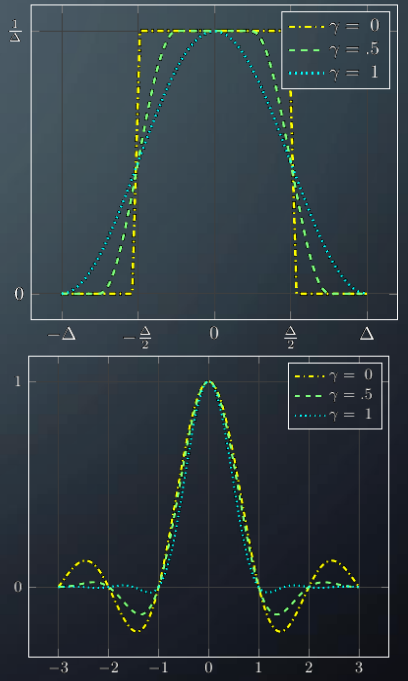
$$x = f(w) = CDF_X^{-1}[\,CDF_W(w)]$$

- No Need for Secret Keys
- Requires good knowledge of statistical distributions
  - NB: here, distributions are referred to overall statistics, not user wise statistics



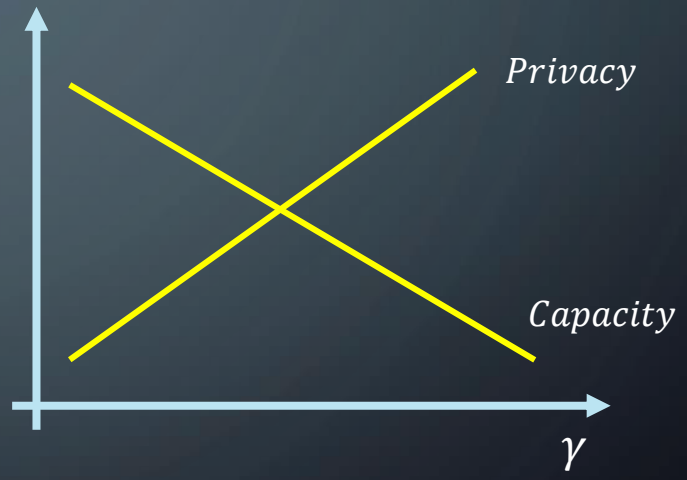
RAISED COSINE DISTRIBUTIONS FAMILY

- $\gamma = 0 \rightarrow rect_{2\pi}$
- ⋮
- $\gamma = 1 \rightarrow cos^2\, rect_{4\pi}$
- Which distribution to choose?



DESIGN PARAMETERS

- The distribution shape affects:
  - Privacy:
    - Feature vector Irreversibility
  - Embedding capacity:
    - Extraction reliability
    - Key-length → Security



PRIVACY VS CHOSEN DISTRIBUTION

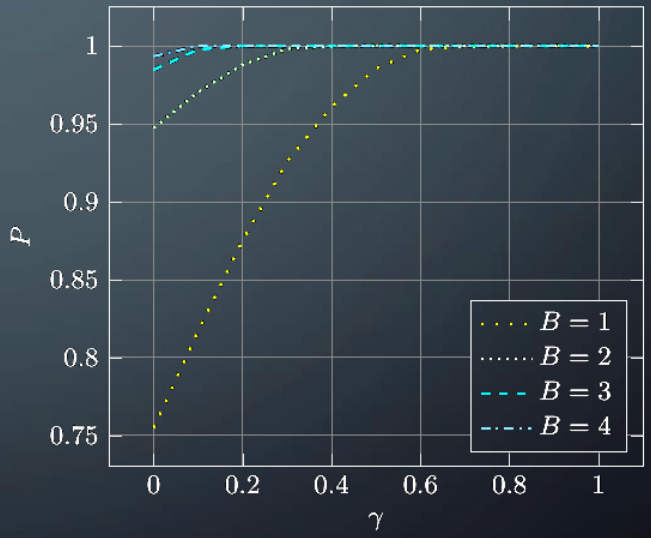
- Optimum estimator

$$\hat{x}(\mathrm{z}) = E_X\{x|z\} = \int x\, p_{X|Z}(x|z)\, dx$$

- Proposed privacy evaluation score

$$P = \frac{E_{X,K}\{(\hat{x}(z) - x)^2\}}{E_X\{x^2\}}$$

- Privacy grows with:
  - gamma parameter
  - number of (encoded) bits embedded to the coefficient



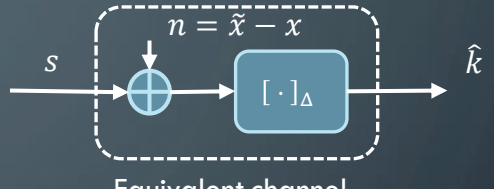
EMBEDDING CAPACITY

- Equivalent channel: PSK + phase noise
- (Shannon) Channel Capacity

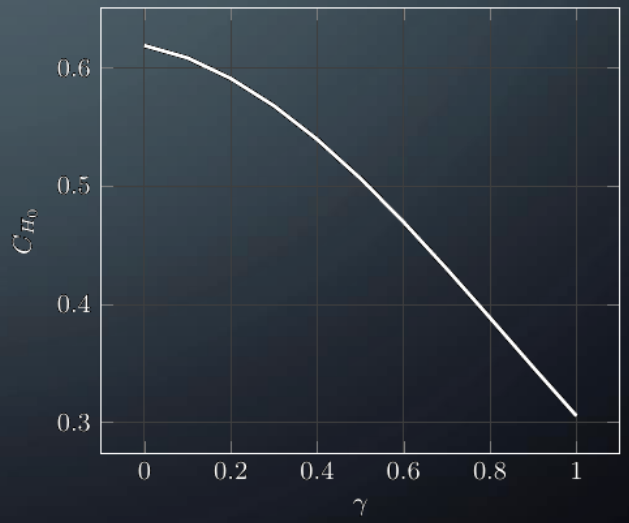
$$C = \log(2\pi) - h(N_{2\pi})$$

- Signal to Noise Ratio very unevenly distributed among coefficients due to whitening:
  - Few coefficients get the majority of the energy
  - Many coefficients are just noise.

→ Adaptive modulation

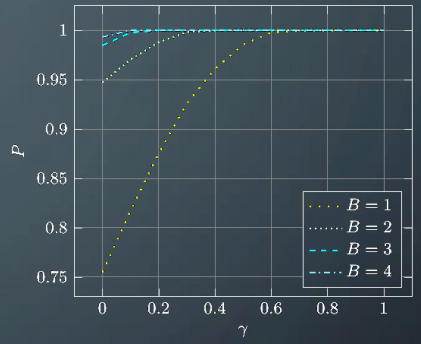


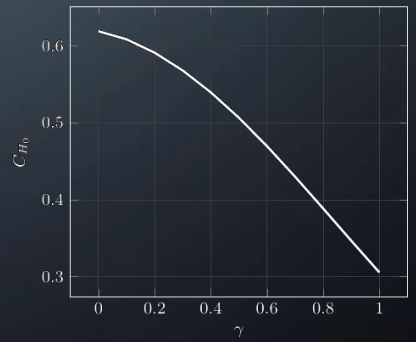
Equivalent channel



DISTRIBUTION TUNING AND ADAPTIVE MODULATION

- Distributions and number of bits to embed are tuned for each coefficient in order to maximise privacy, security, and reliability





NON-GENUINE EMBEDDING CAPACITY

- Thanks to the proposed transformation,
  in the non-genuine hypothesis, C = 0
  - Consequence of I(X,Z) = 0
- As a result, FMR = $2^{-KeyLength} \approx 0$ ☺
- FNMR increses because of the intrinsic trade-off ☹

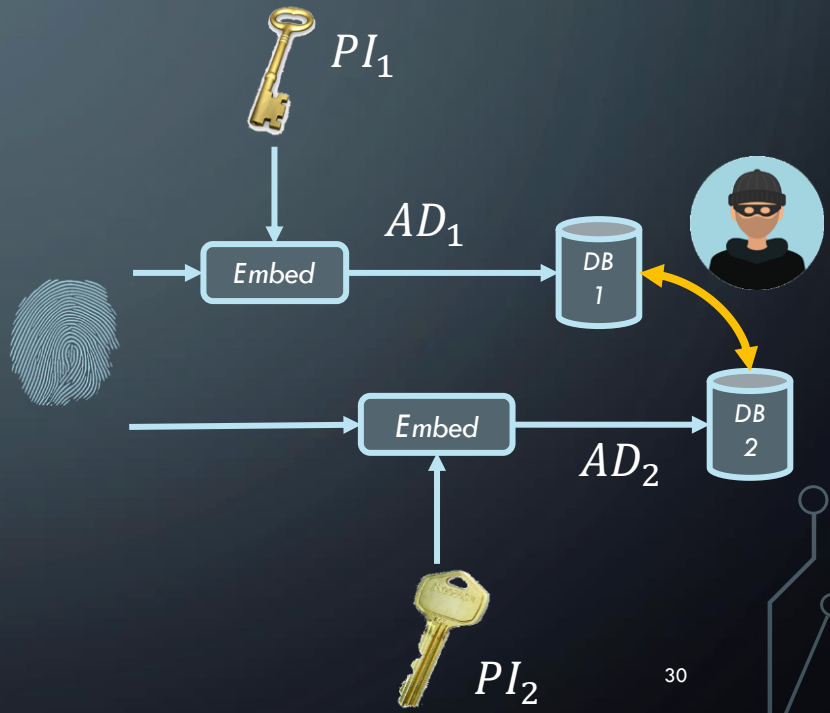
DISCUSSION

Hine et al, *A Zero Leakage Fuzzy-Embedder, From The Theoretical Formulation to Real Data* – IEEE Trans. on Information Forensics & Security

- Provable secure and privacy compliant Fuzzy Embedder
- Framework for system design is provided
  - Many operational points (in contrast with most cryptosystems)
- We haven't spoken about linkability yet (on purpose)

LINKABILITY

- Multiple systems scenario
  - Independent Pseudonymous Identifiers
- Mutual information between Auxiliary data and biometric >0
- ADs may be linked each other
- Probably the hardest issues in biometric protection



CODE-OFFSET LINKABILITY

$$AD_1 = (x_1 + k_1)_{2\pi}$$
$$AD_2 = (x_2 + k_2)_{2\pi}$$

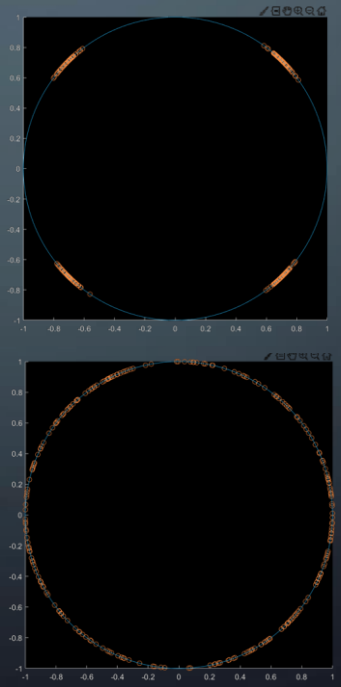
if $x_1 \approx x_2$

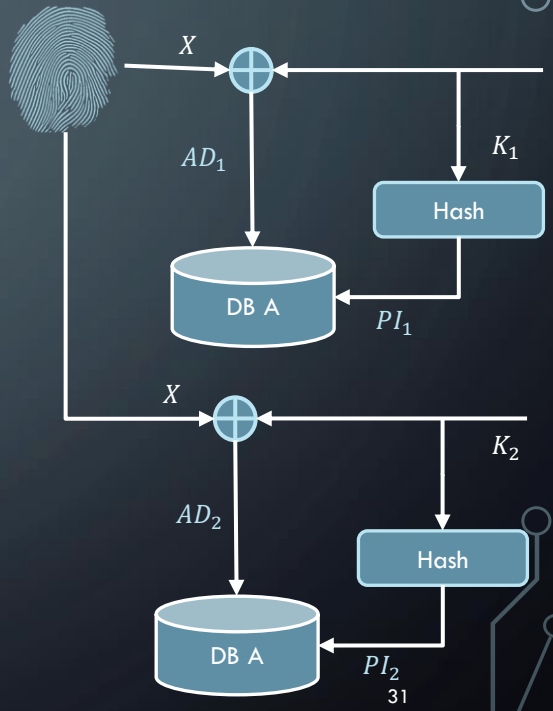
$$[AD_1 - AD_2]_{2\pi} \approx [k_1 - k_2]_{2\pi}$$

(discrete set of values)

else

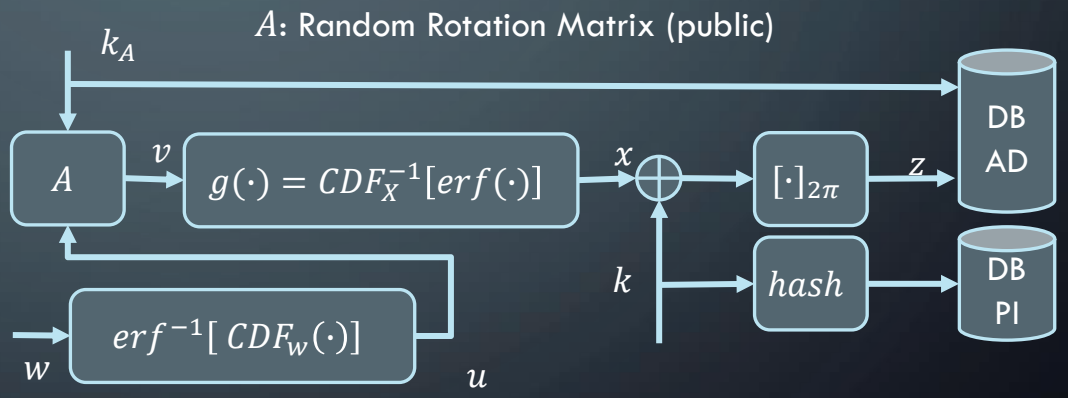
$[AD_1 - AD_2]_{2\pi}$= random in $[0, 2\pi)$





PROPOSED IMPROVED SCHEME

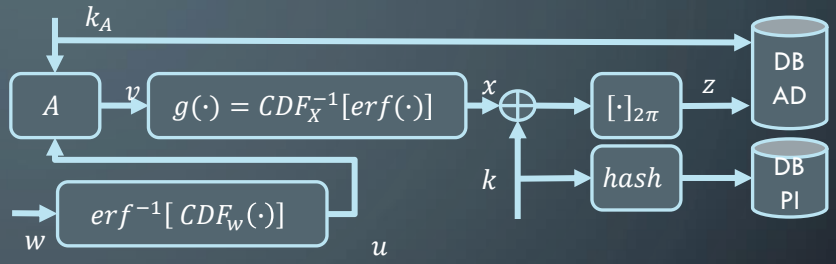
- Non-linear function of all coefficients w
  - Confusion rises
  - Linkability is not trivial
  - ...even if A is public



LINKABILITY ATTACK

Given $\{z_1, A_1\}, \{z_2, A_2\}$

if $u_1 \approx u_2$



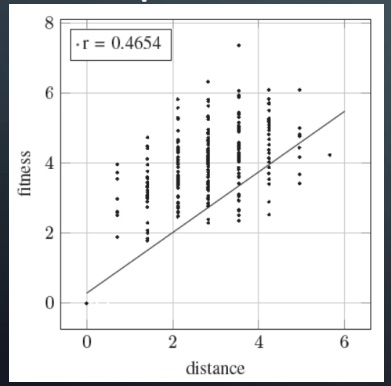
$$\left[g_2\left(A_2{A_1}^T\,{g_1}^{-1}([\,z_1+c_1]_{2\pi}-2\pi l_1)\right)-z_2\right]_{\frac{2\pi}{M}}\approx 0$$

- Non-linear non-smooth system of equations: hard to solve
  - Number of local-minima grows exponentially with the size of the template

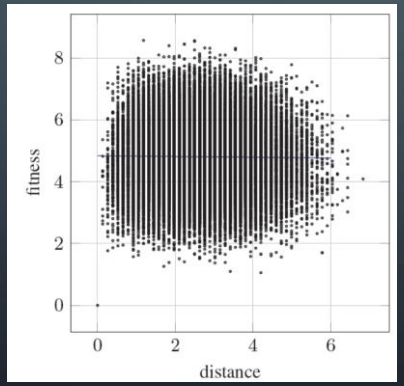
FITNESS-DISTANCE CORRELATION

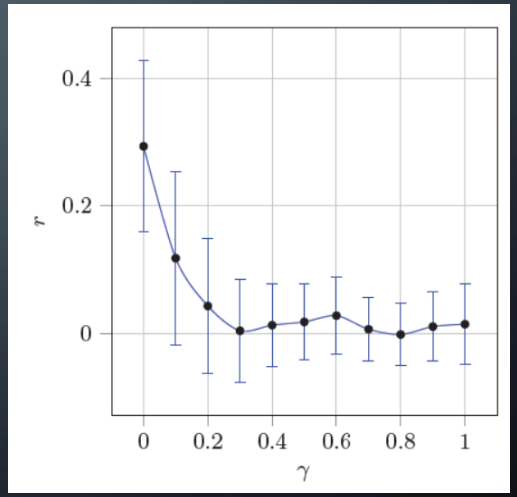
- Method to quantify the problem difficulty for genetic algorithms

$\gamma = 0$



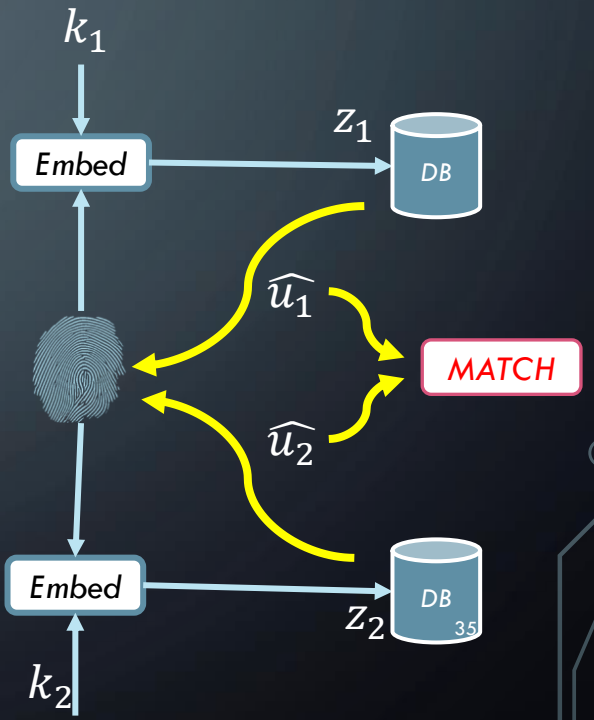
$\gamma = 1$





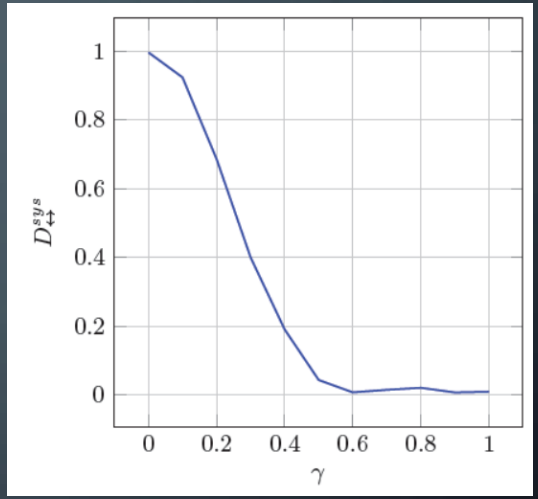
LINKABILITY: ESTIMATION APPROCH

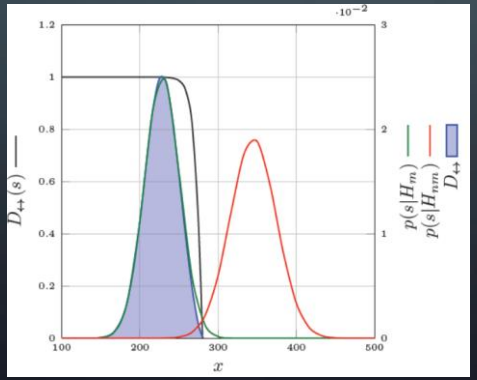
- $\widehat{u_i} = E[A_i{}^T \; g^{-1}([z_i + c_i]_{2\pi} - l_i 2\pi)$
  - (Montecarlo)
- $if\ dist(\widehat{u_1}, \widehat{u_2}) < thr$
  $\rightarrow \; user1 == user2$

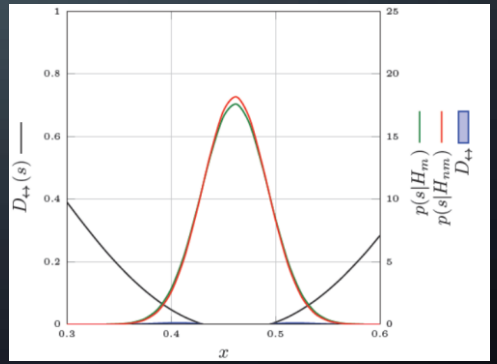


ESTIMATION ATTACK

- $D_{\leftrightarrow}^{sys}$ linkability measure by M. Gomez Barrero et.al
- Compares the distributions of a linkage function for mated and non-mated hypothesis
- $0 \leq D_{\leftrightarrow}^{sys} \leq 1$

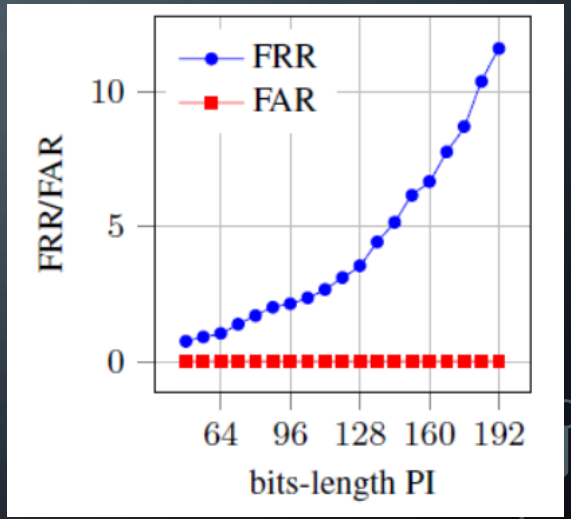






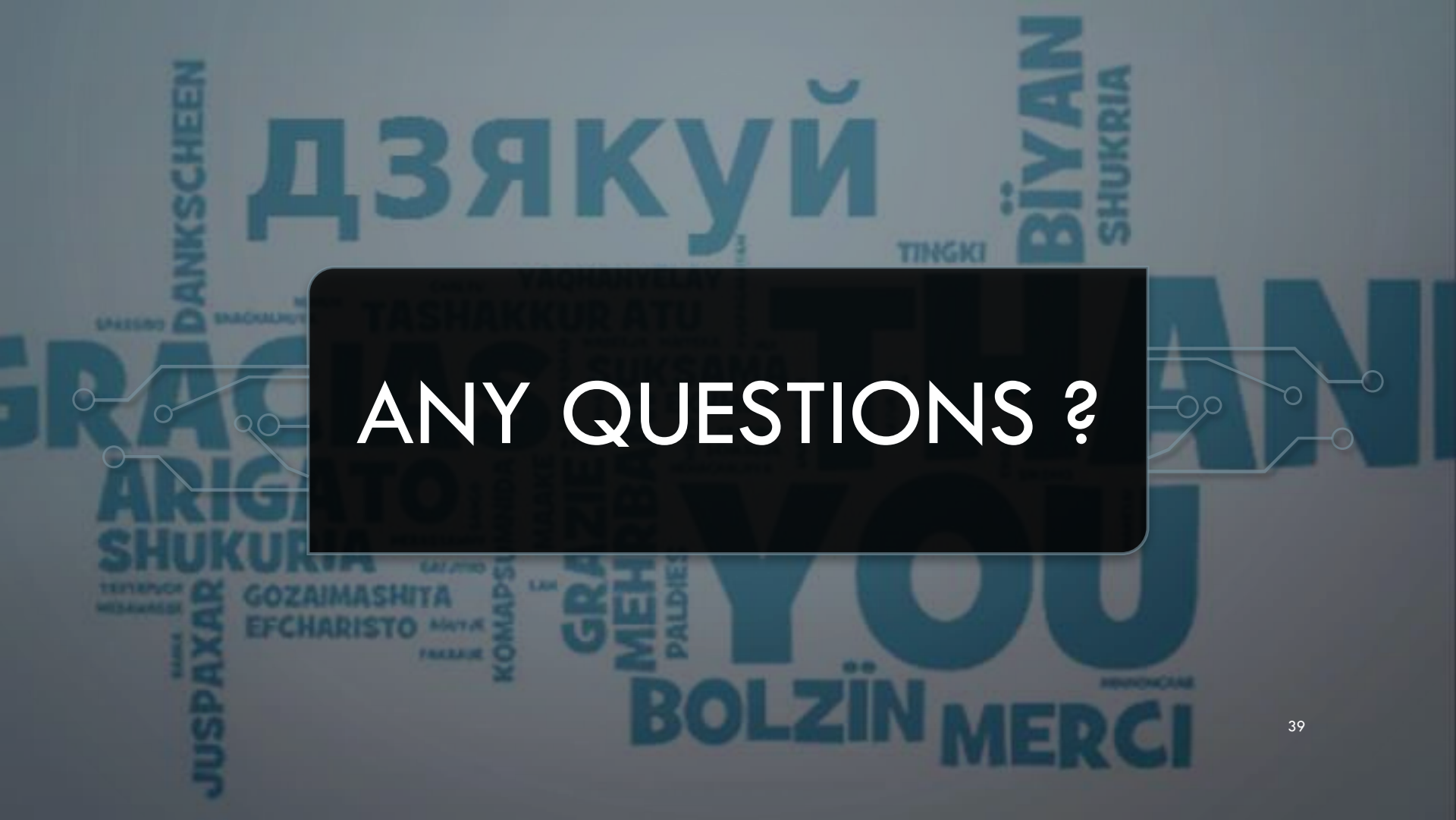
EXPERIMENTAL RESULTS

- finger-vein samples taken from the SDUMLA database
- images of index-, middle- and ring-fingers from 106 subjects.
- Modified Densenet-161 CNN
- Features whitened by means of the RICA algorithm
- 3 x 128 interdependent coefficients.
- the PI is encoded by means of turbo-codes.



DISCUSSION

- GE Hine, RS Kuzu, E Maiorana, P Campisi, “A Privacy-preserving Unlinkable CNN-based Biometric Verification System” (To be submitted)
- The unlinkability is achieved with the support of public parameters and no additionnal secret factor is required in the protocol.
- Strong theoretical demonstration of unlinkability is missing but the strength versus two different types of attacks has been evaluated.
- The experimental results on the finger-vein dataset show the feasibility of the proposed cryptosystem in a real application scenario.



ANY QUESTIONS ?