Constructing Passwords from Biometrical Data

Vladimir B. Balakirsky, Anahit R. Ghazaryan, and A.J. Han Vinck*

Institute for Experimental Mathematics, Ellernstr. 29, 45326 Essen, Germany
v_b_balakirsky@rambler.ru, a_ghazaryan@rambler.ru, vinck@iem.uni-due.de

Abstract. We propose a probabilistic model for constructing passwords on the basis of outcomes of biometrical measurements. An algorithm for the transformation of biometrical data to passwords is given. The performance of the authentication scheme is evaluated by the compression factor, the false acceptance/rejection rates, the probability distribution over the set of passwords, and the probability of a correct guess of the input biometrical data mapped to the known password. An application of the results to DNA measurements is presented.

1 Probabilistic Model for Biometrical Data and Their Use for Authentication

Suppose that there is a non–stationary memoryless source generating a vector $\mathbf{b} = (b_1, \ldots, b_n)$, where $b_t \in \mathcal{B}_t$ and $\mathcal{B}_t = \{0, \ldots, Q_t - 1\}$, $t = 1, \ldots, n$. We also denote $\mathcal{B}^{(n)} = \mathcal{B}_1 \times \ldots \times \mathcal{B}_n$ and write $\mathbf{b} \in \mathcal{B}^{(n)}$. Let the probability distribution (PD) over the set $\mathcal{B}_t$ be given as

$$P_t = \Big(\Pr_{\text{bio}}\{B_t = b\},\, b \in \mathcal{B}_t \Big). \tag{1}$$

Thus, the probability of receiving the vector $\mathbf{b}$ is computed as

$$\Pr_{\text{bio}}\Big\{ B^{(n)} = \mathbf{b} \Big\} = \prod_{t=1}^{n} \Pr_{\text{bio}}\Big\{ B_t = b_t \Big\}, \tag{2}$$

where $B^{(n)} = (B_1, \ldots, B_n)$ denotes the vectors of random variables.

The vector $\mathbf{b}$ will be referred to as the biometric vector. For example, this vector can represent outcomes of $n = 28$ DNA measurements of a person [1]. Some basic characteristics of the PD's $P_1, \ldots, P_n$ can be found in [2] where different additive block coding schemes using the DNA data were developed. We will also use these data to illustrate the main points of the present contribution. In particular, to store the vector $\mathbf{b}$ in the database (DB), one needs $\sum_{t=1}^{n} \lceil \log Q_t \rceil$ bits, and this sum is equal to 140 for the DNA data.

The authentication of people is one of the reasons for processing biometrical data [3]. In this case, names and passwords of a certain group of people, called the users, are stored in the DB. Having received a pair (name, password), the

* This work was partially supported by the DFG.

M. Tistarelli and M.S. Nixon (Eds.): ICB 2009, LNCS 5558, pp. 889–898, 2009.

authentication scheme checks whether the pair is valid or not. In biometrical authentication procedures, passwords are formed on the basis of the outcomes of biometrical measurements of the users. We understand such a transformation as a mapping

$$F : \mathbf{b} \in \mathcal{B}^{(n)} \rightarrow \mathbf{z} \in \mathcal{Z}^{(n)}, \tag{3}$$

where $\mathcal{Z}^{(n)} = \mathcal{Z}_1 \times \ldots \times \mathcal{Z}_n$ and $\mathcal{Z}_t = \{0, \ldots, q_t - 1\}$, $t = 1, \ldots, n$, are fixed sets. We will also assume that $q_1 \leq Q_1, \ldots, q_n \leq Q_n$ and write $\mathbf{z} = F(\mathbf{b})$.

Let us discuss the basic requirements to the strings that can be effectively used as passwords, which are taken into account in our considerations.

1. The length of passwords is usually rather small (about 8 bytes), and passwords of different users "have to look as completely random sequences" to create difficulties to an attacker, who wants to guess the password of a certain user. This requirement to an algorithm of assigning passwords is not easy to formalize, unless a pseudo–random number generator is used. However, biometrical authentication assumes that designers of the system try to extract randomness from user's personal data and do not include any external source, i.e., the function F in (3) is a deterministic function. The problem, which appears in this context, is caused by the point that components of the biometric vector are generated by a non–uniform source. In particular, the probability of the most likely vector of outcomes of the DNA measurements is equal to $2^{-78.8}$, while the vectors have length 140 bits.
2. The biometrical characteristics of a fixed user can be non–typical. Nevertheless, the system has to provide the authentication for such a user as well. Therefore, we also have to study the performance of the authentication scheme for fixed biometrical vectors.
3. The outcomes of the biometrical measurements of the same person can be hardly exactly repeated. The absence of external randomness also implies that basic probabilistic characteristics of the authentication scheme, such as the false rejection rate (the probability that the identity claim of a user is rejected) and the false acceptance rate (the probability that the identity claim of a different user is accepted), have to be computed over the ensemble of biometric vectors and their noisy observations. We will present a probabilistic model for the noise of observations, which is needed for computing the false rejection rate. In our opinion, introduction of a proper model is the key point of processing biometrical data for authentication purposes.
4. The passwords of users are usually stored in a highly protected part of the DB where an attacker is not supposed to have any access. However, the biometrical direction in authentication also generates another types of attackers. Namely, suppose that the attacker knows the password of a certain user and he is interested in the biometrical characteristics of the person instead of passing through a fixed authentication test with the acceptance decision. The reasons for this interest come from the point that some parameters, like the DNA data, are very difficult to receive and, being compromised, these data are compromised forever.

The structure of a general biometrical authentication scheme under our considerations is given in Figure 1. The vector $\mathbf{b}^*$ and the corresponding password $\mathbf{z}^*$ are fixed. Another input to the verifier is a biometric vector $\mathbf{b}$, and we consider two alternatives denoted by "ind" and "dep":

$$\text{"ind"} \Rightarrow \Pr\Big\{B^{(n)} = \mathbf{b} \,|\, B^{(n)*} = \mathbf{b}^*\Big\} = \Pr_{\text{bio}}\Big\{B^{(n)} = \mathbf{b}\Big\}, \tag{4}$$

$$\text{"dep"} \Rightarrow \Pr\Big\{B^{(n)} = \mathbf{b} \,|\, B^{(n)*} = \mathbf{b}^*\Big\} = \Pr_{\text{err}}\Big\{B^{(n)} = \mathbf{b} \,|\, B^{(n)*} = \mathbf{b}^*\Big\}. \tag{5}$$

The conditional probabilities at the left–hand sides specify the source in Figure 1, which either knows the vector $\mathbf{b}^*$ or not. In the "ind" case, when the verifier has to reject the identity claim, the vector $\mathbf{b}$ is formed by a memoryless source introduced in (1), (2). In the "dep", when the verifier has to accept the identity claim, the vector $\mathbf{b}$ is formed by a memoryless source parameterized by the vector $\mathbf{b}^*$ in such a way that

$$\Pr_{\text{err}}\Big\{B^{(n)} = \mathbf{b} \,|\, B^{(n)*} = \mathbf{b}^*\Big\} = \prod_{t=1}^{n} \Pr_{\text{err}}\Big\{B_t = b_t \,|\, B_t^* = b_t^*\Big\},$$

where, for all $t = 1, \ldots, n$,

$$\Big(\Big(\Pr_{\text{err}}\Big\{B_t = b \,|\, B_t^* = b^*\Big\}, b \in \mathcal{B}_t\Big), b^* \in \mathcal{B}_t\Big) \tag{6}$$

is a collection of the conditional PD's satisfying the inequalities

$$\Pr_{\text{err}}\Big\{B_t = b^* \,|\, B_t^* = b^*\Big\} \geq 1 - \varepsilon \tag{7}$$

for all $b^* \in \mathcal{B}_t$ and a fixed $\varepsilon \in (0, 1/2)$.

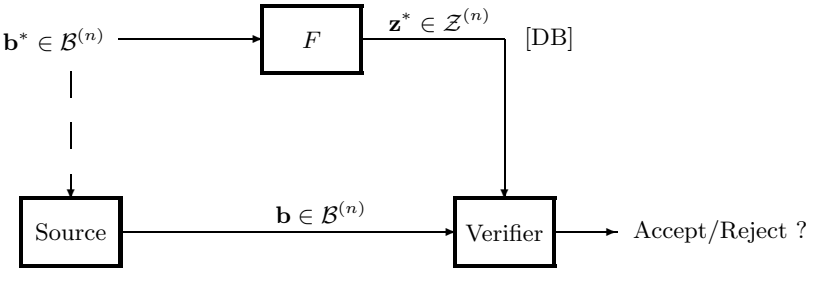

Fig. 1. The structure of a general biometrical authentication scheme

The model for the noise of observations above can be directly used for practical DNA measurements [1]. Some details of this application will be described in Section 5. We also think that this model is general enough in a sense that it can be used for other biometrical data in the following way. Since finding a probabilistic description of the noise of observations is usually very difficult (for example, in

the case when locations of minutiae points of the fingerprint are measured, the errors are caused by shifts and rotations of the finger, the light, the pressure, etc.), in practical biometrical systems, designers include an algorithm that tries to match two outcomes of the measurements. The acceptance/rejection decision is made on the basis of the comparison of the total number of observations and the number of observations that are matched by the program. Our model for the noise, where we only specify the probability that the input symbol is unchanged, assumes that the matching program is included into the channel between the sample and its observation.

In the following considerations, we concentrate on constructing passwords under the model described above. This problem is non–trivial because of the requirement that the strings stored in the DB have to look as randomly chosen strings, while the non–stationary source (1), (2) can be arbitrary. These considerations generalize some ideas developed for the non–stationary continuous sources [4] to the discrete case. In the continuous case, we assumed that there is a memoryless source given by the probability distribution functions $\Phi_1(y), \ldots, \Phi_n(y)$, $y \in \mathcal{R}$, where $\Phi_t(y) \in [0,1]$ is a non–decreasing function with $\Phi_t(-\infty) = 0$ and $\Phi_t(+\infty) = 1$ for all $t = 1, \ldots, n$. Then the transformation of an observed vector $\mathbf{y} = (y_1, \ldots, y_n) \in \mathcal{R}^n$ to $(\Phi_1(y_1), \ldots, \Phi_n(y_n)) \in [0,1]^n$ creates the vector whose components are independent random variables uniformly distributed over the $[0,1]$ interval. The quantization of each component into 2^L levels by using the thresholds $i2^{-L}$, $i = 0, \ldots, 2^L$, where L is a fixed integer, generates a binary vector of length Ln uniformly distributed over the set $\{0,1\}^{Ln}$, and this statement is true for any L. Furthermore, the distance between two arguments y and y' of the function Φ_t defined as $|y - y'|$ is translated into the distance defined as $|\Phi_t(y) - \Phi_t(y')|$, by the properties of the function Φ_t. In the discrete case we cannot use this approach one–by–one, but its ideas are relevant.

2 Restricted Biometrical Authentication Schemes and Their Parameters

Let us restrict the considerations to the following case.

1. Let
$$\max_{b \in \mathcal{B}_t} \Pr_{\text{bio}} \Big\{ B_t = b \Big\} < 1 - \varepsilon. \tag{8}$$
2. Let F be a component–wise mapping determined by functions $f_t : \mathcal{B}_t \to \mathcal{Z}_t$, $t = 1, \ldots, n$, in such a way that
$$F(\mathbf{b}) = (f_1(b_1), \ldots, f_n(b_n))$$
for all $\mathbf{b} \in \mathcal{B}^{(n)}$.
3. Let the verifier make the acceptance/rejection decision on the basis of the pair $(F(\mathbf{b}^*), F(\mathbf{b}))$ instead of the pair $(F(\mathbf{b}^*), \mathbf{b})$.

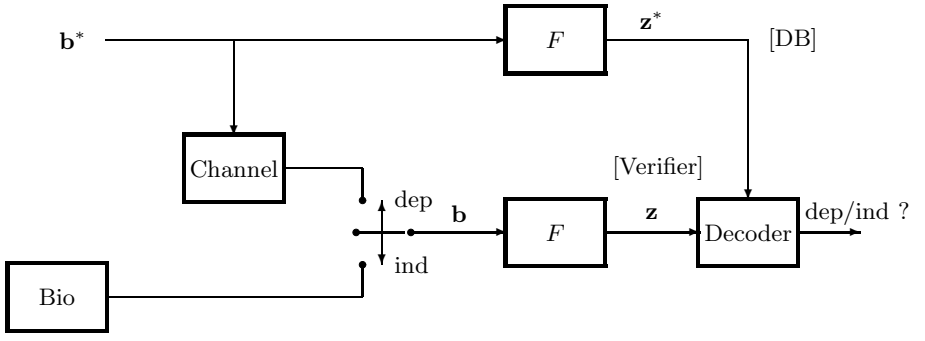

Fig. 2. Representation of the restricted biometrical authentication scheme

The structure of the restricted biometrical authentication scheme is illustrated in Figure 2. We represent the relationships given in (4), (5) as two possibilities: the vector $\mathbf{b}$ is either received as a result of transmission of the vector $\mathbf{b}^*$ over a channel or it is independently generated by the biometric source. The verification is split in two blocks. At first, the password $F(\mathbf{b})$ is computed and then the decoder processes the pair of passwords to make a decision. In general case, the decoding algorithm is specified by the sets $\mathcal{A}(\mathbf{z}^*) \subseteq \mathcal{Z}^{(n)}$, $\mathbf{z}^* \in \mathcal{Z}^{(n)}$, in such a way that the decision is "dep"(accept) if $F(\mathbf{b}) \in \mathcal{A}(\mathbf{z}^*)$ and "ind"(reject) if $F(\mathbf{b}) \notin \mathcal{A}(\mathbf{z}^*)$. Our probabilistic model for the noise assumes that the verifier has to make a decision on the basis of the value of the Hamming distance between received vectors, $\mathcal{A}(\mathbf{z}^*) = \mathcal{A}_\rho(\mathbf{z}^*)$, where

$$\mathcal{A}_\rho(\mathbf{z}^*) = \Big\{ \mathbf{z} \in \mathcal{Z}^{(n)} : \, d(\mathbf{z}, \mathbf{z}^*) \le n\rho \Big\}, \tag{9}$$

where $\rho \in [0, 1]$ is a specified parameter and

$$d(\mathbf{z}, \mathbf{z}^*) \stackrel{\triangle}{=} \Big| \Big\{ t \in \{1, \ldots, n\} : \, z_t \ne z_t^* \Big\} \Big|.$$

Thus,

$$\text{Decision} = \begin{cases} \text{"ind"}, \text{ if } d(\mathbf{z}, \mathbf{z}^*) > n\rho, \\ \text{"dep"}, \text{ if } d(\mathbf{z}, \mathbf{z}^*) \le n\rho, \end{cases}$$

The probabilities of incorrect decisions (the false acceptance and the false rejection rates) for the vector $\mathbf{b}^*$ that is mapped to the vector $\mathbf{z}^*$ are expressed as

$$\text{FAR}(\mathbf{b}^*) = \sum_{\mathbf{b}:\, F(\mathbf{b}) \in \mathcal{A}_\rho(\mathbf{z}^*)} \Pr_{\text{bio}} \Big\{ B^{(n)} = \mathbf{b} \Big\},$$

$$\text{FRR}(\mathbf{b}^*) = \sum_{\mathbf{b}:\, F(\mathbf{b}) \notin \mathcal{A}_\rho(\mathbf{z}^*)} \Pr_{\text{err}} \Big\{ B^{(n)} = \mathbf{b} \,|\, B^{(n)*} = \mathbf{b}^* \Big\}.$$

One of main parameters characterizing the performance of the algorithm is

$$C \stackrel{\triangle}{=} \frac{\sum_{t=1}^{n} \lceil \log Q_t \rceil}{\sum_{t=1}^{n} \lceil \log q_t \rceil},$$

which can be understood as the compression factor. The other parameters introduced below are relevant to the analysis of the false acceptance/false rejection rates and the cryptographic properties. One can easily see that if an attacker wants to guess the biometric vector of a person, then the best estimate is the vector having the maximum probability. The probability that this guess is correct is equal to $\hat{\omega} \stackrel{\triangle}{=} \prod_{t=1}^{n} \hat{\omega}_t$, where $\hat{\omega}_t \stackrel{\triangle}{=} \max_{b \in \mathcal{B}_t} \omega_t(b)$ and

$$\omega_t(b) \stackrel{\triangle}{=} \Pr_{\text{bio}}\Big\{B_t = b\Big\}, \quad b \in \mathcal{B}_t.$$

The transformation of biometric vectors to passwords and changing the attacker's task as guessing the password makes the probability of success equal to $\hat{\pi} \stackrel{\triangle}{=} \prod_{t=1}^{n} \hat{\pi}_t$, where $\hat{\pi}_t \stackrel{\triangle}{=} \max_{z \in \mathcal{Z}_t} \pi_t(z)$ and

$$\pi_t(z) = \sum_{b:\, f_t(b)=z} \Pr_{\text{bio}}\Big\{B_t = b\Big\}, \quad z \in \mathcal{Z}_t, \tag{10}$$

is the PD over the t-th component of the passwords. Furthermore, the best prediction of the input biometric vector, given the password $\mathbf{z}^*$, is the biometric vector having the maximum probability among the vectors mapped to $\mathbf{z}^*$. As it is easy to see, the probability that this prediction is correct can be computed as $\hat{\gamma}(\mathbf{z}^*) \stackrel{\triangle}{=} \prod_{t=1}^{n} \hat{\gamma}_t(z_t^*)$, where $\hat{\gamma}_t(z_t^*) \stackrel{\triangle}{=} \max_{b \in \mathcal{B}_t} \gamma_t(b|z_t^*)$ and

$$\gamma_t(b|z_t^*) \stackrel{\triangle}{=} \frac{1}{\pi_t(z_t^*)} \begin{cases} \omega_t(b), & \text{if } f_t(b) = z_t^*, \\ 0, & \text{if } f_t(b) \neq z_t^*, \end{cases} \quad b \in \mathcal{B}_t,$$

is the z_t^*-conditional PD over the t-th component of the passwords.

The notation above is illustrated in Table 1, where

$$\overline{\gamma}_t \stackrel{\triangle}{=} \sum_{z^* \in \mathcal{Z}_t} \pi_t(z) \hat{\gamma}_t(z^*), \quad \hat{\gamma} \stackrel{\triangle}{=} \max_{z^* \in \mathcal{Z}_t} \hat{\gamma}_t(z^*)$$

denote the average and the maximum probability of the successful guess of the input biometric vector. In the stationary case, when the biometric source is described by the presented PD for all $t = 1, \ldots, n$, the probabilities of the correct guess of the biometric vector and the password are equal to $(0.2401)^n$ and $(0.2596)^n$, i.e., they are close enough. The guessing algorithm for the attacker, who has access to the password and wants to find the biometric vector, depends on the password. The average and the maximum probabilities of success are equal to $(0.7401)^n$ and $(0.9674)^n$, respectively.

The false acceptance/false rejection rates can be expressed using the notation above. These expressions and proofs can be found in [5]. In the present correspondence we restrict ourselves to the description of the algorithm for constructing the functions $f_1, \ldots, f_n$ and to the application of the verification procedure to the biometric systems with the DNA measurements.

Table 1. Example of the mapping $f_t : \{0, \dots, 9\} \to \{0, \dots, 3\}$

$b =$	9	7	3	5	6	0	8	1	2	4
$\omega_t(b) =$	.2401	.0081	.2254	.0342	.1862	.0529	.0882	.0874	.0414	.0361
$z = f_t(b) =$	0		1		2		3			
$\pi_t(z) =$	.2482		.2596		.2381		.2531			
$\gamma_t(b\|z) =$	.9674	.0326	.8682	.1318	.7820	.2180	.3485	.3453	.1636	.1426
$\hat{\gamma}_t, \overline{\gamma}_t =$	.9674,.7401									
$\hat{\pi}_t =$	.2596									
$\hat{\omega}_t =$	.2401									

3 Constructing the Functions $f_1, \dots, f_n$

In the following considerations, we omit the index $t \in \{1, \dots, n\}$ for a formal brevity and extend the ideas presented in [4] for the continuous case to the discrete case.

Let us determine the function f by a partitioning of the set $\mathcal{B}$ by q pairwise disjoint subsets $\mathcal{F}(0), \dots, \mathcal{F}(q-1)$ in such a way that $f(b) = z$ is equivalent to $b \in \mathcal{F}(z)$. For example, the function f with $f(0) = f(2) = 0$, $f(1) = 1$, $f(3) = 2$ is specified by the sets $\mathcal{F}(0) = \{0, 2\}$, $\mathcal{F}(1) = \{1\}$, $\mathcal{F}(2) = \{3\}$. "A greedy algorithm" for constructing the sets $\mathcal{F}(0), \dots, \mathcal{F}(q-1)$ is presented below.

F1: Set $z = 0$.
F2: Set $\mathcal{F}(z) = \emptyset$, $S = 0$, $\Delta = 2^{-q}$.
F3: Denote

$$\mathcal{B}_0 = \mathcal{F}(z) \bigcup \Big[\bigcup_{z'=0}^{z-1} \mathcal{F}(z') \Big],$$

$$\Delta_0 = \min_{b \in \mathcal{B} \setminus \mathcal{B}_0} \Big| 2^{-q} - (S + \omega(b)) \Big|,$$

$$b_0 = \arg \min_{b \in \mathcal{B} \setminus \mathcal{B}_0} \Big| 2^{-q} - (S + \omega(b)) \Big|.$$

If $\Delta_0 > \Delta$, then go to 5.
F4: Include b_0 into the set $\mathcal{F}(z)$, increase S by $\omega(b_0)$, and substitute Δ_0 for Δ.
F5: Increase z by 1. If $z \le q - 2$, then go to 2.
F6: Set

$$\mathcal{F}(q-1) = \mathcal{B} \setminus \Big[\bigcup_{z=0}^{q-2} \mathcal{F}(z) \Big].$$

F7: Output the sets $\mathcal{F}(0), \dots, \mathcal{F}(q-1)$. End.

4 Application of the Authentication Algorithm to the DNA Measurements

We will use the following mathematical model for the DNA measurements [2]. Suppose that there are n sources. Let the t-th source generate a pair of integers according to the PD

$$\Pr_{\text{DNA}}\Big\{ (R_{t,1}, R_{t,2}) = (r_{t,1}, r_{t,2}) \Big\} = p_t(r_{t,1}) p_t(r_{t,2}),$$

where $r_{t,1}, r_{t,2} \in \mathcal{R}_t = \{c_t, \ldots, c_t + k_t - 1\}$ and integers $c_t, k_t > 0$ are given. The outcome of the t-th measurement is defined as

$$R_t \stackrel{\triangle}{=} \Big(\min\{R_{t,1}, R_{t,2}\}, \max\{R_{t,1}, R_{t,2}\} \Big). \tag{11}$$

Hence, for all $i \in \mathcal{R}_t$,

$$\Pr_{\text{DNA}}\Big\{R_t = (i,j)\Big\} = \begin{cases} 0, & \text{if } j \in \{0, \ldots, i-1\}, \\ p_t^2(i), & \text{if } j = i, \\ 2p_t(i)p_t(j), & \text{if } j \in \{i+1, \ldots, c_t + k_t - 1\}. \end{cases}$$

We assume that $R_1, \ldots, R_n$ are mutually independent pairs of random variables, i.e.,

$$\Pr_{\text{DNA}}\Big\{R^{(n)} = \mathbf{r}\Big\} = \prod_{t=1}^{n} \Pr_{\text{DNA}}\Big\{R_t = r_t\Big\},$$

where $R^{(n)} = (R_1, \ldots, R_n)$ and $\mathbf{r} = (r_1, \ldots, r_n)$, $r_t \in \mathcal{R}_t \times \mathcal{R}_t$. To make the notation consistent with the notation of Section 2, let us map $Q_t = k_t(k_t+1)/2$ pairs $r_t = (i_t, j_t)$, where $j_t \geq i_t$, that can occur with positive probability to integers $b \in \mathcal{B}_t = \{0, \ldots, Q_t - 1\}$ in a lexicographic order.

The formalization above appears because the DNA measurements are usually understood as measurements of the numbers of repeats of certain motifs in the paternal and the maternal allele where the measuring device cannot distinguish between data coming from different allele. Therefore, the outcomes $r_{t,1}, r_{t,2}$ can be represented as observations of the sets $\{r_{t,1}, r_{t,2}\}$. This information can be equivalently presented as the value of the random variable R_t defined in (11).

We present parameters of the PDs obtained from the DNA measurements, which are relevant to the evaluation of the performance of the authentication algorithm, for the **TH01** allele and the total sums obtained for $n = 28$ allele in Table 2. A more complete version of this table can be found in [5]. One can see that the storage of biometric vectors requires 140 bits and the PD over these vectors is non–uniform, as the probability of the most likely vector is equal to $2^{-78.8}$. The encoding with parameters $q_t = \lceil \log 1/\hat{\omega}_t \rceil$ creates passwords of length 68 with the PD close to the uniform PD (the probability of the most likely

Table 2. Parameters of three variants of the encoding for the biometrical authentication with the DNA measurements when repeats of certain motifs in $n = 28$ allele, numbered by $t = 1, \ldots, n$, are measured

t	Name	$\log Q_t$	$\log \hat{\omega}_t$	$C = 140/140$			$C = 140/68$			$C = 140/28$		
				$\log q_t$	$\log \hat{\pi}_t$	$\log \hat{\gamma}_t$	$\log q_t$	$\log \hat{\pi}_t$	$\log \hat{\gamma}_t$	$\log q_t$	$\log \hat{\pi}_t$	$\log \hat{\gamma}_t$
12	**TH01**	3.32	−2.07	4	−2.07	0	2	−1.93	−0.04	1	−1.00	−1.07
	$\sum_n$	128.6	−78.8	140	−78.8	0	68	−66.7	−9.0	28	−27.8	−49.0

vector is equal to $2^{-66.7}$). However, the maximum probability of the correct guess of the biometric vector is equal to $2^{-9.0}$. This probability can be decreased to $2^{-49.0}$ by assigning $q_t = 2, t = 1, \ldots, n$, which creates passwords of length 28 bits. However, the false acceptance rate will be also increased, as it is illustrated in the end of the section. A possible implementation of our transformations of input biometrical data is presented by the example below.

Example. (the quantities below describe the **TH01** allele in Table 2, $t = 12$). Let $c_t = 6$, $k_t = 4$, and $(p_t(6), \ldots, p_t(9)) = (0.23, 0.19, 0.09, 0.49)$. Then

$$\Big[p_t(i)p_t(j)\Big]_{i,j=6,\ldots,9} =$$

	$j=6$	$j=7$	$j=8$	$j=9$
$i=6$	.0529	.0437	.0207	.1127
$i=7$	.0437	.0361	.0171	.0931
$i=8$	.0207	.0171	.0081	.0441
$i=9$	.1127	.0931	.0441	.2401

To construct the PD ω_t, we transform this matrix to the right triangular matrix below. The entries above the diagonal are doubled, and the entries below the diagonal are replaced with zeroes,

	$j=6$	$j=7$	$j=8$	$j=9$
$i=6$	.0529	.0874	.0414	.2254
$i=7$		.0361	.0342	.1862
$i=8$			.0081	.0882
$i=9$				.2401

Let $q_t = 4$ and let the sets $\mathcal{F}_t(0), \ldots, \mathcal{F}_t(3)$ be constructed by the F1–F7 algorithm. Then $\mathcal{F}_0(z) = \{9, 7\}$, $\mathcal{F}_1(z) = \{3, 5\}$, $\mathcal{F}_2(z) = \{6, 0\}$, $\mathcal{F}_3(z) = \{8, 1, 2, 4\}$. The PDs obtained using this partitioning of the set $\{0, \ldots, 9\}$ were presented in Table 1.

To implement the encoding, the authentication scheme needs the table below.

$(i,j) =$	$(6,6)$	$(6,7)$	$(6,8)$	$(6,9)$	$(7,7)$	$(7,8)$	$(7,9)$	$(8,8)$	$(8,9)$	$(9,9)$
$b_t =$	0	1	2	3	4	5	6	7	8	9
$z_t =$	3	3	2	1	2	1	3	0	3	0

The outcome of the measurements $(i_{t,1}, j_{t,2})$ has to be found in the first row, and the symbol z in the corresponding column is sent to the output. Notice that the index of the pair $(i_{t,1}, j_{t,2})$ can be easily computed from $c_t = 6$ and $k_t = 4$. Therefore, the storage of the first and the second rows of the table is not necessary.

Some numerical results are given in Table 3 where we show the values of the false acceptance rate as a function of ρ and the length of passwords. These data illustrate the point that our approach is robust in a sense that it can reach a desired trade–off between different parameters characterizing the performance.

Table 3. The DNA measurements: the values of the false acceptance rate and the parameter $\varepsilon(\text{FAR})$ such that $\text{FRR} \leq \text{FAR}$ for all $\varepsilon < \varepsilon(\text{FAR})$

	$C = 140/140$		$C = 140/68$		$C = 140/28$	
$n\rho$	FAR	$\varepsilon(\text{FAR})$	FAR	$\varepsilon(\text{FAR})$	FAR	$\varepsilon(\text{FAR})$
3	1.3e−17		7.0e−15		3.8e−02	0.0376
4	5.9e−16	0.0001	2.4e−13	0.0003	6.7e−02	0.0672
5	2.0e−14	0.0006	6.2e−12	0.0016	1.1e−01	0.1524
6	5.0e−13	0.0024	1.2e−10	0.0053		

References

[1] Korte, U., Krawczak, M., Merkle, J., Plaga, R., Niesing, M., Tiemann, C., Han Vinck, A.J., Martini, U.: A cryptographic biometric authentication system based on genetic fingerprints. Sicherheit, 263–276 (2008)

[2] Balakirsky, V.B., Ghazaryan, A.R., Han Vinck, A.J.: Additive block coding schemes for biometric authentication with the DNA data. In: Schouten, B., Jul, N.C., Drygajlo, A., Tistarelli, M. (eds.) BIOID 2008. LNCS, vol. 5372, pp. 160–169. Springer, Heidelberg (2008)

[3] Bolle, R.M., Connell, J.H., Pankanti, S., Ratha, N.K., Senior, A.W.: Guide to Biometrics. Springer, NY (2004)

[4] Balakirsky, V.B., Ghazaryan, A.R., Han Vinck, A.J.: An Algorithm for Biometric Authentication Based on the Model of Non-Stationary Random Processes. In: Lee, S.-W., Li, S.Z. (eds.) ICB 2007. LNCS, vol. 4642, pp. 319–327. Springer, Heidelberg (2007)

[5] Balakirsky, V.B., Ghazaryan, A.R., Han Vinck, A.J.: Mathematical model for constructing passwords from biometrical data. In: Security and Communication Networks, vol. 2(1), pp. 1–9. Wiley, Chichester (2009)