Fuzzy extractors security under several models of biometric data

Grigory Marshalko Yulia Trufanova

TK 26, Lomonosov Moscow State University

Svetlogorsk, Kaliningrad region
June 4-7, 2019

Biometry usage

Biometric data are unique for each person,
BUT there are certain difficulties while using them.

Difficulties

1. unstable input data
2. low entropy of output bit string
3. how to store the template safely?

Input data instability $\Rightarrow$ errors of type I (False Rejection) and type II (False Acceptance).

Fuzzy extractor

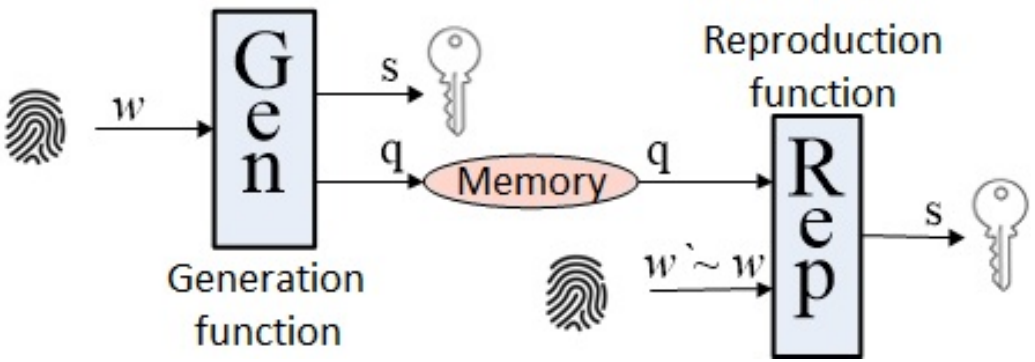


Figure 1: Fuzzy extractor

$\mathcal{M}$ – *metric space with metric* ρ
U_l – *random variable with uniform distribution over* $\{0,1\}^l$
$\mathsf{H}_\infty(A) = -\log_2(\max\limits_{a \in \mathcal{A}} \Pr(A = a))$
$\mathsf{SD}(A, B) =$
$= \frac{1}{2} \sum\limits_{a \in \mathcal{A}} |\Pr(A = a) - \Pr(B = a)|$

$(\mathcal{M}, m, l, t, \varepsilon)$-fuzzy extractor – a pair of functions $(\mathsf{Gen}, \mathsf{Rep})$:

1. $\mathsf{Gen}[w] = \langle s, q \rangle$, where $w \in \mathcal{M}, s \in \{0,1\}^l, q \in \{0,1\}^*$.
2. $\mathsf{Rep}[w', q] = s \Leftrightarrow \rho(w, w') \leq t, w' \in \mathcal{M}$.
3. Security property.
   If for a random variable W over $\mathcal{M}$ it holds true that $\mathsf{H}_\infty(W) \geq m$ and $\mathsf{Gen}[W] = \langle s, q \rangle$, then $\mathsf{SD}(\langle s, q \rangle, \langle U_l, q \rangle) \leq \varepsilon$.

Fuzzy extractor scheme

Fuzzy extractor scheme based on XOR operation:

- $\mathsf{Gen}[w] = \langle s, q \rangle$, where $s \in_R \{0,1\}^m$, $q = w \oplus C[s]$;
- $\mathsf{Rep}[w', q] = D[w' \oplus q] = D[w' \oplus w \oplus C[s]] = s$, if $\rho[w, w'] \leq 1$.

$\mathcal{M} = \{0,1\}^n$, ρ – *Hamming metric*
$C[\cdot]$ – *coding function of Hamming code*
$D[\cdot]$ – *corresponding decoding function*

$w, w' \in \mathcal{M}$ – *biometric parameter feature vectors*
s – *fuzzy extractor secret string*
q – *fuzzy extractor public (helper) string*

Previous results

- Xavier Boyen,"*Reusable cryptographic fuzzy extractors*", 2004.
- Koen Simoens, Pim Tuyls, Bart Preneel, "*Privacy weaknesses in biometric sketches*", 2009.

Real-life security of fuzzy extractors

Main questions

1. What is the distribution of biometric data?
2. How can an attacker use knowledge about biometric data distribution?

Real-life security of fuzzy extractors

Main questions

1. What is the distribution of biometric data?
2. How can an attacker use knowledge about biometric data distribution?

Goal

Study real-life security of a fuzzy extractor scheme based on XOR operation.

Biometric data models

1. Independent non-equiprobable bits with known parameter $p \neq 1/2$.
2. Bits form a stationary Markov chain with known matrix of transition probabilities.

Biometric data models

1. Independent non-equiprobable bits with known parameter $p \neq 1/2$.
2. Bits form a stationary Markov chain with known matrix of transition probabilities.

Statistical analysis of biometric data

Description of biometric vectors database for AT&T Database:
39 persons – from 5 to 15 vectors for each person – 377 vectors – length 130592 bits
Feature extraction with *Local binary patterns (LBP)* algorithm
Vector length after transformation – 75 bits.

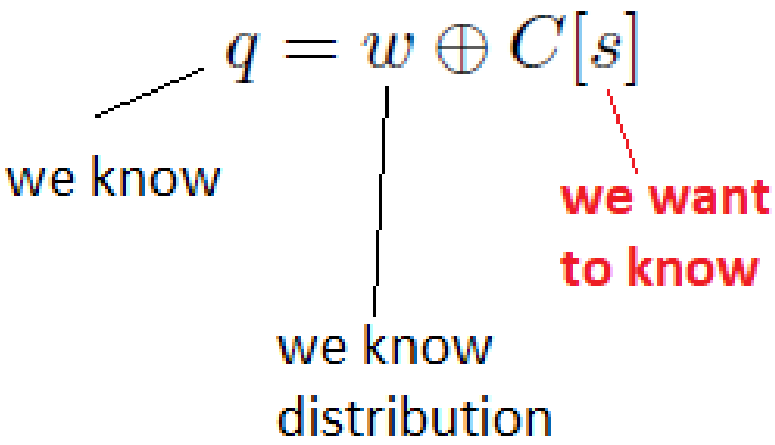
Table 1: Evaluation of probabilities to take the value 1 for several bits.

i	$P(w_i) = 1$	i	$P(w_i) = 1$	i	$P(w_i) = 1$	i	$P(w_i) = 1$
0	0.335677	8	0.514310	16	0.491613	24	0.487495
1	0.529562	9	0.504790	17	0.513379	25	0.483415
2	0.525382	10	0.457895	18	0.500390	26	0.479505
3	0.533092	11	0.493051	19	0.493090	27	0.507518
4	0.403864	12	0.482903	20	0.502562	28	0.485082
5	0.501710	13	0.478931	21	0.498690	29	0.504567
6	0.487790	14	0.479208	22	0.500526	30	0.487244
7	0.488562	15	0.5013541	23	0.473223	31	0.491690

Table 2: Searching for stationary Markov chains.

Person	**Bits sequences**
1	[58-60]
2	[61-63]
7	[63-65]
9	[18-20], [63-65]
13	[28-30], [33-36], [46-48], [49-51], [53-56], [57-60], [61-63], [63-65], [66-68], [71-73]
23	[58-60], [71-73]
24	[36-38], [66-68]
25	[52-54], [72-74]
26	[28-30], [30-32], [41-43], [61-63], [65-67], [68-70], [69-72], [71-73], [72-74]

State of problem



Independent non-equiprobable bits

For the sake of certainty $p < \frac{1}{2}$, $g = 1 - p$.

n – biometric parameter length, $N = 2^n$.

$$\mathsf{Pr}(w^1) \geq \mathsf{Pr}(w^2) \geq \ldots \geq \mathsf{Pr}(w^N)$$

$$W^i = \{w | \mathsf{wt}(w) = i\}$$

$$\left.\begin{array}{l} w^1 = (0,0,\ldots,0,0) \end{array}\right\} W^0$$

$$\left.\begin{array}{l} w^2 = (0,0,\ldots,0,1) \\ w^3 = (0,0,\ldots,1,0) \\ \ldots \\ w^{n+1} = (1,0,\ldots,0,0) \end{array}\right\} W^1$$

$$\ldots$$

Independent non-equiprobable bits

$$w \oplus q = C[s]$$

$$\begin{aligned}
(0,0,\ldots,0,0) \oplus (q_1,q_2,\ldots,q_n) &= (q_1,q_2,\ldots,q_n)\\
(0,0,\ldots,0,1) \oplus (q_1,q_2,\ldots,q_n) &= (q_1,q_2,\ldots,q_n \oplus 1)\\
&\vdots\\
(1,1,\ldots,1,1) \oplus (q_1,q_2,\ldots,q_n) &= (q_1 \oplus 1, q_2 \oplus 1,\ldots,q_n \oplus 1)
\end{aligned}$$

Independent non-equiprobable bits

$$w \oplus q = C[s]$$

$$\begin{aligned}
(0,0,\ldots,0,0) \oplus (q_1,q_2,\ldots,q_n) &= (q_1,q_2,\ldots,q_n) && \equiv s^1\\
(0,0,\ldots,0,1) \oplus (q_1,q_2,\ldots,q_n) &= (q_1,q_2,\ldots,q_n \oplus 1) && \equiv s^2\\
&\vdots && \vdots\\
(1,1,\ldots,1,1) \oplus (q_1,q_2,\ldots,q_n) &= (q_1 \oplus 1, q_2 \oplus 1,\ldots,q_n \oplus 1) && \equiv s^N
\end{aligned}$$

Independent non-equiprobable bits

$$\{s^j\} = \{0,1\}^n \;\Rightarrow\; \exists j_0 : s^{j_0} = C[s]$$

Let's denote:

$$S^i = \{s \in \{s^j\} | s^j = q \oplus w^h, \mathsf{wt}(w^h) = i\},$$

$$|S^i| = \binom{n}{i}, \forall i = 0, \ldots, n.$$

Then

$$S^0 \cup S^1 \cup \ldots \cup S^n = \{s^j\},$$

$$S^i \cap S^j = \oslash, i \neq j.$$

Independent non-equiprobable bits

Consider only first r classes $S^0, S^1, \ldots, S^{r-1} = \widetilde{S}$ of vectors with maximum probabilities. Denote $M = \sum_{i=0}^{r-1} \binom{n}{i}$.

$\widetilde{S}$

$$s^1 = (q_1, q_2, \ldots, q_n)$$

$$s^2 = (q_1, q_2, \ldots, q_n \oplus 1)$$

$$\ldots$$

$$s^M = (q_1 \oplus w_1^M, q_2 \oplus w_2^M, \ldots, q_n \oplus w_n^M)$$

$$s^{M+1} = (q_1 \oplus w_1^{M+1}, q_2 \oplus w_2^{M+1}, \ldots, q_n \oplus w_n^{M+1})$$

$$\ldots$$

$$s^N = (q_1 \oplus 1, q_2 \oplus 1, \ldots, q_n \oplus 1)$$

Independent non-equiprobable bits

K^i – the number of code words in class S^i

Lemma

$$K^i = \frac{\binom{n}{i-1} - K^{i-1} - K^{i-2} \cdot (n - (i-2))}{i}, \ \forall i = \overline{2, n}.$$

Independent non-equiprobable bits

$$R_1(M) = \left(K^0 \cdot \left(g^n + \binom{n}{1} p^1 g^{n-1}\right) + \sum_{j=1}^{r} K^j \cdot \left(p^j g^{n-j} + \binom{j}{1} p^{j-1} g^{n-(j-1)} + \right.\right.$$

$$\left.\left. + \binom{n-j}{1} p^{j+1} g^{n-(j+1)}\right)\right)^{-1} \cdot \left[\left(1 - \left(K^0 \left(g^n + \binom{n}{1} p^1 g^{n-1}\right) + \right.\right.\right.$$

$$\left.\left.\left. + \sum_{j=1}^{r} K^j \left(p^j g^{n-j} + \binom{j}{1} p^{j-1} g^{n-(j-1)} + \binom{n-j}{1} p^{j+1} g^{n-(j+1)}\right)\right)\right) \cdot$$

$$\cdot \left(\sum_{i=0}^{r} K^i (1+n) \right) + \left(K^0 \left(g^n + \binom{n}{1} p^1 g^{n-1}\right) + \sum_{j=1}^{r} \frac{\left(\sum_{i=1}^{j-1} K^i + 1 + \sum_{i=1}^{j} K^i \right)}{2} \cdot \right.$$

$$\left. \cdot K^j \cdot \left(p^j g^{n-j} + \binom{j}{1} p^{j-1} g^{n-(j-1)} + \binom{n-j}{1} p^{j+1} g^{n-(j+1)}\right)\right)\Bigg]. \tag{1}$$

Markov chain model

$P = \begin{pmatrix} p_{00} & p_{01} \\ p_{10} & p_{11} \end{pmatrix}$ – matrix of transition probabilities:

$p_{ij} = \mathsf{Pr}(x_r = i | x_{r-1} = j), i, j \in \{0, 1\}, \forall r \in \{1, 2, \ldots\}.$

$F = \begin{pmatrix} f_{00} & f_{01} \\ f_{10} & f_{11} \end{pmatrix}$ – transition matrix of a chain sample:

f_{ij} – the number of transitions from state i to state j.

Markov chain model

1. Restore all possible transition matrices F for n steps.

$\mathbf{f_{00}}$	$\mathbf{f_{01}}$	$\mathbf{f_{10}}$	$\mathbf{f_{11}}$
f_{00}^1	f_{01}^1	f_{10}^1	f_{11}^1
f_{00}^2	f_{01}^2	f_{10}^2	f_{11}^2
f_{00}^3	f_{01}^3	f_{10}^3	f_{11}^3
...			

2. Evaluate the probability of a Markov chain sample with given transition matrix F.

$\mathbf{f_{00}}$	$\mathbf{f_{01}}$	$\mathbf{f_{10}}$	$\mathbf{f_{11}}$	$\mathbf{Pr}$
f_{00}^1	f_{01}^1	f_{10}^1	f_{11}^1	$\Pr(w_{F^1})$
f_{00}^2	f_{01}^2	f_{10}^2	f_{11}^2	$\Pr(w_{F^2})$
f_{00}^3	f_{01}^3	f_{10}^3	f_{11}^3	$\Pr(w_{F^3})$
...				...

Markov chain model

1. Restore all possible transition matrices F for n steps.

$\mathbf{f_{00}}$	$\mathbf{f_{01}}$	$\mathbf{f_{10}}$	$\mathbf{f_{11}}$
f_{00}^1	f_{01}^1	f_{10}^1	f_{11}^1
f_{00}^2	f_{01}^2	f_{10}^2	f_{11}^2
f_{00}^3	f_{01}^3	f_{10}^3	f_{11}^3
...			

2. Evaluate the probability of a Markov chain sample with given transition matrix F.

$\mathbf{f_{00}}$	$\mathbf{f_{01}}$	$\mathbf{f_{10}}$	$\mathbf{f_{11}}$	**Pr**
f_{00}^1	f_{01}^1	f_{10}^1	f_{11}^1	$\Pr(w_{F^1})$
f_{00}^2	f_{01}^2	f_{10}^2	f_{11}^2	$\Pr(w_{F^2})$
f_{00}^3	f_{01}^3	f_{10}^3	f_{11}^3	$\Pr(w_{F^3})$
...				...

Markov chain model

3. Evaluate the number of Markov chain samples with given transition matrix F.

$\mathbf{f_{00}}$	$\mathbf{f_{01}}$	$\mathbf{f_{10}}$	$\mathbf{f_{11}}$	$\mathbf{Pr}$	$\mathbf{N^{(n)}(F)}$
f_{00}^1	f_{01}^1	f_{10}^1	f_{11}^1	$\Pr(w_{F^1})$	$N^{(n)}(F^1)$
f_{00}^2	f_{01}^2	f_{10}^2	f_{11}^2	$\Pr(w_{F^2})$	$N^{(n)}(F^2)$
f_{00}^3	f_{01}^3	f_{10}^3	f_{11}^3	$\Pr(w_{F^3})$	$N^{(n)}(F^3)$
$\dots$					$\dots$

4. Sort the obtained set by probabilities to get the Markov chain samples with given transition matrix in descending order.

$\mathbf{f_{00}}$	$\mathbf{f_{01}}$	$\mathbf{f_{10}}$	$\mathbf{f_{11}}$	$\mathbf{Pr}$	$\mathbf{N^{(n)}(F)}$
$f_{00}^{t_1}$	$f_{01}^{t_1}$	$f_{10}^{t_1}$	$f_{11}^{t_1}$	$\Pr(w_{F^{t_1}})$	$N^{(n)}(F^{t_1})$
$f_{00}^{t_2}$	$f_{01}^{t_2}$	$f_{10}^{t_2}$	$f_{11}^{t_2}$	$\Pr(w_{F^{t_2}})$	$N^{(n)}(F^{t_2})$
$f_{00}^{t_3}$	$f_{01}^{t_3}$	$f_{10}^{t_3}$	$f_{11}^{t_3}$	$\Pr(w_{F^{t_3}})$	$N^{(n)}(F^{t_3})$
$\dots$				$\dots$	$\dots$

Markov chain model

3. Evaluate the number of Markov chain samples with given transition matrix F.

$\mathbf{f_{00}}$	$\mathbf{f_{01}}$	$\mathbf{f_{10}}$	$\mathbf{f_{11}}$	$\mathbf{Pr}$	$\mathbf{N^{(n)}(F)}$
f_{00}^1	f_{01}^1	f_{10}^1	f_{11}^1	$\Pr(w_{F^1})$	$N^{(n)}(F^1)$
f_{00}^2	f_{01}^2	f_{10}^2	f_{11}^2	$\Pr(w_{F^2})$	$N^{(n)}(F^2)$
f_{00}^3	f_{01}^3	f_{10}^3	f_{11}^3	$\Pr(w_{F^3})$	$N^{(n)}(F^3)$
$\dots$					$\dots$

4. Sort the obtained set by probabilities to get the Markov chain samples with given transition matrix in descending order.

$\mathbf{f_{00}}$	$\mathbf{f_{01}}$	$\mathbf{f_{10}}$	$\mathbf{f_{11}}$	$\mathbf{Pr}$	$\mathbf{N^{(n)}(F)}$
$f_{00}^{t_1}$	$f_{01}^{t_1}$	$f_{10}^{t_1}$	$f_{11}^{t_1}$	$\Pr(w_{F^{t_1}})$	$N^{(n)}(F^{t_1})$
$f_{00}^{t_2}$	$f_{01}^{t_2}$	$f_{10}^{t_2}$	$f_{11}^{t_2}$	$\Pr(w_{F^{t_2}})$	$N^{(n)}(F^{t_2})$
$f_{00}^{t_3}$	$f_{01}^{t_3}$	$f_{10}^{t_3}$	$f_{11}^{t_3}$	$\Pr(w_{F^{t_3}})$	$N^{(n)}(F^{t_3})$
$\dots$				$\dots$	$\dots$

Markov chain model

5. Compile the set of vectors $\{s^j\}$ to be searched in: for each transition matrix from the obtained set and the corresponding initial and final states, restore all possible vectors $\{w\}$ and XOR them with q.
6. Calculate the average amount of work.

Markov chain model

5. Compile the set of vectors $\{s^j\}$ to be searched in: for each transition matrix from the obtained set and the corresponding initial and final states, restore all possible vectors $\{w\}$ and XOR them with q.
6. Calculate the average amount of work.

Markov chain model

M – size of high probable set
$N^{(n)}(F^t)$ – number of Markov chain samples with transition matrix F^t
Z – number of observed classes: $\sum_{t=1}^{Z-1} N^{(n)}(F^t) < M, \sum_{t=1}^{Z} N^{(n)}(F^t) \geq M$

Let's create a list γ:

$$\gamma[t] = N^{(n)}(F^t), \forall t = 1, \ldots, Z-1,$$

$$\gamma[Z] = \begin{cases} N^{(n)}(F^Z) & , \text{if } \sum_{t=1}^{Z} N^{(n)}(F^t) = M \\ M - \sum_{t=1}^{Z-1} N^{(n)}(F^t) & , \text{otherwise.} \end{cases}$$

Markov chain model

M – size of high probable set
$N^{(n)}(F^t)$ – number of Markov chain samples with transition matrix F^t
Z – number of observed classes: $\sum\limits_{t=1}^{Z-1} N^{(n)}(F^t) < M$, $\sum\limits_{t=1}^{Z} N^{(n)}(F^t) \geq M$

Let's create a list γ:

$$\gamma[t] = N^{(n)}(F^t), \forall t = 1, \ldots, Z-1,$$

$$\gamma[Z] = \begin{cases} N^{(n)}(F^Z) & \text{, if } \sum_{t=1}^{Z} N^{(n)}(F^t) = M \\ M - \sum_{t=1}^{Z-1} N^{(n)}(F^t) & \text{, otherwise.} \end{cases}$$

$$R_2(M, p_{00}, p_{11}) = \left(\sum_{i=1}^{Z} \mathsf{Pr}(F^i) \cdot \gamma[i] \right)^{-1} \left(M \cdot \left(1 - \sum_{i=1}^{Z} \mathsf{Pr}(F^i) \cdot \gamma[i] \right) + \right.$$
$$\left. + \sum_{i=1}^{Z} \frac{1}{2} \cdot \left(\sum_{j=1}^{i-1} \gamma[j] + 1 + \sum_{j=1}^{i} \gamma[j] \right) \cdot \gamma[i] \cdot \mathsf{Pr}(F^i) \right). \quad (2)$$

Evaluation of the average amount of work

r – *number of observed vector classes*
p_{00}, p_{11} – *elements of Markov chain transition probabilities matrix*
p_0^{st} – *Bernoulli scheme parameter*
$R_1(r, p_0^{st})$ – *the average amount of work for Bernoulli scheme*
$R_2(r, p_{00}, p_{11})$ – *the average amount of work for Markov chain model*

Table 3: Comparison of the results for two models. Non-truncated search.

$\|w\|$	p_{00}	p_{11}	r	$R_2(r, p_{00}, p_{11})$	p_0^{st}	$R_1(r, p_0^{st})$
255	0.3	0.5	255	$2^{244.87...}$	0.4167	$2^{249.91...}$
255	0.3	0.6	255	$2^{242.99...}$	0.3636	$2^{244.88...}$
255	0.3	0.8	255	$2^{220.58...}$	0.2222	$2^{218.56...}$
255	0.4	0.5	255	$2^{250.91...}$	0.4545	$2^{252.28...}$
255	0.4	0.6	255	$2^{248.57...}$	0.4	$2^{248.56...}$
255	0.4	0.7	255	$2^{241.02...}$	0.3333	$2^{240.98...}$
255	0.4	0.8	255	$2^{225.86...}$	0.25	$2^{225.59...}$

Evaluation of the average amount of work

r – *number of observed vector classes*
p_{00}, p_{11} – *elements of Markov chain transition probabilities matrix*
p_0^{st} – *Bernoulli scheme parameter*
$R_1(r, p_0^{st})$ – *the average amount of work for Bernoulli scheme*
$R_2(r, p_{00}, p_{11})$ – *the average amount of work for Markov chain model*

Table 4: Comparison of the results for two models. Truncated search.

$\|w\|$	p_{00}	p_{11}	r	$R_2(r, p_{00}, p_{11})$	p_0^{st}	$R_1(r, p_0^{st})$
255	0.3	0.5	55	$2^{216.08...}$	0.4167	$2^{224.76...}$
255	0.3	0.6	55	$2^{212.35...}$	0.3636	$2^{210.23...}$
255	0.3	0.8	55	$2^{188.43...}$	0.2222	$2^{188.72...}$
255	0.4	0.5	55	$2^{233.34...}$	0.4545	$2^{237.35...}$
255	0.4	0.6	55	$2^{219.93...}$	0.4	$2^{219.81...}$
255	0.4	0.7	55	$2^{200.68...}$	0.3333	$2^{203.52...}$
255	0.4	0.8	55	$2^{189.56...}$	0.25	$2^{191.20...}$

Evaluation of the average amount of work

R_1 – *the average amount of work for Bernoulli scheme*

R_2 – *the average amount of work for Markov chain model*

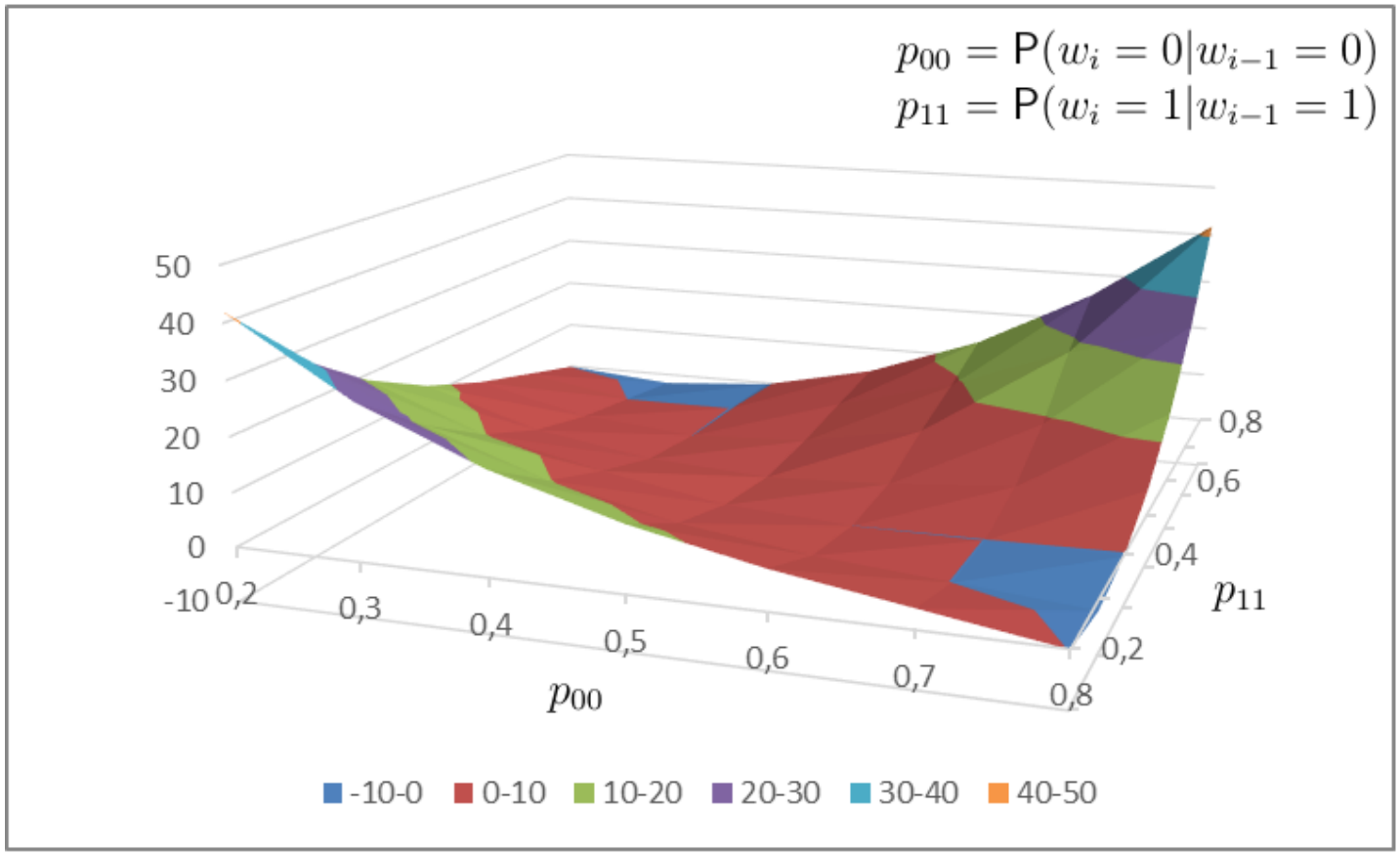


Figure 2: $(log_2 R_1 - log_2 R_2)$ for $n = 255$, non-truncated searching

Conclusion

- We've shown the low entropy of face biometric data. Our results correspond to the previous study of Ushmaev and Kuznetsov for fingerprint biometric data.
- We've studied the impact of statistical properties of biometric data on the security of fuzzy extractors.
- Observed effects should be taken into account while evaluating practical security of fuzzy extractor-based biometric identification systems.

Thank you for your attention!

Any questions?