

# Biometrics Based Identifiers for Digital Identity Management

Abhilasha Bhargav-Spantzel, Anna Squicciarini, Elisa Bertino,  
Xiangwei Kong, Weike Zhang

IDTrust 2010  
April 14th 2010

# Outline

- 1 Identity Concepts Overview
- 2 Biometric Systems Overview
- 3 Biometric Commitments
- 4 Backup

# Digital Identity

- **Digital identity:**

- nyms.
- identity attributes or identifiers:
  - strong identifiers (eg. SSN)
  - weak identifiers (eg. age)

- **Owner of an identity attribute:**

Individual who is

- issued the identity attribute
- authoritative of making the claim

- **Identity verification:**

Claimed attribute is

- owned by the individual
- valid

# Digital Identity

- **Digital identity:**

- nyms.
- identity attributes or identifiers:
  - strong identifiers (eg. SSN)
  - weak identifiers (eg. age)

- **Owner of an identity attribute:**

Individual who is

- issued the identity attribute
- authoritative of making the claim

- **Identity verification:**

Claimed attribute is

- owned by the individual
- valid

# Digital Identity

- **Digital identity:**

- nyms.
- identity attributes or identifiers:
  - strong identifiers (eg. SSN)
  - weak identifiers (eg. age)

- **Owner of an identity attribute:**

Individual who is

- issued the identity attribute
- authoritative of making the claim

- **Identity verification:**

Claimed attribute is

- owned by the individual
- valid

## Digital Identity (cont.)

### Identity assurance and linkability

- **Identity assurance:**

Confidence about

- ownership
- validity

Linkability	Strong	Weak
Assurance		
Strong	SSN is valid and owned by <i>Homer07</i> who is <b>Bob Smith</b>	<i>Homer07</i> is a U.S. Citizen
Weak		

## Digital Identity (cont.)

### Identity assurance and linkability

- **Identity assurance:**  
Confidence about
  - ownership
  - validity

Linkability	Strong	Weak
Assurance		
Strong	SSN is valid and owned by <i>Homer07</i> who is <b>Bob Smith</b>	<i>Homer07</i> is a U.S. Citizen
Weak		

## Digital Identity (cont.)

### Identity assurance and linkability

- **Identity assurance:**

Confidence about

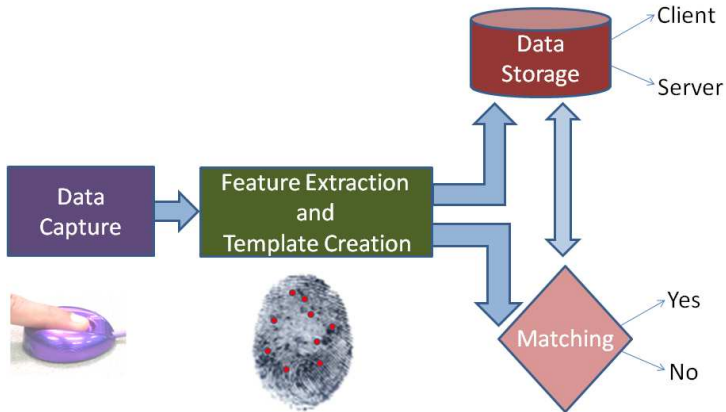
- ownership
- validity

Linkability	Strong	Weak
Assurance		
Strong	SSN is valid and owned by <i>Homer07</i> who is <b>Bob Smith</b>	<i>Homer07</i> is a U.S. Citizen
Weak		

# Outline

- 1 Identity Concepts Overview
- 2 Biometric Systems Overview
- 3 Biometric Commitments
  - Our Approach
  - Main Techniques
  - Experiments and Results
  - Analysis
  - Related Work
- 4 Backup

# Biometric Matching Based Systems



# Biometric Keys: General Idea

- Generating cryptographic keys from biometric measurements:
  - Phase 1:
    - Biometric features → bit string
    - Bit string should have *large inter-class variation* and *small intra-class variation*
  - Phase 2:
    - Bit string  $\xrightarrow{\text{metadata}}$  unique key
    - If two instances of bit strings are 'similar' then the key generated is the same

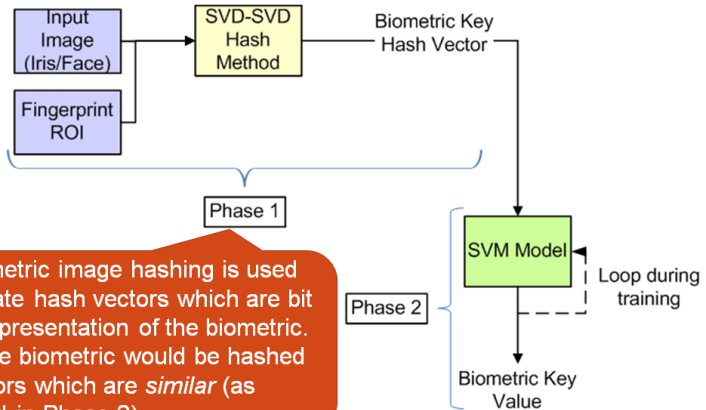
# Biometric Keys: General Idea

- Generating cryptographic keys from biometric measurements:
  - **Phase 1:**
    - Biometric features → bit string
    - Bit string should have *large inter-class variation* and *small intra-class variation*
  - **Phase 2:**
    - Bit string  $\xrightarrow{\text{metadata}}$  unique key
    - If two instances of bit strings are 'similar' then the key generated is the same

# Biometric Keys: General Idea

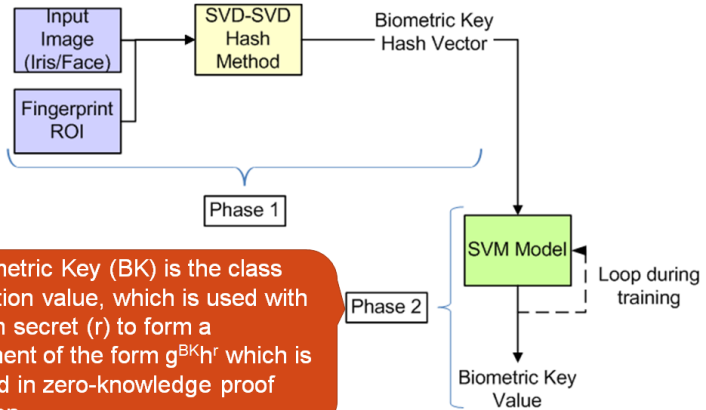
- Generating cryptographic keys from biometric measurements:
  - **Phase 1:**
    - Biometric features → bit string
    - Bit string should have *large inter-class variation* and *small intra-class variation*
  - **Phase 2:**
    - Bit string  $\xrightarrow{\text{metadata}}$  unique key
    - If two instances of bit strings are 'similar' then the key generated is the same

# Two main phases of the biometric key generation



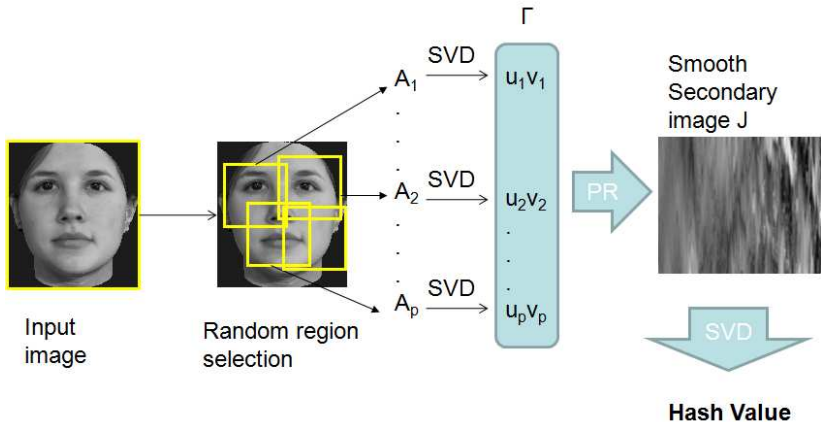
The biometric image hashing is used to generate hash vectors which are bit vector representation of the biometric. The same biometric would be hashed into vectors which are *similar* (as evaluated in Phase 2).

# Two main phases of the biometric key generation



The Biometric Key (BK) is the class combination value, which is used with a random secret ( $r$ ) to form a commitment of the form  $g^{BK}h^r$  which is then used in zero-knowledge proof verification.

# Biometric Hashing Process

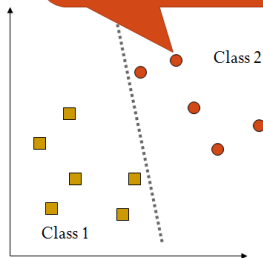


# Key Steps of Biometric Hashing Algorithm

- 1 Random selection of  $A_i$  from biometric image
- 2 **First SVD transform:**  $A_i = U_i S_i V_i^T$   $1 \leq i \leq p$
- 3 Random selection of eigenvectors to create secondary image  $J$
- 4 **Second SVD transform:**  $J = U_J S_J V_J^T$
- 5 **Final hash vector:**  $\vec{H} = \{\vec{u}_J, \vec{v}_J\}$

# SVM Classification

Each point corresponds to a hash vector. The hash vectors of the same person should fall into the same class

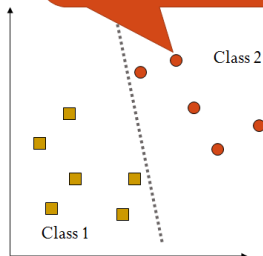


## SVM Usage

- The hash vectors are **ranked** based on confidence degrees from SVM
- Biometric key: **highest confidence class** and **top  $n/2$  classes** (total  $n$  classes)
- Attacker choices for brute force  $n + \binom{n}{i}$
- For  $n > 69$  number of choices is  $> 2^{64}$

# SVM Classification

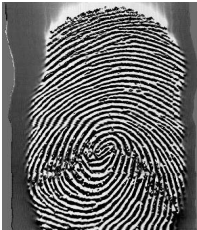
Each point corresponds to a hash vector. The hash vectors of the same person should fall into the same class



## SVM Usage

- The hash vectors are **ranked** based on confidence degrees from SVM
- Biometric key: **highest confidence class** and **top  $n/2$  classes** (total  $n$  classes)
- Attacker choices for brute force  $n + \binom{n}{t}$
- For  $n > 69$  number of choices is  $> 2^{64}$

# Experimental Samples



Thermal (left) and optical (right) sensor fingerprint samples [324 images - FVC]



Iris sample [1695 images - UBIRIS]

## Experimental Samples (cont.)



Yale face samples [100 images]



AT&T face samples [400 images]

## Summary of Experimental Results

Type	# Images	# Persons	CV Accuracy %	FAR %
Fingerprint (Thermal)	204	39	94.96	$5.09 \times 10^{-03}$
Fingerprint (Optical)	120	20	85.83	$7.46 \times 10^{-03}$
Iris	1695	339	92.69	$3.80 \times 10^{-04}$
Face Yale	100	10	99	$1.11 \times 10^{-03}$
Face AT&T	400	40	98.25	$4.49 \times 10^{-04}$

# Uniqueness and Repeatability Analysis

- The metric to measure uniqueness and repeatability is

$$J_2 = \frac{|S_m|}{|S_w|}$$

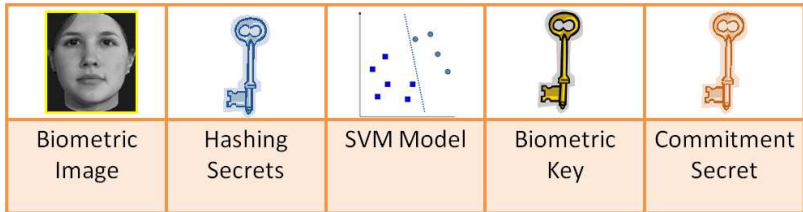
where  $S_m$  is **inter-class** distance and  $S_w$  is **intra-class distance**

- The average values of  $J_2$  calculated were as follows—
  - Fingerprint :  $1.2712 \times 10^{81}$
  - Iris :  $1.5242 \times 10^{303}$
  - Face :  $3.7389^{103}$

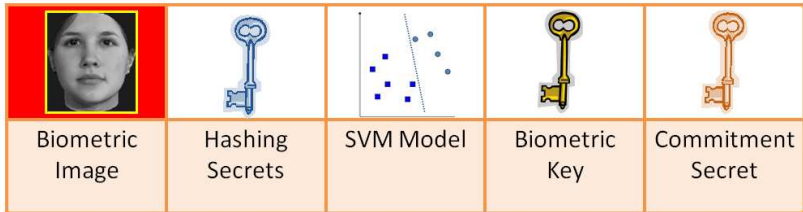
# Biometric Key Analysis

Type	n	Spurious classes	$\eta$	# of BK bits
Fingerprint	69	-	$2.84 \times 10^{19}$	64
Fingerprint	139	69+1	$2.36 \times 10^{40}$	134
Iris	220	-	$4.52 \times 10^{64}$	214
Iris	119	-	$2.43 \times 10^{34}$	114
Face	101	50+1	$1.01 \times 10^{29}$	96

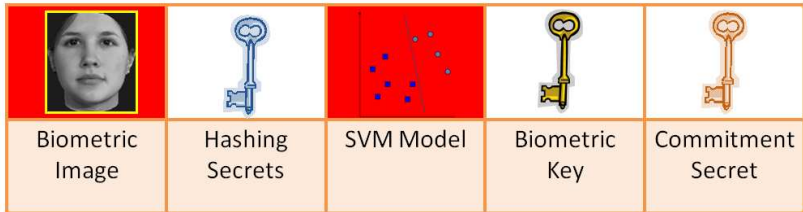
# Biometric Verification System Analysis



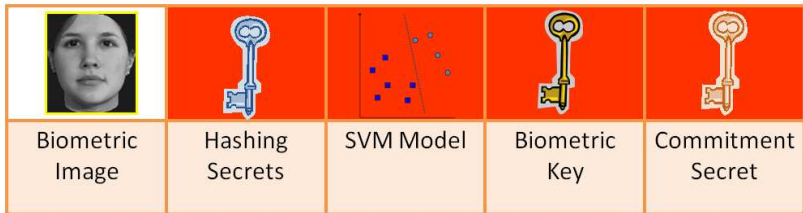
# Biometric Verification System Analysis



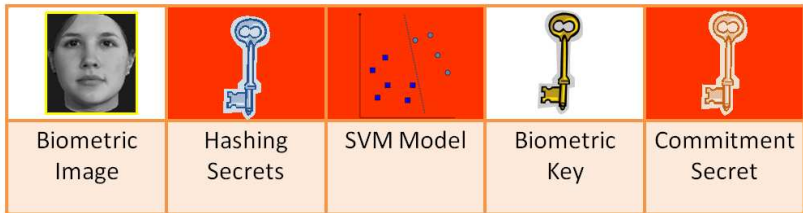
# Biometric Verification System Analysis



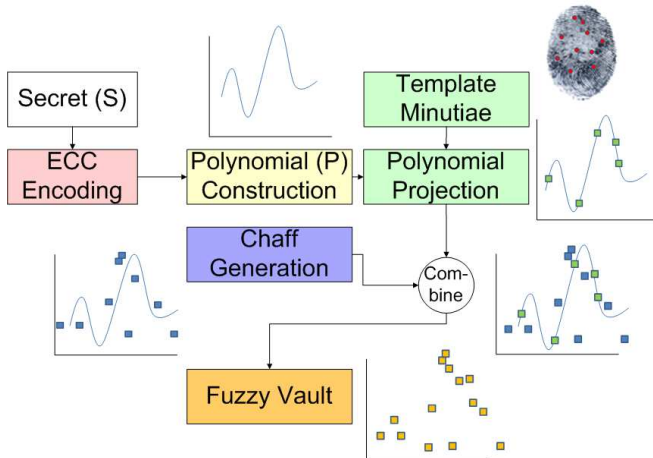
# Biometric Verification System Analysis



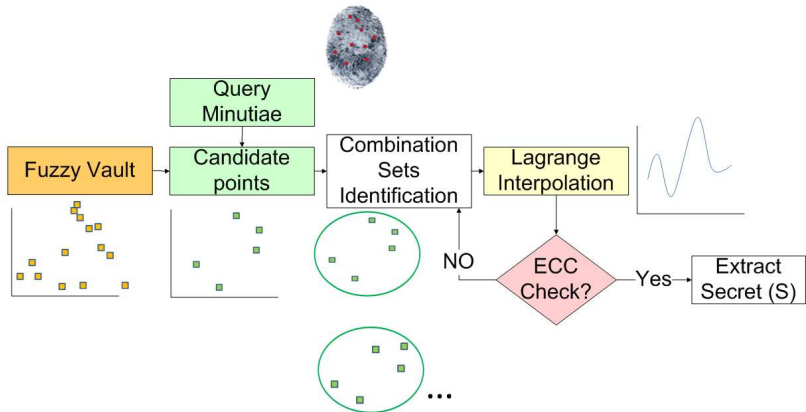
# Biometric Verification System Analysis



# Fuzzy Vault Scheme



# Fuzzy Vault Scheme (cont.)



# Fuzzy Vault Scheme - Shortcomings

- **Intra-class variability:** rotation, translation, # minutia points
  - *'helper data' reduces security*
- Increasing the degree of the **polynomial** increases complexity
  - *require increased number of minutiae points*
- Increasing the number of **chaff points** increases the complexity
  - *empirical bound because of minutiae location*

# Fuzzy Vault Scheme - Shortcomings

- **Intra-class variability:** rotation, translation, # minutia points
  - *'helper data' reduces security*
- Increasing the degree of the **polynomial** increases complexity
  - **require increased number of minutiae points**
- Increasing the number of **chaff points** increases the complexity
  - **empirical bound because of minutiae location**

# Fuzzy Vault Scheme - Shortcomings

- **Intra-class variability:** rotation, translation, # minutia points
  - *'helper data' reduces security*
- Increasing the degree of the **polynomial** increases complexity
  - *require increased number of minutiae points*
- Increasing the number of **chaff points** increases the complexity
  - *empirical bound because of minutiae location*

# BioHashing

Goh *et al.* perform *bio-hashing*

- Based on principal component analysis (PCA)
- Focus only on the first phase of key generation.  
Our approach
  - couples **phase-one** and **phase-two** of key generation
  - analyzes **inter** and **intra-class** variations
  - analyzes **security** and **privacy** of the biometric verification system

# Thank you!

- **Abhilasha Bhargav-Spantzel**  
Intel Corporation  
email: [abhilasha.bhargav-spantzel@intel.com](mailto:abhilasha.bhargav-spantzel@intel.com)

## Tools used

### Singular Value Decomposition (SVD)

If  $A$  is a real  $m$ -by- $n$  matrix, the two orthogonal matrices exist:

$$U = [u_1, \dots, u_m] \in \mathbb{R}^{m \times m} \text{ and } V = [v_1, \dots, v_n] \in \mathbb{R}^{n \times n}$$

such that

$$UAV^T = \text{diag}(\sigma_1, \dots, \sigma_p) \in \mathbb{R}^{m \times n} \quad p = \min\{m, n\}$$

where  $V^T$  is the transpose of matrix  $V$  and  $\sigma_1 \geq \sigma_2 \geq \dots \geq \sigma_p \geq 0$ .  
 $\sigma_i$ 's are the *singular values* of  $A$  and the vectors  $u_i$  and  $v_i$  are the  $i$ th *left singular vector* and the  $i$ th *right singular vector* respectively.

## Tools used (cont.)

### Support Vector Machines (SVM)

- SVM is a classifier based on statistical learning technique developed by Vapnik *et al.*
- It aims at finding optimal hyperplanes to determine the boundaries with the maximal margin separation between every two classes.

SVM applies to classification of vectors, or uni-attribute time series. To classify multi-attribute biometric image data, which are matrices rather than vectors, the multi-attribute data are transformed into uni-attribute data or vectors using SVD.

## Tools used (cont.)

### Support Vector Machines (SVM)

- SVM is a classifier based on statistical learning technique developed by Vapnik *et al.*
- It aims at finding optimal hyperplanes to determine the boundaries with the maximal margin separation between every two classes.

SVM applies to classification of vectors, or uni-attribute time series. To classify multi-attribute biometric image data, which are matrices rather than vectors, the multi-attribute data are transformed into uni-attribute data or vectors using SVD.

# Key Steps of Biometric Hashing Algorithm

- 1: Input biometric image  $I$
- 2: **for** each *random*  $A_i$  where  $1 \leq i \leq p$  **do**
- 3:  $A_i = U_i S_i V_i^T$  **{First SVD Transform}**  
{Collect singular vectors corresponding to the largest singular value}
- 4:  $\vec{u}_i$  = first left singular vector
- 5:  $\vec{v}_i$  = first right singular vector
- 6: **end for**
- 7:  $\Gamma = \{\vec{u}_1, \dots, \vec{u}_p, \vec{v}_1, \dots, \vec{v}_p\}$
- 8: *Randomly* create  $J[m, 2p]$  from  $\Gamma$  **{Second SVD Transform}**
- 9:  $J = U_J S_J V_J^T$  {Collect singular vectors corresponding to the largest singular value}
- 10:  $\vec{u}_J$  = first left singular vector
- 11:  $\vec{v}_J$  = first right singular vector
- 12:  $\vec{H} = \{\vec{u}_J, \vec{v}_J\}$

# Fuzzy Vault Scheme - Shortcomings

## Attacks on Fuzzy Vault

In August 2007, Preda Mihailescu presented a brute force attack in three known implementations of the vault for fingerprints. The vulnerability cannot be avoided by mere parameter selection in the actual frame of the procedure.