Some Numerical Representations on Biometric Identification System

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

Abstract—**In this paper we represent some computations for the bounds of E-capacity of the biometric identification system. This function expresses the dependence of the main characteristics of the system: number of individuals that can be identified based on their biometric data, length of the data sequence, error probability exponent. On a simple example we show that this dependence can be computed.**

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 the biometrics. Biometrics is being 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 automobiles, biometrics is being 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 [2]. 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].
It has been shown recently [1,3] that it is not possible to identify reliably more persons than capacity which is an inherent characteristic of any identification system.
The model of biometric identification system consists of two procedures: enrollment and identification.
In an enrollment phase M individuals are observed and for each individual a noisy version of the biometric data is added to a 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.

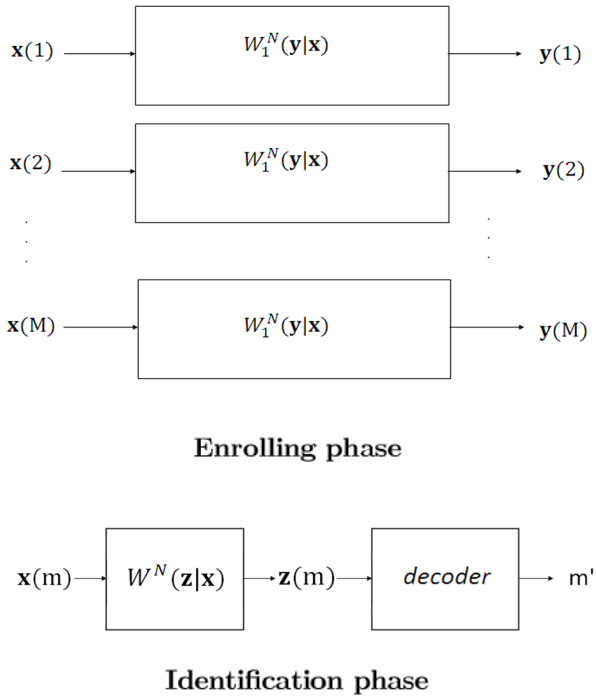

Fig. 1. Model of biometric identification system

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, which is the generalization of the capacity, as it tends to capacity, when E tends to 0. Upper and the lower bounds for identification E-capacity for maximal and average error probabilities are constructed in [4]. When $E \to 0$ we derive upper and lower bounds of

*Manuscript received March 16, 2012. Research was supported by Armenian national grant 11-1b255. Authors are with the Institute for Informatics and Automation Problems (IIAP), Armenia National Academy of Sciences (NAS), E-mail:(armar@ipia.sci.am, mur_art@yahoo.com, lilit@sci.am.)

ISBN: 978-988-19251-0-7
ISSN: 2078-0958 (Print); ISSN: 2078-0966 (Online)

the channel capacity, which coincide with the capacity obtained in [1].
In this paper we give numerical representations of the results to simplify the solutions in view of applications.

2 Notations and Definitions

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 $\mathcal{X}$ we 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 are 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 $m' = g_N(\mathbf{z})$, where 0 stands for the case, when the unknown individual has not been observed by enrollment phase. Following [1] we do not consider the probability that an individual, that 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 also will 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

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

The next parameter is the error probability

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

where

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

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

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

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

We investigate the E-capacity function which for given $E > 0$ is

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

where

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

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

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

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

For the information-theoretic quantities, such as entropy $H_P(Y)$, mutual information $I_{P,V}(Z \wedge Y)$, divergence $D(V||W|P)$ we refer to [5] – [11].

ISBN: 978-988-19251-0-7
ISSN: 2078-0958 (Print); ISSN: 2078-0966 (Online)

3 Formulation of the Results

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

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

$$\triangleq \min_{P,V:D(P\circ V||P^*\circ W)\le E} \Big| I_{P,V}(Z\wedge Y)+$$

$$+D(P\circ V||P^*\circ W)-E\Big|^+.$$

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

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

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

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

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

The proof of theorem can be found in [4].

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 Numerical representations of the results

Similar 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 definition of 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 from N for various E.

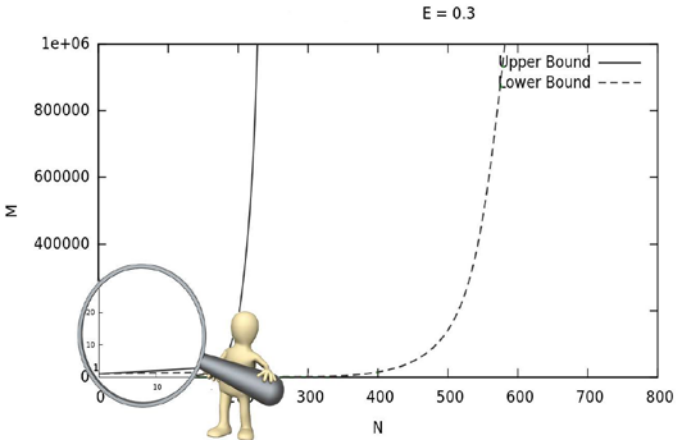

Fig.2. Bounds of E-capacity, when $E = 0.3$

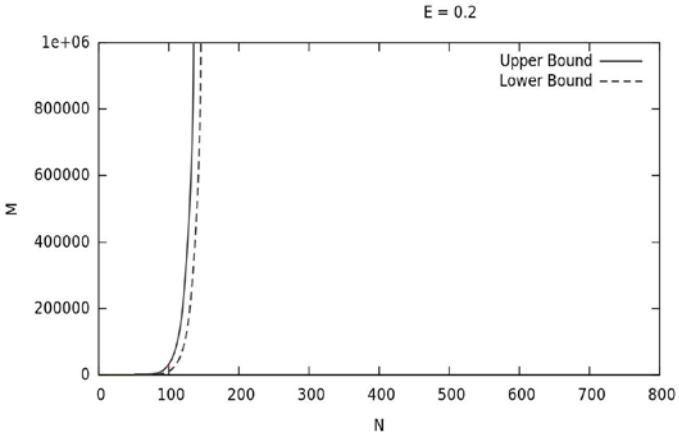

Fig.3. Bounds of E-capacity, when $E = 0.2$

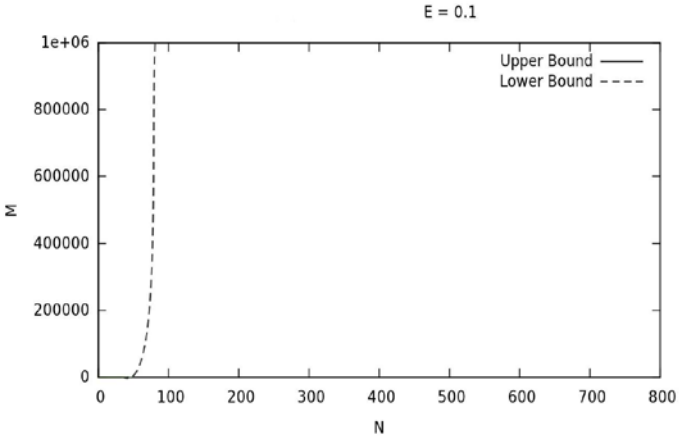

Fig.4. Bounds of E-capacity, when $E = 0.1$

ISBN: 978-988-19251-0-7
ISSN: 2078-0958 (Print); ISSN: 2078-0966 (Online)

We can see that when $E \to 0$ values of M for lower and upper bounds converge.
From the plots the greatest number of individuals can be computed. For example, for $d = 0.1$:
• 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 a practical biometric systems.

References

[1] F. Willems, T. Kalker, J. Goseling, and J.-P. Linnartz, "On the capacity of a biometric identification system", *International Symposium on Information Theory*, Yokohama, Japan, p. 82, June 29 - July 4, 2003.

[2] S. Pankanti, R. M. Bolle and A. Jain, "Biometrics-The Future of Identification", *IEEE Computer*, V33, N 2, pp. 46-49, 2/02

[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] M. Haroutunian, A. Muradyan and L. Ter- Vardanyan, "Upper and lower bounds of biometric identification E- capacity, *Transactions of IIAP of NAS of RA, Mathematical Problems of Computer Science*,V36, pp.1-10, 2012.

[5] E. A. Haroutunian, "On bounds for E-capacity of DMC", *IEEE Transactions on Information Theory*, V53, N11, pp. 4210-4220, 2007.

[6] 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*, V4, N2-3, pp. 97-263, 2008.

[7] M. E. Haroutunian, "Estimates of E-capacity and capacity regions for multiple-access channel with random parameter", *Lecture Notes in Computer Science*, V4123, Springer Verlag, pp. 196-217, 2006.

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

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

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

[11] I. Csiszár, "The method of types", *IEEE Transactions on Information Theory*, V44, N6, pp. 2505-2523, 1998.

ISBN: 978-988-19251-0-7
ISSN: 2078-0958 (Print); ISSN: 2078-0966 (Online)