Biowallet

Cold wallet proposal using a biometric cryptosystem

Master’s Degree in Computer Science

Pasquale Celani (1839634)

Academic Year 2024/2025

SAPIENZA
UNIVERSITÀ DI ROMA

What is Bitcoin?

- Decentralized peer-to-peer cryptocurrency designed to offer an alternative to the highly centralized traditional banking system;
- Pseudo-anonymous;
- Crypographic keys represent a user's identity
  - Managed trough apposite system called wallets.

Current wallet issues

- Currently there are various types of cryptocurrency wallets, including paper wallets, brain wallets, hardware wallets, and software wallets;
- Main challenge with wallets is determining the most secure way to store private keys.
  - **If a private key is lost or stolen the associated funds are permanently inaccessible, with no way to recover them**.

⚠ All current wallets rely either on physical objects, which can be lost or stolen, or on mnemonic phrases, which can be forgotten.

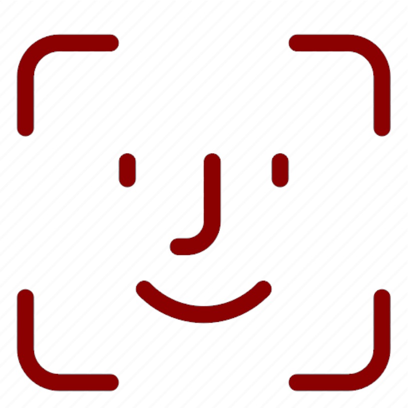
A biometric based solution proposal

- Leverage biometric **facial recognition** to generate and recover cryptographic keys eliminating the need to remember passwords or codes;
- How?
    - Adoption of a biometric key-based cryptographic model referred to as **Key Release** (Prabhjot Kaur et al);
    - In this key construction mechanism, the biometric template and key is stored together as two separate products, and the key is released upon successful **verification**.
    - why?
        - Allows handling of multiple keys which is an essential feature for wallets;
        - Shows lower susceptibility to PIE-A interclass variation from thresholding, especially relative to template-based key methods.

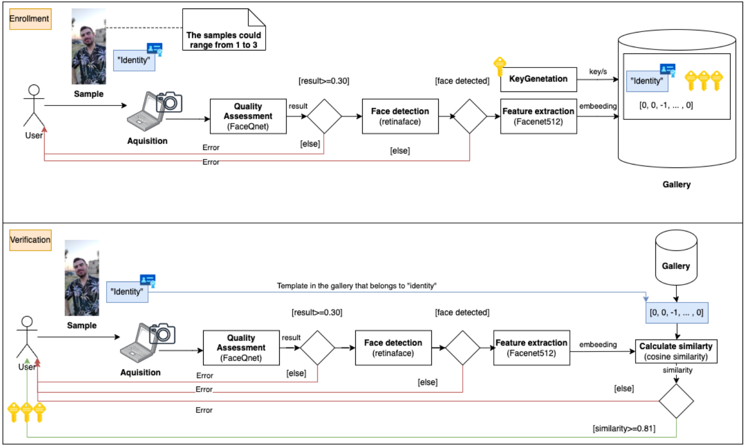
Why facial recognition was chosen

- High universality, collectability and most importantly its strong user **acceptability**;
  - Acceptability is critical, users must feel safe entrusting their funds.;

While affected by PIE-A variations, facial recognition suits our **controlled, user cooperative context**.



Enrollment & Verfication overview



Enrollment

The samples could range from 1 to 3

"Identity"

Sample

User

Aquisition

Quality Assessment (FaceQnet)

result

[result>=0.30]

[else]

Error

Error

Face detection (retinaface)

[face detected]

[else]

KeyGenetation

key/s

Feature extraction (Facenet512)

embeeding

"Identity"

[0, 0, -1, ... , 0]

Gallery

Verification

"Identity"

Sample

User

Aquisition

Template in the gallery that belongs to "identity"

Gallery

[0, 0, -1, ... , 0]

Quality Assessment (FaceQnet)

result

[result>=0.30]

[else]

Error

Error

Error

Face detection (retinaface)

[face detected]

[else]

Feature extraction (Facenet512)

embeeding

Calculate similarity (cosine similarity)

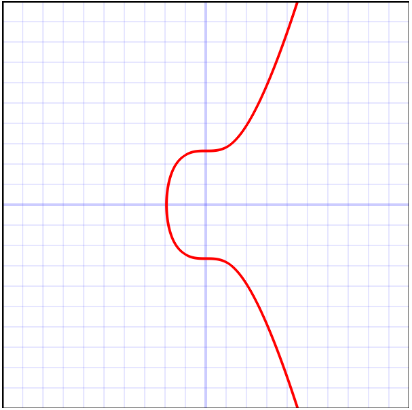
similarity

[else]

[similarity>=0.81]

ECC secp256k1 elliptic curve

- Elliptic Curve is a curve given by an equation of the form $y^2 = x^3 + Ax + B$
  - Secp256k1 is an elliptic curve of the following form $y^2 = x^3 + 7$.
- Public and private keys:
  - A **private key** is defined as $S_k \in \{d \in \mathbb{N} : d \geq 1 \wedge d < n - 1\}$ and is randomly chosen.
  - A **public key** is computed through scalar multiplication $P_k = S_k \cdot G \mod p$ where $G = (G_x, G_y)$ is the generator point.



Biometric quality assessment through FaceQNet (1)

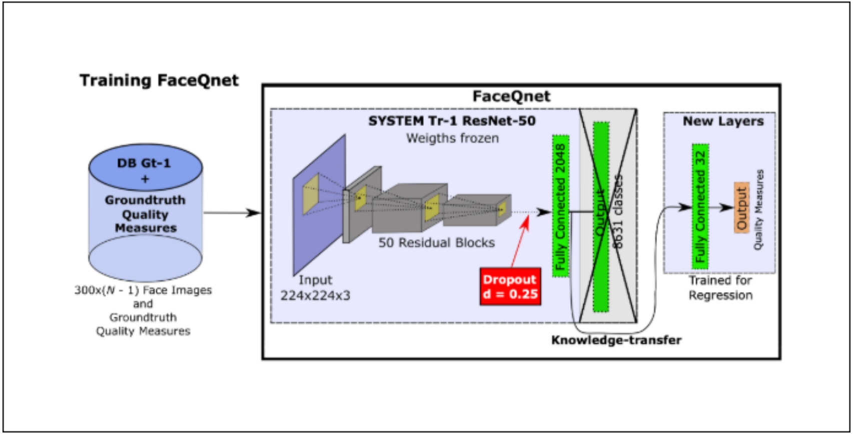
- FaceQNet v1 is a deep learning model that assigns a quality score to an image sample predicting face recognition accuracy;
- How the groundtruth is generated?
  1. Use an automated tool such as the BioLab framework to automatically label the dateset images with their quality.
     - BioLab outputs 30 ICAO test scores (0–100). A weighted subset for face relevant qualities such as blur, illumination, roll/pitch/yaw levels and glasses forms a final averaged global ICAO compliance value.
  2. The used dataset is VGGFace 2;
  3. High-ICAO images are used as references and the other as probes. Probe images are given to FaceNet, DeepSight, and Dlib to get their embeddings. Euclidean distances (converted to similarity scores) quantify quality, lower is the distance implies higher quality. Scores are averaged to create ground truth for training FaceQNet v1.

Biometric quality assessment through FaceQNet (2)

Following the generation of ground truth for the training dateset the subsequent step is the training of the Deep Regression Model to perform an end-to-end regression for quality estimation.

- **Knowledge transfer:** Use a pretrained CNN based on a ResNet-50 on VGGFace2 to do face recognition and then applying knowledge-transfer to change its domain from face recognition to quality assessment;
- **Training:** During training, original ResNet-50 weights are frozen and the last clasification layer is replaced with two FC layers:
    1. 32-dimensional quality embedding;
    2. performs regression;
- **Output:** Takes a face image as input and outputs a quality score ranging from 0 to 1.

FaceQNet Overview



Key aspects of the system evaluation methodology

- **Models:** The models employed in this project include FaceNet, Dlib, DeepIP, RetinaFace, and FaceQNet, with only FaceNet, RetinaFace, and FaceQNet used in the final implementation. These models were used in their **pre-trained** form, where:
  - **RetinaFace:** is trained on the WIDER FACE dataset;
  - **FaceQnet:** is trained on the VGGFace2 dataset;
  - **FaceNet:** is trained on the VGGFace2 and CASIA-WebFace datasets;
  - **DLib:** is trained on the VGGFace2 and FaceScrub datasets;
  - **DeepID:** is trained on the YouTubeFaces dataset.
- **Test datasets:** The **AT&T** dataset along with a custom built dataset (**DB-1**) are used to perform a cross-dataset evaluation.

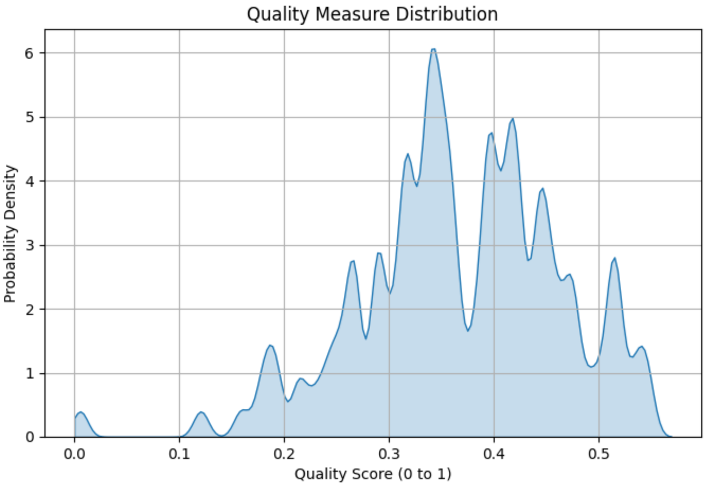
The DB-1 Dataset (1)

- **AT&T limitations:**
  - Complete absence of color informations;
  - low resolution and no background or upper body context;
- **DB-1:**
  - Designed specifically for evaluating biometric cryptosystems in the context of cryptocurrency wallets;
  - It consists of 143 images across 30 subjects with each subject represented by 2 to 5 images;

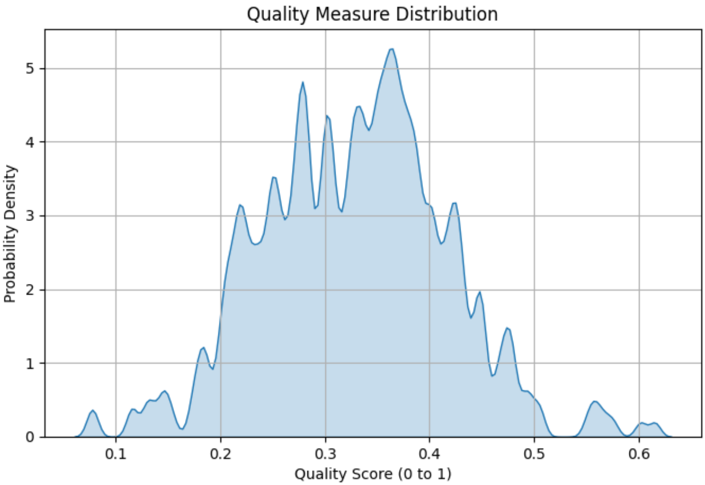
The DB-1 Dataset (2)

- **DB-1 Characteristics:**
  - Images taken with varied indoor illumination, poses, and expressions;
  - Images taken from older and recent videos to capture aging;
  - Demographic representation overview:
    - 60% males and 40% females;
    - ages range approximately from 18 to 60 years old;
    - The dataset includes men and women from diverse backgrounds like Caucasian, African and Asian.

DB-1 FaceQNet quality distribution



AT&T FaceQNet quality distribution



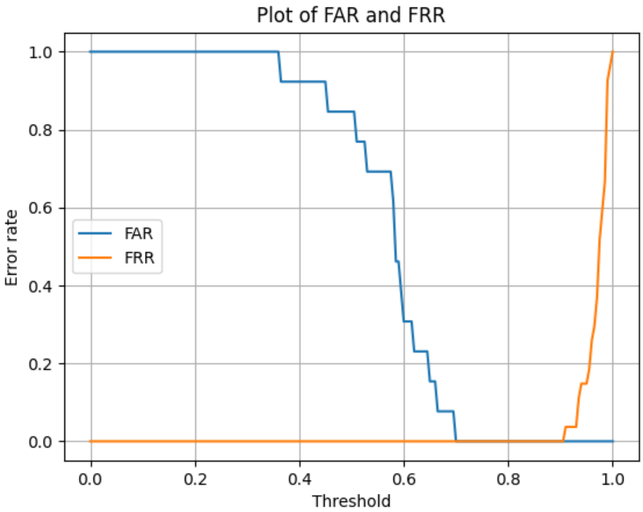
Evaluation methodology (1)

1. Subdivide the dataset test set into probes and gallery.
   - For each subject in the dataset one image sample is randomly selected to serve as the probe and the others to serve as the gallery in the verification context. This random selection is intended to minimize selection bias;
   - The next step involves computing similarities in a $PROBE - against - ALL_{GALLERY}$ context.
2. Define the role of the probe in the verification process. Specifically, whether it should act as a genuine attempt or an impostor.
   - Each similarity matrix is evaluated under different distributions of genuine and impostor attempts. (30-70, 30-70, 50-50, 100-0 and 0-100 of genuine-impostor distributions)

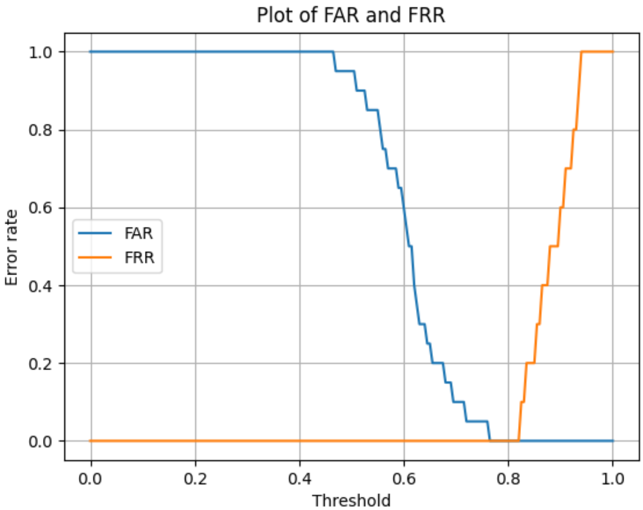
Evaluation methodology (2)

3. For each impostor/genuine setup and for every similarity matrix the actual verification benchmarking is performed. In this process, each probe is compared against the gallery templates of the claimed identity. If multiple templates are present for a given identity, the one with the highest similarity score is selected. Then, for each threshold value t ranging from 0 to 1 in steps of 0.005, the following evaluation is made:
   - $max\_similarity \geq t \land identity = impostor \Rightarrow FA + 1$ (False Acceptance);
   - $max\_similarity < t \land identity = genuine \Rightarrow FR + 1$ (False Rejection);
4. To evaluate the system's performance for each threshold t the False Re- jection Rate (FRR) and False Acceptance Rate (FAR) are computed as follows:
   - $FRR = \frac{FR}{|Genuinie|}$, where $|Genuine|$ is the number of genuine probes attempts;
   - $FAR = \frac{FA}{|Impostors|}$, where $|Impostors|$ is the number of impostors probes attempts;

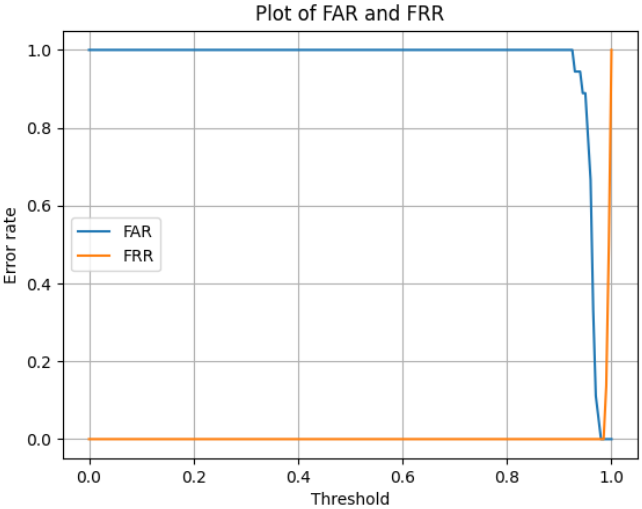
FRR-FAR 50-50 Facenet512 in AT&T



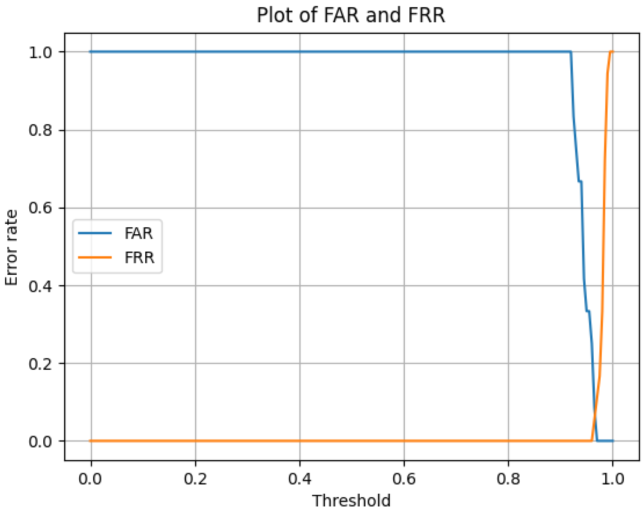
FRR-FAR 50-50 Facenet512 in DB-1



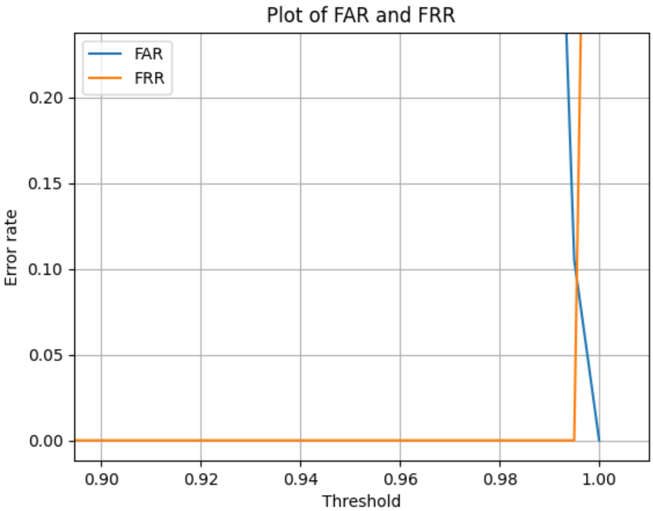
FRR-FAR 50-50 Dlib in AT&T



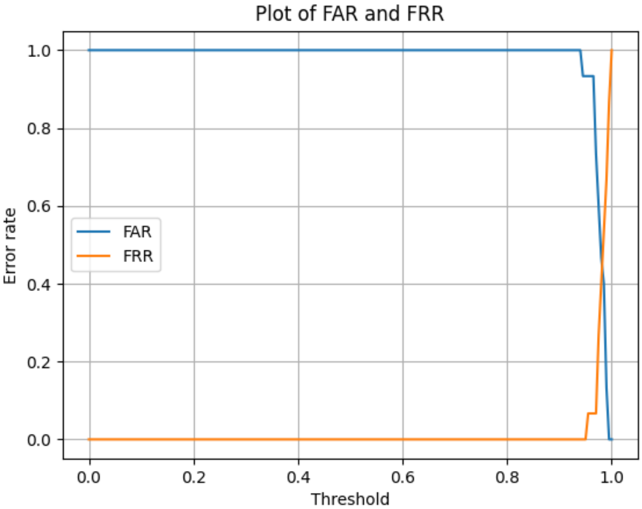
FRR-FAR 50-50 Dlib in DB-1



FRR-FAR 50-50 DeepID in AT&T



FRR-FAR 50-50 DeeeID in DB-1



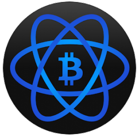
Additional noteworthy result

- Facenet512 combined with RetinaFace consistently achieved an ERR of 0 across all tested impostor to genuine distributions in both the AT&T and DB-1 datasets.
- DeepID achieved an ERR of 0 on the AT&T dataset when the probability of an impostor claim was set to 70%.
- Dlib consistently achieves an ERR of 0 across all scenarios on the AT&T dataset. In other impostor genuine claim distributions it also maintains an ERR of 0, except, in the case where 70% of the claims are impostors, there it achieves a FAR of 0 but with a FRR of 0.375.

Conclusions

- Facenet512 with RetinaFace proved best for verification, achieving 0 EER across all scenarios and datasets AT&T and DB-1.

- Looking ahead, would be good to integrate it with actual blockchain technologies to provide a fully functional cryptocurrency wallet.



Thank you for listening!

Any questions?