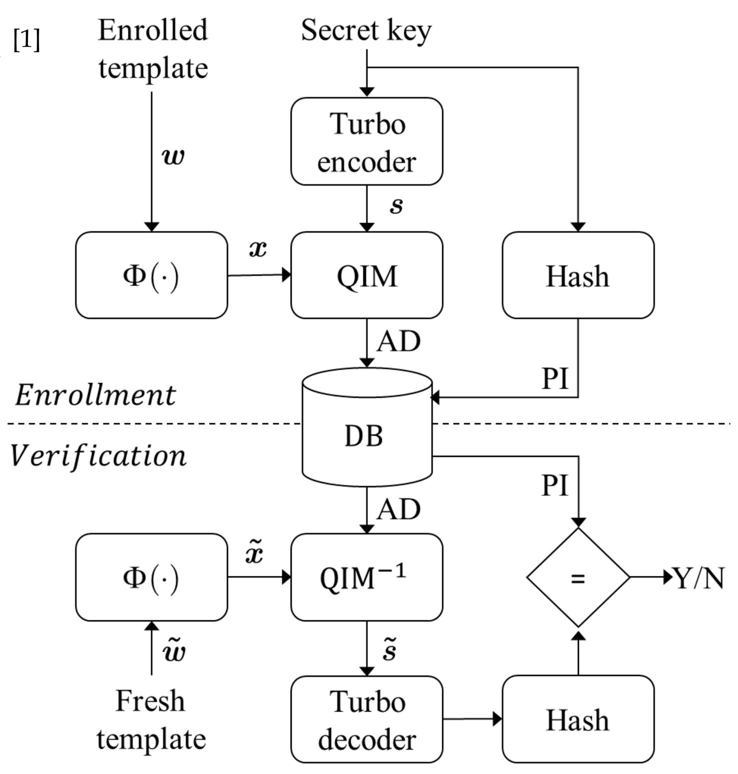
Unlinkable Zero-Leakage Biometric Cryptosystem Theoretical Evaluation and Experimental Validation

Gabriel E. Hine, *Ridvan S. Kuzu , Emanuele Maiorana, and Patrizio Campisi*

I. BACKGROUND [2]

- **Biometric Cryptosystems (BC)** aim at binding the biometric template to a **Pseudonymous Identifier (PI).** [1]
- PI: ***renewable reference*** that represents an individual.
  - It should not contain any information that allows retrieval of the original biometric sample.
- Here, we focus on ***Fuzzy Embedders:***
  - ***PI***: hash of an independent key
  - **Auxiliary Data *(AD):*** information that assists the reconstruct the ***PI*** given a verification biometric sample.
- AD is a (irreversible) function of the enrolled template and the secret key linked to the PI.



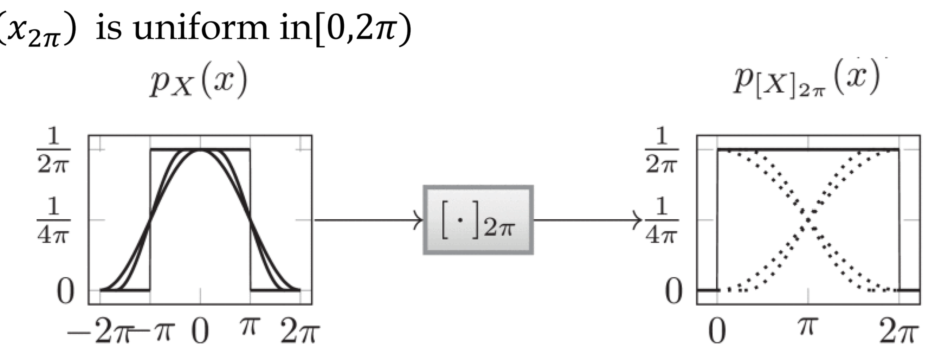
- In [2] we proposed the following ***Zero-Leakage BC:***

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

$$AD \coloneqq z \leftarrow [x+s\,]_{2\pi}$$

where $[\,\cdot\,]_{2\pi}$ is the modulo-2π operator, s is the set of symbols to be embedded encoding the *PI*, and w is the biometric template (continuous values).

- The pointwise mapping function $\Phi(w)$ maps the original (population) distribution W to a target X
- The target distribution X is chosen, and so $\Phi(\,\cdot\,)$, so that:
  1. $I(S,Z) = 0$ *i.e.*, there is no ***mutual information*** between the *AD* and the *PI;*
  2. $P \coloneqq \frac{E_{X,Z}\{[x-\hat{x}(z)]^2\}}{E_X\{x^2\}} \to 1$ *i.e.*, the best estimate of x given z is no different from the best estimate we can have a-priori. That is $z(x,s)$ is ***irreversible***.
- Condition 1. is satisfied if and only if $p_X(x_{2\pi})$ is uniform in$[0,2\pi)$
- Condition 2. is fulfilled by carefully by designing $\Phi(\,\cdot\,)$:
  - Raised Cosine distribution family.



II. THE LINKABILITY THREAT

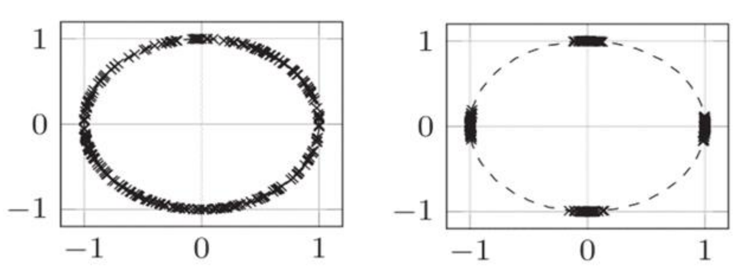
- ***Unlinkability:*** Two or more biometric references should not be linked to each other or to the subject from whom they were derived [1].
- The described Cryptosystem [2] is vulnerable to a Linkability attack.
- In fact, given a pair of instances of ***AD*** $\{z_i = [x_i + s_i\,]_{2\pi} \;\mid\; i \in (1,2)\,\}$,
  - $\Delta z = [z_1 - z_2]_{2\pi} = [(x_1 - x_2) + (s_1 - s_2)\,]_{2\pi}$;
  - If x_1 and x_2 are independent,

    $p_{\Delta Z}(\Delta z)$ is uniform in $[0,2\pi)$;
  - If x_1 and x_2 stand from the same subject,

    $|x_1 - x_2| \to 0$

    $p_{\Delta Z}(\Delta z) \;\to\; p_S(s) = \sum_k P(k)\;\delta(s - s_k)$

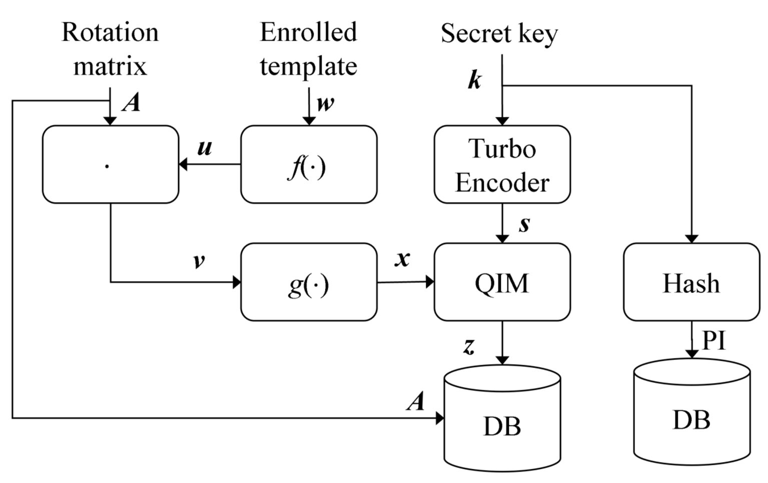
    i.e. it tends to a discrete distribution around the symbols dictionary.



- An attacker may run a simple statistical analysis of the L $z_1 - z_2$ coefficients to decide if the two ***AD*** stem from the same subjects.

III. THE SOLUTION

- Add to the chain a unitary matrix $A \mid A^T A = I$
- $f(\,\cdot\,)$ maps W to U $\sim\mathcal{N}(0, I)$
- Thanks to the rotational symmetry, also V $\sim\mathcal{N}(0, I)$
- $g(\,\cdot\,)$ maps V to the target distribution X
- See sec. I for the rest, except that now $AD \coloneqq \{z, A\}$
- Given a properly designed A matrix, it is not possible to distinguish two independent realizations $(u, u^{*}) \sim N(0, I) \times N(0, I)$ from the couple (u, Au)



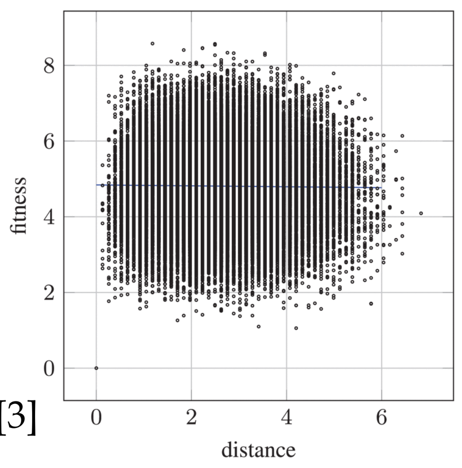
- Any randomly rotated instance is undistinguishable from an independently sampled instance.

IV. UNLINKABILITY ANALISYS

- Given two sets of ***AD*** (z_1, A_1) and (z_2, A_2), if $w_1 = w_2$, we can write:

$$\left[g\{A_2\, A_1^T\;\; g^{-1}([z_1 + s_1]_{2\pi} - 2\pi\,\xi_1)\} - \; z_2\right]_{\frac{2\,\pi}{M}} \;=\; 0$$

- This is a system of non-linear equations whose unknowns are the coefficients of s_i and ξ_i, and whose solution would demonstrate that z_1 and z_2 are linked to the same identity.
- We claim that no algorithm can solve this problem in polynomial time [3]

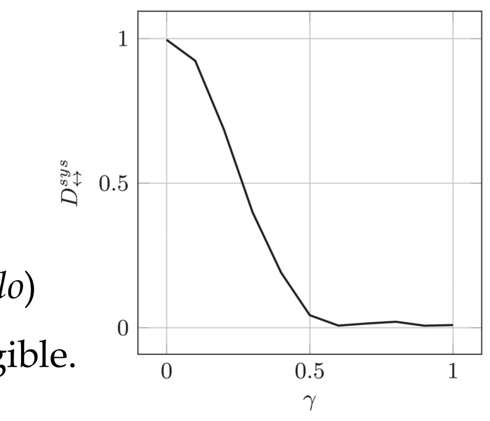


- Another linkability attack consists of estimating the templates from different ADs and trying to match them.

$$\hat{u} = E_{S,\Xi}[A^T \mathrm{g}^{-1}([z+s]_{2\pi} - 2\pi\xi)]$$

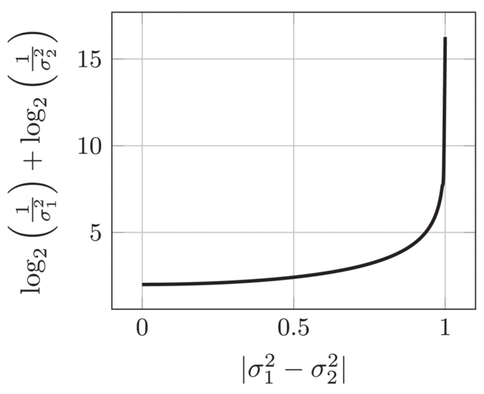
being $E_{S,\Xi}[\,\cdot\,]$ the expected value over all (s,ξ) couples.(*e.g. Monte-Carlo*)

- We showed that we can design g so that the likability [4] of $\hat{u}$ is negligible.

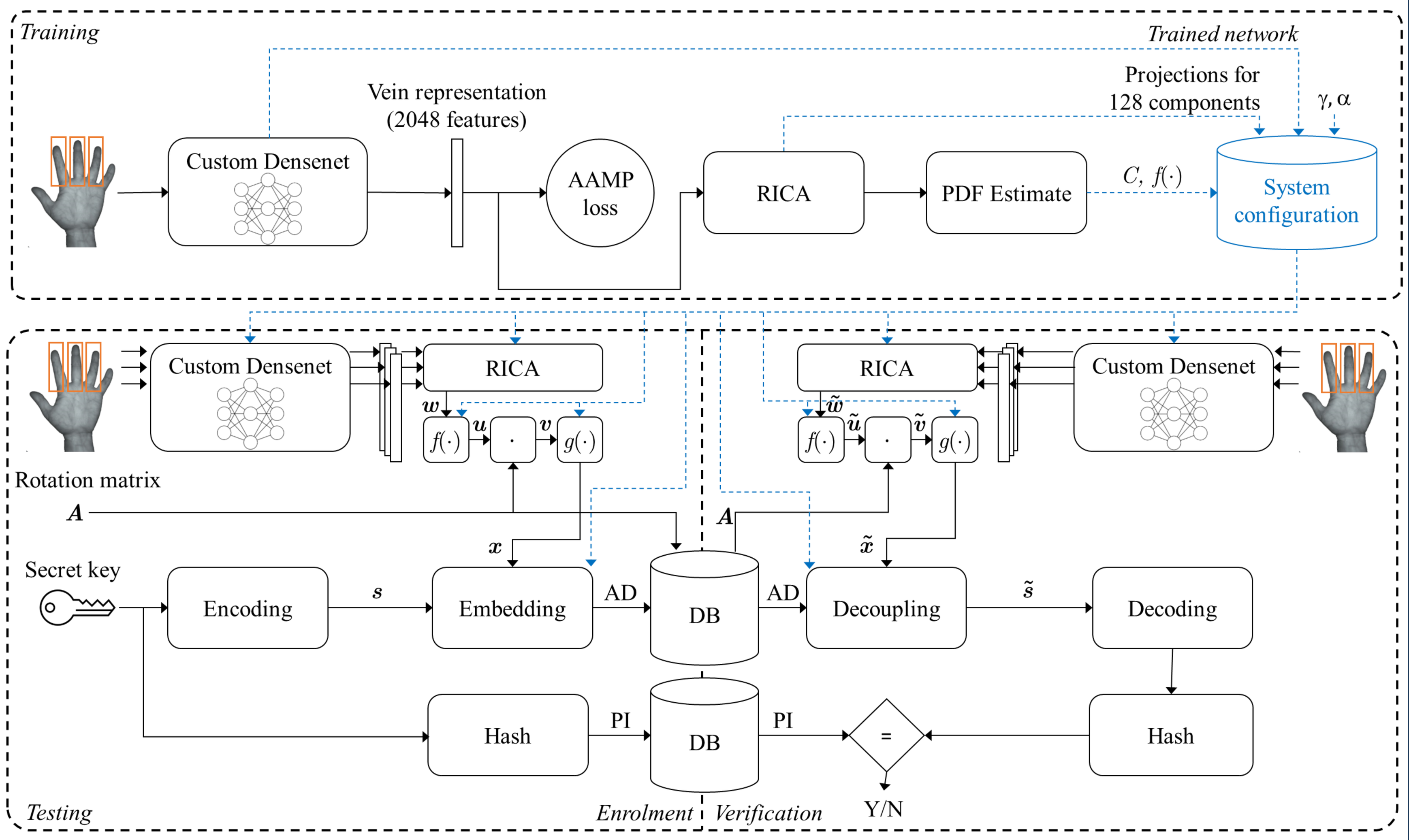


VI. DEALING WITH NON-IDEAL DATA

- The introduction of the rotation matrix degrades performance
- This is because the templates have an SNR that is unevenly distributed across features.
- Mixing good and bad features lowers the total ***Channel capacity***.
- Solution: linearly combine together only features with similar SNR

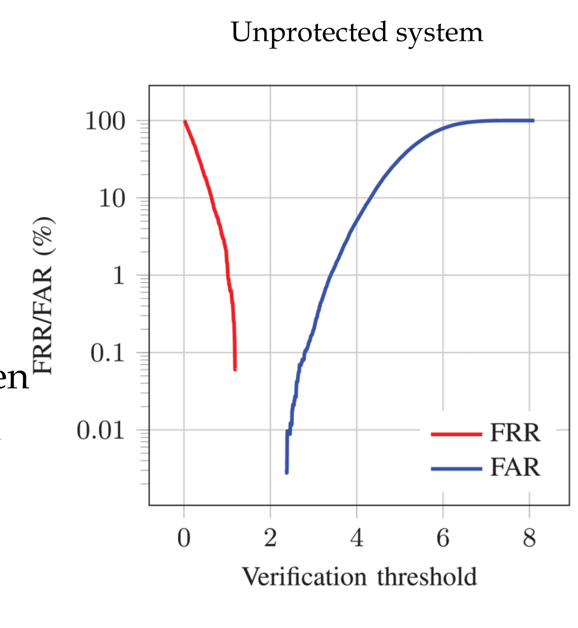


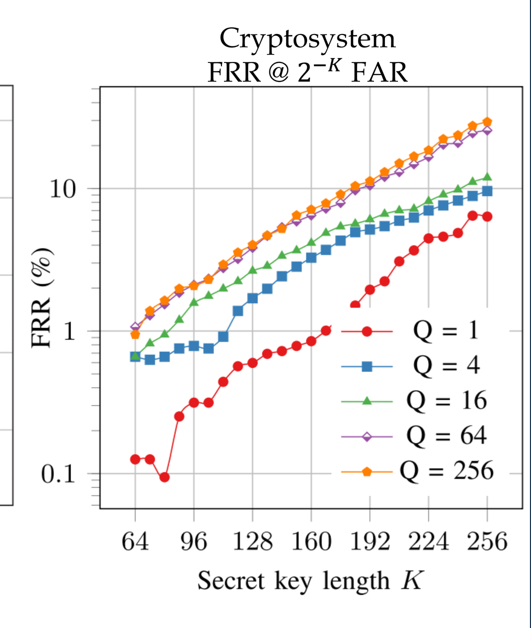
V. THE FULL SYSTEM



VI. EXPERIMENTS ON REAL DATA

- Biometric trait: 3 fingers-veins.
- Dataset: SDUMLA: L/R index-, middle- and ring-fingers of 106 subjects.
- The bandwidth Q of the rotation matrix controls the tradeoff between features reliability and unlikability
- Reasonable long secret keys have been successfully tested





REFERENCES

1. ***ISO/IEC 24745:2022*** Information security, cybersecurity and privacy protection —Biometric infonmation protection
2. ***G. E. Hine****, E. Maiorana and P. Campisi* "A Zero-Leakage Fuzzy Embedder From the Theoretical Formulation to Real Data" in TIFS 2017
3. ***G. E. Hine****, R. S. Kuzu, E. Maiorana and P. Campisi* "Unlinkable Zero-Leakage Biometric Cryptosystem: Theoretical Evaluation and Experimental Validation" in TIFS 2023
4. *M. Gomez-Barrero, J. Galbally, C. Rathgeb and C. Busch* "General Framework to Evaluate Unlinkability in Biometric Template Protection Systems" in TIFS 2017