Mathematical Problems of Computer Science 37, 7–16, 2012.

Upper and Lower Bounds of Biometric Identification E-Capacity

Mariam E. Haroutunian, Lilit A. Ter-Vardanyan and Arthur R. Muradyan

Institute for Informatics and Automation Problems of NAS RA
armar@ipia.sci.am, lilit@sci.am, mur_art@yahoo.com

Abstract

In this paper we introduce a new concept of E-capacity for biometric indentification system, which is the generalization of the capacity studied by Willems et al [1]. We investigate this function by constructing upper and lower bounds. When $E \to 0$ we derive the lower and upper bounds of the channel capacity which coincides with the capacity obtained in [1].
Keywords: Biometric identification system, identification capacity, E-capacity bounds, error exponents.

1. Introduction

Reliable communication and security are very sensitive issues for modern global society with a wide range of application domains. One of those domains is biometrics. Biometrics is now used for physical access control, computer log-in, welfare disbursement, international border crossing and national ID cards, e-passports. It can be used to verify a customer during transactions conducted via telephone and Internet (electronic commerce and electronic banking). In automobile, biometrics is now adopted to replace keys for keyless entry and keyless ignition.

One of the main issues in biometric security is the reliable identification of persons based on their biometric data.

The objective of a biometric identification system is to identify individuals on the basis of physical features. One of the oldest and probably best known of such features is the human fingerprint. Over the last decade other human features have become practical, and there is now an active research community on iris-based recognition, face recognition, voice recognition and others [2].

Willems et al [1,3] investigated the fundamental properties of biomertic identification system. It has been shown that it is not possible to identify reliably more persons than capacity the which is an inherent characteristic of any identification system. They derived the capacity of such system.

We are interested in a number of individuals that can reliably be identified by a biometric identification system in dependence of the quality of the observations. Within this system our question has an answer in terms of an information-theoretical quantity.

We investigate the exponentially high reliability criterion in biometric identification systems. In other words we introduce a new performance concept of biometric identification E-capacity, which takes into account a stronger requirement on identification fault events with extremely small probability (2^{-NE} instead of ε). In terms of practical applications an exponential decrease in error probability (namely, in unwanted identification faults) is more desirable. We investigate the E-capacity function introduced for DMC by E. Haroutunian [4], which is the generalization of the capacity, as it tends to capacity, when E tends to 0. E-capacity has been investigated for a series of channels [4-7].
The model of biometric identification system consists of two procedures: enrollment and identification.

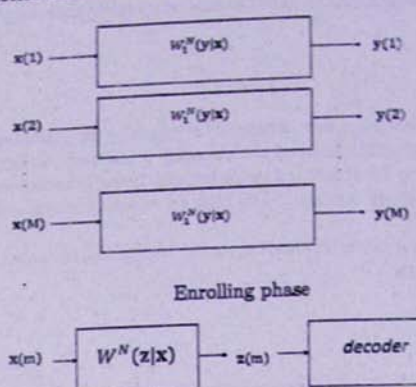

Fig. 1. Model of biometric identification system

In an enrollment phase M individuals are observed and for each individual a noisy version of the biometric data is added to the database. In the identification phase an unknown individual is observed and another noisy version of the biometric data is compared to the enrollment data in the database. The system has to come up with an estimate of the individual.

In this paper the upper and lower bounds of identification E-capacity for maximal and average error probabilities are constructed. When $E \to 0$ we derive the corresponding bounds of the capacity of the biometric identification system, which coincide with the capacity obtained in [1]. We also give numerical representations of the results to simplify the solutions in view of applications.

2. Notations and Definitions

The following conventions are applied within the paper. Capital letters are used for random variables (RV) X, Y, Z taking values in the finite sets $\mathcal{X}, \mathcal{Y}, \mathcal{Z}$, correspondingly, and lower case letters x, y, z - for their realizations. Small bold letters are used for N-length vectors $\mathbf{x} = (x_1, .., x_N) \in \mathcal{X}^N$. The cardinality of the set $\mathcal{X}$ is denote by $|\mathcal{X}|$. The notation $|a|^+$ will be used for $\max(a, 0)$.

There are M individuals and each individual has an index $m = \{1, 2, \cdots, M\}$. A biometric

data sequence $\mathbf{x}(m) = \{x_1, x_2, \cdots, x_N\}$, where $x_n \in \mathcal{X}, n = \overline{1,N}$ corresponds to each individual m. All these sequences are supposed to be generated at random with a given probability distribution

$$Q^N(\mathbf{x}) = \prod_{n=1}^{N} Q(x_n), \mathbf{x} \in \mathcal{X}^N.$$

Enrollment phase. In this phase all biometric data sequences $\mathbf{x}(m)$ are observed via a discrete memoryless enrollment channel $W_1(y|x)$ with finite input alphabet $\mathcal{X}$ and output alphabet $\mathcal{Y}$,

$$W_1^N(\mathbf{y}|\mathbf{x}) = \prod_{n=1}^{N} W_1(y_n|x_n), \quad \mathbf{x} \in \mathcal{X}^N, \quad \mathbf{y} \in \mathcal{Y}^N.$$

The resulting $\mathbf{y}(m)$ enrollment output sequences for all $m = \{1, 2, \cdots, M\}$ are stored in a database (define it as Y_{DB}).

Identification phase. In the identification phase the biometric data sequence of an unknown individual is observed via a memoryless identification channel $W_2(z|x)$ with output alphabet $\mathcal{Z}$,

$$W_2^N(\mathbf{z}|\mathbf{x}) = \prod_{n=1}^{N} W_2(z_n|x_n), \mathbf{z} \in \mathcal{Z}^N, \mathbf{x} \in \mathcal{X}^N.$$

The resulting identification output sequence $\mathbf{z}$ is compared to the sequences $\mathbf{y}(m)$, $m = 1, 2, \cdots, M$, from the database and the identification function

$$g_N : \mathcal{Z}^N \to \{0, 1, 2, \cdots, M\}$$

produces the index of the unknown individual $m' = g_N(\mathbf{z})$, here 0 stands for the case, when the unknown individual has not been observed by the enrollment phase. Following [1], we do not consider the probability who an individual, who did not undergo the enrollment procedure, is identified as one of the individuals that has been enrolled properly.

The following probability distributions are given

$$P^* = \{P^*(y) = \sum_x W_1(y|x)Q(x), \quad x \in X, \quad y \in Y\},$$

$$W(z|y) = \frac{\sum_x W_1(y|x)W_2(z|x)Q(x)}{P^*(y)}.$$

The channel W is assumed to be memoryless:

$$W^N(\mathbf{z}|\mathbf{y}) = \prod_{n=1}^{N} W(z_n|y_n), \quad \mathbf{z} \in \mathcal{Z}^N, \quad \mathbf{y} \in \mathcal{Y}^N.$$

One of the main parameters of the system is the rate, which is supposed to be constant

$$R = \frac{1}{N} \log_2 M.$$

The next parameter is the error probability

$$e(N, m) = W^N(\mathcal{Z}^N \backslash g_N^{-1}(m)|\mathbf{y}(m)),$$

where $$g_N^{-1}(m) = \{z : g_N(z) = m\}.$$

We consider the **maximal** and the **average error probabilities**

$$e(N) = \max_{m\in\{1,2,\cdots,M\}} e(N,m),$$

$$\overline{e}(N) = \frac{1}{M}\sum_{m\in\{1,2,\cdots,M\}} e(N,m).$$

The E-capacity function for the given $E > 0$ is defined as

$$C(E,P^*,W) = \overline{\lim_{N\to\infty}}\frac{1}{N}\log M(E,P^*,W,N),$$

where $$M(E,P^*,W,N) = \sup_{g_N}\{M : e(N) \le \exp(-NE)\}.$$

We denote by $\overline{C}(E,P^*,W)$ the E-capacity for the average error probability.

We shall use the following PD in the formulation of results:

$$P = \{P(y), y \in \mathcal{Y}\},$$

$$V = \{V(z|y), z \in \mathcal{Z}, y \in \mathcal{Y}\}.$$

For the information-theoretic quantities, such as entropy $H_P(Y)$, mutual information $I_{P,V}(Z \wedge Y)$, divergence $D(V||W|P)$ and for the notion of the type we refer to [4]– [9].

3. Formulation of Results

To define the lower bound (*random coding bound*) of the identification E-capacity let us denote:

$$R_r(E,P^*,W) \triangleq$$

$$\triangleq \min_{P,V:D(P\circ V||P^*\circ W)\le E}\left|I_{P,V}(Z\wedge Y) + D(P\circ V||P^*\circ W) - E\right|^+. \quad (1)$$

For the formulation of the upper bound (*sphere packing bound*) of the indentification E-capacity let us introduce the following function:

$$R_{sp}(E,P^*,W) \triangleq \min_{P,V:D(P\circ V||P^*\circ W)\le E} I_{P,V}(Z\wedge Y). \quad (2)$$

Theorem. *For the biometric identification system with the given P^*, W and for all $E > 0$*

$$R_r(E,P^*,W) \le C(E,P^*,W) \le \overline{C}(E,P^*,W) \le R_{sp}(E,P^*,W).$$

For the proof of the theorem we use the method of types [4,5,9].

The proof of the theorem is exposed in sections [4,5].

Corollary. *When $E \to 0$ we derive the lower and upper bounds of the channel capacity, which coincide with the capacity obtained in [1]*

$$C = I_{P^*,W}(Z\wedge Y).$$

4. Proof of Lower Bound

The generation of the biometric data is random, so in database we have a random code Y_{DB}. For decoding the minimum divergence method is used [4-6]. According to this method $\mathbf{z}$ is decoded to such m for which:

$$\mathbf{z} \in \mathcal{T}^N_{P,V}(Z|\mathbf{y}(m))$$

and PDs P, V are such that

$$D(P \circ V||P^* \circ W)$$

is minimal.

The decoder g_N can make an error, if for $\mathbf{y}(m)$ there exists $\mathbf{y}'(m')$ $(m \neq m')$, types P', V' such that

$$\mathbf{z} \in \mathcal{T}^N_{P,V}(Z|\mathbf{y}(m)) \bigcap \mathcal{T}^N_{P',V'}(Z|\mathbf{y}'(m'))$$

and

$$D(P' \circ V'||P^* \circ W) \leq D(P \circ V||P^* \circ W). \tag{3}$$

Let us denote by $\mathcal{D} = \{P, V, P', V' : (3) \text{ is valid}\}$, then

$$e(m) \leq W^N\Big\{\bigcup_{\mathcal{D}} \mathcal{T}^N_{P,V}(Z|\mathbf{y}(m)) \bigcap \bigcup_{m' \neq m} \mathcal{T}^N_{P',V'}(Z|\mathbf{y}'(m'))\Big|\mathbf{y}(m)\Big\}$$

$$\leq \sum_{\mathcal{D}} E\Big|\mathcal{T}^N_{P,V}(Z|\mathbf{y}(m)) \bigcap \bigcup_{m' \neq m} \mathcal{T}^N_{P',V'}(Z|\mathbf{y}'(m'))\Big| \times W^N(\mathbf{z}|\mathbf{y}(m)).$$

The last inequality is true, because for the fixed types of $\mathbf{z}$, $\mathbf{y}(m)$ the probabilities are constant.

We shall use the modification of packing lemma from [4-6].

Packing Lemma. *For the given P^*, W, for any $E > \delta \geq 0$ and*

$$M \geq \exp\Big\{N \min_{P^*,V:D(P \circ V||P^* \circ W) \leq E} \Big|I_{P,V}(Z \wedge Y)$$

$$+D(P \circ V||P^* \circ W) - E - \delta\Big|^+\Big\} \tag{4}$$

for sufficiently large N the following inequality holds for any types P, P', V, V' and for all $m = \overline{1, M}$

$$E\Big|\mathcal{T}^N_{P,V}(Z|\mathbf{y}(m)) \bigcap \bigcup_{m' \neq m} \mathcal{T}^N_{P',V'}(Z|\mathbf{y}'(m'))\Big|$$

$$\leq \Big|\mathcal{T}^N_{P,V}(Z|\mathbf{y}(m))\Big| \exp\Big\{-N\Big|E - D(P' \circ V'||P^* \circ W)\Big|^+\Big\} \exp\{-ND(P||P^*)\}. \tag{5}$$

We omit the proof of lemma as it includes basic steps of the proof for packing lemma.

Taking into account (3) and (5), we can upper estimate the *error probability* of identification of the individual m:

$$e(N, m) \leq \sum_{\mathcal{D}} \Big|\mathcal{T}^N_{P,V}(Z|\mathbf{y}(m))\Big|$$

$$\times \exp\Big\{-N(E-D(P'\circ V''\|P^*\circ W))\Big\}\times\exp\{-ND(P\|P^*)\}$$

$$\times W^N(\mathbf{z}|\mathbf{y}(m))\le\sum_{\mathcal{D}}\exp\{NH_{P,V}(Z|Y)\} \tag{6}$$

$$\times\exp\{-N(E-D(P'\circ V''\|P^*\circ W))\}\times\exp\{-ND(P\|P^*)\}$$
$$\times\exp\{-N(H_{P,V}(Z|Y)+D(V\|W|P))\}$$
$$\le\sum_{P,V,P',V'}\exp\{-NE\}\le\exp\{-N(E-\varepsilon)\}.$$

The last inequality is true as the number of all types P,V,P',V' is not greater than $(N+1)^{2|\mathcal{X}||\mathcal{Y}||\mathcal{Z}|}$. Considering the continuity of all expressions, when $N\to\infty$, the arbitrary probability distributions can be considered instead of the types.

5. Proof of Upper Bound

Let $E>0$ and the average error probability satisfies the condition

$$\bar{e}(N)\le\exp\{-N(E-\delta)\}.$$

It means that

$$\frac{1}{M}\sum_{m\in M}W_1^N(\mathcal{Z}^N-g^{-1}(m)|\mathbf{y}(m))\le\exp\{-NE\}. \tag{7}$$

For each P and V we can write

$$\sum_{m\in\mathcal{T}_P(Y)\cap Y_{DB}}W_1^N(\mathcal{T}_{P,V}(Z|\mathbf{y}(m))-g^{-1}(m)|\mathbf{y}(m))\le M\exp\{-NE\}.$$

Since $W_1^N(\mathbf{z}|\mathbf{y})$ is constant for various $\mathbf{z}$ and $\mathbf{y}$ of fixed P,V, then

$$\sum_{m\in\mathcal{T}_P(Y)\cap Y_{DB}}\Big\{|\mathcal{T}_{P,V}(Z|\mathbf{y}(m))|-|\mathcal{T}_{P,V}(Z|\mathbf{y}(m))\cap g^{-1}(m)|\Big\}W_1^N(\mathbf{z}|\mathbf{y})|$$
$$\le M\exp\{-NE\}$$

or

$$W_1^N(\mathbf{z}|\mathbf{y})\Big[\sum_{m\in\mathcal{T}_P(Y)\cap Y_{DB}}\Big|\mathcal{T}_{P,V}(Z|\mathbf{y}(m))\Big|-\sum_{m\in\mathcal{T}_P(Y)\cap Y_{DB}}\Big|\mathcal{T}_{P,V}(Z|\mathbf{y}(m))\cap g^{-1}(m)\Big|\Big]$$
$$\le M\exp\{-NE\}.$$

As the code is random, then

$$E|\mathcal{T}_P(Y)\cap Y_{DB}|=\sum_{m\in M}P^*(\mathcal{T}_P(Y))=M\cdot\exp\{-ND(P\|P^*)\}. \tag{8}$$

Taking into account that

$$W_1^N(\mathbf{z}|\mathbf{y})=\exp\{-N(D(V\|W_1|P)+H_{P,V}(Z|Y))\},$$

we get

$$\sum_{m\in\mathcal{T}_P(Y)\cap Y_{DB}} |\mathcal{T}_{P,V}(Z|\mathbf{y}(m))| - M\exp\{-NE\}\exp\{N(D(V||W_1|P) + H_{P,V}(Z|Y))\}$$

$$\leq \sum_{m\in\mathcal{T}_P(Y)\cap Y_{DB}} |\mathcal{T}_{P,V}(Z|\mathbf{y}(m)) \cap g^{-1}(m)|.$$

From (8) we have

$$\sum_{m\in\mathcal{T}_P(Y)\cap Y_{DB}} |\mathcal{T}_{P,V}(Z|\mathbf{y}(m))| - E|\mathcal{T}_P(Y)\cap Y_{DB}|\cdot\exp\{ND(P||P^*)\}\cdot\exp\{-NE\}$$

$$\times\exp\{N(D(V||W_1|P)) + H_{P,V}(Z|Y)\} \leq \sum_{m\in\mathcal{T}_P(Y)\cap Y_{DB}} |\mathcal{T}_{P,V}(Z|\mathbf{y}(m)) \cap g^{-1}(m)|,$$

or

$$\sum_{m\in\mathcal{T}_P(Y)\cap Y_{DB}} |\mathcal{T}_{P,V}(Z|\mathbf{y}(m))| - E|\mathcal{T}_P(Y)\cap Y_{DB}|\cdot\exp\{N(D(P\circ V||P^*\circ W_1)$$

$$+H_{P,V}(Z|Y) - E)\} \leq \sum_{m\in\mathcal{T}_P(Y)\cap Y_{DB}} |\mathcal{T}_{P,V}(Z|\mathbf{y}(m)) \cap g^{-1}(m)|.$$

Further,

$$E|\mathcal{T}_P(Y)\cap Y_{DB}|\cdot\exp\{H_{P,V}(Z|Y)\}\Big[(N+1)^{-|\mathcal{Z}||\mathcal{Y}|}$$

$$-\exp\{N(D(PV||P^*W_1) - E)\}\Big] \leq \exp\{NH_{P,V}(Z)\}.$$

As the number of messages M can be presented as a sum

$$M = \sum|\mathcal{T}_P(Y)\cap Y_{DB}|,$$

and there exists a major type $\hat{P}$ such that

$$E|\mathcal{T}_{\hat{P}}(Y)\cap Y_{DB}| \geq M(N+1)^{-|\mathcal{Y}|},$$

then

$$M(N+1)^{-|\mathcal{Y}|}\cdot\exp\{NH_{\hat{P},V}(Z|Y)\}\Big[(N+1)^{-|\mathcal{Z}||\mathcal{Y}|} - \exp\{N(D(\hat{P}\circ V||P^*\circ W_1) - E)\}\Big]$$

$$\leq \exp\{NH_{\hat{P},V}(Z)\},$$

hence

$$M \leq \frac{\exp\{NI_{\hat{P},V}(Z\wedge Y) - \delta_1\}}{(N+1)^{-|\mathcal{Y}||\mathcal{Z}|} - \exp\{N(D(\hat{P}\circ V||P^*\circ W_1) - E)\}}$$

The theorem is proved.

6. Numerical Representations of Results

Similarly to the example described in [1] assume that X is Bernoulli with parameter $p =$ $Pr\{X = 1\} = 0.5$.

Let $Y = X + N_e$ and $Z = X + N_i$, where N_e and N_i are Bernoulli noise variables with parameters d_e and d_i, respectively, addition is modulo 2. Then we get

$$Y = X + N_e = \begin{cases} 0, & \text{with } P = 0.5 \\ 1, & \text{with } P = 0.5 \end{cases}$$

and similarly

$$Z = X + N_i = \begin{cases} 0, & \text{with } P = 0.5 \\ 1, & \text{with } P = 0.5. \end{cases}$$

Then for $W(z|y)$ we have

$$W(1|1) = W(0|0) = de * di + (1 - de) * (1 - di) = 1 - d,$$

$$W(1|0) = W(0|1) = (1 - de) * di + de * (1 - di) = d.$$

Therefore the "channel" between Y and Z is a binary symmetric channel with transition probability d and uniform input on Y. Then the capacity is

$$I(Y \wedge Z) = 1 - h(d).$$

Then from the definition of the rate R we have that the number M of individuals that can be identified reliably is $M = 2^{NR}$. From the bounds of E-capacity we obtain the dependence of M on N for various E.

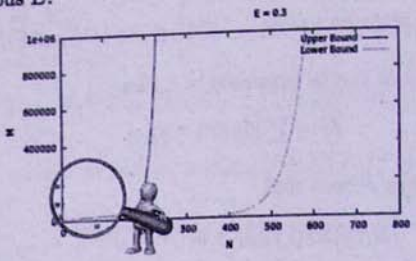

Fig.1. Bounds of E-capacity, when $E = 0.3$.

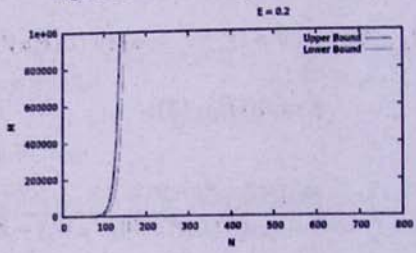

Fig.2. Bounds of E-capacity, when $E = 0.2$.

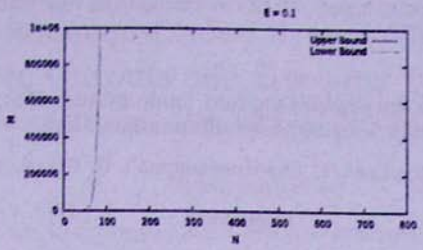

Fig.3. Bounds of E-capacity, when $E = 0.1$.

We can see that when $E \to 0$ values M for lower and upper bounds converge.

From the plots the greatest number of individuals can be computed. For example, for $= 0.1$:

- for small reliability $E = 0.1$ and $N = 100$, we obtain $M = 6357376$,
- for greater reliability $E = 0.2$ and $N = 100$, we obtain $M = 4705$.

Here, the number of individuals is much smaller.

To increase this number consider $N = 110$, then $M = 10960$.

The considered dependence of main characteristics will help to design practical biometric ystems.

References

[1] F. Willems, T. Kalker, J. Goseling, and J.-P. Linnartz, "On the capacity of a biometrical identification system", *International Symposium on Information Theory*, Yokohama, Japan, p. 82, 2003.

[2] S. Pankanti, R. M. Bolle and A. Jain, "Biometrics-the future of identification", *IEEE Computer*, vol. 33, no. 2, pp. 46 49, February, 2002.

[3] T. Ignatenko and F. Willems, "Biometric security from an information-theoretical perspective", *Foundations and Trends in Communications and Information Theory*, vol. 7, no 2-3, pp. 135-316, 2012.

[4] E. A. Haroutunian, "On bounds for E-capacity of DMC", *IEEE Transactions on Information Theory*, vol. 53, no. 11, pp. 4210-4220, 2007.

[5] E. A. Haroutunian, M. E. Haroutunian and A. N. Harutyunyan, "Reliability criteria in information theory and in statistical hypothesis testing", *Foundations and Trends in Communications and Information Theory*, vol. 4, no 2-3, pp. 97-263, 2008.

[6] M. E. Haroutunian, "Estimates of E-capacity and capacity regions for multiple-access channel with random parameter", *Lecture Notes in Computer Science*, vol. 4123, Springer Verlag, pp. 196-217, 2006.

[7] M. E. Haroutunian, S. A. Tonoyan, "Random coding bound of information hiding E-capacity", *Proc. of IEEE International Symposium on Information Theory*, p. 536, USA, Chicago, 2004.

[8] T. M. Cover and J. A. Thomas, *Elements of Information Theory*, Wiley, New York, 1991.

[9] I. Csiszár and J. Körner, *Information Theory: Coding Theorems for Discrete Memoryless Systems*, Academic Press, New York, 1981.

[10] I. Csiszár, "The method of types", *IEEE Transactions on Information Theory*, vol. 44, no. 6, pp. 2505-2523, 1998.

Կենսաչափական նույնականացման համակարգի E-ունակության վերին և ստորին գնահատականները

Մ. Հարությունյան, Լ. Տեր-Վարդանյան և Ա. Մուրադյան

Ամփոփում

Հոդվածում ներմուծվում է կենսաչափական նույնականացման համակարգի E-ունակության նոր հասկացությունը, որն ընդհանրացնում է Վիլեմսի և ուրիշների [1] ուսումնասիրած ունակության գաղափարը: Հետազոտվում է այդ ֆունկցիան՝ վերին և ստորին գնահատականների կառուցման միջոցով: Երբ $E \to 0$, մենք ստանում ենք կապուղու ունակության վերին և ստորին գնահատականները, որոնք համընկնում են [1]-ում ստացված ունակության հետ:

Верхняя и нижняя границы E-пропускной способности биометрической системы идентификации

М. Арутюнян, Л. Тер-Варданян и А. Мурадян

Аннотация

В статье вводится новое понятие E-пропускной способности для биометрической системы идентификации, которая является обобщением пропускной способности, изученной Вилемсом и др. в [1]. Функция исследуется путем построения верхней и нижней границ. Когда $E \to 0$, мы получаем верхнюю и нижнюю границы пропускной способности канала, которые совпадают с пропускной способностью, полученной в [1].