

Information and coding theory in biometrics

TURBINE final Workshop

G rard Cohen

Based on joint works with Julien Bringer, Herv  Chabanne, Bruno Kindarji, Gilles Z mor

- Introduction
- Secure Sketches: definitions
- Secure Sketches: performances and security
- Conclusion

Biometric characteristics are used to identify yourself among a large set of people during your whole lifetime.

Let W be a biometric trait, its different captures are denoted by:

$$w_0, w_1, w_2 \dots \leftarrow W$$

- 1 Enrollment: A good capture, called the template, $w_0 \leftarrow W$ is kept apart.
- 2 Verification: New capture $w_1 \leftarrow W$ is matched against w_0 .

Biometric characteristics are used to identify yourself among a large set of people during your whole lifetime.

Let W be a biometric trait, its different captures are denoted by:

$w_0, w_1, w_2 \dots \leftarrow W$

- 1 Enrollment: A good capture, called the template, $w_0 \leftarrow W$ is kept apart.
- 2 Verification: New capture $w_1 \leftarrow W$ is matched against w_0 .

Paradigm of biometric authentication:

I'm the living person with the biometric data which match those stored as a reference.

Biometric characteristics are used to identify yourself among a large set of people during your whole lifetime.

Let W be a biometric trait, its different captures are denoted by:

$w_0, w_1, w_2 \dots \leftarrow W$

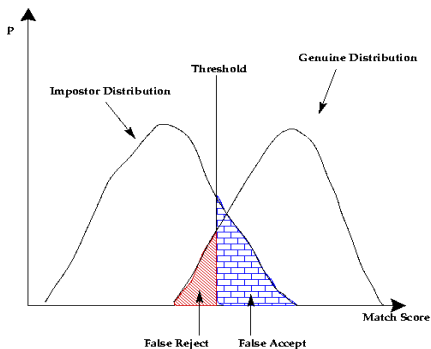
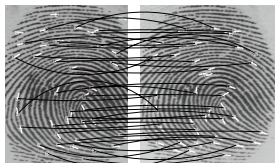
- 1 Enrollment: A good capture, called the template, $w_0 \leftarrow W$ is kept apart.
- 2 Verification: New capture $w_1 \leftarrow W$ is matched against w_0 .

Remark

No confidentiality is required at this point.

Aliveness detection is needed.

How do we evaluate a system?



- Fingerprint:
 - FAR < 0,1 %
 - FRR < 1 %
- Iris:
 - FAR : < 0,01 %
 - FRR : < 1 %

Privacy issue (Why do we need cryptography?)

- The belonging of someone to an application must be protected.

Privacy issue (Why do we need cryptography?)

- The belonging of someone to an application must be protected.

Biometric characteristics are used to identify yourself among a large set of people during your whole lifetime.

Privacy issue (Why do we need cryptography?)

- The belonging of someone to an application must be protected.
- Biometric data must stay encrypted!!!

Adversary's capacity



- Biometric data are public.
- Adversary might have access to different captures of your biometric trait $w_0, w_1, w_2 \dots \leftarrow W$.

Goal: Storing biometric templates in a way

- ① which is easy to renew
- ② which does not leak informations on the underlying biometric data
- ③ but, which still allows matching

Goal: Storing biometric templates in a way

- ① which is easy to renew
- ② which does not leak informations on the underlying biometric data
- ③ but, which still allows matching

How to manage fuzzy biometric authentication with privacy protection?
i.e. how can we reconcile encryption with the variability of different captures?

- Introduction
- **Secure Sketches: definitions**
- Secure Sketches: performances and security
- Conclusion

Definition (Secure sketch [Dodis-Reyzin-Smith 2004])

A (\mathcal{H}, m, m', t) -secure sketch is a pair of functions (SS, Rec) where the sketching function SS takes $w \in \mathcal{H}$ as input, and outputs a sketch in $\{0, 1\}^*$, such that for all random variables W over \mathcal{H} with min-entropy $\mathbf{H}_\infty(W) \geq m$, we have the conditional min-entropy $\overline{\mathbf{H}}_\infty(W \mid SS(W)) \geq m'$.

The recovery function Rec takes a sketch P and a vector $w' \in \mathcal{H}$ as inputs, and outputs a word $w'' \in \mathcal{H}$, such that for any $P = SS(w)$ and $d(w, w') \leq t$, it holds that $w'' = w$.

Juels and Wattenberg's Authentication protocol

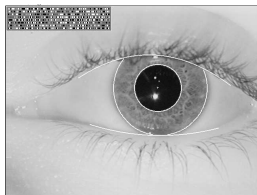
- During the enrolment, we store $P = SS_C(w) = c \oplus w$, where c is a random codeword in C , together with the hash value $Hash(c)$ of c (where $Hash$ is a cryptographic hash function).
- To authenticate someone, we try to correct the corrupted codeword $w' \oplus P = c \oplus (w' \oplus w)$ and if we obtain a codeword c' , we then check: $Hash(c') = Hash(c)$.

- Introduction
- Secure Sketches: definitions
- **Secure Sketches: performances and security**
- Conclusion

Iris as an example of biometric data (1/5)

- Iriscode is made of 2 vectors of 256 bytes:
 - I which carries information and,
 - M a mask which indicates whether information is available or not.
- Matching of $w = (I, M)$ against $w' = (I', M')$ is made by computing the Hamming distance over the portion of the information vectors not erased:

$$\mu(w, w') = \frac{\|(I \oplus I') \cap M \cap M'\|}{\|M \cap M'\|}$$



J. Daugman. How Iris Recognition Works.
IEEE Trans. CSVT 14(1), pp. 21 - 30, 2004.

Iris as an example of biometric data (2/5)

- Feng Hao, Ross Anderson, John Daugman. Combining Crypto with Biometrics Effectively. IEEE Trans. Computers 2006
- First paper published on an implementation of iris secure sketches
- Only information vectors I are used
- Hao, Anderson and Daugman announce very good performances on their own private database
- Unfortunately, we are not able to obtain such good results on a public database (see after)

Iris as an example of biometric data (2/5)

- Can one do better ?
- How much, i.e. are there theoretical limits?
- Can we effectively do better (Which codes to choose)?
- I.e. find a decoding algorithm

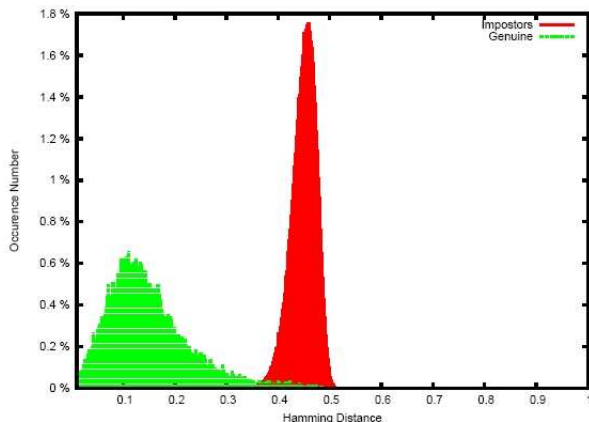
Iris as an example of biometric data (2/5)

Some references:

- Julien Bringer, Hervé Chabanne, Gérard Cohen, Bruno Kindarji, Gilles Zémor. Optimal Iris Fuzzy Sketches. BTAS 2007.
- Julien Bringer, Hervé Chabanne, Gérard Cohen, Bruno Kindarji, Gilles Zémor. Theoretical and Practical Boundaries of Binary Secure Sketches. IEEE Transactions on Information Forensics & Security, 2008.

Iris as an example of biometric data (3/5)

- ICE database: 2953 images from 244 different eyes
- for 5 % of wrongly rejected users, we have: $\mu(w, w') \geq 0.29$
- from 512 to 1977 of information bits are masked



Theorem

Let $k \in \mathbb{N}^*$, C be a binary code of length N and size 2^k , and m a random received message, from a random codeword of C , of length N with w_n errors and w_e erasures. Assume that C is an optimal code with respect to N and k , equipped with an **ML** decoder.

If $\frac{w_n}{N-w_e} > \theta$ then m is only decodable with a negligible probability for a large N , where θ is such that the Hamming sphere of radius $(N - w_e)\theta$ in $\mathbb{F}_2^{N-w_e}$, i.e. the set $\{x \in \mathbb{F}_2^{N-w_e}, d_{\mathcal{H}}(x, \mathbf{0}) = (N - w_e)\theta\}$, contains 2^{N-w_e-k} elements.

Iris as an example of biometric data (4/5)

$n = 2048$

- $k = 42$, FRR $> 2,49$ %
- $k = 64$, FRR $> 3,76$ %
- $k = 80$, FRR $> 4,87$ %
- $k = 128$, FRR $> 9,10$ %

Iris as an example of biometric data (5/5)

- Random interleaver
- Product code $C = RM(1, 6) \otimes RM(1, 5)$ of length $64 \times 32 = 2048$ and dimension 42
- Masked bits are treated as erasures
- Exhaustive decoding of each RM
- Decoding of C with the min-sum iterative decoding algorithm
- $FRR = 5,62 \% > 2,49 \%$, $FAR = 0,0006 \%$

Are Secure Sketches secure enough?

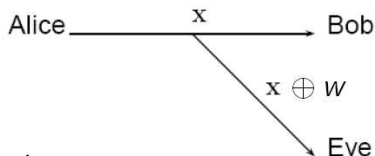
- Let $P = c \oplus w = SS_C(w)$.
- Consider a false acceptance w' .
- From w' , you thus get $P \oplus w'$, can decode c and then obtain w .

- Introduction
- Secure Sketches: definitions
- Secure Sketches: performances and security
- **Conclusion**

Strengthening Secure Sketches

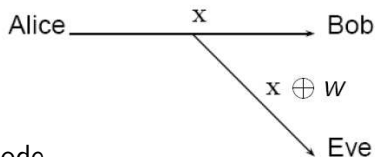
- Gérard Cohen, Gilles Zémor. Generalized coset schemes for the wire-tap channel: application to biometrics. ISIT 2004, Chicago, June 2004
- The distribution of biometric vector w is not uniform.
So $c \oplus w$ leaks information on c .

Wire-tap channels



- A $[n, k]$ C code.
- The syndrome function $\sigma: x \in \{0, 1\}^n \mapsto H^t x \in \{0, 1\}^{(n-k)}$
s.t. C is the set of vectors verifying $\sigma(x) = 0$

Wire-tap channels

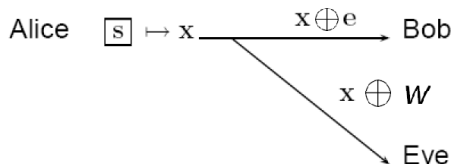


- A $[n, k]$ C code.
- The syndrome function $\sigma: x \in \{0, 1\}^n \mapsto H^t x \in \{0, 1\}^{(n-k)}$
s.t. C is the set of vectors verifying $\sigma(x) = 0$
- Secret s .
Wyner showed how one could obtain perfect secrecy when a receiver enjoys a better channel than does the wire-tapping opponent.
- Noiseless main channel:
 - $s = \sigma(y) = H^t y$ where H is $h(b) \times n = nh(p) \times n$ parity-check matrix
 - Transmit $y \oplus c$ where c is randomly chosen, uniformly in C

Generalisation to noisy case

- Choose H_1 to be $nr_1 \times n$ parity-check matrix of a code C_1 that correct noise e ($r_1 \geq h(p)$)
- Choose H_2 to be $(n(h(p) - r_1) \times n$ matrix
- x is chosen uniformly in C_1 among vectors of syndrome

$$\begin{pmatrix} H_1 \\ H_2 \end{pmatrix} t_x = \begin{pmatrix} 0 \\ s \end{pmatrix}$$



Cohen and Zémor's Authentication protocol

- Two error-correcting codes C_1 and C_2
- Choose a random codeword $c \in C_1$ whose syndrome for C_2 is s
- $P = c \oplus w, \text{Hash}(s)$
- Decode $P \oplus w'$ as z
- Compute the syndrome $\sigma_2(z) = H_2^t z$ for C_2
- Check that $\text{Hash}(\sigma_2(z)) = \text{Hash}(s)$

A condition we need the biometric data to fulfil

In the binary symmetric channel with transition probability p , there exists a set T of typical vectors of cardinality $|T| \approx 2^{nh(p)}$ with a uniform distribution.

Still a very active field of work.

Thank you for your attention. Any questions?