A communication paradigm for biometrics security and privacy

Emanuele Maiorana, Patrizio Campisi

emanuele.maiorana@uniroma3.it

Department of Engineering, University Roma Tre
Via Volterra 62, 00146, Rome, Italy
Biometric Recognition Systems

• Automatic systems performing people recognition using their physical or behavioral characteristics

– applications
  ▪ law enforcement
    o criminal identification
  ▪ access control
    o physical or logical

– advantages
  ▪ cannot be lost or stolen
  ▪ improved security

– desired properties
  ▪ performance
  ▪ circumvention
  ▪ acceptability
Main concerns affecting users’ acceptability

- Security
  - If compromised, biometrics cannot be replaced

- Privacy
  - Misuse of data (function creep)

Need for securing the employed biometric templates

- Cryptography
  - Templates vulnerable during authentication

- Hashing
  - Not error-tolerant, biometrics can vary

- Template protection schemes
  - Templates modified for
    - Providing protection
    - Providing renewability
    - Guaranteeing performance
Most employed biometric protection scheme

- binarization of $x \rightarrow$ binary template $b$
- error correcting code $c$ XORed with $b$ to manage variability
- storage of helper data $v$

Issues

- binding process biometrics/keys limited to the XOR
- binary block codes have low error correction capability (ECC)
  - high False Rejection Rates (FRRs) due to biometrics variability
  - low security, limited by the length $k$ of the binary key $m$
• General construction
  – \( c \): codeword selected from a code \( C \)
  – \( x \): biometric template
  – \( v = f(x, c) \): code-offset
    ▪ example: \( v = f(x, c) = x + c \)
    ▪ should not reveal information on neither \( x \) nor \( c \)
  – \( \tilde{c} = g(\tilde{x}, v) \): revert binding operation
    ▪ user recognized iff \( \tilde{x} \approx x \rightarrow \tilde{c} \approx c \)
• Modulation
  – \( \mathbf{c} \): symbol selected from a constellation (ex.: QAM, PSK)
  – \( \mathbf{n} \): noise added by the channel
    ▪ \( \mathbf{n} = \mathbf{x} - \hat{\mathbf{x}} \) in the code-offset scheme
  – \( \hat{\mathbf{c}} \): received corrupted symbol to be demodulated

• Proposed approach
  – use of a modulation-like scheme for FC generalization
    ▪ improvement in both verification rates and security
• Characteristics
  – use of turbo-codes for managing intra-class variability
  – codes modulated into $s$ symbols of an $L$-points constellation
  – binding expressed through a generic function $f(\cdot)$
    ▪ values in $x$ may belong to an alphabet different than $c$’s one
    ▪ added noise characterized as $n = \tilde{c} - c = g(\tilde{x}, f(x, c)) - f(x, c)$
  – joint demodulation and decoding of $\tilde{c} = g(\tilde{x}, v)$
    ▪ use of turbo-codes in soft-decoding modality
    ▪ ECC improved without user-specific information (privacy leakage)
Practical Implementation

- Constellation
  - Phase Shift Keying (PSK) modulation
    - symbols in \( c \) as \( L \) possible points in a complex circle

- Binding function
  - hard to say which symbols may generate a given value in \( v \)
    - quantization of each element of \( x \) in \( D \) possible values (\( D > L \))
    - linear mapping of the values to the interval \([-\pi; \pi]\)
    - binding
      \[
      \begin{align*}
      v &= f(x,c) = c \cdot e^{ix} \\
      \tilde{c} &= g(\tilde{x}, v) = v \cdot e^{-i\tilde{x}}
      \end{align*}
      \]
• Measured through the entropy $H(x/v)$
  – knowing $v$, having $m$ directly provides $x$: $H(m/v) = H(x/v)$
    ▪ using a $(n, k)$ turbo code, if $v$ is known $x$ can be recovered from only $z = \frac{k}{\log_2 L}$ symbols (instead of $s = \frac{n}{\log_2 L}$)
    ▪ $H(x^z|v)$ estimated with second-order dependency approximation

$$\hat{P}(x^z) = \prod_{i=1}^{z} P\left(x_{u_i}^z|x_{t(u_i)}^z\right), \ 1 \leq t(u_i) < u_i, \ u = \{u_i\}, \ 1 \leq i \leq z$$

$$\hat{H}(x^z|v) = \min_{z \in Z} \left\{ \min_{\{u\}} \left\{ \sum_{i=1}^{z} \left\{ H\left(x_{u_i}^z, x_{t(u_i)}^z|v\right) - H\left(x_{t(u_i)}^z|v\right) \right\} \right\} \right\}$$

– ideal biometric representations $x$ should possess
  ▪ low intra-class variability, for managing it with feasible ECCs
  ▪ a large number of features, for using codes with large $k$
  ▪ independent features, for the $z$ most-correlated ones determine the security
Application: On-line Signatures

- Biometric representation
  - use of Universal Background Models (UBMs)
    - users’ templates obtained adapting a person-independent model
      - use of Hidden Markov Models (HMMs) as global model
        - adaptation of the Gaussians mean values with a user-independent matrix \( P \)
          \[ m_u = m_{UBM} + P \cdot x_u \]
        - theoretically uncorrelated features
        - possible large feature number: 4800
          - significant intra-class variability
            - \( s \) most stable features selected in a training phase
              - no information-leakage
  - characteristics
    - theoretically uncorrelated features
    - possible large feature number: 4800
      - significant intra-class variability
        - \( s \) most stable features selected in a training phase
          - no information-leakage

- Database
  - samples taken from the MCYT on-line signature DB
    - 25 genuine and 25 skilled forgeries for each of 100 users
Experimental results

- **Fuzzy commitment**
  - BCH codes → unacceptable performance
    
    | $s$  | 511 | 1023 | 2047 | 4095 |
    |------|-----|------|------|------|
    | FRR  | 50.0| 61.5 | 77.8 | 93.6 |
    | FAR  | 0.3 | 0.2  | 0.0  | 0.0  |
    | $H(x^2|v)$ | 8.7 | 9.6  | 10.4 | 11.3 |

  - turbo codes on binary data ($L=D=2$) → still high FRR
    
    - no soft-decoding
      
    | $s$  | 511 | 1023 | 2047 | 4095 |
    |------|-----|------|------|------|
    | FRR  | 4.6 | 7.4  | 15.0 | 32.4 |
    | FAR  | 8.5 | 3.9  | 1.6  | 0.4  |
    | $H(x^2|v)$ | 22.9 | 50.7 | 108.1 | 221.5 |

- **Proposed approach**
  
  - increasing $L$ improves FAR and security, worsening FRR
  
  - increasing $D$ improves FRR, worsening FAR and security

<table>
<thead>
<tr>
<th>$s$</th>
<th>2047</th>
<th>4095</th>
</tr>
</thead>
<tbody>
<tr>
<td>$L$</td>
<td>2</td>
<td>4</td>
</tr>
<tr>
<td>$D=4$</td>
<td>FRR</td>
<td>2.1</td>
</tr>
<tr>
<td></td>
<td>FAR</td>
<td>22.3</td>
</tr>
<tr>
<td></td>
<td>$H(x^2</td>
<td>v)$</td>
</tr>
<tr>
<td>$D=8$</td>
<td>FRR</td>
<td>1.3</td>
</tr>
<tr>
<td></td>
<td>FAR</td>
<td>67.2</td>
</tr>
<tr>
<td></td>
<td>$H(x^2</td>
<td>v)$</td>
</tr>
</tbody>
</table>

E. Maiorana, P. Campisi - A communication paradigm for biometrics security and privacy
• Biometric representation
  – use of Daugman’s rubber sheet model
    ▪ iris segmented and normalized to create a rectangular template
      o phase information of the Gabor filtering retained as features
  – characteristics
    ▪ highly correlated features (severe issue for security)
    ▪ possible large feature number: $I \times J = 240 \times 20 = 4800$ features
    ▪ significant intra-class variability
      o intra-class variability reduction with a user-independent mask
        □ keeps regions where occlusions (eyelids) are not encountered ($s = 2048$)

• Database
  – samples taken from the CASIA-v4 iris DB
    ▪ 2251 images taken from 395 irises, from 249 subjects
Experimental results

- Fuzzy commitment vs Proposed approach

|        | FRR | FAR | k  | $H(x^2|v)$ |
|--------|-----|-----|----|------------|
| BCH Codes | 48.8 | 0.0 | 13 | 5.7        |
|         | 53.7 | 0.0 | 31 | 13.9       |
|         | 60.4 | 0.0 | 47 | 20.7       |
|         | 71.3 | 0.0 | 71 | 31.2       |
| Turbo Codes | 13.7 | 0.0 | 268 | 117.7     |
|         | 13.8 | 0.0 | 310 | 140.4     |
|         | 13.9 | 0.0 | 368 | 163.5     |
|         | 14.3 | 0.0 | 450 | 200.4     |
| Proposed Approach | L=2 | $D=2$ | 13.4 | 0.0 | 132 | 67.4 |
|      | $D=4$ | 4.84 | 0.1 | 132 | 18.6 |
|      | $D=8$ | 3.00 | 0.3 | 132 | 11.6 |
|      | $L=4$ | $D=2$ | - | - | - |
| Proposed Approach | $D=4$ | - | - | - | - |
|      | $D=8$ | 13.6 | 0.0 | 268 | 117.7 |
|      | $L=4$ | $D=8$ | 8.92 | 0.0 | 268 | 46.7 |

- Recognition accuracy comparison
Conclusions

• Proposal of a template protection scheme inspired by the digital communication paradigm
  – modulation constellations for symbol representation
  – turbo-decoding employed in soft-modality
    ▪ generalization of the fuzzy commitment scheme

• Tests with on-line signatures and iris biometrics
  – great flexibility in selecting the operating conditions
  – improved performance in verification rates and security
    ▪ achieving low FRR and high security are conflicting requisites
      o proper selection of parameters $L$ and $D$

• Future developments
  – applications to other biometrics
  – analysis of multi-biometrics systems
Bibliography


Thanks for your attention!

Questions?

emanuele.maiorana@uniroma3.it