Investigation of E-achievability region for identification and secret-key generation system

Mariam E. Haroutunian, Lilit A. Ter-Vardanyan *

Abstract— **The identification and secret-key generation system is investigated. Bounds of E-achievable rates region of this system are obtained.**

Keywords: ***Biometric identification system, identification capacity, E-capacity bounds, error exponents, secret-key generation.***

1 Introduction

Biometrics is often used by companies, governments, military, border control, hospitals, banks etc. to either verify a person's identity, i.e. for physical access control, computer log-in, welfare disbursement, international border crossing and national ID cards, e-passports, allowing access to a certain building area or to identify individuals to retain information about them, i.e. criminals, forensics, etc. In automobiles, biometrics is being adopted to replace keys for keyless entry and keyless ignition. [1].

Biometric identification systems were studied by O'Sullivan and Schmid [2] and Willems et al. [3]. They assumed storage of biometric enrollment sequences in the clear and determined the corresponding identification capacity. Later Turcel [4] analyzed the trade-off between the capacity of a biometric identification system and the storage space (compression rate) required for the biometric templates. It should be noted that Turcel's method realizes a kind of privacy protection scheme. Recall that secrecy capacity introduced by Ahlswede and Csiszar [5] can be regarded as the amount of common secret information that can be obtained in an authentication system in which helper data are (publicly) available. Interestingly this secrecy capacity, which is equal to the mutual information between enrollment and authentication biometric sequences in the biometric setting, equals the identification capacity found by O'Sullivan and Schmid [2] and Willems et al. [3].

In this paper we consider the model of identification and secret-key generation system studied by T. Ignatenko and F. Willems [6, 7].

In that system two terminals observe the enrollment and identification biometric sequences of a group of individuals. The first terminal forms a secret key for each enrolled individual and stores the corresponding helper data in a public database. These helper data on one hand facilitate reliable reconstruction of the secret key and on the other hand allow determination of the individuals identity for the second terminal, based on the presented biometric identification sequence. All helper data in the database are assumed to be public. Since the biometric secrets produced by the first terminal are used e.g. to encrypt data, the helper data should provide no information on these secret keys.

T. Ignatenko and F.Willems determined what identification and secret-key rates can be jointly realized by such a biometric identification system.

The E-capacity of that model was investigated in [8, 9]. The notion of E-capacity is studied for various communication systems [10, 11].

In this paper, we construct the bounds of E-achievable rates region for model of identification and secret-key generation system. The obtained result is the generalisation of bounds for biometric identification E-capacity [8]. We also got the result of bounds for the secrecy E-capacity, which is the generalization of the secrecy capacity [5]. Another obtained result is the generalization of bounds for achievable rate region [7].

2 Notations and Definitions

Following conventions are applied within the paper. Capital letters are used for random variables (RV) X, Y taking values in the finite sets $\mathcal{X}, \mathcal{Y}$, correspondingly, and lower case letters x, y 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 $\max(a, 0)$. Lower case letters k, h are used for secret key and for helper data.

The model of biometric identification and secret-key generation system consists of enrollment and identification procedures (Fig.1).

*Authors are with the Institute for Informatics and Automation Problems (IIAP), Armenian National Academy of Sciences (NAS), E-mail:(armar@ipia.sci.am, lilit@sci.am.)

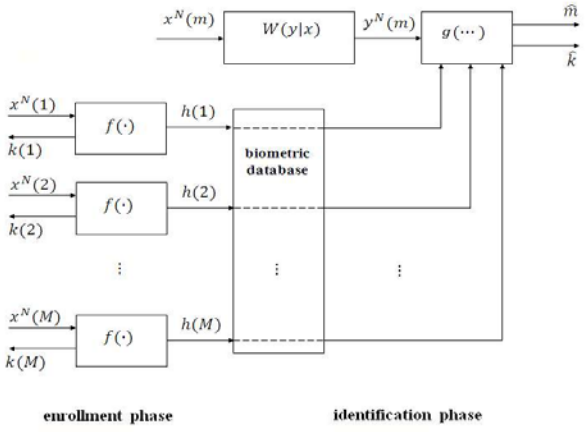

Fig.1. Model of identification and secret-key generation system

In an **enrollment phase** M individuals are observed. For each individual $m \in \{1,2,\cdots,M\}$ in the system the biometric source produces a biometric enrollment sequence $\mathbf{x}(m) = \{x_1, x_2, \cdots, x_N\}$, where $x_n \in \mathcal{X}$, $n = \overline{1,N}$. All these sequences are supposed to be generated at random with a given probability distribution

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

that is the symbols x_n; $n = 1,2,...,N$, are independent and identically distributed. During the enrollment procedure the biometric sequence $\mathbf{x}(m)$ of individual $m \in \{1,2,...,M\}$ is encoded into helper data $\mathbf{h}(m)$ and a secret key $\mathbf{k}(m)$, hence

$$f(\mathbf{x}(m)) = (\mathbf{h}(m); \mathbf{k}(m)); \quad \text{for} \quad m \in \{1,2,...,M\},$$

where $f(\mathbf{x}(m))$ is the **encoder mapping**. The helper data $\mathbf{h}(m)$ is then stored in a (public) database at position m. The generated secret key $\mathbf{k}(m)$ is handed over to the individual.

The helper data that are stored in the database make reliable identification possible. They should only contain a negligible amount of information about the corresponding secret key.

During the **identification procedure** a biometric identification sequence $\mathbf{y}(m) = (y_1, y_2, \cdots, y_N)$, consisting of N symbols from the finite alphabet $\mathcal{Y}$, is observed. This sequence is the output of the biometric channel whose input was the enrollment sequence $\mathbf{x}(m)$ of the unknown individual m. If individual m was observed, the sequence $\mathbf{y}(m)$ occurs with probability

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

since the biometric channel $W^N(\mathbf{y}|\mathbf{x})$ is memoryless. We assume here that all individuals are equally likely to be observed for identification, hence

$$P\{m\} = \frac{1}{M}; \quad \text{for all} \quad m \in \{1,2,...,M\}.$$

During identification, upon observing the biometric identification sequence $\mathbf{y}$; the decoder forms an estimate $\hat{m}$ of the identity of the observed individual as well as an estimate of his secret key $\widehat{\mathbf{k}(m)}$,

$$(\hat{m}, \widehat{\mathbf{k}(m)}) = g(\mathbf{y}, \mathbf{h}(1), \mathbf{h}(2), ..., \mathbf{h}(M)),$$

where g is the **decoder mapping**.

The estimate of the individual's identity $\hat{m}$ takes on values from the set of individuals, i.e. $\hat{m} \in \{1,2,...,M\}$. Moreover, the decoder's estimate of the secret key $\widehat{\mathbf{k}(m)}$ assumes values from the same alphabet as the secret key generated during enrollment, hence $\widehat{\mathbf{k}(m)} \in \{1,2,...,K\}$.

Definition (E-achievability). For $E > 0$ an identification and secret-key rate pair $(R_I; R_K)$ with $R_I \geq 0$ and $R_K \geq 0$ is E-achievable in a biometric identication setting, if for all $\delta > 0$ for all N large enough, there exists encoder and decoder such that

$$Pr\{(\hat{m}, \hat{k}) \neq (m,k)\} \leq \exp\{-N(E-\delta)\},$$

$$\frac{1}{N} \log M \geq R_I - \delta,$$

$$\frac{1}{N} \log K \geq R_K - \delta,$$

and helper data gives negligible information about secret key. The base of log is taken 2.

Let us denote by $\mathcal{R}(E,Q,W)$ the region of all E-achievable rate pairs.

For the formulation of result we use the following PD:

$$P = \{P(x), x \in \mathcal{X}\},$$

$$V = \{V(y|x), y \in \mathcal{Y}, x \in \mathcal{X}\}.$$

For the information-theoretic quantities, such as entropy $H_P(X)$, mutual information $I_{P,V}(X \wedge Y)$, divergence $D(V||W|P)$ we refer to [10-14].

3 Formulation of the Result

For the formulation of the inner and outer bounds of $\mathcal{R}(E,Q,W)$ region let us denote:

$$\mathcal{R}_r(E,Q,W) = \{(R_I, R_K):$$

$$R_I + R_K \leq \min_{P,V: D(P\circ V||Q\circ W) \leq E} \Big| I_{P,V}(X \wedge Y) +$$

$$+ D(P \circ V||Q \circ W) - E \Big|^+,$$

$$\mathcal{R}_{sp}(E,Q,W)=\{(R_I,R_K)):$$

$$R_I+R_K\leq \min_{P,V:D(P\circ V||Q\circ W)\leq E} I_{P,V}(X\wedge Y).$$

Theorem. *For the biometric identification system with secret-key generation for the given Q,W and for all $E>0$*

$$R_r(E,Q,W)\subseteq\mathcal{R}(E,Q,W)\subseteq R_{sp}(E,Q,W).$$

Corollary 1. *When $E\to 0$ we get the inner and outer bounds of achievable rate region which coincide with result obtained in [7]*:

$$\mathcal{R}(Q,W)=\{(R_I,R_K):0\leq R_I+R_K\leq I_{Q,W}(X\wedge Y)\}.$$

Corollary 2. *When $R_K\to 0$ we obtain the inner and outer bounds of the identification E-capacity which coincide with result obtained in [8]*:

$$\mathcal{R}_r(E,Q,W)=\{R_I:$$

$$R_I\leq \min_{P,V:D(P\circ V||Q\circ W)\leq E}\Big|I_{P,V}(X\wedge Y)+$$

$$+D(P\circ V||Q\circ W)-E\Big|^+\},$$

$$\mathcal{R}_{sp}(E,Q,W)=\{R_I:$$

$$R_I\leq \min_{P,V:D(P\circ V||Q\circ W)\leq E} I_{P,V}(X\wedge Y)\}.$$

Corollary 3. *When $R_I\to 0$ we come to the inner and outer bounds of the secrecy E-capacity*

$$\mathcal{R}_r(E,Q,W)=\{R_K:$$

$$R_K\leq \min_{P,V:D(P\circ V||Q\circ W)\leq E}\Big|I_{P,V}(X\wedge Y)+$$

$$+D(P\circ V||Q\circ W)-E\Big|^+\},$$

$$\mathcal{R}_{sp}(E,Q,W)=\{R_K:$$

$$R_K\leq \min_{P,V:D(P\circ V||Q\circ W)\leq E} I_{P,V}(X\wedge Y)\}.$$

which is the generalization of the secrecy capacity [5], as when $E\to 0$ this bounds coincide with the last one

$$R_K=I_{Q,W}(X\wedge Y).$$

References

[1] S. Pankanti, R. M. Bolle and A. Jain, "Biometrics-The Future of Identification", *IEEE Computer*, V33, N 2, pp. 46-49, 2002.

[2] J. A. OSullivan and N. A. Schmid, "Performance prediction methodology for biometric systems using a large deviations approach", *IEEE Trans. on it Signal Proc.*, vol. 52, no. 10, pp. 3036-3045, 2004.

[3] 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, 2003.

[4] E. Tuncel, "Capacity/storage tradeoff in high-dimensional identification systems", *IEEE International Symposium on Information Theory*, Washington, USA, pp. 1929-1933, 2006.

[5] R. Ahlswede and I. Csiszár, "Common randomness in information theory and cryptography - Part I : Secret sharing", *IEEE Trans. Inform. Theory*, vol. IT - 39, pp. 1121 - 1132, July 1993.

[6] 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.

[7] Ignatenko, T., Willems, F.M.J. "Secret-key and identification rates for biometric identification systems with protected templates". *Proceedings of the 31st Symposium on Information Theory in the Benelux*, Rotterdam, The Netherlands, pp. 121-128, 2010.

[8] 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*, vol. 37, pp. 7-16, 2012.

[9] M. E. Haroutunian, A. R. Muradyan, L. A. Ter-Vardanyan, "Some numerical representations on biometric identification system", *World Congress on Engineering 2012, The 2012 International Conference of Information Engineering*, London.

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

[11] 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.

[12] T. M. Cover and J. A. Thomas, *Elements of Information Theory*, Second Edition, John Wiley and Sons, 2006.

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

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