# Security Evaluation of Biometric Privacy Enhancing Techniques

Xuebing Zhou[a], Bian Yang[b] and Christoph Busch[a]

[a]Fraunhofer Institute for Computer Graphics Research IGD, Germany

[b]Gjovik University College, Norway

# Content

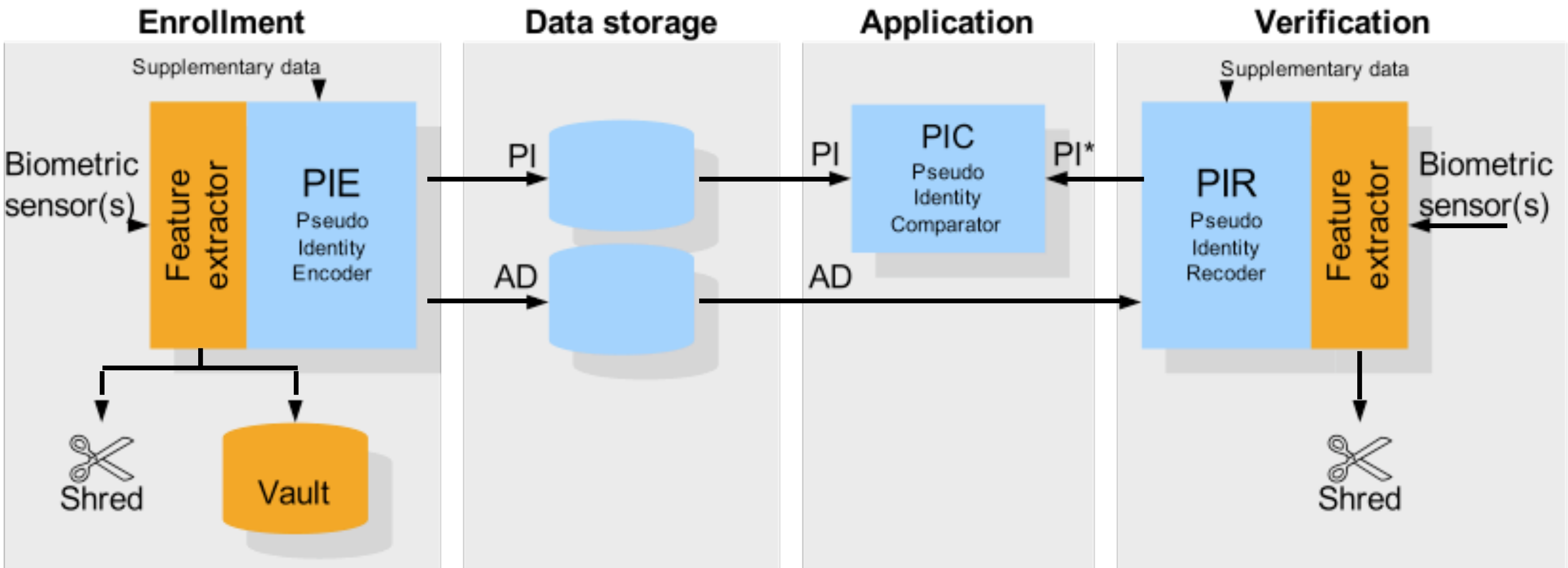
- Motivation
- Overview of Template Protection Techniques
- Protection Goal and Threat Model
- Examples of Security Evaluation
  - Fuzzy Commitment Scheme
  - Biohashing
- Evaluation Framework
- Conclusions

# Motivation

- Privacy and Security Risks in Common Biometric System
  - Identity theft
  - Cross matching
  - Unchangeability
  - Privacy and legislation
- Template protection technique
  - It converts biometric data into **multiple various** references, from which it is **infeasible** *and* **hard** to retrieve the original information.
  - Biometric template protection is a very important supplement to improve security and enhance privacy protection of biometric systems.
- New challenge: how to evaluate *security* and *privacy enhancing ability* of a template protection technique?

•ISO/IEC CD 24745 Information technology - Security techniques - Biometric template protection

# Reference Architecture for Template Protection Technique



Enrollment

Supplementary data

Biometric sensor(s)

Feature extractor

PIE Pseudo Identity Encoder

Shred

Vault

Data storage

PI

AD

Application

PI

PIC Pseudo Identity Comparator

PI*

AD

Verification

Supplementary data

PIR Pseudo Identity Recoder

Feature extractor

Biometric sensor(s)

Shred

- Capture biometric reference
- Create Pseudo Identity (PI)
- Create Auxiliary Data (AD)
- Delete biometric reference
- Publish PI and AD

- PI & AD storage
- May be any type of storage; smart-card, bar-code, central database
- Provide protected identity to application

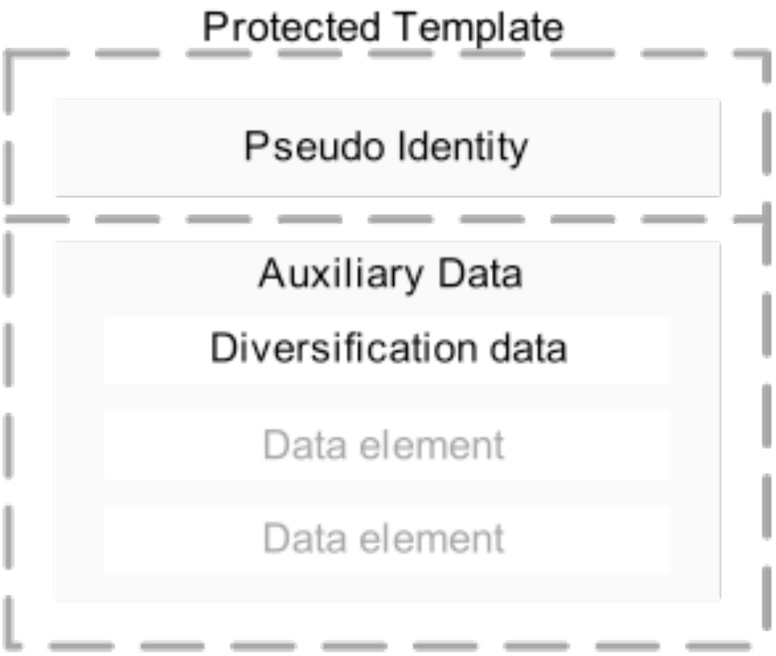
- Match PI with PI*
- Off-line or on-line
- Central or local
- Watch-list functionality
- Identity management

- Capture biometric sample
- Regeneration of PI*
- Publish PI* to application
- Delete biometric sample

- J. Breebaart, C. Busch, J. Grave, E. Kindt: A Reference Architecture for Biometric Template Protection based on Pseudo Identities, in Proceedings BIOSIG 2008

# Template Protection Techniques*

- Protected template
  - Pseudonymous Identifier : PI
  - Auxiliary Data: AD
- Supplementary Data: SD
- Pseudonymous Identifier Encoder:
  - [PI, AD] = PIE(X, SD)
- Pseudonymous Identifier Recorder:
  - [PI’] = PIR(X’, AD, SD)



•ISO/IEC CD 24745 Information technology - Security techniques - Biometric template protection

# State of the Art of Template Protection Techniques

- Transformation algorithms
  - Biometric Salting:
    - *Biohashing*: randomly converts features into bit strings
    - *Biometric Encryption*: whitening features in frequency space
  - Cancelable Biometrics: transformation of features or morphing of samples with non-invertible function

- Biocrypto algorithms
  - *Fuzzy commitment*: cryptographic hashing (ordered set)
  - *Fuzzy vault*: secret sharing protocol (non-ordered set like fingerprint minutiae)

# State of the Art of Template Protection Techniques

| | | Transformation | Biocrypto |
|---|---|---|---|
| Protected template (public) | PI | Transformed feature/ sample | Protected binary identifier |
| | AD | -- | e.g. helper data or point set |
| SD (secret) | | Specific parameters used in tranformation funtion | Optional personal PW |
| Comparison result | | Similarity score / Exact match | Exact match |

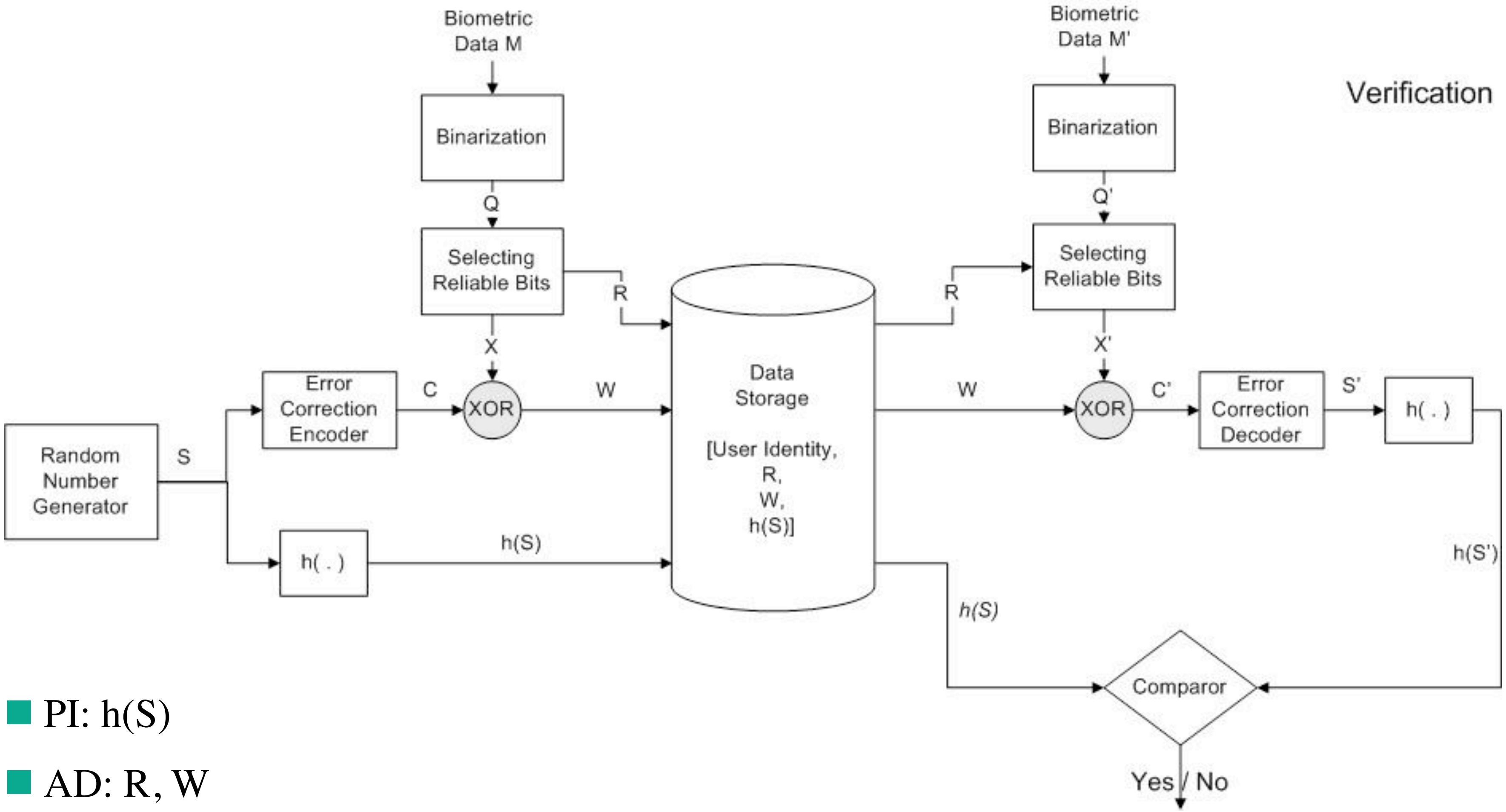
# Protection Goals

- Sufficient security of PI:
  - It is hard to guess the true binary identifier for an enroled subject in the biocrypto system
  - It is hard to reconstruct a sample or feature set yielding an identical transformation, which can successfully pass pseudonymous identifier verification
- Low leakage of biometric information in PI or AD
  - Little biometric information can be learned from PI or AD
- Good randomness of PI
  - The true binary identifiers should be independent
  - No correlation between PIs (the transformed samples/features)
- No personally identifiable information in AD
  - AD is at best random

# Threat Models (TM)

- TM 1: Storage - A protected template of a subject is known
- TM 2: Signal processing - Template protection algorithm is public (Kerckhoffs' principle)
  - Linkage between databases - Stored protected templates of the same subject from different applications are known
  - System parameters are known
  - (Secret) supplementary data $SD$ are known
  - The statistical properties of the biometric data are known
- TM 3: Decision - A large biometric database is available

# An Example of Fuzzy Commitment Scheme



- PI: h(S)
- AD: R, W
- System information: coding algorithm, coding parameters, e.g. $L_s$, $L_c$

# Security Evaluation of the Fuzzy Commitment Scheme

- TM 1: A protected template is known
  - Cryptographic security of $\mathrm{PI} = \mathrm{L}_s$ bits (secret size)
  - Leakage of biometric information in $\mathrm{PI}$ and $\mathrm{AD}$ is neglectable
- TM 2: Template protection algorithm is public
  - Coding algorithm and coding parameters are known:
    - Cryptographic security of $\mathrm{PI} = \mathrm{L}_s$ bits
    - Leakage of biometric information in $\mathrm{AD} = \mathrm{L}_c\text{-}\mathrm{L}_s$ bits
  - More entries of the same subject in different databases are known:
    - Identification ability of $\mathrm{AD}$: $\mathrm{FMR}, \mathrm{FNMR}$ or probability of identification for numerous ranks given $\mathrm{R}$ ($\mathrm{R}$ is indicating reliable bits in the binary vector)

# Security Evaluation of the Fuzzy Commitment Scheme

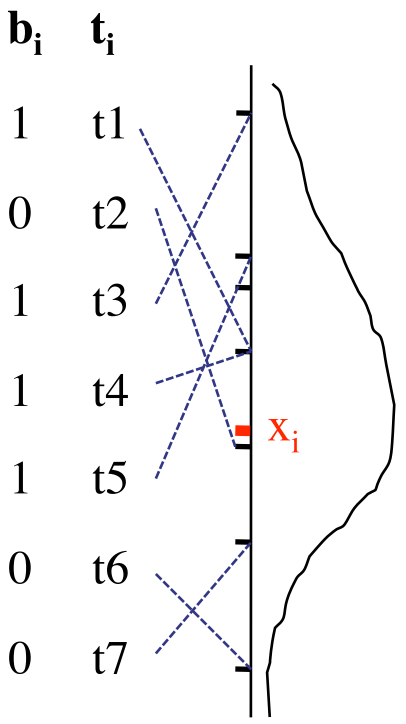
- TM 2 (continued): Template protection algorithm is public
  - Coding algorithm, parameters and more entries of the same subject in different databases (combining leakage)
    - Cryptographic security of PIs: $\hat{H}(S1|W1,W2)$, $\hat{H}(S2|W1,W2)$
    - Security leakage: $L_s$-$\hat{H}(S1|W1,W2)$, $L_s$-$\hat{H}(S2|W1,W2)$
    - Leakage of biometric information in AD: $\hat{I}(X1;W1,W2)$, $\hat{I}(X1;W1,W2)$
  - Statistical properties of binary biometric feature vector are known or partially known as well as the coding algorithm and coding parameters:
    - Cryptographic security of PI: $\hat{H}(S|W)$
    - Security leakage: $L_s$- $\hat{H}(S|W)$
    - Leakage of biometric information in AD: $\hat{I}(X;W)$

# Security Evaluation of the Fuzzy Commitment Scheme

- Remarks:
  - A stream cipher is used in the fuzzy commitment scheme
    - Cryptographic security of PI = Uncertainty about X in W (H(X|W))
    - The assumption that the public helper data W has zero leakage of S is true only in the case that X is identically independently distributed
    - The security leakage is allowed, if PI is cryptographically secure
    - Leakage of biometric information in AD can lead to leakage of PI: $\hat{H}(S1|W1,W2) \le \hat{H}(S1|W1)$. Because W2 is dependent on biometric information X2 and randomly generated S2, increasing leakage is caused by the correlation between W2 and X1

# An Example of Biohashing based on Scalar Randomization

- Any feature $x_i$ in $\mathbf{x} = \{x_i \mid i\in[1,\cdots,n]\}$ can be binarized with $m$ randomly generated thresholds $\mathbf{t} = \{t^j_i \mid j\in[1,\cdots,m], i\in[1,\cdots,n]\}$ into $\mathbf{b} = \{b^j_i \mid j\in[1,\cdots,m], i\in[1,\cdots,n]\}$, where $t^j_i$ has the same distribution as $x_i$.
- Comparison based on Hamming distance between Biohashes.
- Pseudonymous Identifier $\mathrm{PI}$ : $\mathbf{b}$ containing $m\times n$ bits
- Supplementary Data $\mathrm{SD}$ : $\mathbf{t}$
- System parameter: $m, n$



| $\mathbf{b_i}$ | $\mathbf{t_i}$ |
|---|---|
| 1 | t1 |
| 0 | t2 |
| 1 | t3 |
| 1 | t4 |
| 1 | t5 |
| 0 | t6 |
| 0 | t7 |

# Security Evaluation of the Biohashing Algorithm

- TM 1: A protected template (Biohashes) is known
    - --: without information about Biohashing algorithm, Biohashes mean nothing
- TM 2: Template protection algorithm is public
    - The parameters of the algorithm are known:
        - Leakage of biometric information in PI: $\hat{f}(x_i) = \frac{\sum_j b_i^j}{m}$, where $\hat{f}(x_i)$ is the cumulative probability of with e.g. the 95% confident interval of
          $$\hat{f}(x_i) \pm 1.96\sqrt{\frac{\hat{f}(x_i)(1-\hat{f}(x_i))}{m}}$$
        - Randomness of PI $= \sum_{i=1}^{n} m \cdot H\left(f(x_i)\right)$; for all users, the average information rate of Biohashes is 0.72

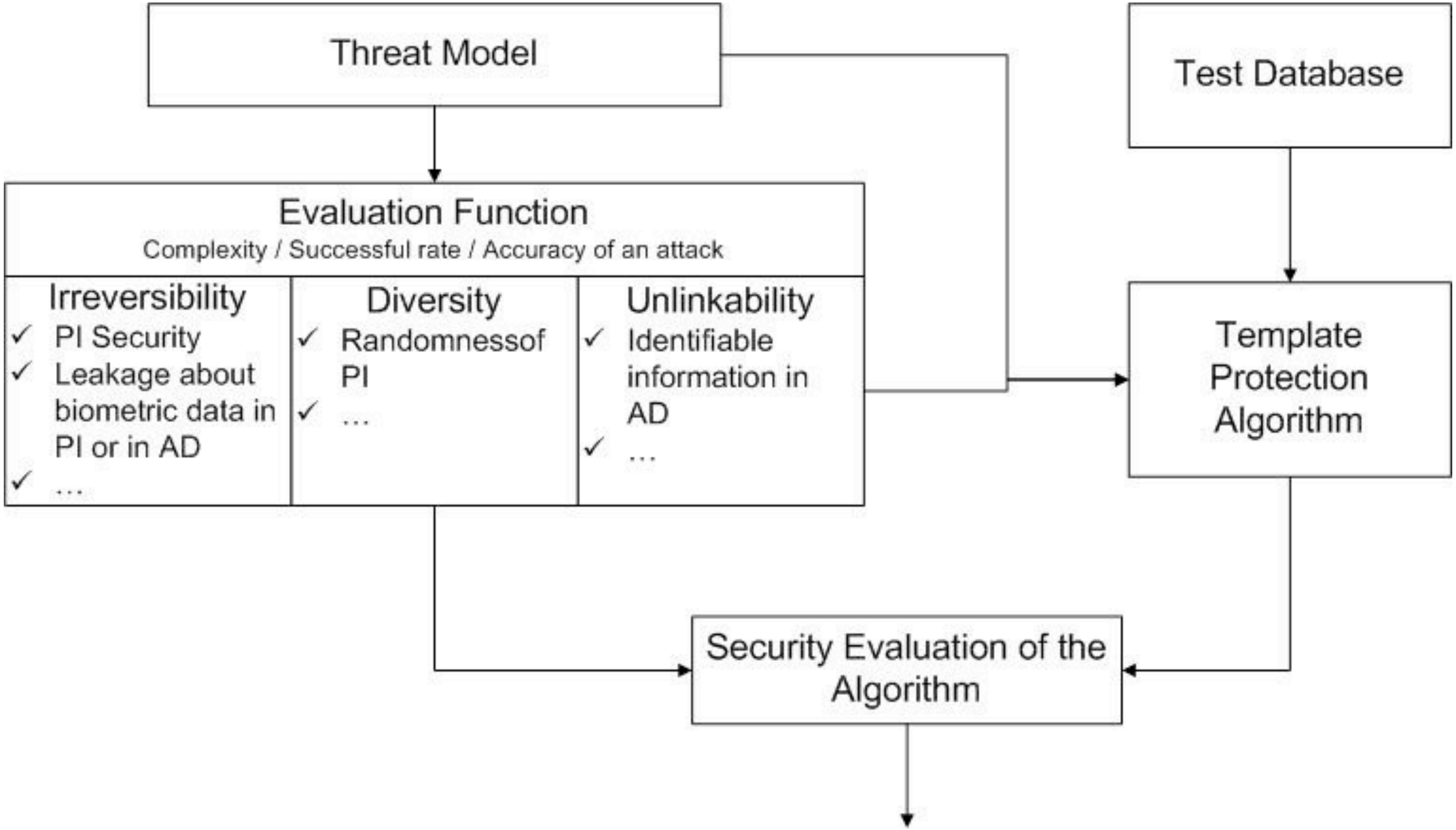
# Security Evaluation of the Biohashing Algorithm

- TM 2 (continued):
  - The parameters of the algorithm and $\mathrm{SD}$ (binarization thresholds are compromised) are known
    - Leakage of biometric information
      - $$\frac{\left|\min_{b_i=1}\{t_i\} - \max_{b_i=0}\{t_i\}\right|}{\left|\max\{t_i\} - \min\{t_i\}\right|}$$
      - $$\hat{p}(x_i) = \hat{p}(t_i)$$
    - If more entries of the same subject in different databases are given, more accurate estimation can be done
  - The parameters of the algorithm and the distribution of $\mathrm{x}$ are known
    - Leakage of biometric information in $\mathrm{PI}$: $\hat{x}_i = f^{-1}\left(\frac{\sum_j b_i^j}{m}\right)$

# Threat Model 3

- False acceptance attack: $P\left(PI_X = PIR(X', AD, SD)\right) = FMR$, where X and X’ are from different subjects.
- FMR shows the average collision probability of biometric data
- The probability of successful false acceptance attack might be different than FMR. For example in Doddington’s biometric menagerie:
  - Lambs: $FMR_L > FMR$
  - Wolves: $FMR_W < FMR$
  - In contrast to randomly generated identifiers, where all the identifiers are equally probable, pseudonymous identifiers generated from lambs/wolves characteristics are not equally probable
- False acceptance is inherent prosperity of biometric system and false acceptance attack is not avoidable. However it can be controled with e.g. time out policy

# Evaluation Framework


Threat Model
Test Database
Evaluation Function
Complexity / Successful rate / Accuracy of an attack
Irreversibility
✓ PI Security
✓ Leakage about biometric data in PI or in AD
✓ …
Diversity
✓ Randomnessof PI
✓ …
Unlinkability
✓ Identifiable information in AD
✓ …
Template Protection Algorithm
Security Evaluation of the Algorithm

# Remarks on Evaluation Framework

- All the requirements on template protection should be measured quantitatively with evaluation function
- From the security and privacy point of view, leakage on biometric information should be limited
  - to $\mathrm{AD}$ of protected template in biocrypto algorithms
  - even when the $\mathrm{SD}$ of transformation algorithms is compromised
- The likelihood to reconstruct true biometric data from protected template is dependent on $\mathrm{SD}$ chosen. An efficient template protection algorithm should have homogenous security for all the subjects and possible $\mathrm{SD}$.

# Conclusions

- A general evaluation framework according to 3 main threat models was developed.
- Depending on the threat model, the **measurable** security evaluation functions are defined
- It is necessary to measure privacy leakage of protected template including leakage of biometric information and any other personal identifiable information. Privacy leakage causes linkability and can give rise to further security problem
- In order to improve confidentiality of authentication and integrity of protected templates, cryptographic technique such as digital signature and encryption can be combined with template protection algorithms.

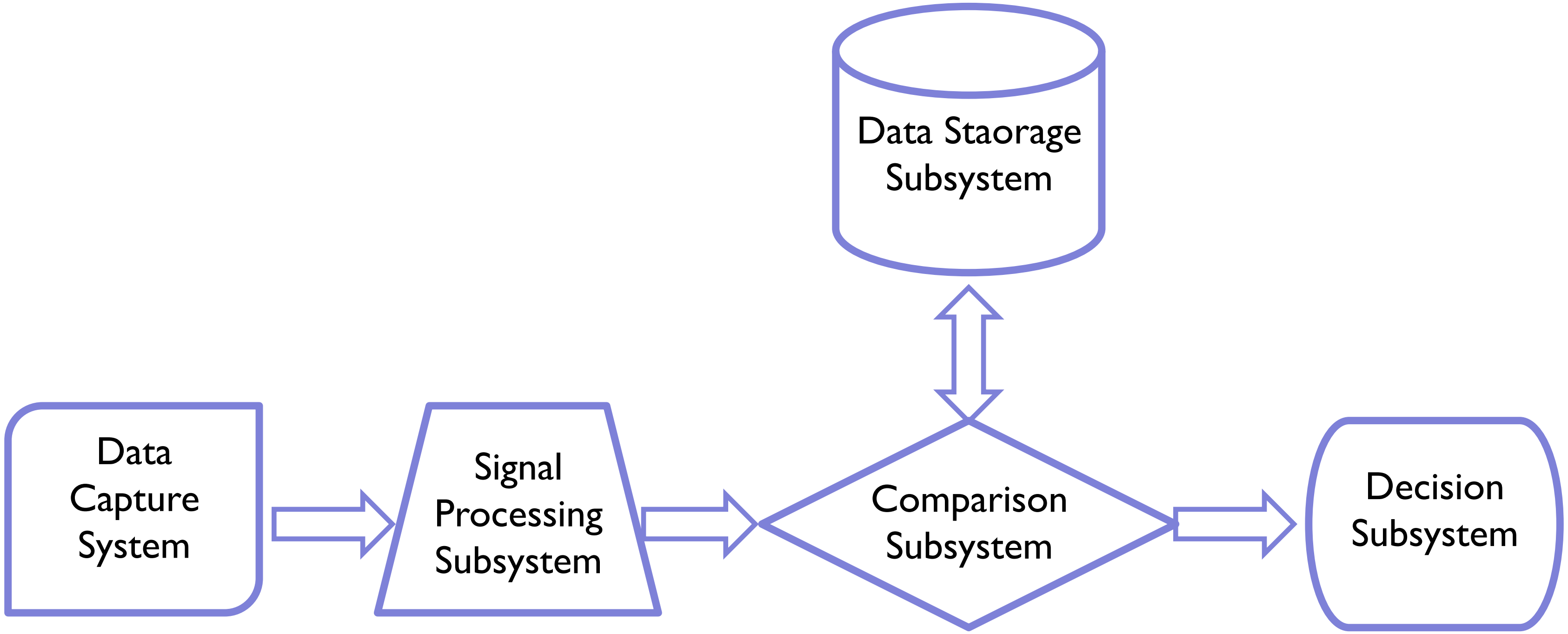
# Thank you for your attention

Xuebing Zhou

Fraunhofer Institute for Computer Graphics
Research IGD
Fraunhoferstraße 5
64283 Darmstadt
Germany

Tel +49 6151 155 – 537
Fax +49 6151 155 – 499
xuebing.Zhou@igd.fraunhofer.de
www.igd.fraunhofer.de

# Threats in the Biometric System



•ISO/IEC CD 24745 Information technology - Security techniques - Biometric template protection