Mathematical Problems of Computer Science 26, 2006, 33–37.

Computation of Information Hiding Capacity and E-Capacity Lower Bounds*

Smbat A. Tonoyan

Institue for Informatics and Automation Problems of NAS of RA
e-mail smbatt@ipia.sci.am

Abstract

The binary–Hamming case of the E-capacity and capacity results for information hiding system [1, 2] is evaluated for practical interests. A special parallel algorithm is elaborated and the computational software utilities are developed. The graphs, describing dependences of information hiding rate from the reliability and allowed distortion levels for the information hider and the attacker are obtained and presented. Also the graphical view of the capacity, depending from the allowed distortion levels is plotted.

1 Capacity and E-capacity of Information Hiding System

We consider the model of information hiding system, studied by M. Haroutunian and S. Tonoyan [1, 2]. Various setups of information hiding systems are studied by P. Moulin and J. O'Sullivan [3], N. Merhav [4], N. Merhav and A. Somekh-Baruch [5] and others.

Description of the system is the following: an information hider (or an encoder) embeds the message $m \in \mathcal{M}$ within the host data set (covertext) $\mathbf{s} \in \mathcal{S}^N$ using the side information $\mathbf{k} \in \mathcal{K}^N$. Codewords $\mathbf{x} \in \mathcal{X}^N$ are transmitted via attack channel with the finite input and output alphabets $\mathcal{X}$ and $\mathcal{Y}$. The *attacker* trying to exploit the message m, transforms the data blocks $\mathbf{x} \in \mathcal{X}^N$ into $\mathbf{y} \in \mathcal{Y}^N$. The decoder, using the side information, decodes $\mathbf{y} \in \mathcal{Y}^N$, deriving the message m. Random variables S and K, describing the host data and the side information sources, are distributed jointly by fixed probability distribution $Q = \{Q(s,k),\ s \in \mathcal{S}, k \in \mathcal{K}\}$.

Information hider introduces certain distortion in the host data set for the data embedding. The attacker, trying to change or remove this hidden information, introduces some other distortion. Also it is assumed that:

a) the attacker uses only discrete memoryless channels (DMC),
b) the covertext is available at the information hider only,
c) the decoder knows the attack strategy but not the DMC chosen by the attacker,
d) the attacker knows the information-hiding strategy but not the side information,
e) information hiding process is transparent (the distortion introduced by information hider does not exceed the allowable level)

*The work was supported by Armenian Target Programm 04.10.31.

f) the system is robust (the distortion introduced by the attacker should be restricted by corresponding level).

Information hider designs a memoryless covert channel $P = \{P(u, x|s, k),\ u \in \mathcal{U},\ x \in \mathcal{X},\ s \in \mathcal{S},\ k \in \mathcal{K}\}$. The set of the covert channels, subject to the distortion level Δ_1, allowed for the information hider is:

$$\mathcal{P}(Q, \Delta_1) = \{P : \sum_{u,x,s,k} d_1(s, x)P(u, x|s, k)Q(s, k) \le \Delta_1\},$$

where $d_1 : \mathcal{S} \times \mathcal{X} \to [0, \infty)$ the distortion function for information hider.

Attacker designs an attack channel $A = \{A(y|x),\ y \in \mathcal{Y},\ x \in \mathcal{X}\}$. The set of the attack channels, subject to the distortion level Δ_2, allowed for the information hider, under the condition of any covert channel P is:

$$\mathcal{A}(Q, P, \Delta_2) = \{A : \sum_{u,x,y,s,k} d_2(x, y)A(y|x)P(u, x|s, k)Q(s, k) \le \Delta_2\},$$

where $d_2 : \mathcal{X} \times \mathcal{Y} \to [0, \infty)$ the distortion function for attacker.

M is the cardinality of the message set $\mathcal{M}$. The nonnegative number

$$R = \frac{1}{N} \log M$$

is called *the information hiding code rate.*

The maximal error probability of the code, maximal over all attack channels from $\mathcal{A}(Q, P, \Delta_2)$ is

$$e(f, g, N, Q, P, \Delta_2) = \max_{m \in \mathcal{M}} \max_{A \in \mathcal{A}(Q,P,\Delta_2)} \sum_{(\mathbf{s},\mathbf{k}) \in \mathcal{S}^N \times \mathcal{K}^N} Q^N(\mathbf{s}, \mathbf{k})A^N\{\mathcal{Y}^N - g^{-1}(m|\mathbf{k})|f(m, \mathbf{s}, \mathbf{k})\},$$

and the average error probability of the code, maximal over all attack channels from $\mathcal{A}(P, \Delta_2)$ is:

$$\overline{e}(f, g, N, Q, P, \Delta_2) = \frac{1}{M} \max_{A \in \mathcal{A}(Q,P,\Delta_2)} \sum_{m \in \mathcal{M}} \sum_{(\mathbf{s},\mathbf{k}) \in \mathcal{S}^N \times \mathcal{K}^N} Q^N(\mathbf{s}, \mathbf{k})A^N\{\mathcal{Y}^N - g^{-1}(m|\mathbf{k})|f(m, \mathbf{s}, \mathbf{k})\}.$$

Information hiding E-capacity is defined as:

$$R(Q, E, \Delta_1, \Delta_2) = C(Q, E, \Delta_1, \Delta_2) = \varliminf_{N \to \infty} \frac{1}{N} \log M(Q, E, N, \Delta_1, \Delta_2),$$

where $M(Q, E, N, \Delta_1, \Delta_2)$ is the highest volume of the code, the maximal error probability of which exponentially decreases with the given exponent $E > 0$.

The random coding bound of information hiding system is defined as:

$$R_r(Q, E, \Delta_1, \Delta_2) = \max_{P \in \mathcal{P}(Q,\Delta_1)} \min_{A \in \mathcal{A}(Q,P,\Delta_2)} \min_{Q', V : D(Q' \circ P \circ V \| Q \circ P \circ A) \le E} |I_{Q',P,V}(Y \wedge U|K) -$$

$$-I_{Q',P}(S \wedge U|K) + D(Q' \circ P \circ V \| Q \circ P \circ A) - E|^+. \qquad (1)$$

In [1, 2] the following theorem was proved.

Theorem. *For all $E > 0$, for information hiding system with distortion levels Δ_1, Δ_2*

$$R_r(Q, E, \Delta_1, \Delta_2) \le C(Q, E, \Delta_1, \Delta_2) \le \overline{C}(Q, E, \Delta_1, \Delta_2),$$

where by $\overline{C}(Q, E, \Delta_1, \Delta_2)$ the information hiding E-capacity for average error probability is denoted.

Corollary. *When $E \to 0$ (1) presents the lower bound of information hiding capacity:*

$$R_\tau(Q, \Delta_1, \Delta_2) = \max_{P \in \mathcal{P}(Q, \Delta_1)} \min_{A \in \mathcal{A}(Q, P, \Delta_2)} \{I_{Q,P,A}(Y \wedge U|K) - I_{Q,P}(S \wedge U|K)\}, \qquad (2)$$

which coincides with the information hiding capacity [1, 2, 3].

2 Binary-Hamming Case

We consider the case, coming from the applications, when all sets are binary: $\mathcal{U}, \mathcal{X}, \mathcal{Y}, \mathcal{S}, \mathcal{K} = \{0, 1\}$. And distortion functions are Hamming distances

$$d_1(s, x) = d_H(s, x) = \begin{cases} 1, & \text{if } s = x, \\ 0, & \text{if } s \neq x, \end{cases}, \quad d_2(x, y) = d_H(x, y) = \begin{cases} 1, & \text{if } x = y, \\ 0, & \text{if } x \neq y. \end{cases}$$

The joint distribution of host data and side information sources let be

$$Q = \begin{pmatrix} 0.3 & 0.25 \\ 0.1 & 0.35 \end{pmatrix}.$$

Now consider the random coding bound (1) dependence from the distortion levels. In the figure 1 information hiding rate dependences from the reliabilities are presented. The graph "1" satisfies to the distortion pair $\Delta_1 = 0.4$, $\Delta_2 = 0.7$, the graph "2" to $\Delta_1 = 0.7$, $\Delta_2 = 0.7$ and the graph "3" satisfies to the pair $\Delta_1 = 0.4$, $\Delta_2 = 0.8$.

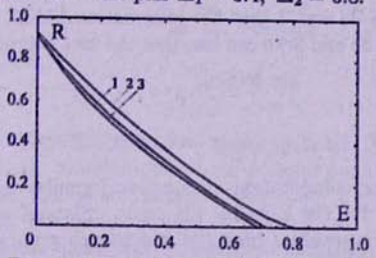

Figure 1. E-capacity dependence from the reliability for the different distortion pairs.

As we can see from figure 1, the rate curve "1" is going up, when the allowed distortion level of information hider Δ_1 grows from the value 0.4 to 0.7 and we obtain the curve "2". It takes place, because when the allowed level is greater, the information hider can hide more information. And when the allowed distortion level of attacker Δ_2 grows from the value 0.7 to 0.8 the curve "1" is going down and we obtain the curve "3".

In the next figure the capacity (2) dependence from the allowed distortion levels is presented.

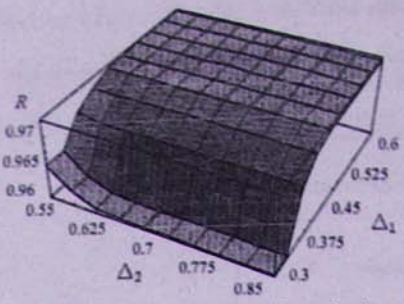

Figure 2. The capacity dependence form the distortion levels.

On the figures 3a and 3b the intersections of the previous surface are shown. 3a presents the capacity dependence from the Δ_1, when Δ_2 is fixed and equal to 0.816. 3b presents the capacity dependence from the Δ_2, if Δ_1 is fixed and equals 0.333.

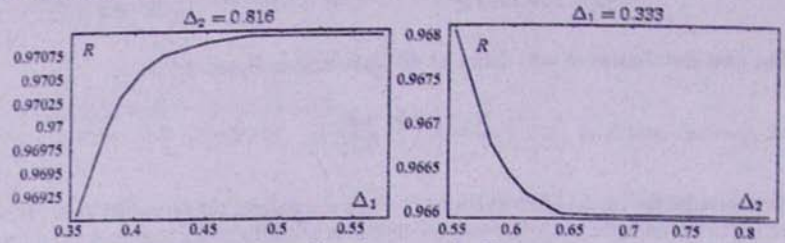

Figure 3. a. The capacity dependence form Δ_1. b. The capacity dependences form Δ_2.

We can see from the figures 3a and 2 that the information hiding rate grows with the hider's allowed level. And from 3b and 2 we can see, that the rate decreases, when attacker's distortion level grows.

3 Computation Overview

A special parallel algorithm for computation of complex formulas was evaluated for E-capacity (1) lower bound and for the capacity (2) cases. Sixteen parallel branches are realized in the algorithm. In the separate branches and branch groups the following tasks are realized.

a) the creation of the covert and the attack channels,
b) the supporting of probability distribution sets (two and four dimensional),
c) the sorting of the channels by distortion levels,
d) the computation of information theoretic terms of the mutual informations and divergences,
f) other computational subtasks.

The distributed mechanism for maximin derivation is applied.

Special utilities for computations were developed in C++ language on the base of evaluated algorithm, using the Message Passing Interface. Utilities were run in the high perfor-

mance environment of ArmCluster, on 16 separate processors. The graphs on the base of computed data sets are plotted using the Wolfram Research Mathematica package.

References

[1] M. E. Haroutunian and S. A. Tonoyan, "Random coding bound of information hiding *E*-capacity", *Proc. of IEEE Intern. Symp. Infrom. Theory*, p. 536, USA, Chicago, 2004.

[2] M. E. Haroutunian and S. A. Tonoyan, "On estimates of rate-reliability-distortion function for information hiding system", *Transactions of the Institute for Informatics and Automation Problems of the NAS of RA. Mathematical Problems of Computer Scence 23*, pp. 20-31, 2004.

[3] P. Moulin and J. A. O'Sullivan, "Information-theoretic analysis of information hiding", *IEEE Trans. Inform. Theory*, vol. 49, no. 3, pp. 563-593, Mar. 2003.

[4] N. Merhav, "On random coding error exponents of watermarking systems", *IEEE Trans. Inform. Theory*, vol. 46, no. 2, pp. 420-430, Mar. 2000.

[5] N. Merhav and A. Somekh-Baruch, "On the error exponent and capacity games of private watermarking systems", *IEEE Trans. Inform. Theory*, vol. 49, no. 3, pp. 537-562, Mar. 2003.

Տվյալներ թաքցնող համակարգի ունակության և *E*-ունակության ստորին գնահատականների հաշվումը

Ս. Ա. Տոնոյան

Ամփոփում

Երկուական Հեմմինգի դեպքի համար դիտարկվել են [1,2]-ում հետազոտված տվյալներ թաքցնող համակարգի ունակության և *E*-ունակության ստորին գնահատականները, ելնելով կիբառական նշանակությունից: Կառուցվել և ծրագրավորվել է հաշվարկների կատարման համար զուգահեռ ալգորիթմ: Ստացված և ներկայացված են տվյալներ թաքցնելու արագության գրաֆիկները՝ կախված հուսալիությունից և տվյալներ թաքցնողի ու հարձակվողի համար թույլատրելի շեղման մակարդակներից: Նաև կառուցված է ունակության գրաֆիկական պատկերը՝ կախված թույլատրելի շեղման մակարդակներից: