On Reversible Information Hiding System

Mariam Haroutunian
IIAP of NAS RA
Yerevan, Armenia
Email: armar@ipia.sci.am

Smbat Tonoyan
IIAPof NAS RA
Yerevan, Armenia
Email: smbatt@ipia.sci.am

Oleksiy Koval
University of Geneva
Geneva, Switzerland
Email: Oleksiy.Koval@cui.unige.ch

Svyatoslav Voloshynovskiy
University of Geneva
Geneva, Switzerland
Email: svolos@cui.unige.ch

***Abstract*—In this paper we consider the problem of reversible information hiding in the case when the attacker uses only discrete memoryless channels (DMC), the decoder knows only the class of channels, but not the DMC chosen by the attacker, the attacker knows the information-hiding strategy, probability distributions of all random variables, but not the side information.**

We introduce the notion of reversible information hiding E-capacity, which expresses the dependence of the information hiding rate on the error probability exponent E and the distortion levels for the information hider, for the attacker and for the host data approximation. The random coding bound for reversible information hiding E-capacity is found. We obtain the lower bound for reversibility information hiding capacity for $E \to 0$.

In particular, we have analyzed two special cases of the general problem formulation, pure reversibility and pure message communications.

I. Introduction

Problem of information transmission over state dependent channels rises in the situation when the transmitter has a certain prior knowledge about the environment or channel. Such a situation is typical in data hiding where the main goal is to communicate information message embedded into the body of a media file over a certain channel [1]. It is also relevant to simultaneous transmission of digital and analog information in audio broadcasting [2].

A corresponding research area of communications over channels with side information available at the transmitter has attracted considerable attention in information theory. The main problem was to establish the highest possible rates of reliable communications in such channels [3]-[6] that is related to the optimal solution to the problem of channel interference cancellation.

However, in certain cases one is rather interested not in maximizing the rate of pure information transmission [7] but in the most accurate estimate of the channel interference (channel state). Such a problem arises in simultaneous broadcasting of analog and digital audio [2] where digital data designed in a way to enhance the overall reception quality can be considered as interference degrading the communication of analog information.

The problem of accurate channel state estimation at the output of the state-dependent channel (reversibility) was studied in communication and data hiding formulation. Sutivong et al. [7] considered the problem of simultaneous channel state transmission in addition to the pure information. They present the optimal protocol design and rate distortion region for pure information transmission / partial channel state recovery for the state-dependent channel. Similar results were obtained in [8] for multimedia authentication formulation of the problem. The maximum achievable rates for pure information transmission in the case of complete recovery of the channel state at the decoder are analyzed in [9]. One should also mention the work of Eggers et al. [10], who considered the reversibility of quantization-based data hiding as structured codebook approximation of random binning. Finally, the analysis of [11] considers a formulation where the communication protocol is specifically optimized to a particular pure information communication regime while reversibility is analyzed as a by-product of this design. Similarly to the previous cases, the rate-distortion region of pure information transmission and channel state recovery is defined.

In this paper we would like to make one step forward with respect to the existing results justifying joint information transmission rates and channel state estimation accuracy at the output of the data hiding channel and to establish the error exponents that can be attained in the reversible information hiding protocols in terms of E-capacity [12]-[14].

II. Problem Formulation

We use capital letters X to denote random variables (RV) and corresponding small letters x for their realizations. Small bold letters $\mathbf{x}$ designate length-N vectors $\mathbf{x} = [x_1, x_2, ..., x_N]$ with n^{th} element x_n. Calligraphic fonts $\mathcal{X}$ denote sets and $|\mathcal{X}|$ denotes the cardinality of set $\mathcal{X}$. All logarithms and exponents in the paper are of the base 2. We use the following notation $m = \overline{1, |\mathcal{M}|}$ for $m = 1, 2, \ldots, |\mathcal{M}|$. Information-theoretic quantities, such as conditional entropy of RV Y relative to RV X with probability density (PD) P_0, $V = \{V(y|x), x \in \mathcal{X}, y \in \mathcal{Y}\}$ will be denoted as $H_{P_0,V}(Y|X)$; the conditional mutual information of the RV S and $\hat{S}$ relative to RV by K is $I_{Q,Q_2}(S \wedge \hat{S}|K)$; the informational divergence of the PD Q^* and Q is denoted by $D(Q\|Q^*)$ and the conditional informational divergence of joint PD $Q^* \circ Q_2 \circ P \circ V$ and $Q^* \circ Q_2 \circ P \circ A$ by $D(Q^* \circ Q_2 \circ P \circ V \| Q^* \circ Q_2 \circ P \circ A) = D(V\|A|Q^*, Q_2, P)$. We denote the types or empirical distributions by small letters. The set of all vectors $\mathbf{k}$ of type q_0 we denote by $\mathcal{T}^N_{q_0}(K)$. The set of all vectors $\mathbf{s} \in \mathcal{S}^N$ of conditional type q_1 for given $\mathbf{k} \in \mathcal{T}^N_{q_0}(K)$ we denote by $\mathcal{T}^N_{q_1}(S|\mathbf{k})$. It is called also q-shell of vector $\mathbf{k}$. The notation $|a|^+$ will be used for $\max(a, 0)$.

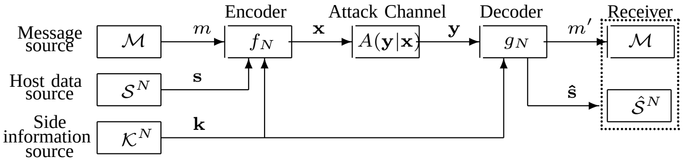

Figure 1. Reversible information hiding system

A reversible information hiding system is presented in Figure 1. It is supposed that a message m to be transmitted to the receiver is uniformly distributed over the message set $\mathcal{M}$. Host data source is described by the RV S, which takes values in the discrete finite set $\mathcal{S}$ and generates N-length sequences of independent and identically distributed (i.i.d.) components. The side information source is described by the RV K, which takes values in the discrete finite set $\mathcal{K}$, and in the most general case has the given joint PD $Q^* = \{Q^*(s,k),\ s \in \mathcal{S},\ k \in \mathcal{K}\}$ with the RV S. When the side information is a cryptographic key, S and K are independent. The side information in the form of i.i.d. N-length sequences is available to the encoder and decoder. It is assumed that $Q^{*N}(\mathbf{s},\mathbf{k}) = \prod\limits_{n=1}^{N} Q^*(s,k)$.

The *information hider* (encoder) embeds the message $m \in \mathcal{M}$ in the host data blocks $\mathbf{s} \in \mathcal{S}^N$ using the side information $\mathbf{k} \in \mathcal{K}^N$. The resulting codeword $\mathbf{x} \in \mathcal{X}^N$ is transmitted via attack channel $A = \{A(y|x), x \in \mathcal{X}, y \in \mathcal{Y}\}$ with the finite input and output alphabets $\mathcal{X}$ and $\mathcal{Y}$. The *attacker*, trying to modify or remove the message m, produces corrupted blocks $\mathbf{y} \in \mathcal{Y}^N$ based on $\mathbf{x} \in \mathcal{X}^N$, respectively. The decoder, possessing side information, derives the message m' and the approximation $\hat{\mathbf{s}}$ of the original data block, within the fixed distortion level using $\mathbf{y}$.

Let the mappings $d_0 : \mathcal{S} \times \hat{\mathcal{S}} \to [0,\infty)$, $d_1 : \mathcal{S} \times \mathcal{X} \to [0,\infty)$, $d_2 : \mathcal{X} \times \mathcal{Y} \to [0,\infty)$, be single-letter distortion functions. The distortion functions are supposed to be symmetric: $d_0(s,\hat{s}) = d_0(\hat{s},s), d_1(s,x) = d_1(x,s), d_2(x,y) = d_2(y,x)$ for all $s \in \mathcal{S}, \hat{s} \in \hat{\mathcal{S}}, x \in \mathcal{X}, y \in \mathcal{Y}$ and assume that $d_0(s,\hat{s}) = 0$, if $s = \hat{s}, d_1(s,x) = 0$, if $s = x, d_2(x,y) = 0$, if $x = y$. Distortion functions for the N-length vectors are defined as $d_0^N(\mathbf{s},\hat{\mathbf{s}}) = \frac{1}{N}\sum\limits_{n=1}^{N} d_0(s_n,\hat{s}_n)$, $d_1^N(\mathbf{s},\mathbf{x}) = \frac{1}{N}\sum\limits_{n=1}^{N} d_1(s_n,x_n)$, $d_2^N(\mathbf{x},\mathbf{y}) = \frac{1}{N}\sum\limits_{n=1}^{N} d_2(x_n,y_n)$.

Definition 1. The information hiding N-length code is a pair of mappings (f_N, g_N) subject to distortions Δ_0, Δ_1, where $f_N : \mathcal{M} \times \mathcal{S}^N \times \mathcal{K}^N \to \mathcal{X}^N$ is the encoder, mapping host data block $\mathbf{s}$, the message m and side information $\mathbf{k}$ to $\mathbf{x} = f_N(\mathbf{s},m,\mathbf{k})$, which satisfies the following distortion constraint:

$$d_1^N(\mathbf{s}, f_N(\mathbf{s},m,\mathbf{k})) \leq \Delta_1, \tag{1}$$

and $g_N : \mathcal{Y}^N \times \mathcal{K}^N \to \mathcal{M} \times \hat{\mathcal{S}}^N$ is the decoding, mapping the received sequence $\mathbf{y}$ and side information $\mathbf{k}$ to the decoded message m' and $\hat{\mathbf{s}}$, which satisfies the following distortion constraint: $d_0^N(\mathbf{s},\hat{\mathbf{s}}) \leq \Delta_0$.

Note that the definition of the distortion constraint (1) means that the maximum distortion constraint with respect to $\mathbf{s},\mathbf{k}$ and m is used, as distinct from [17], where the average distortion constraint is considered, and the maximum distortion constraint is mentioned as a more difficult case.

The attack channel, defined by $A^N(\mathbf{y}|\mathbf{x}) = \prod\limits_{n=1}^{N} A(y_n|x_n)$, subject to distortion Δ_2, satisfies the following constraint:

$$\sum_{\mathbf{x}\in\mathcal{X}^N}\sum_{\mathbf{y}\in\mathcal{Y}^N} d_2^N(\mathbf{x},\mathbf{y})A^N(\mathbf{y}|\mathbf{x})p^N(\mathbf{x}) \leq \Delta_2.$$

Definition 2. The nonnegative number $R = \frac{1}{N}\log|\mathcal{M}|$ is called *the information hiding code rate*.

For any $Q = Q_0 \circ Q_1 = \{Q(s,k) = Q_0(k)Q_1(s|k), s \in \mathcal{S}, k \in \mathcal{K}\}$ and Δ_0, denote by $\mathcal{Q}_2(Q,\Delta_0)$ the set of all conditional PDs $Q_2(\hat{s}|s,k)$, for which the following inequality takes place:

$$\sum_{s,\hat{s},k} Q(s,k)Q_2(\hat{s}|s,k)d_0(s,\hat{s}) \leq \Delta_0. \tag{2}$$

Definition 3. A memoryless covert channel P, subject to distortion Δ_1, is a PD $P = P_0 \circ P_1 = \{P(u,x|s,\hat{s},k) = P_0(u|s,\hat{s},k)P_1(x|u,s,\hat{s},k),\ u \in \mathcal{U},\ x \in \mathcal{X},\ s \in \mathcal{S}, \hat{s} \in \hat{\mathcal{S}},\ k \in \mathcal{K}\}$ such that for any Q and $Q_2 \in \mathcal{Q}_2(Q,\Delta_0)$:

$$\sum_{u,x,s,\hat{s},k} Q(s,k)Q_2(\hat{s}|s,k)P(u,x|s,\hat{s},k)d_1(s,x) \leq \Delta_1, \tag{3}$$

where U is an auxiliary RV taking values in the finite set $\mathcal{U}$ and forming the following Markov chain $(K,S,\hat{S},U) \to X \to Y$.

Denote by $\mathcal{P}(Q,Q_2,\Delta_1)$ the set of all covert channels $P^N(\mathbf{u},\mathbf{x}|\mathbf{s},\hat{\mathbf{s}},\mathbf{k}) = \prod\limits_{n=1}^{N} P(u_n,x_n|s_n,\hat{s}_n,k_n)$, subject to distortion Δ_1.

Definition 4. A memoryless attack channel A, subject to distortion Δ_2, under the condition of covert channel $P \in \mathcal{P}(Q,Q_2,\Delta_1)$, is defined by a PD A, for which for Q and $Q_2 \in \mathcal{Q}_2(Q,\Delta_0)$

$$\sum_{u,x,s,\hat{s},k,y} Q(s,k)Q_2(\hat{s}|s,k)P(u,x|s,\hat{s},k)A(y|x)d_2(x,y) \leq \Delta_2.$$

Denote by $\mathcal{A}(Q,Q_2,P,\Delta_2)$ the set of all attack channels, under the condition of covert channel $P \in \mathcal{P}(Q,Q_2,\Delta_1)$ and subject to distortion level Δ_2. The sets $\mathcal{Q}_2(Q,\Delta_0), \mathcal{P}(Q,Q_2,\Delta_1)$ and $\mathcal{A}(Q,Q_2,P,\Delta_2)$ are defined by linear inequality constraints and hence are convex.

Denote by $g_{N,\mathbf{k}}^{-1}(m,\hat{\mathbf{s}})$ the set of all $\mathbf{y}$ which for a given $\mathbf{k}$ are decoded into $(m,\hat{\mathbf{s}})$: $g_{N,\mathbf{k}}^{-1}(m,\hat{\mathbf{s}}) = \{\mathbf{y} :\ g_N(\mathbf{y},\mathbf{k}) = (m,\hat{\mathbf{s}})\}$.

Definition 5. The probability of erroneous reconstruction of the message $m \in \mathcal{M}$ and the approximation of data block $\mathbf{s} \in \mathcal{S}^N$ for $\mathbf{k} \in \mathcal{K}^N$ obtained at the output of the channel A is:

$$e(m,\mathbf{s},\mathbf{k},A) = 1 - A^N\left\{\bigcup_{\hat{\mathbf{s}}:\ d(\mathbf{s},\hat{\mathbf{s}})\leq\Delta_0} g_{N,\mathbf{k}}^{-1}(m,\hat{\mathbf{s}})|f_N(m,\mathbf{s},\mathbf{k})\right\}.$$

The error probability of the message m averaged over all $(\mathbf{s},\mathbf{k}) \in \mathcal{S}^N \times \mathcal{K}^N$ equals to:

$$e(m,A) = \sum_{(\mathbf{s},\mathbf{k})\in\ \mathcal{S}^N\times\mathcal{K}^N} Q^{*N}(\mathbf{s},\mathbf{k})e(m,\mathbf{s},\mathbf{k}).$$

Denote by $\mathbf{\Delta} = [\Delta_0, \Delta_1, \Delta_2]$ the collection of distortion levels, fixed for the current system.

The error probability of the code, for any message $m \in \mathcal{M}$, maximal over all attack channels from $\mathcal{A}(Q, Q_2, P, \Delta_2)$ is denoted by:

$$e(m) = \max_{A \in \mathcal{A}(Q,Q_2,P,\Delta_2)} e(m, A).$$

The *maximal error probability* of the code over all attack channels from $\mathcal{A}(Q, Q_2, P, \Delta_2)$ is equal to: $e = \max\limits_{m \in \mathcal{M}} e(m)$, and the *average error probability* of the code, maximal over all attack channels from $\mathcal{A}(Q, Q_2, P, \Delta_2)$ equals to: $\overline{e} = \frac{1}{|\mathcal{M}|} \sum\limits_{m \in \mathcal{M}} e(m)$.

III. Reversible Information Hiding E-Capacity

Consider the codes whose maximal error probability exponentially decreases with the given exponent $E > 0$, (called *reliability*), i.e., $e \leq \exp\{-NE\}$.

Denote by $M(Q^*, E, N, \mathbf{\Delta})$ the highest volume of the code, satisfying this condition for the given reliability E and the distortion levels $\mathbf{\Delta}$.

The rate-reliability-distortion function, which we call *reversible information hiding E-capacity* by analogy with the E-capacity of ordinary channel [12], is defined as:

$$R(Q^*, E, \mathbf{\Delta}) = C(Q^*, E, \mathbf{\Delta}) \triangleq \overline{\lim_{N \to \infty}} \frac{1}{N} \log M(Q^*, E, N, \mathbf{\Delta}).$$

By $C(Q^*, E, \mathbf{\Delta})$ and $\overline{C}(Q^*, E, \mathbf{\Delta})$ we denote the reversible information hiding E-capacity defined for maximal and average error probabilities respectively.

In this paper the lower bound of reversible information hiding E-capacity for maximal and average error probabilities is constructed.

Consider the following function, which we call *the random coding bound*

$$R_r(Q^*, E, \mathbf{\Delta}) = \min_{Q} \max_{Q_2 \in \mathcal{Q}_2(Q,\Delta_0)} \max_{P \in \mathcal{P}(Q,Q_2,\Delta_1)}$$
$$\min_{A \in \mathcal{A}(Q,Q_2,P,\Delta_2)} \min_{V: D(Q \circ Q_2 \circ P \circ V \| Q^* \circ Q_2 \circ P \circ A) \leq E}$$
$$\Big| I_{Q,Q_2,P,V}(Y \wedge U|K) - I_{Q,Q_2,P_0}(S \wedge U, \hat{S}|K)$$
$$+ D(Q \circ Q_2 \circ P \circ V \| Q^* \circ Q_2 \circ P \circ A) - E\big|^+ . \quad (4)$$

Theorem. *For any $E > 0$, for reversible information hiding system with distortion levels $\mathbf{\Delta}$*

$$R_r(Q^*, E, \mathbf{\Delta}) \leq C(Q^*, E, \mathbf{\Delta}) \leq \overline{C}(Q^*, E, \mathbf{\Delta}).$$

Corollary 1. *When $E \to 0$, we obtain the lower bound of reversible information hiding capacity based on (4) :*

$$R_r(Q^*, E, \mathbf{\Delta}) = \max_{Q_2 \in \mathcal{Q}_2(Q^*,\Delta_0)} \max_{P \in \mathcal{P}(Q^*,Q_2,\Delta_1)} \min_{A \in \mathcal{A}(Q^*,Q_2,P,\Delta_2)}$$
$$\left\{ I_{Q^*,Q_2,P,A}(Y \wedge U|K) - I_{Q^*,Q_2,P_0}(S \wedge U, \hat{S}|K) \right\}.$$

Corollary 2: pure reversibility. *If $\Delta_0 = 0$ from (4) we have*

$$R_r(Q^*, E, \mathbf{\Delta}) = \min_{Q} \max_{P \in \mathcal{P}(Q,\Delta_1)} \min_{A \in \mathcal{A}(Q,P,\Delta_2)}$$
$$\min_{V: D(Q \circ P \circ V \| Q^* \circ P \circ A) \leq E} |I_{Q,P,V}(Y \wedge U|K)$$
$$- H_Q(S|K) + D(Q \circ P \circ V \| Q^* \circ P \circ A) - E|^+ . \quad (5)$$

Corollary 3: pure message communications. *If $\Delta_0 \to \infty$ then*

$$R_r(Q^*, E, \mathbf{\Delta}) = \min_{Q} \max_{P \in \mathcal{P}(Q,\Delta_1)} \min_{A \in \mathcal{A}(Q,P,\Delta_2)}$$
$$\min_{V: D(Q \circ P \circ V \| Q^* \circ P \circ A) \leq E} |I_{Q,P,V}(Y \wedge U|K)$$
$$- I_{Q,P}(S \wedge U|K) + D(Q \circ P \circ V \| Q^* \circ P \circ A) - E|^+ . \quad (6)$$

In (5) and (6), $P = \{P(u, x|s, k),\ u \in \mathcal{U}, x \in \mathcal{X}, s \in \mathcal{S}, k \in \mathcal{K}\}$ and $\mathbf{\Delta} = (\Delta_1, \Delta_2)$.

IV. Proof of the Theorem

The theorem is proved using Shannon's random coding argument, the method of types, covering lemma and a generalization of packing lemma [12], [15], [16].

To prove the random coding bound, we must show the existence of a code with R satisfying (4) and $e \leq \exp\{-N(E - \varepsilon)\}$, for any $0 < \varepsilon < E$.

We will construct the encoding and the decoding and explore the errors caused by each.

For encoding we use the idea of Gelfand-Pinsker [5].

The decoding is based on *minimum divergence* method, first introduced by E. Haroutunian [12] and developed in [13], [14]. The extension of this method adopted to data hiding can be considered as *semi-universal decoding*, since the decoder needs to know only the specific class of channels, instead of a particular one, used by the attacker.

Encoding.

Step 1. Denote by $\mathcal{Q}(Q^*, E) = \{q : \ D(q\|Q^*) \leq E\}$ and

$$\mathcal{T}^E_{Q^*}(S, K) = \bigcup_{q \in \mathcal{Q}(Q^*,E)} \mathcal{T}^N_q(S, K). \quad (7)$$

We will construct the code only for $(\mathbf{s}, \mathbf{k})$ from $\mathcal{T}^E_{Q^*}(S, K)$, because for sufficiently large N, the probability of $(\mathbf{s}, \mathbf{k}) \notin \mathcal{T}^E_{Q^*}(S, K)$ is exponentially small:

$$Q^{*N} \left\{ \bigcup_{q \notin \mathcal{Q}(Q^*,E)} \mathcal{T}^N_q(S, K) \right\} = \sum_{q \notin \mathcal{Q}(Q^*,E)} Q^{*N}\{\mathcal{T}^N_q(S, K)\}$$
$$\leq \sum_{q \notin \mathcal{Q}(Q^*,E)} \exp\{-ND(q\|Q^*)\}$$
$$< (N+1)^{|\mathcal{S}||\mathcal{K}|} \exp\{-NE\} \leq \exp\{-N(E - \varepsilon_1)\}, \quad (8)$$

where $\varepsilon_1 > 0$.

Step 2. Denote $\hat{q}_1(\hat{s}|k) = \sum\limits_{s} q_2(\hat{s}|s, k) q_1(s|k)$.

Covering lemma. *For every type q, conditional type q_2, vector $\mathbf{k} \in \mathcal{K}^N$, there exists a collection of vectors $\{\hat{\mathbf{s}}_j \in \mathcal{T}^N_{\hat{q}_1}(\hat{S}|\mathbf{k}),\ j = \overline{1, J_1}\}$, where*

$$J_1 = \exp\left\{ N \left(I_{q,q_2}(S \wedge \hat{S}|K) + \delta/2 \right) \right\},\ \delta > 0$$

such that the set $\left\{\mathcal{T}^N_{q,q_2}(S|\hat{\mathbf{s}}_j,\mathbf{k})\ j=\overline{1,J_1}\right\}$ *covers* $\mathcal{T}^N_{q_1}(S|\mathbf{k})$ *for* N *large enough:*

$$\mathcal{T}^N_{q_1}(S|\mathbf{k}) \subset \bigcup_{j=1}^{J} \mathcal{T}^N_{q,q_2}(S|\hat{\mathbf{s}}_j,\mathbf{k}).$$

For the proof of covering lemma see [15].

For type $q\in\mathcal{Q}(Q^*,E)$ and conditional type $q_2\in\mathcal{Q}_2(q,\Delta_0)$ denote

$$\mathcal{S}(q,q_2,j)=\mathcal{T}^N_{q,q_2}(S|\hat{\mathbf{s}}_j,\mathbf{k})\backslash \bigcup_{j'<j}\mathcal{T}^N_{q,q_2}(S|\hat{\mathbf{s}}_{j'},\mathbf{k}),$$

therefore for the vectors $\mathbf{s}$ from $\mathcal{S}(q,q_2,j)$, we put into correspondence the vector $\hat{\mathbf{s}}_j, j\in[1,J_1]$.

Taking into account the inequality (2), we can show that for types $q\in\mathcal{Q}(Q^*,E), q_2\in\mathcal{Q}_2(q,\Delta_0)$ and any $j=\overline{1,J_1}$, $\mathbf{s}\in\mathcal{S}(q,q_2,j)$, $\hat{\mathbf{s}}_j$

$$d_0(\mathbf{s},\hat{\mathbf{s}}_j)=N^{-1}\sum_{s,\hat{s}} n(s,\hat{s}|\mathbf{s},\hat{\mathbf{s}}_j)d_0(s,\hat{s})$$

$$=\sum_{s,\hat{s},k} q(s,k)q_2(\hat{s}|s,k)d_0(s,\hat{s})\le\Delta_0,\ \ j=\overline{1,J_1}.$$

Step 3. Fix the type $p=p_0\circ p_1\in\mathcal{P}(q,q_2,\Delta_1)$. For fixed p_0,E, for each type $q\in\mathcal{Q}(Q^*,E)$, $q_2\in\mathcal{Q}(q,\Delta_0)$ and vectors $\hat{\mathbf{s}}$, $\mathbf{k}$, we choose independently, at random from $\mathcal{T}^N_{q_0,p_0}(U|\hat{\mathbf{s}},\mathbf{k})$ $|\mathcal{M}|$ collections $\mathcal{J}_2(m), m=\overline{1,|\mathcal{M}|}$, of vectors $\mathbf{u}_j(m),\ j=\overline{1,J_2}$, where

$$J_2=\exp\left\{N\left(I_{q,q_2,p_0}(S\wedge U|\hat{S},K)+\delta/2\right)\right\}\ \ (\delta>0).$$

Then, for each $\mathbf{s}\in\mathcal{T}^N_q(S|\mathbf{k})$ we select such $\mathbf{u}_j(m)$ from $\mathcal{J}_2(m)$, that $\mathbf{u}_j(m)\in\mathcal{T}^N_{q,q_2,p_0}(U|\mathbf{s},\hat{\mathbf{s}},\mathbf{k})$. Denote this vector by $\mathbf{u}(m,\mathbf{s},\hat{\mathbf{s}},\mathbf{k})$.

If for some $\mathbf{s}$ there is no such a vector in $\mathcal{J}_2(m)$, we choose $\mathbf{u}(m,\mathbf{s},\hat{\mathbf{s}},\mathbf{k})$ at random from the $\mathcal{T}^N_{q,q_2,p_0}(U|\mathbf{s},\hat{\mathbf{s}},\mathbf{k})$. Denote the probability of such an event by $\Pr\{b_{q,q_2,p_0}(m,\mathbf{s},\hat{\mathbf{s}},\mathbf{k})\}$.

$$\Pr\{b_{q,q_2,p_0}(m,\mathbf{s},\hat{\mathbf{s}},\mathbf{k})\}$$

$$=\Pr\left\{\bigcap_{j=1}^{J_2}\mathbf{u}_j(m)\notin\mathcal{T}^N_{q,q_2,p_0}(U|\mathbf{s},\hat{\mathbf{s}},\mathbf{k})\right\}$$

$$\le\prod_{j=1}^{J_2}[1-\Pr\{\mathbf{u}_j(m)\in\mathcal{T}^N_{q,q_2,p_0}(U|\mathbf{s},\hat{\mathbf{s}},\mathbf{k})\}]$$

$$\le\left[1-\frac{|\mathcal{T}^N_{q,q_2,p_0}(U|\mathbf{s},\hat{\mathbf{s}},\mathbf{k})|}{|\mathcal{T}^N_{q_0,p_0}(U|\hat{\mathbf{s}},\mathbf{k})|}\right]^{J_2}$$

$$\le[1-\exp\{-N(I_{q,q_2,p_0}(S\wedge U|\hat{S},K)$$

$$+\delta/4)\}]^{\exp\{N(I_{q,q_2,p_0}(S\wedge U|\hat{S},K)+\delta/2)\}}.$$

Using the inequality $(1-t)^n\le\exp\{-nt\}$, which holds for any n and $t\in(0,1)$, we can see that

$$\Pr\{b_{q,q_2,p_0}(m,\mathbf{s},\hat{\mathbf{s}},\mathbf{k})\}\le\exp\{-\exp\{N\delta/4\}\}. \qquad (9)$$

Notice that for each m, the code contains $J=J_1\times J_2=\exp\left\{N\left(I_{q,q_2,p_0}(S\wedge U,\hat{S}|K)+\delta\right)\right\}$ $(\delta>0)$ vectors $\mathbf{u}$.

Step 4. The codeword $\mathbf{x}$ is constructed in the following way. For each $m=\overline{1,|\mathcal{M}|}$, $\hat{\mathbf{s}}$, $\mathbf{s}$ and $\mathbf{k}$ we choose at random a vector $\mathbf{x}(m,\mathbf{s},\hat{\mathbf{s}},\mathbf{k})$ from $\mathcal{T}^N_{q,q_2,p}(X|\mathbf{u}(m,\mathbf{s},\hat{\mathbf{s}}_j,\mathbf{k}),\mathbf{s},\hat{\mathbf{s}},\mathbf{k})$.

It is easy to demonstrate that such an encoding satisfies the distortion constraint. Indeed, for types $p\in\mathcal{P}(q,q_2,\Delta_1)$, $q\in\mathcal{Q}(Q^*,E), q_2\in\mathcal{Q}_2(q,\Delta_0)$, taking into account the inequality (3), we have

$$d_1^N(\mathbf{s},\mathbf{x}(m,\mathbf{s},\hat{\mathbf{s}},\mathbf{k}))=N^{-1}\sum_{s,x}n(s,x|\mathbf{s},\mathbf{x})d_1(s,x)$$

$$=\sum_{u,x,s,\hat{s},k}q(s,k)q_2(\hat{s}|s,k)p(u,x|s,\hat{s},k)d_1(s,x)\le\Delta_1.$$

Denote by $e_E(m)$ *the encoding error probability* for any $m\in\mathcal{M}$:

$$e_E(m)\le\sum_{(\mathbf{s},\mathbf{k})\notin\mathcal{T}^E_{Q^*}(S,K)}Q^{*N}(\mathbf{s},\mathbf{k})$$

$$+\sum_{(\mathbf{s},\mathbf{k})\in\mathcal{T}^E_{Q^*}(S,K)}Q^{*N}(\mathbf{s},\mathbf{k})\Pr\{b_{q,q_2,p_0}(m,\mathbf{s},\hat{\mathbf{s}},\mathbf{k})\}.$$

Now taking into account (7), (8) and (9):

$$e_E(m)\le\exp\{-N(E-\varepsilon_1)\}$$

$$+\sum_{q\in\mathcal{Q}(Q^*,E)}Q^{*N}\left\{\mathcal{T}^N_q(S,K)\right\}\exp\{-\exp\{N\delta/4\}\}.$$

As the number of types q in $\mathcal{Q}(Q^*,E)$ does not exceed $(N+1)^{|\mathcal{S}||\mathcal{K}|}$ according to type counting lemma [15], [16] and $Q^{*N}\left\{\mathcal{T}^N_q(S,K)\right\}\le1$, we can write

$$e_E(m)\le\exp\{-N(E-\varepsilon_1)\}+\exp\{-\exp\{N\delta/4\}+\varepsilon_1\}, \quad (10)$$

for N large enough.

The attacker chooses the attack channel A from the set $\mathcal{A}(q,q_2,p,\Delta_2)$ as he knows probability distributions of all random variables. It is clear, that in this case the average distortion constraint is satisfied, since:

$$\sum_{\mathbf{x}\in\mathcal{X}^N}\sum_{\mathbf{y}\in\mathcal{Y}^N}d_2^N(\mathbf{x},\mathbf{y})A^N(\mathbf{y}|\mathbf{x})p^N(\mathbf{x})$$

$$=\mathbf{E}d_2^N(X^N,Y^N)=\frac{1}{N}\sum_{n=1}^{N}\mathbf{E}d_2(x_n,y_n)$$

$$=\sum_{u,x,s,\hat{s},k,y}q(s,k)q_2(\hat{s}|s,k)p(u,x|s,\hat{s},k)A(y|x)d_2(x,y)\le\Delta_2.$$

Decoding. For brevity the vector pair $\mathbf{u}(m,\mathbf{s},\hat{\mathbf{s}},\mathbf{k}),\mathbf{x}(m,\mathbf{s},\hat{\mathbf{s}},\mathbf{k})$ is denoted by $\mathbf{u},\mathbf{x}(m,\mathbf{s},\hat{\mathbf{s}},\mathbf{k})$.

We use the following decoding rule: every pair of $\mathbf{y}$ and $\mathbf{k}$ is decoded to such m and $\hat{\mathbf{s}}$ that $\mathbf{y}\in\mathcal{T}^N_{q,q_2,p,v}(Y|\mathbf{u},\mathbf{x}(m,\mathbf{s},\hat{\mathbf{s}},\mathbf{k}),\mathbf{s},\hat{\mathbf{s}},\mathbf{k})$, where q,q_2,p,v are such that $\min\limits_{A\in\mathcal{A}(q,q_2,p,\Delta_2)}D(q\circ q_2\circ p\circ v\|Q^*\circ q_2\circ p\circ A)$ is minimal.

The decoder can make an error when the message $m\in\mathcal{M}$ is transmitted and $(\mathbf{s},\mathbf{k})\in\mathcal{T}^E_{Q^*}(S,K)$. However, there exist such types q',q_2',p',v', vector $\mathbf{s}'$ and pair $(m',\hat{\mathbf{s}}')$ that $m'\neq m$ or $m'=m$, $d_0(\mathbf{s},\hat{\mathbf{s}}')>\Delta_0$, with

$$\mathbf{y}\in\mathcal{T}^N_{q,q_2,p,v}(Y|\mathbf{u},\mathbf{x}(m,\mathbf{s},\hat{\mathbf{s}},\mathbf{k}),\mathbf{s},\hat{\mathbf{s}},\mathbf{k})$$

$$\bigcap \mathcal{T}_{q',q_2',p',v'}(Y|\mathbf{u}',\mathbf{x}'(m',\mathbf{s}',\hat{\mathbf{s}}',\mathbf{k}),\mathbf{s}',\hat{\mathbf{s}}',\mathbf{k})$$

$$\text{and} \min_{A\in\mathcal{A}(q',q_2',p',\Delta_2)} D(q'\circ q_2'\circ p'\circ v'\|Q^*\circ q_2'\circ p'\circ A)$$

$$\leq \min_{A\in\mathcal{A}(q,q_2,p,\Delta_2)} D(q\circ q_2\circ p\circ v\|Q^*\circ q_2\circ p\circ A). \tag{11}$$

Denote by $\mathcal{D} = \{q, q', p, p', q_2, q_2', v, v' : (11) \text{ is valid}\}$ and

$$F = \mathcal{T}^N_{q,q_2,p,v}(Y|\mathbf{u},\mathbf{x}(m,\mathbf{s},\hat{\mathbf{s}},\mathbf{k}),\mathbf{s},\hat{\mathbf{s}},\mathbf{k})$$

$$\bigcap \bigcup_{(m',\hat{\mathbf{s}}'):\left\{\substack{m'\neq m \ or \\ m'=m,\ d_0(\mathbf{s},\hat{\mathbf{s}}')>\Delta_0}\right\}} \bigcup_{\mathbf{s}'\in\mathcal{T}^N_{q',q_2'}(S|\hat{\mathbf{s}}',\mathbf{k})}$$

$$\mathcal{T}^N_{q',q_2',p',v'}(Y|\mathbf{u}',\mathbf{x}'(m',\mathbf{s}',\hat{\mathbf{s}}',\mathbf{k}),\mathbf{s}',\hat{\mathbf{s}}',\mathbf{k}) \ .$$

The decoding error probability $e_D(m)$ of message $m \in \mathcal{M}$, maximal over all attack channels $A \in \mathcal{A}(q, q_2, p, \Delta_2)$, can be estimated in the following way:

$$e_D(m) \leq \max_{A\in\mathcal{A}(q,q_2,p,\Delta_2)} \sum_{(\mathbf{s},\mathbf{k})\in\mathcal{T}^E_{Q^*}(S,K)} Q^{*N}(\mathbf{s},\mathbf{k})$$

$$\times A^N \left\{\bigcup_{\mathcal{D}} F|\mathbf{x}(m,\mathbf{s},\hat{\mathbf{s}},\mathbf{k})\right\} \leq \sum_{\mathcal{D}} |F|$$

$$\times \max_{A\in\mathcal{A}(q,q_2,p,\Delta_2)} \sum_{(\mathbf{s},\mathbf{k})\in\mathcal{T}^E_{Q^*}(S,K)} Q^{*N}(\mathbf{s},\mathbf{k})A^N(\mathbf{y}|\mathbf{x}).$$

The last inequality is true, because for fixed types of $\mathbf{x}$ and $\mathbf{y}$ the probability $A^N(\mathbf{y}|\mathbf{x})$ is constant.

For the estimation of decoding error probability we use the statement of the following lemma, which is the modification of packing lemma from [14].

Packing Lemma. *For any* $E > 2\delta \geq 0$, *fixed* $q \in \mathcal{Q}(Q^*, E)$, $q_2 \in \mathcal{Q}_2(q, \Delta_0)$ *and covert channel* $p \in \mathcal{P}(q, q_2, \Delta_1)$, *there exists a code with*

$$|\mathcal{M}| = \exp\left\{N \min_{A\in\mathcal{A}(q,q_2,p,\Delta_2)} \min_{v:D(q\circ q_2\circ p\circ v\|Q^*\circ q_2\circ p\circ A)\leq E}\right.$$

$$|I_{q,q_2,p,v}(Y\wedge U|K) - I_{q,q_2,p_0}(S\wedge U,\hat{S}|K)$$

$$\left. +D(q\circ q_2\circ p\circ v\|Q^*\circ q_2\circ p\circ A) - E - 2\delta|^+\right\},$$

such that

1) *for each* $\mathbf{k}$, $\mathbf{s}$ *and* $\hat{\mathbf{s}}$, *the vector pairs* $\mathbf{u}, \mathbf{x}(m, \mathbf{s}, \hat{\mathbf{s}}, \mathbf{k})$ *are distinct for different* $m \in \mathcal{M}$,
2) *for sufficiently large* N *the following inequality holds for any types* $q' \in \mathcal{Q}(Q^*, E), q_2' \in \mathcal{Q}_2(q', \Delta_0)$, $p' \in \mathcal{P}(q', q_2', \Delta_1)$, v, v', *and for all* $m = \overline{1, |\mathcal{M}|}$, $(\mathbf{s}, \mathbf{k}) \in \mathcal{T}^N_q(S, K)$, $\hat{\mathbf{s}} \in \mathcal{T}^N_{q_2}(\hat{S}|\mathbf{s}, \mathbf{k})$

$$|F| \leq |\mathcal{T}^N_{q,q_2,p,v}(Y|\mathbf{u},\mathbf{x}(m,\mathbf{s},\hat{\mathbf{s}},\mathbf{k}),\mathbf{s},\hat{\mathbf{s}},\mathbf{k})|\times\exp\left\{-N\left|E\right.\right.$$

$$\left.\left. - \min_{A\in\mathcal{A}(q',q_2',p',\Delta_2)} D(q'\circ q_2'\circ p'\circ v'\|Q^*\circ q_2'\circ p'\circ A)\right|^+\right\}. \tag{12}$$

The lemma guarantees the existence of a good code, the codewords of which must be far from each other in a sense that all q, v-shells have possibly small intersections.

Using (11) and (12) it is easy to see that for N large enough

$$e_D(m) \leq \exp\{-N(E-\varepsilon_2)\}, \tag{13}$$

where $\varepsilon_2 > 0$.

From (10) and (13) one can see that the error probability of the message $m \in \mathcal{M}$ is small enough.

Taking into account the continuity of all expressions, when $N \to \infty$, arbitrary probability distributions can be considered instead of types. The theorem is proved.

V. Acknowledgments

This paper was partially supported by SNF PB grant 114613 and SNF 200021–111643, by the EU projects IST-2002-507932-ECRYPT, FP6-507609-SIMILAR and Swiss IM2 project.

References

[1] I. J. Cox, M. L. Miller, J.A. Bloom, ”Digital Watermarking”, *Morgan Kaufmann Publishers, Inc.*, San Francisco, 2001.

[2] H. C. Papadopoulos and C.-E. W. Sundberg, ”Simultaneous broadcasting of analog FM and digital audio signals by means of adaptive precanceling techniques”, *IEEE Trans. on Communicationsy*, vol. 46, num. 9, p. 1233-1242, 1998.

[3] C. E. Shannon, ”Channels with side information at the transmitter”, *IBM J. Res. Develop*, vol. 2, pp. 289–293, 1958.

[4] A. V. Kusnetsov and B. S. Tsybakov, ”Coding in a memory with defective cells”, translated from *Prob. Peredach. Inform.*, vol. 10, no. 2, pp. 52–60, April-June 1974.

[5] S. I. Gel’fand and M.S. Pinsker, ”Coding for channel with random parameters”, *Prob. Cont. and Inf. Theory*, vol. 9, no. 1, p. 19-31, 1980.

[6] C. Heegard and A. El Gamal, ”On the capacity of computer memories with defects”, *IEEE Trans. Inform. Theory*, vol. 29, pp. 731–739, September 1983.

[7] A. Sutivong, M. Chiang, T.M. Cover, and Y.-H. Kim, ”Channel Capacity and State Estimation for State-Dependent Gaussian Channels”, *IEEE Trans. on Inform. Theory*, vol. 51, no. 4, pp. 1486–1495, 2005.

[8] E. Martinian, G. W. Wornell, and B. Chen, ”Authentication with Distortion Criteria”, *IEEE Trans. on Inform. Theory*, pp. 2523–2542, July 2005.

[9] T. Kalker and F. M. Willems, ”Capacity bounds and code constructions for reversible data-hiding”, *SPIE Proceedings, Security and Watermarking of Multimedia Contents V*, vol. 5020, Santa Clara, USA, 2003.

[10] J. J. Eggers and R. Beuml and R. Tzschoppe and B. Girod, ”Inverse Mapping of SCS-Watermarked Data”, *Eleventh European Signal Processing Conference* EUSIPCO’2002, Toulouse, France, 2002.

[11] S. Voloshynovskiy, O. Koval, E. Topak, T. Pun, ”Message Communications and Channel State Estimation for State Dependent Channels”, *27th Symposium on INFORMATION THEORY in the BENELUX*, June 8-9, 2006, Noordwijk, The Netherlands.

[12] E. A. Haroutunian, ”On Bounds for E-capacity of DMC,” *IEEE Trans. on Information Theory*, vol. 53, no.11, pp. 4210-4220, 2007.

[13] 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, 2006.

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

[15] I. Csiszár and J. Körner, *Information Theory: Coding theorems for discrete memoryless systems*, Academic Press, New York, 1981.

[16] I. Csiszár, ”The method of types”, *IEEE Trans. Inform. Theory*, vol. 44, no. 6, pp. 2505–2523, 1998.

[17] 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.